Plant Biotechnology Journal

Plant Biotechnology Journal (2012) **10**, pp. 555–568 doi: 10.1111/j.1467-7652.2012.00689.x

In-depth molecular and phenotypic characterization in a rice insertion line library facilitates gene identification through reverse and forward genetics approaches

Mathias Lorieux[1,2,†], Mélisande Blein[3,†], Jaime Lozano[4], Mathieu Bouniol[1], Gaétan Droc[5], Anne Diévart[5], Christophe Périn[5], Delphine Mieulet[5], Nadège Lanau[5], Martine Bès[5], Claire Rouvière[5], Céline Gay[6], Pietro Piffanelli[5], Pierre Larmande[2], Corinne Michel[3], Isabelle Barnola[6], Corinne Biderre-Petit[6], Christophe Sallaud[7], Pascual Perez[7], Fabienne Bourgis[2], Alain Ghesquière[2], Pascal Gantet[8], Joe Tohme[4], Jean Benoit Morel[3,*] and Emmanuel Guiderdoni[5,*]

[1]*IRD, UMR DIADE, CIAT, Agrobiodiversity and Biotechnology Project, Cali, Colombia*
[2]*IRD, UMR DIADE, Montpellier Cedex, France*
[3]*INRA, UMR BGPI, Campus International de Baillarguet, Montpellier Cedex, France*
[4]*CIAT, Agrobiodiversity and Biotechnology Project, CIAT, Cali, Colombia*
[5]*CIRAD, UMR AGAP, Montpellier Cedex, France*
[6]*INRA, UMR AGAP, Montpellier Cedex, France*
[7]*Biogemma, Site de La Garenne, Chappes, France*
[8]*Université Montpellier II, UMR DIADE, Agricultural Genetics Institute, National Key Lab for Plant Cell Biotechnology, Tuliem, Hanoi, Vietnam*

Received 18 August 2011;
revised 21 December 2011;
accepted 21 December 2011.
[*]*Correspondence* (Tel +33 4 67 61 56 29;
fax +33 4 67 61 56 05;
email jbmorel@cirad.fr;
guiderdoni@cirad.fr)
[†]These authors equally contributed to the work.

Keywords: forward genetics, insertion mutagenesis, reverse genetics, rice.

Summary

We report here the molecular and phenotypic features of a library of 31 562 insertion lines generated in the model japonica cultivar Nipponbare of rice (*Oryza sativa* L.), called *Oryza Tag Line* (OTL). Sixteen thousand eight hundred and fourteen T-DNA and 12 410 *Tos*17 discrete insertion sites have been characterized in these lines. We estimate that 8686 predicted gene intervals—i.e. one-fourth to one-fifth of the estimated rice nontransposable element gene complement—are interrupted by sequence-indexed T-DNA (6563 genes) and/or *Tos*17 (2755 genes) inserts. Six hundred and forty-three genes are interrupted by both T-DNA and *Tos*17 inserts. High quality of the sequence indexation of the T2 seed samples was ascertained by several approaches. Field evaluation under agronomic conditions of 27 832 OTL has revealed that 18.2% exhibit at least one morphophysiological alteration in the T1 progeny plants. Screening 10 000 lines for altered response to inoculation by the fungal pathogen *Magnaporthe oryzae* allowed to observe 71 lines (0.7%) developing spontaneous lesions simulating disease mutants and 43 lines (0.4%) exhibiting an enhanced disease resistance or susceptibility. We show here that at least 3.5% (four of 114) of these alterations are tagged by the mutagens. The presence of allelic series of sequence-indexed mutations in a gene among OTL that exhibit a convergent phenotype clearly increases the chance of establishing a linkage between alterations and inserts. This convergence approach is illustrated by the identification of the rice ortholog of *AtPHO2,* the disruption of which causes a lesion-mimic phenotype owing to an over-accumulation of phosphate, in nine lines bearing allelic insertions.

