the plant journal
SEB Society for Experimental Biology

The Plant Journal (2011) 66, 241–246 doi: 10.1111/j.1365-313X.2011.04492.x

Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data

François Sabot[1,†], Nathalie Picault[1,†], Moaine El-Baidouri[1], Christel Llauro[1], Cristian Chaparro[1], Benoit Piegu[1], Anne Roulin[1], Emmanuel Guiderdoni[2], Mélissa Delabastide[3], Richard McCombie[3] and Olivier Panaud[1,*]

[1]*Laboratoire Génome et Développement des Plantes, UMR CNRS/IRD/UPVD, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, F-66860 Perpignan Cedex, France,*

[2]*CIRAD, UMR DAP, Biological Systems Department, 2477 Avenue Agropolis, F-34398 Montpellier Cedex 5, France, and*

[3]*Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA*

Received 10 August 2010; revised 13 December 2010; accepted 4 January 2011; published online 1 March 2011.
[*]For correspondence (fax +33 468668499; e-mail panaud@univ-perp.fr).
[†]These authors contributed equally to this work.

SUMMARY

Transposable elements (TEs) are mobile entities that densely populate most eukaryotic genomes and contribute to both their structural and functional dynamics. However, most TE-related sequences in both plant and animal genomes correspond to inactive, degenerated elements, due to the combined effect of silencing pathways and elimination through deletions. One of the major difficulties in fully characterizing the molecular basis of genetic diversity of a given species lies in establishing its genome-wide transpositional activity. Here, we provide an extensive survey of the transpositional landscape of a plant genome using a deep sequencing strategy. This was achieved through paired-end mapping of a fourfold coverage of the genome of rice mutant line derived from an *in vitro* callus culture using Illumina technology. Our study shows that at least 13 TE families are active in this genotype, causing 34 new insertions. This next-generation sequencing-based strategy provides new opportunities to quantify the impact of TEs on the genome dynamics of the species.

Keywords: rice, transposable elements, next-generation sequencing, genomics, mutant line, paired-end mapping.

INTRODUCTION

Transposable elements (TEs) are mobile entities that are the major constituents of most eukaryotic genomes. Several reports have shown that TEs have a significant impact on the structure (SanMiguel *et al.*, 1996; Morgante, 2006; Feschotte and Pritham, 2007; Feschotte, 2008; Cordaux and Batzer, 2009) and function (Kobayashi *et al.*, 2004; Xiao *et al.*, 2008) of both plant and animal genomes. TE-associated structural variations have been widely documented by recent genome-level studies (Korbel *et al.*, 2007; Huang *et al.*, 2008). These variations have been detected based on comparative genomic studies among individuals in natural populations (Korbel *et al.*, 2007) or, in the case of crops, between varieties from the same species (Huang *et al.*, 2008). Most TE families are under the strict control of silencing pathways at both the transcriptional and post-transcriptional levels (Slotkin and Martienssen, 2007), and subsequently become non-functional due to accumulation of mutations (Vitte *et al.*, 2007). Moreover, TE insertion polymorphisms can be retained within gene pools for long periods, either through lineage sorting (particularly in the case of domesticated species) or because they are neutral (in the case of large populations). Therefore, the polymorphisms observed today only reflect the past transpositional activity of TE families that may since have become extinct. However, identification of active families in a given genome is necessary in order to fully understand the molecular basis of biodiversity. Moreover, in the case of crops, active TEs not only represent a source of such variation, but can also be exploited to develop genetic markers for genetic mapping and in marker-assisted breeding (Flavell *et al.*, 1998).

Asian rice (*Oryza sativa*) is grown worldwide and provides the basic food supply for billions of people in Asia, Africa and America. Completion of its genomic sequence (International Rice Genome Sequencing Project, 2005), which consists of 12 chromosomes with a total size of 430 Mbp, has enabled it to become a model species in plant genomics. Despite its relatively small size compared to other cereal crops such as sorghum (*Sorghum bicolor*; 900 Mbp), maize

