

## Large-scale characterization of Tos17 insertion sites in a rice T-DNA mutant library

Pietro Piffanelli · Gaétan Droc · Delphine Mieulet · Nadège Lanau ·  
Martine Bès · Emmanuelle Bourgeois · Claire Rouvière · Frédérick Gavory ·  
Corinne Cruaud · Alain Ghesquière · Emmanuel Guiderdoni

Received: 14 May 2007 / Accepted: 8 August 2007 / Published online: 15 September 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** We characterized the insertion sites of newly transposed copies of the tissue-culture-induced ty1-copia retrotransposon Tos17 in the Oryza Tag Line (OTL) T-DNA mutant library of rice cv. Nipponbare. While Nipponbare contains two native copies of Tos17 the number of additional copies, deduced from Southern blot analyses in a subset of 384 T-DNA lines and using a reverse transcriptase probe specific to the element, ranged from 1 to 8 and averaged 3.37. These copies were shown to be stably inherited and to segregate independently in the progenies of insertion lines. We took advantage of the absence of *EcoRV* restriction sites in the immediate vicinity of the 3' LTR of the native copies of Tos17 in the genome sequence of cv. Nipponbare, thereby preventing amplification of corresponding PCR fragments, to efficiently and selectively amplify and sequence flanking regions of newly transposed Tos17 inserts. From 25,286 T-DNA plants, we recovered 19,252 PCR products (76.1%), which were sequenced yielding 14,513 FSTs

anchored on the rice pseudomolecules. Following elimination of redundant sequences due to the presence of T-DNA plants deriving from the same cell lineage, these FSTs corresponded to 11,689 unique insertion sites. These unique insertions exhibited higher densities in subtelomeric regions of the chromosomes and hot spots for integration, following a distribution that remarkably paralleled that of Tos17 sites in the National Institute for Agrobiological Sciences (NIAS) library. The insertion sites were mostly found in genic regions (77.5%) and preferably in coding sequences (68.8%) compared to unique T-DNA insertion sites in the same materials (49.1% and 28.3%, respectively). Predicted non-transposable element (TE) genes prone to a high frequency of Tos17 integration (i.e. from 5 to 121 inserts) in the OTL T-DNA collection were generally found to be also hot spots for integration in the NIAS library. The 9,060 Tos17 inserts inserted into non TE genes were found to disrupt a total of 2,773 genes with an average of 3.27 inserts per gene, similar to that in the NIAS library (3.28 inserts per gene on average) whereas the 4,472 T-DNA inserted into genes in the same materials disrupted a total of 3,911 genes (1.14 inserts per gene on average). Interestingly, genes disrupted by both Tos17 and T-DNA inserts in the library represented only 14.9% and 10.6% of the complement of genes interrupted by Tos17 and T-DNA inserts respectively while 52.1% of the genes tagged by Tos17 inserts in the OTL library were found to be tagged also in the NIAS Tos17 library. We concluded that the first advantage in characterizing Tos17 inserts in a rice T-DNA collection lies in a complementary tagging of novel genes and secondarily in finding other alleles in a same genetic background, thereby greatly enhancing the library genome coverage and its overall value for implementing forward and reverse genetics strategies.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11103-007-9222-3) contains supplementary material, which is available to authorized users.

P. Piffanelli · G. Droc · D. Mieulet · N. Lanau · M. Bès ·  
E. Bourgeois · C. Rouvière · E. Guiderdoni (✉)  
Biological Systems Department TAA96/03, CIRAD, UMR DAP  
1098, 2477 Avenue Agropolis, Montpellier cedex 5 34398,  
France  
e-mail: guiderdoni@cirad.fr

F. Gavory · C. Cruaud  
Génoscope, Centre National de Séquençage, 2 rue Gaston  
Crémieux, CP5706, Evry cedex 91057, France

A. Ghesquière  
IRD, UMR GDP 5096, 911, Avenue Agropolis, BP 64501,  
Montpellier cedex 5 34394, France

**Keywords** Insertion mutagenesis · Rice · T-DNA · Tos17

## Introduction

With the completion of a high quality sequence of the rice genome (cv Nipponbare) (Matsumoto et al. 2005), this crop species has become the model for deciphering the molecular control of agronomical traits in cereals, motivating the concurrent development of large libraries of mutant lines using chemical, physical or insertion mutagenesis (reviewed in (Hirochika et al. 2004). Like in *Arabidopsis*, generation of insertion libraries in rice has largely relied on the use of the T-DNA (reviewed in (Guideroni et al. 2007) and of class II transposable elements, such as the maize Ac/Ds and En/Spm transposon systems (recently updated in (Zhu et al. 2007). Another powerful mutagen in rice is the class I ty1-copia retrotransposon Tos17, which naturally resides in rice cultivars in a low number of copies and the transpositional activity of which has been evidenced in the mid-90s (Hirochika et al. 1996). Tos17 activity is specifically induced by the stress resulting from tissue culture and, as the retroelement transposes through a copy and paste mechanism, its copy number gradually increases with the tissue culture duration (Hirochika 2001). This property has been profitably used for generating a library of 55,000 insertion lines, regenerated from cell aggregates in suspension at the National Institute for Agrobiological Sciences (NIAS) (Tsukuba, Japan). These lines have been evaluated in the field (Miyao et al. 2007) while more than 18,000 Tos17 insertion sites have been characterized (Miyao et al. 2003) (data available at <http://tos.nias.affrc.go.jp/>). Due to the cell culture procedure, each line harbours an average of 10 new Tos17 copies. Based on the frequency of tagging by Tos17 of alterations observed in field evaluated insertion lines, it was estimated that the retroelement causes 5–10% of the observed somaclonal variation (Hirochika 2001), a phenomenon extensively reported in rice plants deriving from callus and cell cultures.

As the *Agrobacterium*-mediated transformation procedure of rice involves tissue culture steps, T-DNA insertion lines may contain additional copies of the Tos17 element. This actually depends on the presence of active copies of the retroelement in the cultivar used to generate the T-DNA insertion library. In rice, Nipponbare, which was used to produce a reference high-quality sequence and to generate insertion libraries, is known to harbour two copies of Tos17, one of which is active and located on chromosome 7. On the other hand, the 3 native copies of Tos17 residing in another *japonica* cultivar, Tainung 67 exhibit a very low activity with an average of 0.2 new copy per T-DNA plant (Hsing et al. 2007). Though the presence of an active

element in the T-DNA plants may first appear a drawback in term of generation of unwanted mutations, this may turn in an advantage once the insertion sites of the element have been characterized. Tos17 may for instance create alleles in genes already tagged by a T-DNA insert or tag new genes in the same genetic background. Tos17 and T-DNA alleles in the POSTECH insertion line library (Jeong et al. 2006) have for instance been used to ascertain the function of genes such as *FLORAL ORGAN NUMBER 1* (Moon et al. 2006) and *UNDEVELOPED TAPETUM 1* (Jung et al. 2005). Moreover, the preference of insertion of the two mutagens appear complementary: while the two mutagens are both more frequently found in gene-rich regions, T-DNA inserts are more frequent in promoter and 3'UTRs regions while Tos17 inserts in coding sequences (An et al. 2003; Hsing et al. 2007; Jeong et al. 2006; Miyao et al. 2003; Sallaud et al. 2004; Zhang et al. 2006).

In the frame of the French genomics initiative Génoplante, we have generated a library of 30,000 T-DNA insertion lines in cv. Nipponbare, which has been characterized for T-DNA integration sites (Sallaud et al. 2004), and is being seed increased and evaluated under agronomic conditions at the Centro Internacional de Agronomía Tropical (CIAT) in Colombia (Lorieux et al unpublished). The resulting phenotypic information -now collected from about 14,000 insertion lines and gathered in the Oryza Tag Line (OTL) database (<http://urgi.versailles.inra.fr/OryzaTagLine> and Larmande et al., submitted)- has evidenced that 22% of the lines exhibit alteration in at least one trait. Aside on going establishment of tagging by DNA blot analyses using T-DNA- and Tos17- specific probes in mutant lines altered for specific traits, large scale sequencing of the Tos17 insertion sites in the library may provide new clues on causative links between the predicted functions of the disrupted gene and phenotype alterations observed in the field.

To complement the characterization of the OTL library, we first estimated the range and average number as well as inheritance of new copies of Tos17 in the regenerated T-DNA plants using DNA blot analysis and a reverse transcriptase probe. In a first attempt to characterize insertion sites, we developed a protocol, called TOSTRAP, based on adapter-ligation PCR (Siebert et al. 1995) for both readily and specifically amplifying genomic regions flanking insertion sites of new Tos17 copies. Applying this protocol to a large part of the library, we generated 19,252 PCR products and 14,513 good sequences, which, once anchored on the rice pseudomolecules, corresponded to 11,689 unique sites. We then investigated the preference of insertion of Tos17 and compared it to that of T-DNA in the same materials and that of Tos17 in the NIAS Nipponbare lines. We also assessed the number and the functional classification of genes tagged both by T-DNA and Tos17 inserts in the library.