Introduction

The release of the high-quality genome sequence of cultivated rice of Asian origin (*Oryza sativa* L.), the main cereal of human consumption (Matsumoto *et al.*, 2005), has motivated large international efforts for setting up tools and resources aiming at inactivating gene function in this model graminaceous system. These tools encompass either sequence-specific inactivation through post-transcriptional (Warthmann *et al.*, 2008; Wesley *et al.*, 2001) or transcriptional (Li *et al.*, 2011) gene silencing and gene targeting (Iida and Terada, 2005), or random disruption through physical (Bruce *et al.*, 2009; Wu *et al.*, 2005), chemical (Till *et al.*, 2007) or insertional mutagenesis (for a review, see Krishnan *et al.*, 2009). In rice, efficient insertional mutagens allowing genome-wide mutagenesis include the endogenous ty-1 copia retroelement *Tos*17 (Miyao *et al.*, 2003), the introduced maize *Ac/Ds* (Chin *et al.*, 1999) and *En/Spm* (Kumar *et al.*, 2005) transposon systems, and the T-DNA of *Agrobacterium tumefaciens* (Jeon *et al.*, 2000).

Random insertion of a known DNA fragment—insertion mutagenesis—allows high-throughput PCR-based recovery and sequencing of genomic regions flanking the insertion sites and therefore to precisely determine their position on the chromosome pseudomolecules, which can be displayed in web-accessible databases. Implementation of *in silico* reverse genetics has been an extremely powerful tool to decipher gene function in *Arabidopsis,* where 385 000 flanking sequence tags (FSTs) are available online (http://signal.salk.edu/cgi-bin/tdnaexpress) (O'Malley and Ecker, 2010). An important requisite for that approach is an accurate indexation of the lines by their FST and availability of seeds in sufficient quantities for distribution. However, both FST characterization and seed multiplication are complex, error-prone

processes involving multiple steps that can each be a source of contamination or mislabelling. It is therefore important to assess the quality of the library by carrying out quality checks as well as by benefiting of return experience from the users. So far, the insertion mutagenesis effort in rice has allowed the generation of 500K insertion lines and the release of 200K indexed insertion sites in public databases (Krishnan *et al.*, 2009). Although still insufficient with regard to the rice genome size, the current insertion coverage nevertheless permits to find several alleles already seen in the same gene sequence, notably thanks to the strong insertion bias of *Tos*17-generating series of allelic insertions (Miyao *et al.*, 2003; Piffanelli *et al.*, 2007). One can take advantage of this property in determining whether convergent phenotypes are observed among the lines carrying allelic insertions in a given sequence. This would give an additional hint that the phenotype is attributable to the disruption of a gene and worth further 'wet' verification.

Large-scale field evaluation of rice insertion lines has allowed to observe altered phenotypes at a high frequency in *Tos*17 (Miyao *et al.*, 2007), *Ac/Ds* (Kolesnik *et al.*, 2004), and T-DNA (Chern *et al.*, 2007) lines also documented in public databases (http://tos.nias.affrc.go.jp/~miyao/pub/tos17/index.html.en, http://rmd.ncpgr.cn/, http://trim.sinica.edu.tw, http://oryzatagline.cirad.fr). The forward genetics approach could be extremely powerful for the discovery of novel gene functions, but unfortunately the frequency of linkage between the phenotype and the insertion mutagen generally remains extremely low (5%) (Nonomura *et al.*, 2003). In rice, this approach has allowed to identify genes, mainly in *Tos17* mutant lines and for traits relatively easy to screen such as viviparity (Agrawal *et al.*, 2001) or panicle fertility (Nonomura *et al.*, 2003, 2004). Successful examples of the disruptive action of the T-DNA conducting a conspicuous phenotype are however scarce, but the presence of a promoter trap and/or an activation tag carried by the T-DNA may additionally lead to reporter-mediated gene detection (Jung *et al.*, 2003; Lee *et al.*, 2004a,b) and observation of dominant phenotypes (Chern *et al.*, 2007; Jeong *et al.*, 2002; Mori *et al.*, 2007), respectively. Nevertheless, the recent development of next-generation sequencing technologies may offer a 'second life' to interesting but still untagged mutations, because sequencing the genome of a mutant of particular interest has now became affordable for a research laboratory (Sabot *et al.*, 2011).