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd

(*Zea mays*; 2500 Mbp) or wheat (*Triticum aestivum*; 18 000 Mbp), the rice genome harbours a significant proportion of TE-related sequences (at least 33%) that belong to most of the classes described in eukaryotes (International Rice Genome Sequencing Project, 2005; Wicker *et al.*, 2007). The most predominant are the long terminal repeat (LTR) retrotransposons, with nearly 400 families ranging from one to several hundred copies, which constitute over 90 Mbp of the genome. Other types of TEs (long interspersed repetitive elements (LINEs), short interspersed repetitive elements (SINEs), transposons, miniatures inverted repeat transposable element (MITEs) and helitrons) represent an additional 40 Mbp. Altogether, several hundred TE families have been characterized in the rice genome, some of which are curated in dedicated databases (Jurka *et al.*, 2005; Chaparro *et al.*, 2007). At present, only eight of these TE families are known to exhibit transpositional activity *in planta*. These are the LTR retrotransposons *Tos17* (Hirochika *et al.*, 1996), *Lullaby* (Picault *et al.*, 2009), *osr7* and *osr23* (Wang *et al.*, 2009), the LINE *Karma* (Komatsu *et al.*, 2003), the MITEs *dTok* (Moon *et al.*, 2006), *nDart* (Tsugane *et al.*, 2006) and *mPing* (Jiang *et al.*, 2003), and the transposon *Pong* (Jiang *et al.*, 2003). Altogether, these eight TE families represent <100 kbp, i.e. 0.1% of the TE-related sequences in the rice genome.

A number of rice mutant collections have been developed for functional genomics studies (Rice Annotation Project, 2008). These have been produced either through *in vitro* callus culture or transformation using a T-DNA vector (which also requires *in vitro* cultivation for regeneration of the plant). Four of the eight families mentioned above are activated during *in vitro* culture of rice calli (*Tos17*, *Lullaby*, *mPing* and *Karma*). Consequently, it is expected that rice mutant lines harbour several TE-associated mutations that remain to be characterized in order to fully exploit these resources. Here we provide an extensive survey of the transpositional landscape of a plant genome using next-generation sequencing (NGS) technology. This was achieved through paired-end mapping (PEM) of a fourfold genome coverage of a rice mutant derived from *in vitro* callus culture using Illumina technology. Our study reveals that at least 13 TE families are active in a single mutant plant, and thus constitutes the first step towards understanding of the impact of these mobile elements on the genome dynamics of the species.

RESULTS AND DISCUSSION

A genome-wide survey of TE activity was performed using NGS technology. As a proof of concept, we selected a line showing a high number of *Tos17* insertions in order to use this element as a positive control of TE activity in the analysis. The line AB156365 was thus selected from those available in the mutant collections because it harbours at least 11 such insertions, as indicated by a Southern hybridization experiment using *Tos17* as a probe (Figure S1). A total of 44 061 780 36-mer paired-end reads, representing a 4.17-fold genome coverage (base count), were generated using Illumina technology. These data are freely available upon request. PEM (Korbel *et al.*, 2007) was performed in order to identify TE-associated homology breakpoints compared to the reference genome sequence (International Rice Genome Sequencing Project rice pseudomolecules version 4; http://rapdb.dna.affrc.go.jp/): amplicons for which one end mapped to an unambiguously unique, non-TE-related sequence and the other corresponded to a known rice TE located at a distant site in the reference pseudomolecules were selected (see Figure 1 and Experimental procedures). In this process, many of the homology breakpoints that were identified in the first screening of the paired-end reads (i.e. abnormally mapped pairs) were not fully characterized, in particular those caused by insertions in repeated sequences. In addition, R_3 plants are expected to harbour a significant proportion of the mutations that originated from the callus culture at the heterozygous state. For such loci, half of the amplicons (corresponding to the wild-type) would exhibit no abnormal pairing, while the other half (corresponding to the mutated allele) would exhibit abnormal paired-end mapping. This leads to a discrepancy in the mapping phase of the analysis pipeline and subsequently to their elimination from TE insertion candidate loci. For these reasons, the present study provides a partial, non-exhaustive view of the overall transpositional activity of the rice genome. Nevertheless, our method allowed identification of 34 TE insertions caused by the activity of 13 TE families in the mutant line (Tables 1 and S1). A set of 27 randomly chosen breakpoints, representing the 13 active families, were confirmed using PCR amplification, cloning and sequencing of the region for a set of ten selfing progenies of the original mutant line (Figure 2 and Appendix S1). Our

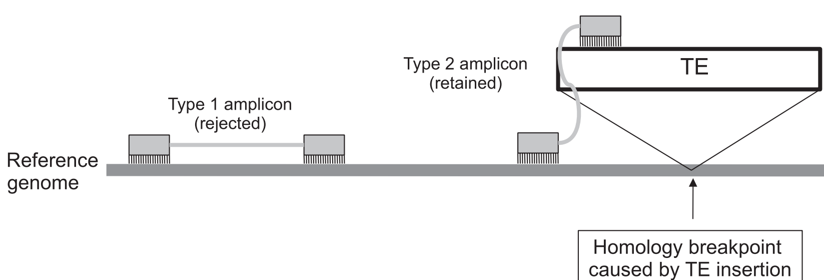

Figure 1. Schematic view of the procedure followed for PEM.