## Materials and methods

### Plant materials

The Génoplante insertion line library was generated by *Agrobacterium*-mediated transformation of mature seed embryo-derived callus of rice (*Oryza sativa* L.), *japonica* cultivar Nipponbare following the procedure detailed in (Sallaud et al. 2003). Seeds were kindly provided by M. Yano (RGP, NIAS Tsukuba, Japan). We hereafter briefly summarize the time course of the procedure and the resulting cell lineage of T-DNA plants (Fig. 1): Three weeks following inoculation of mature seeds on a 2,4-D containing induction medium, several (5 on average) globular somatic embryos are generally isolated from the scutellum-derived primary callus, and allowed to grow on fresh medium for 10–14 days, forming Embryogenic Nodular Units (ENUs) (Bec et al. 1998). ENUs are immersed in an EHA105 or LBA4404 *Agrobacterium tumefaciens* suspension, then blot dried and transferred to a solid co-cultivation medium for 3 days. Following 5 weeks of selection, several hygromycin-resistant cell lines may arise from a single co-cultivated ENU and further regenerate T-DNA plants (4 plants on average in cultivar Nipponbare) which are generally independent transformation events. T-DNA plants may therefore share a same ENU and/or primary callus origin.

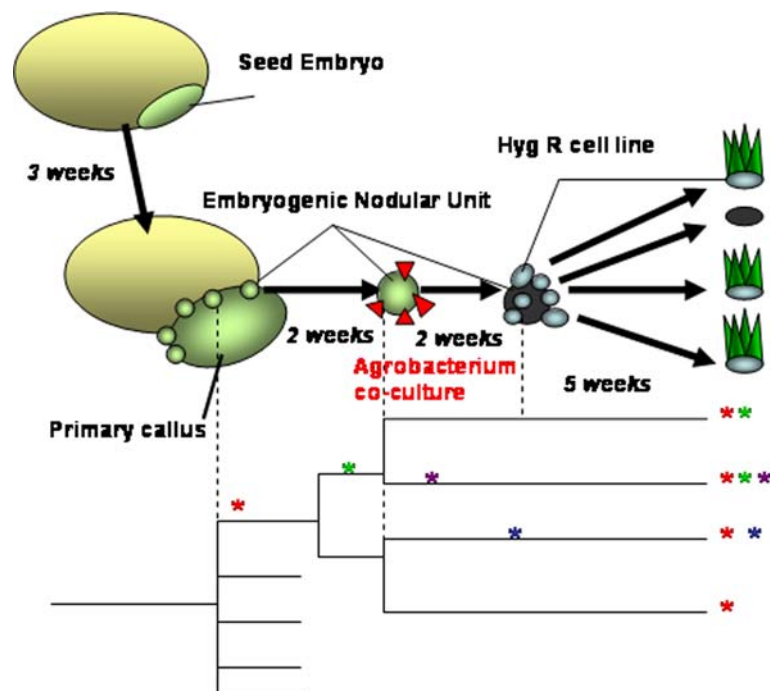
About 25,286 T-DNA plants of the library were used in this study. DNA was isolated from leaves of regenerated T-DNA plants (T0) at their transfer to test tubes (for PCR

analysis) (Sallaud et al. 2004) and, for 384 lines randomly selected in the pilot study, from the antepenultimate leaf of the several tillers of adult T0 plants (for DNA blot analysis). To ascertain transmission of Tos17 inserts in progenies, leaf samples of 20 T1 progeny seedlings were collected and pooled, in 65 randomly selected insertion lines belonging to the 384 subset. The segregation of the inserts was further investigated by collecting leaves on 8–16 individual T1 progeny plants in 15 other T1 lines exhibiting a high Tos17 copy number.

### DNA blot analysis

To investigate the integration and transmission of Tos17 copies in T-DNA lines, total genomic DNA was extracted from leaf tissue using the MATAB method. Five  $\mu\text{g}$  of DNA were digested with XbaI endonuclease which cuts once within the Tos17 sequence of both resident copies and DNA fragments were separated on 0.8% agarose gels in TBE buffer, and transferred to nylon membranes (Hybond N+ Amersham) through alkaline transfer. A 0.3 kbp PCR fragment of the Tos17 reverse transcriptase (5' GCTTTTC TTCATGGTGA 3' and 5' CACGGGAGGAAGTATGA 3') served as template for synthesizing [ $\alpha$ - $^{32}\text{P}$ ] labelled probes through random priming. Following hybridization, the membrane was washed twice (SSC 2 $\times$ , SDS 1% and SSC 0.1 $\times$ , SDS 0.1% both at 65°C for 10 min) and analyzed by autoradiography.

**Fig. 1** Scheme summarising the tissue culture procedure used in the rice transformation protocol of the OTL T-DNA collection and illustrating the possibility for newly transposed Tos17 inserts (shown as stars) to sequentially accumulate in the cell lineages giving rise to independent transformation events and also explaining the observed partial redundancy in the sequences obtained of Tos 17 inserts



**Table 1** Conditions employed to selectively amplify the newly-transposed Tos17 elements in the OTL T-DNA rice collection

Amplification from the 5' LTR	Adapters	5' CAC TGA ATC TTG CTG ACT AGG TCT GGG GAG GT 3' 5' P-ACC TCC CCA GAC-NH2 3'
	Adapter-specific primers	PCR1: 5' CTG AAT CTT GCT GAC T 3' PCR2: 5' ATC TTG CTG ACT AGG T 3' PCR 3 and sequencing: 5' ACT AGG TCT GGG GAG GT 3'
	5' LTR specific primer	PCR1: 5' GACATGGGCCAACTA 3' PCR2: 5'ATGGGCCAACTATACAGTACA 3'
Amplification from the 3' LTR	Adapters	5' CTA ATA CGA CTC ACT ATA GGG CTC GAG CGG CCG CCC GGG GAG GT 3' 5' P-ACC TCC CC-NH2 3'
	Adapter-specific primer	PCR1: 5' GGA TCC TAA TAC GAC TCA CTA TAG GGC 3' PCR2: 5' CTA TAG GGC TCG AGC GGC 3'
	3' LTR specific primer	PCR1: 5' AGT CGC TGA TTT CTT CAC CAA GG 3' PCR2: 5' GAG AGC ATC ATG GGT TAG ATC TTC TC 3' Sequencing: 5' GTA CTG TAT AGT TGG CCCATG TCC 3'
PCR mix	PCR1 (final concentration)	PCR 2 (final concentration)
	16.5 ng of digested-ligated DNA (5× dilution)	2 µl 50× diluted PCR1
	75 µM dNTP	75 µM dNTP
	0.2 µM Adapter primer	0.2 µM Adapter primer
	0.2 µM Right LTR specific primer	0.2 µM Right LTR specific primer
	0.05 U/µl Taq	0.05 U/µl Taq
	1× Euro Mg Buffer	1× Euro Mg
	H <sub>2</sub> O (final volume 20 µl)	H <sub>2</sub> O (final volume 100 µl)
94°C 3 mn	30 cycles	35 cycles
92°C 30 s		
67°C 45 s		
72°C 150 s		
72°C 5 mn		

Upper panel: Adapter and primer oligonucleotides used in the walk PCR protocol. Lower panel: PCR1 and PCR2 conditions

#### Amplification of flanking regions of Tos17 inserts by adapter-ligation PCR

Amplification of regions flanking the 5' and the 3' LTR of Tos17 inserts was performed following an adapter ligation-PCR protocol (Siebert et al. 1995), adapted to plant DNA (Balzergue et al. 2001; Devic et al. 1997), and modified for Tos17. Briefly, in a 10µl reaction volume, 62,5 ng of genomic DNA was digested with 1U *EcoRV* (or *SspI*) blunt end restriction enzyme and ligated with 0.24 U of T4 ligase to 375 µM of asymmetric adapters (Table 1) in a single step at 25°C overnight. To determine the number of amplification products, two successive amplifications with nested primers specific to the 3' (or 5') Long Terminal Repeat (LTR) of Tos17 and to the adapters were performed according to conditions detailed in Table 1. PCR2 products generated from the 3' LTR in *EcoRV* digests were sequenced by Genome Express company (Grenoble, France) and the National Sequencing Centre Génomscope (Evry, France) and were uploaded to the EMBL nucleotide

sequence database under accession numbers ER893126 to ER893861 and CU313326 to CU326334 respectively.

#### Analysis of Tos17 sequences

##### Data resources

The rice genome and its annotation were downloaded from the TIGR Rice Genome Database (Version 4.0, January 2006). Genomic DNA sequences corresponding to entire chromosomes were formatted into FASTA-format nucleotide database files for BLAST (Basic Local Alignment Search Tool). The genome annotations were also used for the insertion mapping and as input files integrated into the OryGenesDB database <http://OryGenesDB.cirad.fr> (Droc et al. 2006).