In the frame of the plant genomics collaborative programme Génoplante, we have generated through *Agrobacterium*-mediated transformation of seed embryo calluses of the temperate japonica cultivar Nipponbare (Sallaud *et al.*, 2003) a library of 31 562 insertion lines carrying T-DNA (Sallaud *et al.*, 2004) and *Tos*17 (Piffanelli *et al.*, 2007) inserts. The inserted T-DNA contained either a *gusA* enhancer trap (C. Gay and E. Guiderdoni, unpublished data) or a GAL4:GFP enhancer trap (Johnson *et al.*, 2005), acting as a gene detector when inserted inside or in the vicinity of a gene. In a long-term collaborative effort of the International Center for Tropical Agriculture (CIAT, Cali, Colombia), T1 progenies of the insertion lines have been field-evaluated under agronomic conditions in Cali, Colombia, for the collection of phenotypes and seed propagation. Most of the T1 lines have been successfully harvested in a bulk manner, and their T2 seeds are available upon request. Individual harvest of plants exhibiting mutant phenotype has also been carried out and can be delivered as well (http://oryzatagline.cirad.fr).

A subset of the collection has been evaluated more thoroughly for detecting specific alterations in grain development and in response to inoculation by the fungal pathogen *Magnaporthe oryzae*. All that phenotypic information based on the trait ontology (TO) nomenclature is now gathered in the Oryza Tag Line database (Larmande *et al.*, 2008). The Oryza Tag Line database is linked through sequence information to the genome navigator OrygenesDB (Droc *et al.*, 2006, 2009). The direct link from the phenotype to the sequence through the line ID allows back and forth searches and to implement *in silico* reverse and forward strategies to focus the effort of gene discovery. Lines exhibiting convergent TO phenotypes can be investigated for the genes interrupted by characterized insertions, while allelic series of insertions in the same gene sequence can be analysed to highlight convergent phenotypes.

The objective of this study was threefold: (i) to estimate the genome coverage by the indexed inserts in the library and assess their quality, (ii) to have a first assessment of the frequency of linkage between the mutagens and the observed mutations and (iii) to determine whether *in silico* forward and reverse genetics could enhance the chance of identifying linkage between genes and phenotypes for agronomic traits.

Results

Molecular features of the inserts in the library

Number of T-DNA and Tos17 inserts

DNA blot of 384 T0 plants harbouring the p4978 T-DNA (Figure 1) hybridized with the *hpt* and *gusA* probes revealed that they contain an average of 2.7 T-DNA copies. These insertions often present a complex T-DNA organization resulting from sequence rearrangements (Figure S1). Forty-six per cent of the transformation events were found to also harbour vector backbone sequence. The probability for a transformation event to possess vector sequence increases as a function of the number

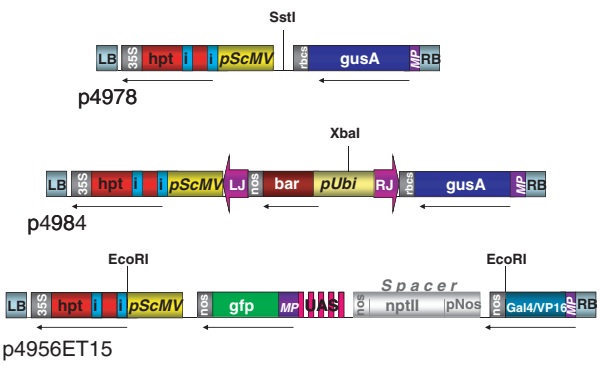

Figure 1 Structural elements of the T-DNAs used to generate the OTL library. Left (LB) and right (RB) borders of the T-DNA; left (LJ) and right (RJ) junctions of the Ac transposable element of maize) *E.coli* β-glucuronidase reporter gene (GUSA); composite transcription activator containing the yeast GAL4 binding domain fused to a modified VP16 activation domain (GAL4/VP16); multimerized yeast 17-bp upstream activation sequences (UAS); enhanced version of the *Aequorea victoria* green fluorescent protein reporter gene (eGFP); minimal promoter (−90 or −48 bp: MP); promoter sequence (35S P) and terminator region (35S T) of the CaMV 35S, subterranean clover mosaic virus promoter (ScMV P) fused to the rice Actin 1 first intron (Ai); hygromycin phosphotransferase (HPH) selectable gene with or without a castor bean catalase intron (i); *Agrobacterium* nopaline synthase terminator (nos T). Arrows point transcription direction.

of T-DNA copies integrated: 80% of the single T-DNA copy lines do not harbour vector backbone, while only 20% of the three copy lines have no backbone sequence (Figure S1). The Southern blot analysis of 47, 37, and 27 T1 progeny lines of plants harbouring the p4978, p4984, and p4956ET15 T-DNA showed that they contain an average of 2.8, 2.4, and 3.2 copies integrated at an average of 1.5, 1.2, and 1.4 loci, respectively. As the analysis of 384 T0 plants by Southern blot has previously shown that the T-DNA lines also contain an average of 3.3 unlinked new *Tos*17 inserts (Piffanelli *et al.*, 2007), one can conclude that the library of 31 562 lines carries approximately 45 000 and 100 000 T-DNA and *Tos*17 inserts, respectively.