© 2011 The Authors

Table 1 TE insertions identified in the mutant line

TE name	TE type	Number of insertions	Number of copies	Expression
BAJIE	LTR retrotransposon	1	663	0
osr10	LTR retrotransposon	1	248	0
osr37	LTR retrotransposon	3	128	0
RIRE2	LTR retrotransposon	2	451	–
RIRE3	LTR retrotransposon	2	587	–
RN363	LTR retrotransposon	1	1	0
RN216	LTR retrotransposon	1	135	0
Tos17	LTR retrotransposon	11	2	+
Mite#1	MITE	3	1696	
Mite#2	MITE	1	2554	
Mite#3	MITE	2	59	
mPing	MITE	3	8	
Tami2	MITE	3	42	

The expression column refers to previously published data (Picault *et al.*, 2009). 0, no difference in expression between callus and imbibed embryos; –, down-regulation in callus; +, over-expression in callus. Expression data are not available for the MITEs.

strategy allowed us to confirm and map *in silico* the 11 insertions of the LTR retrotransposon *Tos17* that were first identified based on the Southern hybridization experiment (Figure S1). The 23 additional insertions were caused by the activity of 12 other distinct TE families (seven LTR retrotransposons and five MITEs). Among these, only the MITE *mPing* was previously shown to be active in the rice genome (Jiang *et al.*, 2003), while the remaining 11 families were only identified through annotation of the rice pseudomolecules (Rice Annotation Project, 2008). Therefore, the PEM strategy implemented here allowed identification of active TEs in the rice genome at an unprecedented scale, increasing the number of known active families from eight to 19 in a single experiment. As indicated by the number of insertions that we identified, the most active type of TEs in rice callus appears to be the LTR retrotransposons, with eight distinct families causing 22 mutations. The second most active type of TEs in our rice mutant line are MITEs (Wicker *et al.*, 2007), which are short degenerated forms of transposons. Our study did not show transposition of two of the four TE families previously shown to be active in callus, i.e. the LTR retrotransposon *Lullaby* and the LINE *Karma*. As mentioned above, our screening for TE insertions is biassed towards gene-rich regions, and we cannot conclude that these two families are not active in the re-sequenced mutant. In addition, transposition is a stochastic process, and analysis of a single genotype is not sufficient to draw conclusions on the activity of all rice TE families. Additional studies of this type may further extend the list of active TEs in the species.

Tos17 has been used to develop large mutant collections derived from *in vitro* culture of rice calli (Miyao *et al.*, 2007). Over 100 000 mutant lines are now available to the scientific community for both forward and reverse genetics studies in rice. Until now, identification of the insertion sites of *Tos17* was achieved through direct sequencing of the