Tos17 FSTs (Miyao et al. 2003) and FL-cDNA sequences (The Full Length cDNA consortium, 2003) were obtained from GenBank at the National Center for

Biotechnology Information (NCBI) with query “txid4530 [orgn] AND GSS [PROP] AND Tos17 AND NIAS” and <http://cdna.01.dna.affrc.go.jp/cDNA/> respectively.

### *FST Mapping and annotation*

Each FST was mapped to the rice genome by BLAST, and annotated with the TIGR gene code when the insertion qualifies as a ‘gene hit’. If a flanking sequence was mapped to multiple genomic loci, only the best mapping was retained. We define ‘gene hit’ as the insertion site being located between 1000 bp upstream of the ATG and 300 bp downstream of stop codon of a gene. Based on the genomic coordinate, Flanking sequences were determined to be identical if they had less than a 3 bp mismatch at the insertion site.

### *GC content*

We evaluated the GC content of 1 kb regions extending 5′ and 3′ from the mutagen insertion point corresponding to different sets of FSTs (see results section). We then determined the distribution of the GC content for each population. (Fig. 5).

### *Gene Ontology annotation*

Using the latest release of the Gene Ontology Slim versions ([ftp://ftp.geneontology.org/pub/go/GO\\_slims/goslim\\_plant.obo](ftp://ftp.geneontology.org/pub/go/GO_slims/goslim_plant.obo)), which give a high level view of the three GO ontologies, we classified genes which are annotated to the term itself, or to any of the children of the GO slim term, are included in the corresponding GO slim category.

### *Expression of Tos17 target genes in callus tissue*

To determine the frequency of expression of Tos17 target genes in callus tissues, we based our search on the number of occurrences of Massive Parallel Signature Sequencing (MPSS) cDNA-derived tags in the NCA (callus) library at <http://mpss.udel.edu/rice/>, through bulk query using subsets of 150 gene accessions.

### *Software*

The data are stored in a MySQL database management system in order to provide fast and flexible queries as well as to facilitate the analysis. The back end sequences

analysis codes were written in the Perl programming language. R (R development Core Team, 2005) and Bioconductor packages (i.e. limma for Venn’s diagram, Fig. 6) were used for statistical computing and graphics.

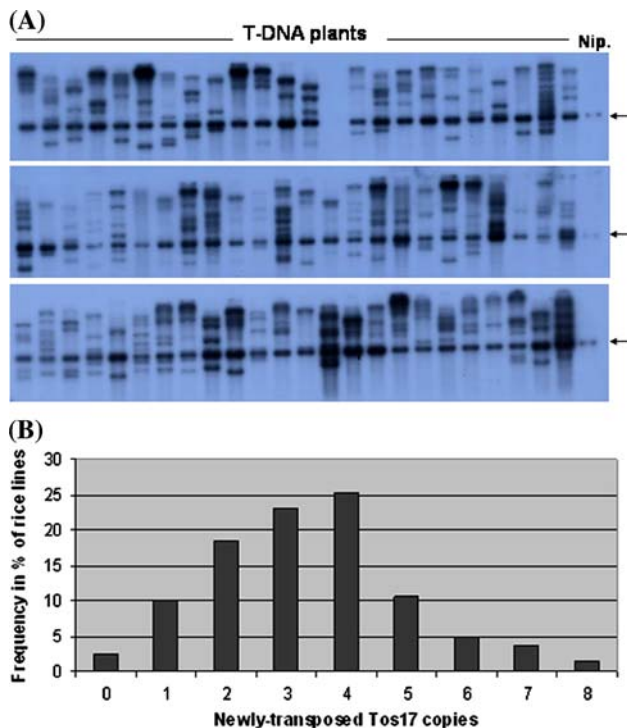
### *Availability of plant materials*

Seeds and related phenotype information of the insertion lines used to generate the FSTs in this study are publicly available. The T2 seeds are distributed once seed increase of the insertion lines is achieved. A Material Transfer Agreement and an order form are downloadable at <http://urgi.versailles.inra.fr/OryzaTagLine>. Lines can be identified from their sequence tags at <http://OryGenesDB.cirad.fr>, which now provides a direct link to the Oryza Tag Line database through the line ID.

## **Results**

Southern analysis of a subset of 384 lines for number of newly transposed Tos17 copies

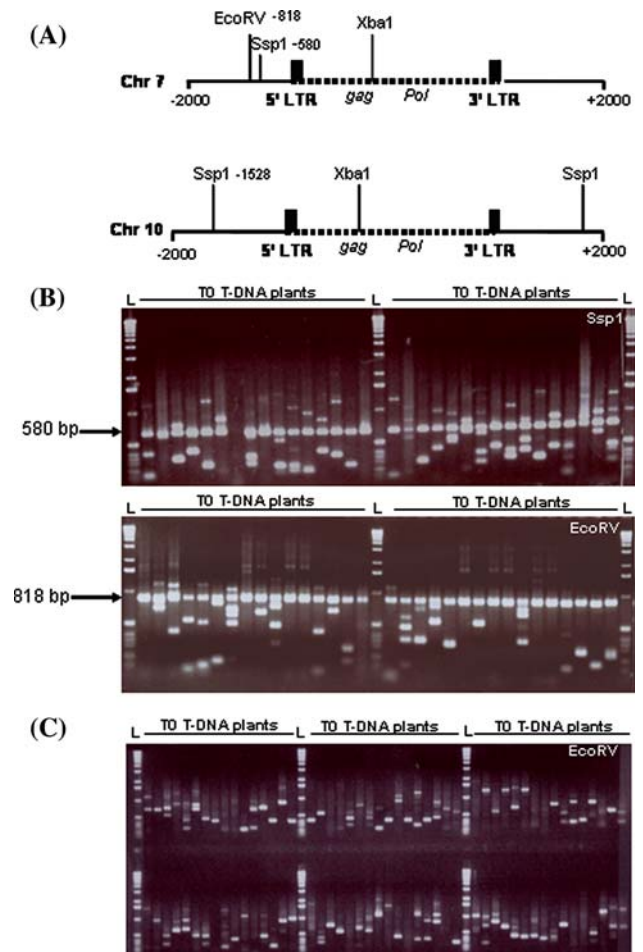
DNA blots of the 384 T-DNA plants and untransformed Nipponbare hybridized to the Tos17 specific reverse transcriptase probe shared a strong hybridization signal at 4 kbp which actually consists of two thinner bands of similar molecular weight (3.9 and 4.1 kbp) (Fig. 2A). These bands correspond to the two Tos17 native copies residing on chromosomes 10 and 7 respectively in cv. Nipponbare. In addition, the T-DNA plants exhibited 0–8 additional bands corresponding to putative newly-transposed Tos17 copies, the transposition of which was induced by the tissue culture steps of the rice transformation procedure. Distribution of these additional bands in the 384 subset is shown in Fig. 2B. A total of 1,295 putative new copies were found in the 384 subset of analysed lines, representing an average of 3.37 copies per line. Stability and transmission of these copies in the T1 progeny of primary transformants was ascertained by parallel DNA blot analysis of T0 plants and pooled samples of T1 plants in 65 lines (Supplementary Fig. 1A). While all the bands observed in T0 were found in T1 samples, no additional band, which would have resulted from new transposition events, was observed. Furthermore, T1 progeny plants of 15 T0 transformants harbouring a high number of Tos17 newly-transposed copies were individually analyzed by DNA blot and demonstrated that the copies observed in T0 segregated, as expected, independently in the T1 progeny (Supplementary Fig. 1B). From this analysis, we extrapolated that the 25,286 T-DNA lines regenerated in this study overall harbour 83,686 newly-transposed Tos17 copies.



**Fig. 2** (A) DNA blot analysis of Tos17 integration patterns in a set of 72 T-DNA plants (XbaI digests). The arrow points to the two hybridizing fragments of approx. 4 kbp corresponding to the hybridization signals of the two native (resident) copies of Tos17 in the genome of the *Oryza sativa* cv. Nipponbare. Nip.: Nipponbare control. (B) Distribution of frequency of additional hybridization signals corresponding to newly-inserted Tos17 copies, among the 384 analysed T-DNA plants

#### Amplification of Tos17 flanking regions through adapter ligation PCR in the 384 subset

To determine the fraction of these putative newly-transposed Tos17 copies from which a flanking genomic fragment can be recovered using adapter-ligation PCR, we used Tos17- and adapter- specific primers and genomic DNA digested by the blunt end restriction enzymes Ssp1 and *EcoRV*. Amplification with primers specific to the 5' LTR of Tos17 was first investigated. As shown in Fig. 3A,B, all of the PCR2 products in Ssp1 and *EcoRV* digests, except the negative controls, exhibited a common fragment at 580 bp and 818 bp, respectively and indicative of the amplification of the region flanking the chromosome 7 Tos17 “native” copy residing in the Nipponbare cultivar (Fig. 3A). In Ssp1 digests, 80.2% of the samples exhibited additional PCR fragments (51.5% a single fragment and 48.5% multiple (2–5) additional fragments). A total of 523 PCR fragments, putative flanking regions of new Tos17 inserts, were amplified from the 384 subset, representing 40.4% of the 1,295 additional signals identified by DNA blot analysis.