Following elimination of redundant sequences isolated from the border of the same T-DNA inserts but generated by using different restriction enzymes, we determined that the library of 31 562 lines contains 16 814 discrete T-DNA integration sites indexed through the genome sequence extending from the LB (11 708), the RB (3743) or both the LB and RB (1363) at insertion points. Analysis of the 1363 T-DNA inserts characterized at both the LB and RB, through alignment of these sequences to the Nipponbare pseudomolecule template, permitted to estimate that T-DNA insertion created an average sequence gap of 75 bp. The 14 284 sequences flanking *Tos*17 inserts were resolved into 12 410 discrete insertion sites, the redundancy being attributable to sequencing inserts from distinct T-DNA lines sharing the same cell lineage (Piffanelli *et al.*, 2007).

These 16 814 T-DNA insertion sites are carried by a core of 17 667 lines in the library (Table 1). The 12 410 unique *Tos*17 insertion sites are similarly found in a core of 14 058 lines. Eight thousand five hundred and forty-nine lines carry no characterized insert. Failure of FST recovery from an insert can be due to the absence of restriction enzyme site used for FST capture in its vicinity, unfavourable organization (T-DNA vector backbone, tandem repeats, etc.), presence of multiple T-DNA or *Tos*17 inserts at different loci in a line, or quality of the DNA preparation.

Rice genome coverage: number of genes interrupted by a T-DNA and/or a Tos17 insert

Distribution of these inserts in genic and intergenic regions shows that of more than 29 224 (16 814 + 12 410) sequence-indexed inserts, 18 303 fall in predicted gene intervals (−1000 upstream ATG to 300 bp downstream the stop codon) (Tables 2 and S1). Integration of *Tos*17 occurs more frequently in genic regions (84% vs. 47%) and in coding sequences than that of T-DNA, whereas T-DNA inserts are more frequent in 5′ and 3′ regions than *Tos*17. Seven thousand nine hundred and nineteen and 10 384 discrete insertion sites of T-DNA and *Tos*17, respectively, fall in predicted gene intervals. However, as the number of inserts per gene averages 1.2 and 3.8 for T-DNA and *Tos*17, respectively, the T-DNA and *Tos 17* inserts eventually disrupted 6563 and 2755 discrete non-TE gene intervals, respectively. Six hundred and forty-three genes have allelic insertion of both T-DNA and *Tos*17. From this study, we can therefore deduce that about 8686 rice genes are interrupted by at least one sequence-indexed insert in the library.

FST quality assessment

FSTs were generated from DNA isolated from young regenerated T0 plants. Plants were later harvested, T1 seeds were sown on the field, and then T2 seeds were harvested to serve for distribution. To determine whether the FSTs are accurately assigned to a T2 seed sample, 69 T2 progeny lines were analysed by DNA blot using a gene-specific probe. Insertion creates a shift of position of a hybridization signal, allowing a discrimination of homozygous, heterozygous, and azygous plants and confirming the presence of a T-DNA insert at the expected integration site. These lines were analysed in the frame of several scientific projects and can be considered, in first instance, as a random set. The reconfirmation rate for the T-DNA was of 55 lines over 69 (79.7%). As a complementary approach, genomic regions flanking T-DNA and *Tos*17 inserts were also amplified and sequenced from a random set of 40 T2 seed samples, and the same FSTs than those obtained from the original T0 plants were retrieved in 36 samples (90%). The 4 remaining FSTs were different from the one sequenced at first, but not previously existing in our database. They therefore likely result from the sequencing of genomic regions flanking a second insert in the line.