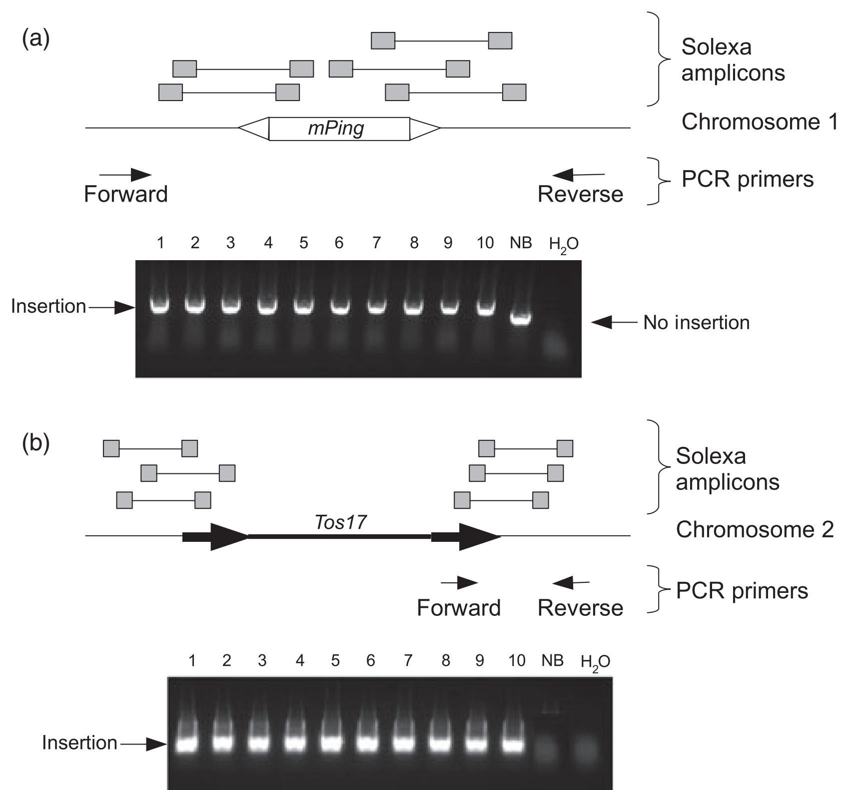

Figure 2. Validation of the insertions of *mPing* on chromosome 1 (a) and *Tos17* on chromosome 2 (b).
The Illumina amplicons mapped *in silico* are indicated above the genome sequence. The primers used for the PCR are shown below. The gels show PCR amplifications for 10 selfing progenies.

© 2011 The Authors

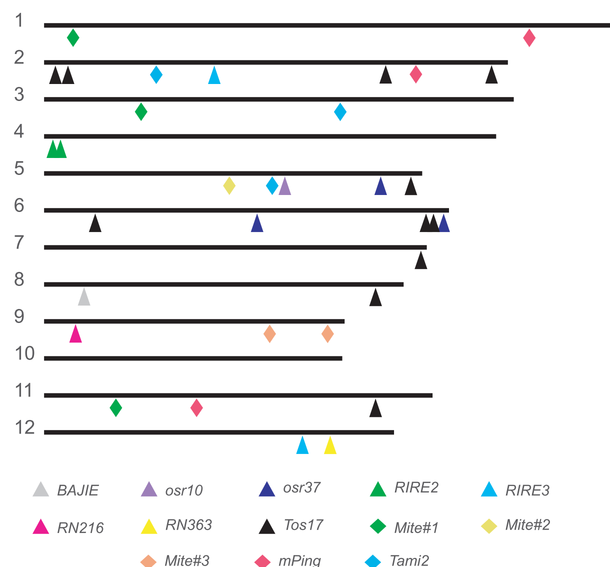

Figure 3. *In silico* mapping of the 34 TE insertions detected using our method on each chromosome (1–12, solid black lines).
Arrows and diamonds represent LTR retrotransposons and MITEs, respectively.

genomic region flanking the element (i.e. the flanking site tag; Miyao *et al.*, 2007). However, in the case of mutant lines harbouring multiple insertions of this element, flanking site tags cannot be obtained in a single sequencing experiment. The PEM strategy that we used allowed precise and unambiguous localization of all the *Tos17* insertions present in the line AB156365. Moreover, it allowed mapping of the 23 other mutations caused by the insertion of TEs that were not known to be active and for which flanking site tag mapping was therefore not applicable. These mutations are evenly distributed in the genome of the mutant line (Figure 3). Moreover, the *in silico* mapping of all the TE insertions was examined with regard to the position of the nearest rice gene. A complete list of the genes for which a TE insertion was identified within 1 kb in both the upstream and downstream directions in line AB156365 is given in Table 2. Fifteen of the 34 identified insertions occurred in the vicinity of genes, with the majority (ten) being caused by the activity of LTR retrotransposons. Eight of the 15 gene sequences were modified by insertion in either an exon, an intron or one UTR, which is twice the number caused by insertion of a *Tos17* element. Only 10% of the phenotypes observed in the rice mutant collections are tagged by flanking site tags of this element (Miyao *et al.*, 2007). As shown in the present study, PEM enables fast characterization of all genes mutated by TE insertions, and therefore increases the probability of gene discovery through forward genetics in mutants exhibiting altered phenotypes.