**Fig. 3** (A) Positions of *EcoRV* and Ssp1 restriction sites on 2 kbp regions extending from the 5' and 3' LTRs of the Tos17 elements located on chromosome 7 and 10 in the genome of the *Oryza sativa* variety Nipponbare (B) Results obtained using the walk PCR mediated amplification of regions flanking the 5' LTR of Tos17 in Ssp1 and *EcoRV* digests of genomic DNA of T-DNA plants. Positions of the fragments corresponding to the flanking region of the chromosome 7 native copy are pointed by arrows. (C) Results obtained using the walk PCR mediated selective amplification of regions flanking the 3' LTR of newly-transposed Tos17 copies in *EcoRV* digests of genomic DNA of T-DNA plants

In *EcoRV* digests, a lower frequency of 64.3% of the samples exhibited additional PCR fragments compared to the control Nipponbare (64.3% a single fragment and 35.7% multiple (2–4) fragments), likely reflecting the lower density of *EcoRV* sites in the rice genome. A total of 321 PCR fragments, representing 24.8% of the DNA blot signals were amplified from the 384 subset. Thirty seven samples with no additional fragments in Ssp1 digests had additional fragments in *EcoRV* digests while 121 samples with no additional fragment in *EcoRV* digests had additional fragments in Ssp1 digests. Manual picking and individual amplification in a third PCR of the additional bands observed in PCR2, followed by sequencing of the products proved to

be possible (Supplementary Fig. 2), but with a fluctuating efficiency (60% on average) and found to be very time- and labour- consuming. We concluded that an automation of the process was needed before pursuing the recovery of flanking regions from the 5′LTR of Tos17 inserts in the whole T-DNA rice collection, although the recovery rate of PCR2 fragments from this border appeared satisfactory.

Contrastingly, amplification of flanking regions from the 3′ LTR of Tos17 in DNA samples digested with *EcoRV*, did not conduct to the amplification of PCR fragments from the Nipponbare control due to the absence of *EcoRV* sites in the immediate vicinity (2 Kbp) of the native copies residing on chromosomes 7 and 10 (Fig. 3A). Out of the 384 analysed TO T-DNA plants 265 (69%) of them (Fig. 3C) exhibited PCR2 fragments (of which 50.2% with a single fragment and 49.8% with multiple (2–4) fragments). A total of 415 PCR fragments representing 32% of the DNA blot signals were amplified from the 384 subset of rice T-DNA lines. These 265 PCR products were readily sequenced and yielded 237 (89.4%) good quality sequences which were anchored on the rice pseudomolecules and represented a success rate of 0.62 compared to the original 384 lines under examination. From these results we extrapolated that the analysis of the whole collection of 25,286 T-DNA plants should yield 15,677 sequences, representing the 18.7 % of the estimated 83,686 newly-transposed Tos17 copies in the library.

#### Sequence and analysis of Tos17 flanking regions

Amplification of flanking regions from the 3′LTR of Tos17 inserts in the *EcoRV* digests of the 25,286 DNA samples yielded 19,252 PCR products (76.1%) of which 55.9% with single and 44.1% with multiple fragments, respectively.

Direct sequencing of single fragment conducted to a higher frequency of good quality sequences than direct sequencing of multiple PCR fragments. On the other hand, when the sequence was long enough, we were able to retrieve two distinct sequences anchoring in different points of the rice genome sequence in 4% of the multiple fragment PCR products. This indicates that the dominant sequence signal is that of the shorter sequence corresponding to the low molecular weight PCR fragment. Overall 98.6% of the obtained sequences were anchored to the rice pseudomolecules totalling 14,513 sequences positioned in the rice genome which represent a slightly lower number than anticipated from the 384 subset data.

#### FST redundancy

We found that 2,824 of the 14,513 sequences were redundant with at least one other sequence, with a FST start

pointing exactly to a same interval of  $\pm 3$  nucleotides on the rice genome sequence. Even though Tos17 insertion is prone to hot spots (see below), it is rather unlikely that two Tos17 integrations randomly occurred precisely at the same nucleotide position. Redundancy of sequences may therefore result from either cross contamination of DNA samples and PCR products or from sequencing the flanking regions of Tos17 inserts in plants with a shared cell lineage. As described in the Materials and Methods section, the co-cultivation procedure uses embryogenic nodular units (ENUs), which may derive from the same primary seed embryo, scutellum callus, and later yield an average of 4 regenerating, hygromycin-resistant cell lines (Fig. 1). As the ENU lineage information was available in our materials, we examined whether the redundancy of sequences in several T-DNA plants can be explained by a shared cell lineage. Manual curation of the redundant sequences indicated that in most cases (88.2%), redundant Tos17 FSTs were shared between different T-DNA plants regenerated from the same ENU. The rest (11.8%) could be explained by either PCR contamination or ENUs sharing cell lineage in originating from the same primary callus. The former result indicates that Tos17 transposition occurred early during ENU formation and resulted in a Tos17 insert present in most cells of the co-cultivated ENU. These redundant sequences were further eliminated in the subsequent set of analyses which were therefore carried out with 11,689 unique Tos17 newly-transposed insertion points.

#### Distribution of Tos17 inserts among and along rice chromosomes

The preference of insertion of Tos17 (Miyao et al. 2003) as well as that of the T-DNA (An et al. 2003; Hsing et al. 2007; Jeong et al. 2006; Sallaud et al. 2004) were extensively documented in previous reports. On the other hand, parallel characterization of Tos17 and T-DNA insertional behaviours in the same regenerated plant material had yet to be examined. Comparative analysis revealed a similar distribution and chromosomal insertion preference for Tos17 inserts in the T-DNA OTL and Tos17 NIAS collections (Table 2). In both collections chromosomes 2, 1, 3 and 6 exhibited the highest insertion densities while chromosomes 10 and 11 showed lower insertion densities of Tos17 (Table 3). T-DNA plants with characterized Tos17 insertion sites were found to harbour 9,113 anchored T-DNA insertion sites (Table 2). These T-DNA inserts also exhibited the highest density on chromosomes 1, 2 and 3. The largest chromosomes 1, 2 and 3 were also found to bear the highest densities of non-TE predicted genes and of expressed genes (Table 3). The distribution along the rice

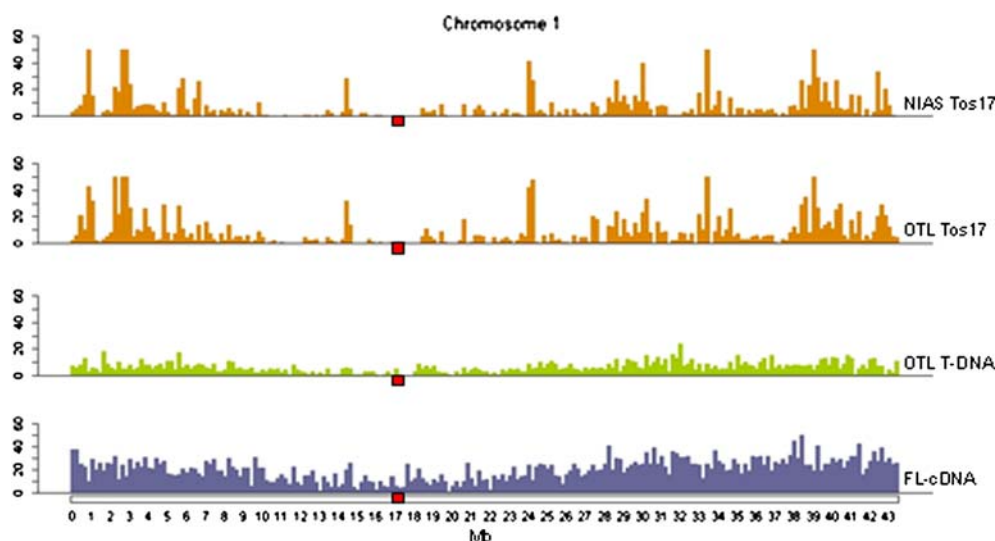
**Table 2** Distribution of newly-transposed Tos 17 inserts from the OTL insertion line library on the 12 rice chromosomes compared to those of Tos17 and T-DNA inserts in the NIAS and OTL libraries, respectively

Chromosome	Tos17 OTL		Tos17 NIAS		T-DNA OTL		Non TE genes	TE genes	FL cDNA
	All sequences	Non redundant insertion points	All sequences	Non redundant insertion points	All sequences	Non redundant insertion points			
1	1753	1544	2215	1957	1899	1338	5433	1161	4460
2	1961	1597	2065	1800	1455	1045	4372	969	3622
3	1538	1272	1997	1735	1600	1162	4651	907	4169
4	1314	1114	1528	1340	1058	737	3815	1525	2733
5	1135	966	1459	1258	961	655	3407	1191	2584
6	1421	1049	1615	1413	1044	743	3577	1160	2396
7	1374	848	1280	1105	994	676	3410	1078	2281
8	1103	913	1290	1141	936	627	3104	1087	2044
9	798	634	1120	974	761	545	2482	909	1702
10	664	539	1036	742	703	492	2532	907	1570
11	635	540	1159	997	802	559	3106	1092	1513
12	817	673	1191	1031	733	534	2764	1251	1585
Total	14513	11689	17955	15493	12946	9113	42653	13237	30659