Another insight is to determine whether lines harbouring sequence-indexed inserts in genes at position anticipated to create knockouts (K.O.) with known phenotype exhibit the actual phenotype. In that aim, we searched the OTL insertion library for allelic inserts in genes described in the literature and observed the plant phenotype in the corresponding OTL at the T2 generation. Among these genes is *SLENDER1* (*SLR1*) (Os03g49990), a plant-specific transcription factor of the GRAS family. SLR1 is a negative regulator of Gibberelin response, and its disruption induces a giant phenotype first visible at seedling stage and later on at flowering (Ikeda *et al.*, 2001). Another example is *MULTIPLE SPOROCYTE1 (MSP1)*, (Os01g68870) an

Table 1 A summary of lines of the OTL library harbouring 0, 1, or 2 sequence-indexed T-DNA and *Tos*17 inserts

		Locus *Tos*17			
		0	1	2	No. of lines with X sequence-indexed T-DNA inserts
Locus T-DNA	0	8549	5301	45	13 895
	1	5643	4919	54	10 616
	2	3312	3612	127	7051
	No. of lines with X sequence-indexed *Tos*17 inserts	17 504	13 832	226	31 562

Table 2 Distribution of discrete T-DNA and *Tos*17 insertion loci in genic and intergenic regions in the OTL library

Region	Locus T-DNA	%	Locus *Tos*17*	%	Total	%
−1000/ATG	2844	16.9	747	6.0	3591	12.3
Intron	2504	14.9	3960	31.9	6464	22.1
Exon	1760	10.5	5197	41.9	6957	23.8
Stop/+300	811	4.8	480	3.9	1291	4.4
Intergenic	8895	52.9	2026	16.3	10 921	37.4
Total	16 814	100.0	12 410	100.0	29 224	100.0

*From the study of Piffanelli *et al.* (2007).

LRR-RLK involved in restricting the number of cells entering into male and female sporogenesis, the K.O. of which being sterile (Nonomura *et al.*, 2003) A T-DNA and a *Tos*17 insertion line in *SLR1* and *MSP1* were, respectively, identified in the OTL library, and the resulting homozygous lines displayed the anticipated phenotypes (Figure S2). In addition, we recently showed that insertions in five genes, which are previously described to be involved in disease resistance to rice blast pathogen, corresponded to lines indeed affected in disease resistance (Delteil *et al.*, 2012).

Phenotypic variation in the collection

Variant trait classification

A total of 27 832 lines were observed at seedling stage in the screenhouse and then under field conditions till harvesting during six seasons from 2003 to 2008 at the CIAT headquarters experimental station, Cali, Colombia. The rest of the lines (3730) had a very limited set of T1 seeds and were kept at Cirad, Montpellier, France, for on demand seed increase under greenhouse conditions. The field observations were recorded in the *Paddy Genes Book* database, using the T0 classification. To compare our phenotypic observations with the two previously published—the NIAS *Tos*17 collection (Miyao *et al.*, 2007) and the TRIM activation tagging T-DNA collection (Chern *et al.*, 2007)—we established a classification very similar to those adopted in these studies with a few modifications (Table 3). This led to the identification of 12 trait categories, namely germination, growth, leaf colour, leaf morphology, plant morphology, mimic response, tillering, heading date, flower, panicle morphology, seed fertility, seed morphology, and root system. The final trait category list comprised 71 subcategory traits (vs. 61 in the study of (Chern *et al.*, 2007) and 53 in the study of (Miyao *et al.*, 2007) (see Table S2).

Trait variation frequencies

The observed frequencies of variation of traits were displayed following categories appearing in the studies of Miyao *et al.* (2007) and Chern *et al.* (2007) and according to the TO classification in Tables S2 and S3 respectively. The most commonly observed phenotypes were low germination rate (21.95%), sterility (4.71%), dwarfism (3.04%), albinism (2.69%), yellow leaf (2.35%), late heading (1.44%), horizontal leaf (1.19%), abnormal hull (1.06%), and weak plant (1.03%). The frequency of lines exhibiting low germination rate appears much higher than those in previous published studies. However, a large part of the low germination rate lines observed in our trials might be due to the conditions of seed drying and storage before sowing as this was dependent on the batch of lines. Up to 12 altered traits have been observed simultaneously on a given variant line (Figure S3). It is noteworthy that the frequencies may vary significantly between the three studies. For example, the frequency of lines segregating an at least 95% sterility phenotype over the total number of lines under observation was of 4.71% in the OTL collection. 7.65% and 2.52% of sterile lines were observed in the *Tos*17 and TRIM collections, respectively. Figure 2 illustrates some of the remarkable phenotypes observed in the OTL collection.