The LTR retrotransposon *Tos17* is present in two copies in the genome of the rice cultivar Nipponbare, on chromosomes 7 and 10. However, only the copy on chromosome 7 is active, the other being silenced by DNA methylation (Cheng *et al.*, 2006). No such information is available for the other seven LTR retrotransposon families that we identified in this study, as this report is the first evidence of their transpositional activity. Except for *RN363* (of which there is only one copy in the Nipponbare genome), these elements are present at a high copy number, which prevents identification of the active copy. As mentioned above, TEs are under the strict control of epigenetic silencing (Slotkin and Martienssen, 2007). Several reports have described the role of major pathways in plants, such as DNA methylation, histone methylation, RNA polymerase IV/V and siRNAs, by

Table 2 Genes for which a TE insertion was found within 1 kbp

Chromosome	Position	Gene ID	Gene name	Relative position	Location of insertion	TE type
2	1001426	Os02g0118900	NBS-LRR disease resistance protein, putative	0	Exon	*Tos17*
2	1657933	Os02g0131800	Similar to root-specific metal transporter	1000		*Tos17*
2	29511802	Os02g0696500	Glycosyl hydrolases family 16, putative, expressed	0	3′ UTR	*mPing*
2	35472538	Os02g0809401	Conserved hypothetical protein	−400		*Tos17*
4	909838	Os04g0115500	Zinc finger, C2H2-type containing domain protein	−1000		*RIRE2*
5	14832678	Os05g0319900	Similar to (1.4)-β-xylan endohydrolase, isoenzyme X-II (EC 3.2.1.8) (fragment)	0	Exon	*Mite#2*
5	18323796	Os05g0378800	Methyltransferase, putative, expressed	0	Exon–intron	*Tami2*
5	29177343	Os05g0584600	AAA family ATPase, putative, expressed	0	Exon–intron	*Tos17*
6	31913095	Os06g0728766	Glycosyl transferase, family 48 protein	−900		*osr37*
7	30050157	Os07g0691100	Similar to pectin methyltransferase 6 (fragment)	300		*Tos17*
8	3226371	Os08g0155700	Similar to RNA polymerase II largest subunit (fragment)	0	Exon	*BAJIE*
8	26383373	Os08g0528100	Conserved hypothetical protein	0	Exon–intron	*Tos17*
9	18074709	Os09g0460500	Gibberellin receptor GID1L2, putative, expressed	200		*Mite#3*
11	5770880	Os11g0211300	Similar to NBS-LRR disease resistance protein homologue (fragment)	−100		*Mite#1*
11	26064634	Os11g0621300	DUF1399-containing protein, putative, expressed	0	Exon	*Tos17*

The relative position is given in bp.

analysis of mutants impeding one or several of the enzymatic activities involved in these pathways (e.g. *met1*, *kyp* or *nrpd1a*; Mirouze *et al.*, 2009). The line AB156365 analyzed here was regenerated from *in vitro* culture of calli obtained from the wild-type cv. Nipponbare. The exact mechanisms of transpositional activation in cultured tissues are not fully understood, except for the LTR retrotransposon *Tos17*. Thus, whether demethylation is involved in activation of the other 12 TE families identified in this study remains to be elucidated. PEM of high-throughput sequence data could help to decipher the silencing pathways in rice on a genome-wide scale by analysis of the above-mentioned mutants.

In a previous study, we performed a genome-wide analysis of the transcriptional activity of rice LTR retrotransposons in calli using a microarray hybridization strategy (Picault *et al.*, 2009). All transpositionally active LTR retrotransposons are expected to be present in the form of transcripts in the tissues in which they are activated. The results of this transcriptomic survey are given in Table 1. Surprisingly, of the eight LTR retrotransposon families found to be active transpositionally, only *Tos17* showed over-expression in calli, while two (*RIRE2* and *RIRE3*) were down-regulated in these tissues. The remaining five did not show any significant difference in expression level between calli and the control. *RIRE2* and *RIRE3* contrast with *Tos17* in terms of copy number. There are only two copies of *Tos17*, but *RIRE2* and *RIRE3* are among the most highly repeated families in the rice genome. As suggested previously (Picault *et al.*, 2009), the expression level of LTR retrotransposons in rice may be controlled by a copy number-dependent pathway. However, the present results show a lack of correlation between expression level and transpositional activity, which suggests that retrotransposition occurs in calli regardless of transcriptional or post-transcriptional control, although this requires confirmation in further studies.