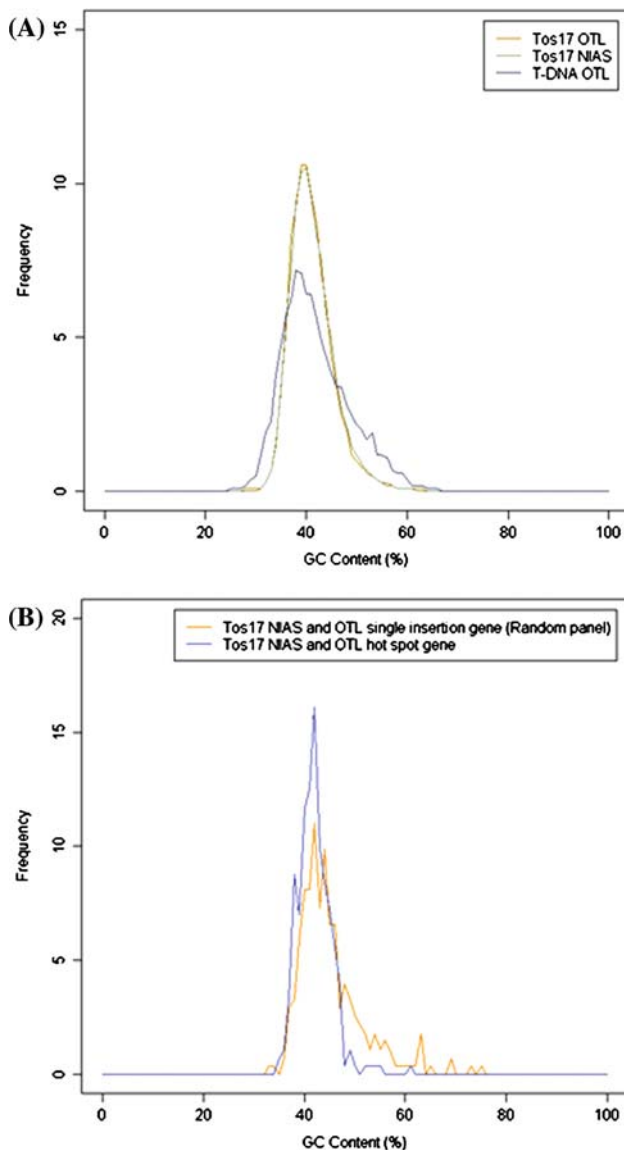
The TIGR release 4 of the rice genome sequence was used. Tos17 FSTs from the NIAS were downloaded from Genbank. For Tos17, non redundant sequences excluded sequences starting exactly at the same nucleotide ( $\pm 3$  nucleotides) in the FST data set, due to characterization of inserts in plants sharing a common cell lineage. For T-DNA, non redundant sequences excluded either the left or the right border flanking sequence of a same T-DNA insert characterized at both ends and redundant sequences due to presence of T-DNA plants clonal for the T-DNA insert (same transformation event). TE = transposable element; FL cDNA = full length cDNA

chromosomes of the 11,689 Tos17 inserts and of the 9,113 unique T-DNA inserts characterized in lines also containing a characterized Tos17 insert is illustrated in Fig. 4 (for Chromosome 1) and Supplementary Fig. 3 (for all 12 chromosomes). Density graphs of Tos17 inserts of the OTL library paralleled those of Tos17 inserts characterized at NIAS with a higher frequency in subtelomeric,

euchromatic regions and a lower frequency in pericentromeric, heterochromatic regions. Hot spots for integration were also generally lying in the same sequence intervals. On the other hand, T-DNA inserts were more evenly distributed along the chromosomes, being more frequently recovered in pericentromeric regions and less prone to hot spots for integration than Tos17.

**Fig. 4** Density graphs of (Up to Down tracks) NIAS Tos17 insertions, OTL Tos17 insertions, OTL T-DNA insertions and of FL-cDNA along chromosome 1. Position of the centromere is shown by a red square





**Fig. 5** (A) Distribution of GC content of 1 kb regions extending 5' and 3' from OTL Tos17, NIAS Tos17 and OTL T-DNA insertion points. (B) Distribution of GC content of the 274 Tos17 hot spot genes and of a random set of 274 non hot spot Tos17 target genes

#### Preference for insertion in genes and gene regions

The GC contents of Tos17 and T-DNA FSTs were similar averaging 41.4% and 40.6% respectively, slightly lower than the average GC content of the rice genome (43.6%). The GC range at Tos17 insertion sites was narrower than that of T-DNA insertion sites in the OTL lines, and found to parallel the GC content distribution of Tos17 insertion sites in the NIAS library (Fig. 5A). Tos17 insertion sites were mostly found in genic vs. intergenic regions (84.1%), in non-TE predicted genes (77.5%) and preferably in coding sequences (68.8%), compared to the unique T-DNA

insertion sites characterized in the same plant material (53.5% in genic, 49.1% in non-TE predicted genes and 28.3% in coding sequences, respectively) (Table 4). As previously reported, T-DNA inserts were more frequently retrieved in promoter and 3'UTR regions than in coding sequences, strikingly differing in that from Tos17 inserts recovered in the same lines.

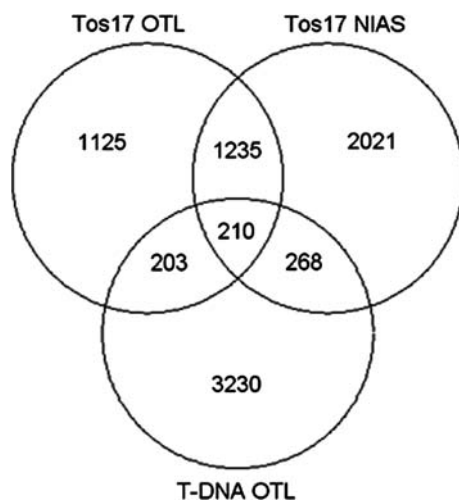
#### Complementary feature of Tos17 and T-DNA insertions

The 9,060 Tos17 newly-transposed elements inserted into non-TE genes were found to disrupt a total of 2,773 genes with an average of 3.27 inserts per gene, similar to that observed in the NIAS library (3.28 inserts per gene on average) whereas the 4,472 T-DNA elements inserted into non-TE genes in the same analyzed lines disrupted 3,911 genes (1.14 inserts per gene on average). The overlap between genes targeted by Tos17 and T-DNA in the OTL library appear to be limited, since genes interrupted by both mutagens in the library represented only 14.9% and 10.6% of the sets of genes interrupted by Tos17 and T-DNA inserts, respectively. On the other hand, a large fraction (48%) of the genes tagged by Tos17 inserts in the OTL T-DNA library were also found to be tagged by Tos17 in the NIAS collection (Fig. 6). This indicates that characterizing Tos17 inserts in a T-DNA rice collection allows a rapid increase in genome coverage through a highly-complementary tagging of additional genes. On the other hand, the chance of finding Tos17 mutant alleles in a gene already tagged by a T-DNA insert in the same genetic background may prove to be lower, as long as the overall genome coverage by both categories of inserts is limited. Genes tagged by Tos17 were classified through gene ontology (GO) and their distribution in different classes compared to that of the T-DNA inserts, the NIAS Tos17 inserts and the whole classified, non-TE rice gene complement. We did not detect any striking bias for a given broad GO category which would be specific to Tos17. However, some categories tended to be consistently over- and under-represented among Tos17 tagged genes when compared to T-DNA tagged ones in both the OTL and NIAS libraries (Supplementary Table 1). As Tos17 was reported to often integrate into defence response genes, due to their GC content falling into the target site narrow GC range and their frequent clustered organization along the chromosomes (Miyao et al. 2003), we investigated whether this class of genes was also over-represented in the genes interrupted by Tos17 inserts in the OTL library. Though representing 4.75% of the non-TE gene complement successfully classified through GO ( $n = 18,584$ ) and 4.15% of the genes interrupted by T-DNA inserts in the OTL library,

**Table 3** Densities of insertion in the 12 rice chromosomes of the newly-transposed Tos17 elements in the OTL library compared to those of Tos17 and T-DNA inserts in the NIAS and OTL libraries, respectively