Specific screens for grain phenotypes

Upon laboratory examination of the harvested panicles, manual dehulling, and observation of the seeds under a stereomicroscope, 3.5% of the 7187 T0 plants observed were found to exhibit a T1 seed development phenotype. The most frequent among the observed phenotypes illustrated in Figure 3a was aborted seeds (abs) followed by wrinkled (wr), shrunken (shr), round kernel (rd), defective (def), and brown seeds (Br). As the phenotype observed on a primary regenerant may be prone to extensive nonheritable variation, observations were also made on panicles bearing T2 seeds from field- or greenhouse-grown plants. The phenotypes were generally confirmed in these more advanced generations. To determine whether the grain phenotype observed in the most interesting 34 lines was linked to a T-DNA or a *Tos*17 insert, DNA of 20–30 T3 progeny plants segregating the phenotype was isolated and subjected to DNA hybridization using a T-DNA- or a *Tos*17-specific probe to examine linkage between phenotype and the mutagens. For the aborted seeds' phenotype which is supposed to be a consequence of reaching homozygous status for the mutation, all the heterozygous plants showing 25% seed lethality should contain the insert. However, none of the observed phenotypes appeared to be due to a mutagen upon linkage analysis.

The *gusA*- and *Gfp* reporter gene-based enhancer trap detection systems borne by the T-DNA inserted in the library may increase the chance of identifying genes expressed in seed tissues that might turn to play an important role in grain development and filling. In that aim, T1 seed sections of the 7187 primary transformants harbouring either the p4978 or the p4956ET15 T-DNA were observed for GUS activity and fluorescence, respectively. Nine per cent and 4% of the p4978 and p4956ET15 lines, respectively, exhibited reporter gene expression in at least one of the seed tissue (embryo, aleurone, and/or endosperm). These lines can be further characterized for isolating genes expressed or promoters active in seed tissues.

Specific screen for response to Magnaporthe oryzae

In a recent paper, we demonstrated the usefulness of the OTL collection for the analysis by reverse genetics of disease resistance (Delteil *et al.*, 2012). Here, we tested whether this collection could be used for forward genetic analysis of rice blast resistance. Approximately 10 000 T1 lines were screened for altered disease resistance to *M. oryzae* under controlled, growth room conditions. The virulent FR13 isolate was chosen to enable the identification of either lines showing enhanced disease resistance (EDR) or lines showing enhanced disease susceptibility (EDS). The observed phenotypes were checked either after sowing again the T1 progeny or by selfing the putative mutants and testing their progeny. Only the lines that showed reproducible phenotypes were further analysed. A total of 26 and 17 lines were detected as displaying an EDR and an EDS phenotype, respectively. Selected examples of these phenotypes are shown in Figure 4a. Fourteen of these phenotypes were found independently in two lines derived from the same callus, suggesting that in seven cases the same mutation was responsible for the observed phenotype. These shared mutations are resulting from the line generation process where several lines can be regenerated independently from a single cocultivated callus (Piffanelli *et al.*, 2007; Sallaud *et al.*, 2004). Therefore, it is likely that the 43 phenotypes identified represent in fact 31 phenotypes. Different degrees of EDS and EDR phenotypes could be observed. For example, the AAMF11 line showed the most severe EDS phenotype, whereas the AFHG05 line the strongest EDR phenotype. Except for the mutation existing in both lines AFHG04 and AFHG05 lines—derived from the same

Table 3 Classification of observed phenotypes and their associated frequencies in the OTL library