The development of high-throughput sequencing techniques has opened new perspectives in the field of genomics. They offer a fast and low-cost alternative to the classical Sanger-based technology. PEM has been used to detect structural variants in various organisms, such as human (Korbel *et al.*, 2007), *Bombyx* (Xia *et al.*, 2009) and Arabidopsis (Mirouze *et al.*, 2009). It was also used recently to perform high-density mapping in rice (Huang *et al.*, 2008). In that study, we used PEM to determine the transpositional activity in the rice genome, and showed that this approach allows fast identification of TE insertions in a mutant line, a task that until now required production of several generations of back-crossing, followed by positional cloning of the genes of interest. More generally, PEM could be used to detect TE activity in any species for which the complete genome sequence is available, and in any genetic background or physiological state. Use of this technology should clarify many aspects of TE biology in complex genomes, such as their epigenetic control and their genomic impact on stress responses (whether biotic or abiotic).

EXPERIMENTAL PROCEDURES

Sequencing

Rice mutant line AB156365 (from the Tos17 library at the National Institute of Agrobiological Sciences, Tsukuba, Japan) was selected for this study based on the number of *Tos17* insertions that its genome harbours (see Results and discussion). This line was obtained by A. Miyao and collaborators through regeneration of callus culture. Callus from strain NF7023 of the cultivar Nipponbare (from which the reference rice genome sequence was obtained) was obtained by induction in MS medium supplemented with 2 mg L^{-1} 2,4-dichlorophenoxyacetic acid (2,4-D) and 0.3% gellan gum, followed by transfer into liquid medium (N6 + 1 mg L^{-1} 2,4-D) after 21 days. After 19 weeks, calli were transferred into pre-regeneration medium (N6 + 0.3% casamino acid + 0.1% proline + 1 mg L^{-1} 2,4-D) for 12 days, followed by 10 days in regeneration medium (N6 + 0.1 mg L^{-1} benzyladenine (BA) + 0.01 mg L^{-1} naphtalene-acetic acid (NAA) + 1% agarose). Plants (R_0) were then transferred into the field. Total genomic DNA was extracted from a single R_2 progeny of the regenerated plants (corresponding to two generations of selfing). Sequencing was performed in two lanes of the Illumina GA (http://www.illumina.com/) for 36-cycle paired-end sequencing.

Data analyses

The 36 bp paired-end reads were mapped against the International Rice Genome Sequencing Project rice genome sequence (IRGSP version 4, http://rgp.dna.affrc.go.jp/E/IRGSP/Build4/build4.html), using bowtie software (Langmead *et al.*, 2009) with non-stringent parameters authorizing 3 mismatches (option -v 3) and an insertion size for pair ends of 300 bases (option -X 300) as amplicons had a theoretical size of 250 bases. We discarded all matching pairs and continued the analysis with the non-matching pairs using bowtie software to map them against a collection of transposable elements. The unmatched pairs from the previous step were mapped individually against the transposable elements database, and those not producing hits were BLAST searched against the genome. A list was constructed, indexed by chromosome position, for all non-mapped pairs, containing the position and best match of each member of the pair. Then the chromosome matching reads were grouped using a window of 600 bp, and those that presented more than 60% (from a minimum of three) of the corresponding pair reads pertaining to the same element or to the same chromosomal location were filtered. From this filtered set, a 1 kb region surrounding the candidate locations was subjected to a low-stringency BLAST search against the transposable elements database in order to eliminate candidates that showed resemblance to their corresponding pair hit. The remaining candidates were subjected to a manual curation step whereby dot plots were generated from the chromosomic regions, the inserted element and the paired ends. Those candidates that showed a clear pattern of paired-end mapping on the chromosomal region and the transposable element were retained for further analysis. This step allowed elimination of false positives that passed our previous filters.

Experimental validation

Validation of the insertions was achieved through PCR amplification followed by sequencing of the PCR products on an Applied Biosystems 3130xl genetic analyzer (http://www.appliedbiosystems.com/). Amplification was performed for 35 cycles (2 min at

© 2011 The Authors

94°C, 35 cycles of 45 sec at 94°C, 45 sec at 58°C, 1 min at 72°C) with an annealing temperature of 58°C. PCR products were separated by 1% agarose gel electrophoresis. After purification using a Geneclean® Turbo kit (MP Biomedicals, http://www.mpbio.com/) according the manufacturer's instructions, each product was sequenced using corresponding primers in an ABI377 sequencer (Applied Biosystems). The sequences of the primers used for PCR and sequencing are available upon request.