Chromosome	Chromosome size (Mb)	Density (per Mb)			Insertion density (per Mb)		
		non-TE genes	TE	FL cDNA	Tos17 OTL	Tos17 NIAS	T-DNA OTL
1	<b>43.6</b>	<b>124.62</b>	26.63	<b>102.30</b>	<b>35.42</b>	<b>44.89</b>	<b>30.69</b>
2	<b>35.9</b>	<b>121.70</b>	26.97	<b>100.82</b>	<b>44.45</b>	<b>50.10</b>	<b>29.09</b>
3	<b>36.3</b>	<b>127.97</b>	24.95	<b>114.70</b>	<b>35.00</b>	<b>47.74</b>	<b>31.97</b>
4	35.2	<i>108.24</i>	<b>43.27</b>	77.54	31.61	38.02	<i>20.91</i>
5	29.9	114.05	<b>39.87</b>	86.50	32.34	42.11	21.93
6	31.2	114.48	37.12	76.68	33.57	45.22	23.78
7	29.7	114.86	36.31	76.83	28.56	37.22	22.77
8	28.3	109.65	38.40	72.20	32.25	40.30	22.15
9	23.0	<i>107.86</i>	39.50	73.96	27.55	42.33	23.68
10	22.9	110.68	39.65	<i>68.63</i>	<i>23.56</i>	32.43	21.51
11	28.5	109.13	38.37	<i>53.16</i>	<i>18.97</i>	<i>35.03</i>	<i>19.64</i>
12	27.5	<i>100.52</i>	<b>45.50</b>	<i>57.64</i>	<i>24.48</i>	37.49	<i>19.42</i>
Total	372.1	Av. 114.63	Av. 35.58	Av. 82.40	Av. 31.42	Av. 41.64	Av. 24.49

Densities of predicted, non TE and TE genes and of FL-cDNA are also shown. The three highest and lowest values per data set appear in bold and in italic, respectively. TE = transposable element; FL cDNA = full length cDNA

**Fig. 6** Venn's diagram of rice genes tagged by OTL Tos17, NIAS Tos17 and OTL T-DNA inserts. 210 genes contain at least one insert in each of the three insert populations

the defence response genes indeed represented 10.1% and 9.7% of the genes interrupted by Tos17 in the OTL and NIAS libraries respectively.

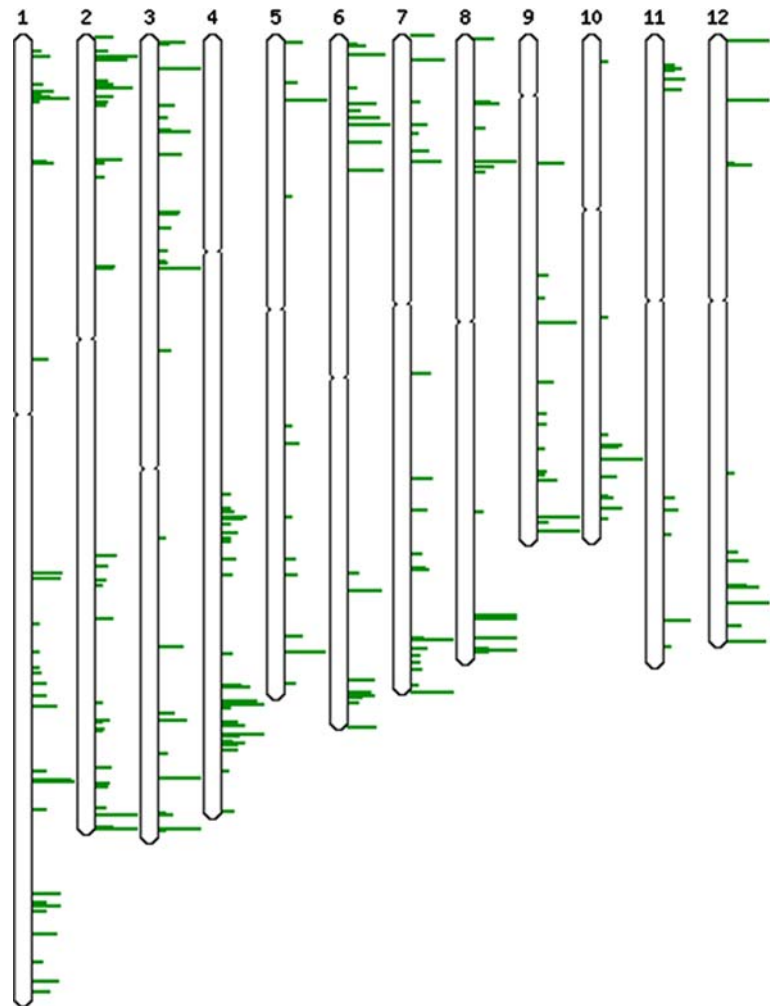
#### Comparison of hot spots for integration in the OTL and NIAS libraries

Over the 2,773 genes interrupted by Tos17 inserts in the OTL T-DNA library, 444 genes were categorized as “hot spots” as they exhibited a range of 5 to 121 integrations of

newly-transposed Tos17 elements in the analysed regenerated lines (Fig. 6). Interestingly, a majority of 274 (61.7%) of these 444 genes were also found to exhibit more than 5 inserts (precisely from 5 to 106) in the NIAS Tos17 library (Supplementary Table 2), while only 23 “hot spot genes” were not represented at all in the NIAS Tos17 FST database. On the other hand, over the 2,329 genes exhibiting a range of 1–4 Tos17 inserts in the OTL T-DNA collection, only 189 (8.1%) exhibited more than 5 integrations in the NIAS Tos17 library, whereas 1095 (47%) were not found in the NIAS FST collection. These findings indicate that the “hot spot genes” are generally found in both libraries and represent integration sites highly accessible to the Tos17 integration even in different culture and regeneration conditions. Native Tos17 copies residing on chromosomes 7 and 10 proved to be themselves hot spot for Tos17 integration since, applying the same filter than for non-TE genes, we observed 10 and 21 insertions respectively in the OTL library (11 and 25 respectively in the NIAS library).

High frequency insertions of Tos17, represented in both libraries, appear to occur in genes exhibiting a lower and narrower GC range than that of a random panel of non hot spot genes ( $42.8 \pm 4.4$  vs.  $45.2 \pm 6.6$ ) (Fig. 5B). Positioning the 274 hot spot genes, represented in both OTL and NIAS library, on the rice genome, tends to accentuate the trends observed in the overall distribution of insertion sites notably the preference for subtelomeric regions of the chromosomes and low accessibility of short arms of chromosomes 4, 9 and 10 found to be heterochromatic regions (Fig. 7). Comparison of the density of hot spots

**Fig. 7** Distribution of Tos17 hot spot genes, found in both the OTL and NIAS libraries of insertion lines, on the rice chromosomes (i.e. exhibiting at least 5 insertions in each of the library). The number of cumulated insertions (bars) ranges from 10 to 227 but has been voluntarily framed to a 60 insertion threshold for better resolution



genes compared to the density of the overall Tos17 insertion sites (Supplementary Table 3) allows to distinguish three classes of chromosomes where i. Both hot spot and insertion site densities are consistently either higher (Chromosomes 1, 2 and 3 and to a lesser extent Chromosome 6) or lower (Chromosomes 10, 11 and 12) than the genome average, following in that the density of FL-cDNA (see Table 2) ii. The density of hot spot genes is lower than that of the genome average whereas the insertion site density matches that of the genome average (Chromosome 5) iii. The density of hot spot genes is higher than that of the genome average whereas the insertion site density is lower than that of the genome average (Chromosome 4 -with all the hot spots located on the long arm- and Chromosome 7).

We successfully categorized through Gene Ontology 201 genes among the 274 which were prone to hot spots for Tos17 integration in both the OTL and NIAS libraries. Though the overall distribution of hot spot genes did not differ significantly from that of the 18,584 non-TE rice genes classified according to the same procedure

(Supplementary Table 1), a detailed analysis of these genes showed a strong representation of specific gene families such as genes with NB-ARC domain (494 genes in rice, mainly disease response related such as the NBS LRR) or encoding LRR family protein. A closer examination of the distribution of hot spot genes in relation with their function shows that genes sharing similar function and organized in clusters can be preferred targets of high frequency insertion of Tos17. Genes belonging to receptor kinase family, LRR family, NB-ARC, D-mannose binding lectin family, Protein phosphatase, aminotransferase Terpene synthase family, Ubiquitin carboxyl-terminal hydrolase, B3 DNA binding domain, TPR domain containing protein notably fall in this category (Supplementary Table 4).

To determine whether genes expressed at a high level in callus tissue are more prone to high frequency integration of Tos17 than other genes, we made a bulk query with the 274 Tos17 hot spot genes, a random panel of 274 Tos17 non hot spot genes and of random panel of 750 non TE rice genes against the NCA (callus) MPSS library. Percentages of genes exhibiting at least one signature in the NCA

**Table 4** Frequencies of insertion of Tos17 inserts from the OTL insertion line collection in the genic and intergenic regions of the genome of *Oryza sativa* variety Nipponbare

Genome compartment	Rice genome		OTL Tos17 inserts		NIAS Tos17 inserts		OTL T-DNA inserts	
	Size (bp)	%	Nb	%	Nb	%	Nb	%
Genic (non-TE)	152797284	41.06	9060	77.51	11899	76.80	4472	49.07
Promoter -1000	43324266	11.64	649	7.16	1029	8.65	1468	32.83
Exon	46186480	12.41	4303	47.49	5147	43.26	871	19.48
Intron	54621261	14.68	3744	41.32	5200	43.70	1711	38.26
Stop codon + 300	8665277	2.33	364	4.02	523	4.40	422	9.44
Genic (TE)	61584245	16.55	829	7.09	1090	7.04	408	4.48
Intergenic	157774420	42.39	1800	15.40	2504	16.16	4233	46.45
Total	372155949	100	11689	100.00	15493	100.00	9113	100.00

These frequencies are compared to those of Tos17 and T-DNA inserts in the NIAS and OTL libraries, respectively. Nb, number of inserts

library for the 3 sets of genes were of 11.6, 33.4 and 20.1 in the 3 respective sets (Supplementary Table 5). Whereas Tos17 target genes are more frequently expressed in callus than random genes, their levels of expression in callus based on the occurrence of signatures does not seem higher than those of random genes. Tos17 hot spot genes did not exhibit either higher frequency or higher level of expression in callus than random Tos17 target genes.