Category	Subcategory	No. lines	Frequency
Growth	Low germination rate*	6108	21.95
	Lethal	560	2.01
	Abnormal plants	210	0.75
	Weak	286	1.03
Leaf colour	Albino	186	0.67
	Yellow leaf	653	2.35
	Dark green leaf	29	0.10
	Pale green leaf	19	0.07
	Virescent	4	0.01
	Stripe	254	0.91
	Zebra	17	0.06
	Stay green	2	0.01
	Others	3	0.01
Leaf morphology	Wide leaf	92	0.33
	Narrow leaf	48	0.17
	Long leaf	87	0.31
	Short leaf	32	0.11
	Drooping leaf	32	0.11
	Rolled leaf	88	0.32
	Spiral leaf	18	0.06
	Brittle leaf/culm	14	0.05
	Horizontal leaf	330	1.19
	Erect leaf	69	0.25
	Others	64	0.23
Plant morphology	Semidwarf	847	3.04
	Dwarf	250	0.90
	Long culm	136	0.49
	Erect	5	0.02
	Spread-out	104	0.37
	Compact	90	0.32
	Thin culm	21	0.08
	Thick culm	24	0.09
	Lazy	28	0.10
	Culm branching	7	0.03
Mimic response	Lesion mimic	273	0.98
Tiller	Monoculm	35	0.13
	Low tillering	167	0.60
	High tillering	79	0.28
	Red culm	1	0,00
Heading date	Early heading	0	0.00
	Late heading	401	1.44
	No heading	0	0.00
Flower	Abnormal hull	223	0.80
	Abnormal floral organ	0	0.00
	Awned	106	0.38
	Abnormal hull colour	52	0.19
	Others	3	0.01
Panicle morphology	Long panicle	4	0.01
	Short panicle	4	0.01
	Sparse panicle	0	0.00
	Dense panicle	28	0.10
	Shattering	0	0.00
	Neck leaf	0	0.00
	Abnormal panicle shape	120	0.43
	Incomplete exertion	27	0.10
	Few spikelets	40	0.14
	Terminal spikelet	8	0.03
	Others	49	0.18
Seed fertility	Sterile†	1311	4.71
Seed morphology	Large grain	0	0.00
	Small grain	71	0.26
	Slender grain	0	0.00
	Others	0	0.00
Root system	Weak root	22	0.08

*Likely overestimated and attributable to seed preservation problems.

†Means segregation of an at least 95% spikelet sterility phenotype.

transformation event—all phenotypes were segregating as recessive mutations (data not shown).

Linkage analysis was performed for the T-DNA using the *hpt* probe on the 43 lines (Table 4). The absence of tagging was demonstrated for 33 lines (mutant plants showing no common T-DNA band on Southern blot). In contrast, we could not exclude potential tagging by the T-DNA for 12 of 43 phenotypes analysed. For example, the link between the T-DNA and the EDS phenotype was found in 19 of 19 T1s in the case of AKJH07 (data not shown), suggesting a strong linkage between this phenotype and the T-DNA.

For two lines (AKJH07 and ALKE03), the genes potentially mutated by the T-DNA were mutated in another, independent line (AEWH07 and AJVE08, respectively). We thus tested whether these independent alleles were displaying an EDS phenotype. This was the case for the AEWH07 line. This observation and the strong linkage between the T-DNA and the EDS phenotype in the AKJH07 line suggest that the insertion in the Os08g06140 gene encoding an NAC transcription factor is likely responsible for the observed phenotype. Similarly, in the case of the ALKE03 and AJVE08 lines (Figure 4b,c), the same phenotype could be associated with an insertion in the promoter of the Os01g52500 gene encoding an NADP-malic enzyme. At this level of analysis, the data indicate that at least two of the 31 EDR/EDS phenotypes have a high probability to be tagged.

Specific screen for lesion-mimic phenotypes

During the *M. oryzae* screen, we also look for mutants that would display spontaneous lesions resembling disease before inoculation or lesions not typical of *M. oryzae* infection. Such mutants are well known in rice and other plant species (Lorrain *et al.*, 2003). Of approximately 10 000 T1 lines screened, we identified 41 lines showing discrete lesions simulating disease (LSD; see for example ALMD12, ALTG10, and APIE05 illustrated in Figure S4) and 30 lines showing necrotic sectors (NEC; see for example AKOG07, AMWG09, and AICG07; Figure S4). Quite interestingly, two mutant lines (AGGC08 and AGOD03) showed lesions that expanded, resembling propagation mutants like the Sekiguchi mutant in

Figure 2 Selection of phenotypes for trait ontology–referenced traits observed during field evaluation of the OTL collection in CIAT, Colombia. Only mutations segregating according to a 1 : 3 ratio are shown.

rice (Fujiwara *et al.*, 2010). The AGGC08 phenotype was inducible by *M. oryzae* inoculation (Figure S5). The analysis of the APIE05 T-DNA insertion site revealed that the known *SPL7* gene (Yamanouchi *et al.*, 2002) was mutated in this line (Delteil *et al.*, 2012). We thus conclude that in this case the lesion-mimic phenotype was probably due to this insertion.