ACKNOWLEDGEMENTS

This work was supported by CNRS, IRD and the University of Perpignan, France. The authors thank Giuseppe Casacuberta (CSIC Barcelona, Spain) and Scott Jackson (Purdue University, USA) for their critical review of the manuscript. The authors thank A. Miyao and collaborators (NIAS, Tsukuba, Japan) for providing AB156365 seeds and information related to the tissue culture pedigree of the line.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Southern hybridization of mutant line AB156365 progeny using the *Tos17* probe.

Table S1. Detailed description of the TE insertions.

Appendix S1. Validation of the TE insertions.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Chaparro, C., Guyot, R., Zuccolo, A., Piégu, B. and Panaud, O. (2007) RetrOryza: a database of the rice LTR-retrotransposons. *Nucleic Acids Res.* **35**, D66–D70.

Cheng, C., Daigen, M. and Hirochika, H. (2006) Epigenetic regulation of the rice retrotransposon Tos17. *Mol. Genet. Genomics*, **276**, 378–390.

Cordaux, R. and Batzer, M.A. (2009) The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **10**, 691–703.

Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405.

Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368.

Flavell, A.J., Knox, M.R., Pearce, S.R. and Ellis, T.H. (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.* **16**, 643–650.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H. and Kanda, M. (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. USA*, **93**, 7783–7788.

Huang, X., Lu, G., Zhao, Q., Liu, X. and Han, B. (2008) Genome-wide analysis of transposon insertion polymorphisms reveals intraspecific variation in cultivated rice. *Plant Physiol.* **148**, 25–40.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R. and Wessler, S.R. (2003) An active DNA transposon family in rice. *Nature*, **421**, 163–167.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O. and Walichiewicz, J. (2005) Repbase update: a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467.

Kobayashi, S., Goto-Yamamoto, N. and Hirochika, H. (2004) Retrotransposon-induced mutations in grape skin color. *Science*, **304**, 982.

Komatsu, M., Shimamoto, K. and Kyozuka, J. (2003) Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. *Plant Cell*, **15**, 1934–1944.

Korbel, J.O., Urban, A.E., Affourtit, J.P. *et al.* (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science*, **318**, 420–426.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J. and Mathieu, O. (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature*, **461**, 427–430.

Miyao, A., Iwasaki, Y., Kitano, H., Itoh, J.I., Maekawa, M., Murata, K., Yatou, O., Nagato, Y. and 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.

Moon, S., Jung, K.H., Lee, D.E., Jiang, W.Z., Koh, H.J., Heu, M.H., Lee, D.S., Suh, H.S. and An, G. (2006) Identification of active transposon dTok, a member of the hAT family, in rice. *Plant Cell Physiol.* **47**, 1473–1483.

Morgante, M. (2006) Plant genome organisation and diversity: the year of the junk!. *Curr. Opin. Biotechnol.* **17**, 168–173.

Picault, N., Chaparro, C., Piegu, B. *et al.* (2009) Identification of an active LTR retrotransposon in rice. *Plant J.* **58**, 754–765.

Rice Annotation Project (2008) The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res.* **36**, D1028–D1033.

SanMiguel, P., Tikhonov, A., Jin, Y.K. *et al.* (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, **274**, 765–768.

Slotkin, R.K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285.

Tsugane, K., Maekawa, M., Takagi, K., Takahara, H., Qian, Q., Eun, C.H. and Iida, S. (2006) An active DNA transposon nDart causing leaf variegation and mutable dwarfism and its related elements in rice. *Plant J.* **45**, 46–57.

Vitte, C., Panaud, O. and Quesneville, H. (2007) LTR retrotransposons in rice (*Oryza sativa* L.): recent burst amplifications followed by rapid DNA loss. *BMC Genomics*, **8**, 218.

Wang, H., Chai, Y., Chu, X., Zhao, Y., Wu, Y., Zhao, J., Ngezahayo, F., Xu, C. and Liu, B. (2009) Molecular characterization of a rice mutator-phenotype derived from an incompatible cross-pollination reveals transgenerational mobilization of multiple transposable elements and extensive epigenetic instability. *BMC Plant Biol.* **9**, 63.

Wicker, T., Sabot, F., Hua-Van, A. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982.

Xia, Q., Guo, Y., Zhang, Z. *et al.* (2009) Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science*, **326**, 433–436.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J. and van der Knaap, E. (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science*, **319**, 1527–1530.

© 2011 The Authors