## Discussion

In a first attempt to complement the characterization of Tos17 inserts existing in a T-DNA insertion library of rice, we developed a protocol aiming at efficiently and selectively recovering and sequencing the regions flanking the newly-transposed Tos17 elements, whose transposition was specifically induced by the tissue culture steps of the T-DNA transformation procedure. Our strategy took advantage of the absence of *EcoRV* restriction sites in the regions flanking the 3' end of the two Tos17 copies resident in the genome of the rice variety Nipponbare. The selective amplification protocol was based on the use of *EcoRV* digests of genomic DNA followed by ligation of adapters and two rounds of PCR using nested Tos17 specific primers. The scaling up of this protocol proved to be highly efficient and it allowed the processing of 25,286 DNA samples in a short time span with limited manpower and led to the anchoring of 14,513 Tos17 inserts to the rice pseudomolecules. However, though satisfactory, this number yet represents only 17.3% of the total of 83,686 new inserts estimated in the library according to Southern blot data. To complement the coverage of the Tos17 inserts, further effort should therefore be dedicated to widen the range of enzymes used for digesting DNA in the walk PCR protocol—as shown in this study, but with the likely additional burden of picking the PCR fragments and

running a third PCR- or the use of a different protocol such as TAIL-PCR (Miyao et al. 2003).

The average number of 3.37 new Tos17 inserts found in the OTL T-DNA collection appears lower and therefore more tractable for insert purification than the average of 10 inserts found in plants regenerated from cell aggregate suspension in the NIAS library. This number is equivalent to the average 4 inserts per line estimated in T-DNA lines of another japonica cultivar, Dong Jin, which has been used to generate the POSTECH T-DNA insertional mutant library (G.An, unpublished). On the other hand, the japonica cultivar Tainung 67, used to generate the TRIM T-DNA insertion library (Hsing et al. 2007) proved to contain poorly active Tos17 copy (ies) with an average of 0.2 new insert per T-DNA plant.

Some (19.5 %) of the 14,513 Tos17 FSTs were found to be redundant with at least one other sequence in the population. We showed that this redundancy can be mainly explained by a shared cell lineage of some lines from which the FSTs were recovered. In our transformation protocol, an average of 4, mostly independent T-DNA lines are indeed generally derived from a single, co-cultivated ENU in cv. Nipponbare (Sallaud et al. 2004; Sallaud et al. 2003). Though the characterization of Tos17 FSTs is far from exhaustive in the library, this indicates that a fraction of Tos17 insertions occurred at an early step of callus induction from the seed embryo scutellum, e.g. during divisions of the somatic proembryo which give rise to the ENU (Bec et al. 1998), while other insertions may have occurred later e.g. during the growth of hygromycin-resistant cell lines. Whether these successive insertions cause functional alterations, this finding is reminiscent of the phenomenon of sequential accumulation of somaclonal mutations reported during callus growth in cv. Nipponbare (Fukui 1983). The fact that the common cell lineage in T-DNA plants can be tracked back in our library may also readily provide valuable information on whether the

phenotypic alterations observed during forward screenings are likely due to the T-DNA or have more likely to be related to a somaclonal variation source, including Tos17 (J.B. Morel, Pers. Com.). Extent of redundancy of FSTs due to a shared cell lineage has been estimated to be of 35% in the NIAS Tos17 insertion line library (Miyao et al. 2003).

Distributions of Tos17 inserts among and along the chromosomes in the OTL T-DNA collection were found to remarkably parallel those observed in the NIAS library. Hot spots for integration were notably found to lie in the same sequence intervals and often in the same genes. This shows that though two independent and different procedures were used to generate the libraries in cv. Nipponbare, the overall behaviour and target site specificity of the retroelement deduced from their respective FST populations was highly conserved. It would be nevertheless interesting to investigate whether these features are conserved in another cultivar harbouring an active copy, which might be different from that of chromosome 7, the active copy in cv. Nipponbare. An early concern was also that a bias might have been introduced by the flanking region recovery method itself, which notably involved the use of a relatively rare cutter unique restriction enzyme, *EcoRV*, in our study. The consistency of the results of distribution, target site specificity and preferred genes of Tos17 in the OTL library compared to those of Miyao and collaborators, who used a different protocol of recovery of flanking regions based on TAIL-PCR, proves that the findings can be considered robust and not biased.

Overall, we showed that the genes tagged by Tos17 inserts in the OTL population contain an average of 3.27 inserts per gene, consistent with what is observed in the NIAS library, whereas genes tagged by T-DNA contain an average of 1.14 insert. This means that an effort of insert characterization aiming at covering the whole gene complement has to be higher using Tos17 than the T-DNA. Genome saturation may prove not tractable with Tos17 alone not associated with another mutagen. On the other hand, Tos17 is more frequently inserted into coding sequences, thereby often creating putative KOs, than the T-DNA, which is more frequently recovered in promoter and 3'UTR regions. A consequence of the tendency of Tos17 to insert in preferred genes should also result in a higher occurrence of allelic series composed of Tos17 inserts rather than of both Tos17 and T-DNA inserts in the OTL library. On the other hand, the complementary gene regions preferred by T-DNA and Tos17 elements may result in a wider variety of changes in expression in a given gene tagged by both insertional mutagens (e.g; aside from KOs, creation of dominant negative mutants or mutants with changes in gene expression patterns).

We found many disease response and signal perception and transduction related genes which are known to be often organized in clusters along the chromosomes, are among the Tos17 preferred target genes, and that this organization in tandem series creates hot spot regions along the chromosomes. A striking example are the NB-ARC domain containing genes, with 14 representatives among the 274 very hot spots of the genome (5.1%) though they represent only 1.16 % of the non TE genes (494:42,653). Results presented here confirmed that insertion sites highly accessible to Tos17 exhibit a narrow GC content range. Taken together these results are consistent with the findings of Miyao and collaborators (Miyao et al. 2003), who attributed this preference to the overall clustering of these genes and the overlap of their GC content with the narrow GC content distribution of Tos17 target sites. Another hypothesis for explaining high frequency integration in certain categories of genes is that these are preferentially expressed in dedifferentiated cells during tissue culture and that Tos17 benefits from these windows of open chromatin structure to insert specifically in these genes. We showed that indeed Tos17 target genes exhibit a 2 to 3 fold higher expression frequency in callus tissues than random rice genes. However, Tos17 preferred genes do not appear to have a significantly higher level of expression in callus tissues than random genes. On the other hand, frequency and level of expression of genes hot spots for integration in callus did not differ from those of genes prone to single Tos17 insertions suggesting that high frequency insertion might be more related to specific GC content or DNA structure than to particular expression in callus tissues.

The large allelic series existing in genes which are Tos17 preferred targets (see Supplementary Table 2) also offers, when the phenotypic information of the corresponding lines is available, a straightforward means for establishing *in silico* links between gene disruption and phenotypes and provide a clue on gene function. Phenotypes of these lines are accessible at <http://urgi.versailles.inra.fr/OryzaTagLine>.

In conclusion, this first characterization of Tos17 inserts in the T-DNA library of rice using the TOSTRAP protocol proved very cost effective since it rapidly increased the range of novel genes tagged, not only because of the still low coverage of the genome by the inserts in the population but also because of an insertion preference in different types of genes than those tagged by the T-DNA. This indicates that advantage in characterizing Tos17 inserts in a T-DNA collection lies more in a highly complementary tagging of additional genes than in finding alleles in a same genetic background in genes already tagged by the T-DNA. This greatly enhances the library genome coverage and its overall value for implementing forward and reverse genetics strategies targeting specific traits.

**Acknowledgments** We wish to thank Jean benoît Morel for his help in MPSS analysis. This project was supported by the Génoplante programme (M1 project) and benefited from a granted project at the French National Sequencing Centre (Génoscope). This work was carried out under the infrastructures of the Génopole Montpellier Languedoc Roussillon.

## References

- An SY, Park S, Jeong DH, Lee DY, Kang HG, Yu JH, Hur J, Kim SR, Kim YH, Lee M, Han SK, Kim SJ, Yang JW, Kim E, Wi SJ, Chung HS, Hong JP, Choe V, Lee HK, Choi JH, Nam JM, Kim SR, Park PB, Park KY, Kim WT, Choe S, Lee CB, An GH (2003) Generation and analysis of end sequence database for T-DNA tagging lines in rice. *Plant Physiol* 133:2040–2047
- Balzergue S, Dubreucq B, Chauvin S, Le-Clainche I, Le Boulaire F, de Rose R, Samson F, Biauudet V, Lecharny A, Cruaud C, Weissenbach J, Caboche M, Lepiniec L (2001) Improved PCR-walking for large-scale isolation of plant T-DNA borders. *Biotechniques* 30:496–498
- Bec S, Chen L, Michaux Ferriere N, Legavre T, Fauquet C, Guiderdoni E (1998) Comparative histology of microprojectile-mediated gene transfer to embryogenic calli in japonica rice (*Oryza sativa* L.): influence of the structural organization of target tissues on genotype transformation ability. *Plant Sci* 138:177–190
- Devic M, Albert S, Delseny M, Roscoe TJ (1997) Efficient PCR walking on plant genomic DNA. *Plant Physiol Biochem* 35:331–339
- Droc G, Ruiz M, Larmande P, Pereira A, Piffanelli P, Morel JB, Dievart A, Courtois B, Guiderdoni E, Perin C (2006) OryGenesDB: a database for rice reverse genetics. *Nucl Acids Res* 34:D736–D740
- Fukui K (1983) Sequential occurrence of mutations in a growing rice callus. *Theor Appl Genet* 65:225–230
- Guiderdoni E, An G, Yu SM, Hsing YI, Wu C (2007) T-DNA insertion mutants as a resource for rice functional genomics. In: Upadhyaya NM (ed) *Rice functional genomics- challenges, progress and prospects*, Springer, NY, 187–227
- Hirochika H (2001) Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr Opin Plant Biol* 4:118–122
- Hirochika H, Guiderdoni E, An G, Hsing YI, Eun MY, Han CD, Upadhyaya N, Ramachandran S, Zhang QF, Pereira A, Sundarasan V, Leung H (2004) Rice mutant resources for gene discovery. *Plant Mol Biol* 54:325–334
- Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *PNAS* 93:7783–7788
- Hsing YI, Chern CG, Fan MJ, Lu PC, Chen KT, Lo SF, Sun PK, Ho SL, Lee KW, Wang YC, Huang WL, Ko SS, Chen S, Chen JL, Chung CI, Lin YC, Hour AL, Wang YW, Chang YC, Tsai MW, Lin YS, Chen YC, Yen HM, Li CP, Wey CK, Tseng CS, Lai MH, Huang SC, Chen LJ, Yu SM (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol Biol* 63:351–364
- Jeong DH, An S, Park 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* 45:123–132
- Jung KH, Han MJ, Lee YS, Kim YW, Hwang IW, Kim MJ, Kim YK, Nahm BH, An GH (2005) Rice Undeveloped Tapetum1 is a major regulator of early tapetum development. *Plant Cell* 17:2705–2722
- Matsumoto T, Wu JZ, Kanamori H, Katayose Y, Fujisawa M, Namiki N, Mizuno H, Yamamoto K, Antonio BA, Baba T, Sakata K, Nagamura Y, Aoki H, Arikawa K, Arita K, Bito T, Chiden Y, Fujisuka N, Fukunaka R, Hamada M, Harada C, Hayashi A, Hijishita S, Honda M, Hosokawa S, Ichikawa Y, Idonuma A, Iijima M, Ikeda M, Ikeno M, Ito K, Ito S, Ito T, Ito Y, Ito Y, Iwabuchi A, Kamiya K, Karasawa W, Kurita K, Katagiri S, Kikuta A, Kobayashi H, Kobayashi N, Machita K, Maehara T, Masukawa M, Mizubayashi T, Mukai Y, Nagasaki H, Nagata Y, Naito S, Nakashima M, Nakama Y, Nakamichi Y, Nakamura M, Meguro A, Negishi M, Ohta I, Ohta T, Okamoto M, Ono N, Saji S, Sakaguchi M, Sakai K, Shibata M, Shimokawa T, Song JY, Takazaki Y, Terasawa K, Tsugane M, Tsuji K, Ueda S, Waki K, Yamagata H, Yamamoto M, Yamamoto S, Yamane H, Yoshiki S, Yoshihara R, Yukawa K, Zhong HS, Yano M, Sasaki T, Yuan QP, Shu OT, Liu J, Jones KM, Gansberger K, Moffat K, Hill J, Bera J, Fadrosch D, Jin SH, Johri S, Kim M, Overton L, Reardon M, Tsitrin T, Vuong H, Weaver B, Cieccko A, Tallon L, Jackson J, Pai G, Van Aken S, Utterback T, Reidmuller S, Feldblyum T, Hsiao J, Zismann V, Iobst S, de Vazeille AR, Buell CR, Ying K, Li Y, Lu TT, Huang YC, Zhao Q, Feng Q, Zhang L, Zhu JJ, Weng QJ, Mu J, Lu YQ, Fan DL, Liu YL, Guan JP, Zhang YJ, Yu SL, Liu XH, Zhang Y, Hong GF, Han B, Choise N, Demange N, Orjeda G, Samain S, Cattolico L, Pelletier E, Couloux A, Segurens B, Wincker P, D'Hont A, Scarpelli C, Weissenbach J, Salanoubat M, Quetier F, Yu Y, Kim HR, Rambo T, Currie J, Collura K, Luo MZ, Yang TJ, Ammiraju JSS, Engler F, Soderlund C, Wing RA, Palmer LE, de la Bastide M, Spiegel L, Nascimento L, Zutavern T, O'Shaughnessy A, Dike S, Dedhia N, Preston R, Balija V, McCombie WR, Chow TY, Chen HH, Chung MC, Chen CS, Shaw JF, Wu HP, Hsiao KJ, Chao YT, Chu MK, Cheng CH, Hour AL, Lee PF, Lin SJ, Lin YC, Liou JY, Liu SM, Hsing YI, Raghuvanshi S, Mohanty A, Bharti AK, Gaur A, Gupta V, Kumar D, Ravi V, Vij S, Kapur A, Khurana P, Khurana P, Khurana JP, Tyagi AK, Gaikwad K, Singh A, Dalal V, Srivastava S, Dixit A, Pal AK, Ghazi IA, Yadav M, Pandit A, Bhargava A, Sureshbabu K, Batra K, Sharma TR, Mohapatra T, Singh NK, Messing J, Nelson AB, Fuks G, Kavchok S, Keizer G, Llaça ELV, Song RT, Tanyolac B, Young S, Il KH, Hahn JH, Sangsakoo G, Vanavichit A, de Mattos LAT, Zimmer PD, Malone G, Dellagostin O, de Oliveira AC, Bevan M, Bancroft I, Minx P, Cordum H, Wilson R, Cheng ZK, Jin WW, Jiang JM, Leong SA, Iwama H, Gojobori T, Itoh T, Niimura Y, Fujii Y, Habara T, Sakai H, Sato Y, Wilson G, Kumar K, McCouch S, Juretic N, Hoen D, Wright S, Bruskiewicz R, Bureau T, Miyao A, Hirochika H, Nishikawa T, Kadowaki K, Sugiura M, Project IRGS, (2005) The map-based sequence of the rice genome. *Nature* 436:793–800
- Miyao A, Iwasaki Y, Kitano H, Itoh J, Maekawa M, Murata K, Yatou O, Nagato Y, Hirochika H (2007) A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol Biol* 63:625–635
- Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H (2003) Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* 15:1771–1780
- Moon S, Jung KH, Lee DE, Lee DY, Lee J, An K, Kang HG, An G (2006) The rice FON1 gene controls vegetative and reproductive development by regulating shoot apical meristem size. *Mol Cells* 21:147–52
- Sallaud C, Gay C, Larmande P, Bes M, Piffanelli P, Piegu B, Droc G, Regad F, Bourgeois E, Meynard D, Perin C, Sabau X, Ghesquiere A, Glaszmann JC, Delseny M, Guiderdoni E

- (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. *Plant J* 39:450–464
- Sallaud C, Meynard D, van Boxtel J, Gay C, Bes M, Brizard JP, Larmande P, Ortega D, Raynal M, Portefaix M, Ouwerkerk PB, Rueb S, Delseny M, Guiderdoni E (2003) Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. *Theor Appl Genet* 106:1396–408
- Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucl Acids Res* 23:1087–1088
- Zhang J, Li C, Wu C, Xiong L, Chen G, Zhang Q, Wang S (2006) RMD: a rice mutant database for functional analysis of the rice genome. *Nucleic Acids Res* 34:D745–D748
- Zhu QH, Eun MY, Han CD, Kumar CS, Pereira A, Ramachandran S, Sundaresan V, Eamens AL, Upadhyaya NM, Wu R (2007) Transposon insertional mutants: a resource for rice functional genomics. In: Upadhyaya NM (ed) *Rice functional genomics—challenges, progress and prospects*. Springer, NY, pp 229–280