Similar to EDS and EDR mutants, we looked for linkage between the *Tos*17/T-DNA insertion elements and the LSD/NEC phenotypes. We found very similar potential tagging frequencies for T-DNA between the EDS/EDR phenotypes and LSD/NEC phenotypes (Table 4). Like EDS/EDR mutants, when this was possible we looked for independent mutations for genes that were potentially mutated based on linkage analysis. This could be done for nine genes corresponding to nine LSD/NEC lines. At least two lines per gene were tested, and the presence of the putative mutation was checked (data not shown). We could show that in one case (initial line AICG07), the phenotype was attributable to a mutation in the gene initially identified. This gene corresponds to the *PHO2* gene (Figure 5a) for which we found nine of 13 lines displaying necrosis (see example Figure 5b). All the lines that were knocked out for the *PHO2* gene displayed necrosis and, as expected, an over-accumulation of inorganic phosphate (Figure 5c,d).

Discussion

Genome coverage

The rice community has the joint objective of deciphering the function of most of the agronomically important genes by year

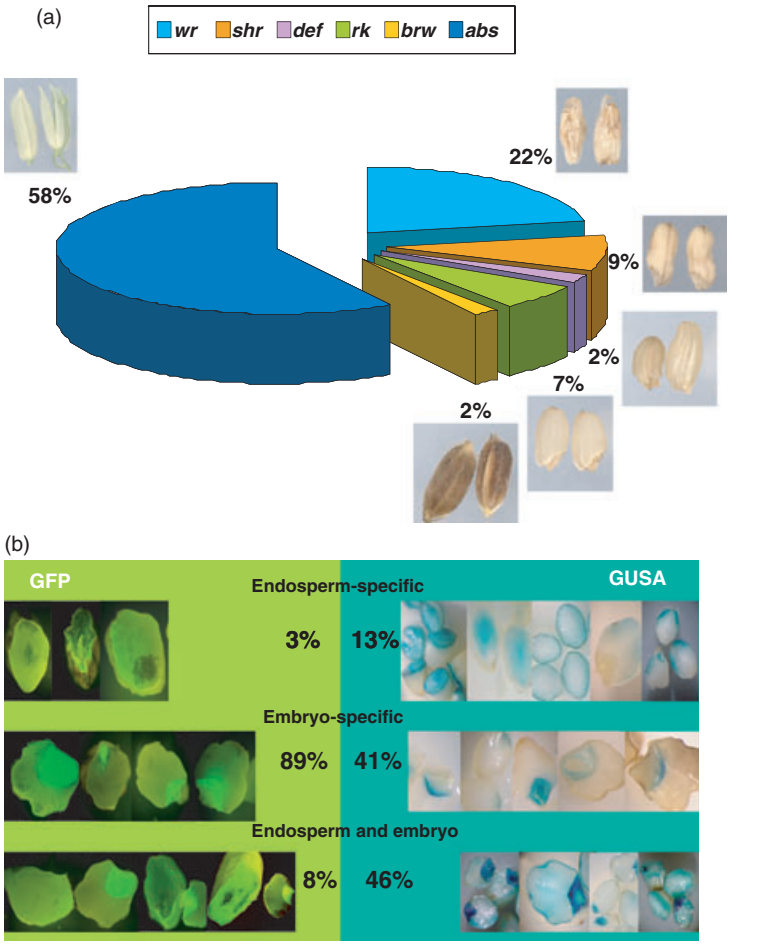

Figure 3 Grain phenotypes observed in the OTL mutant collection. (a) Distribution of occurrence of grain filling and development phenotypes: aborted seeds (abs), wrinkled (wr), shrunken (shr), round kernel (rd), defective (def), and brown seeds (Br). (b) A summary of frequencies of GUSA activity or GFP fluorescence observed in T1 seed sections of enhancer trap lines harbouring the p4978 and p4956ET15 T-DNA construct, respectively.

2020 (Zhang *et al.*, 2008), an effort much needed to unravel the evolution of trait determinism and the molecular bases of monocot- and crop-specific traits. In that aim, sequence-indexed insertion libraries represent a helpful resource. To determine whether the effort engaged more than a decade ago has been sufficient, it is necessary to assess the quality and the genome coverage of the sequence-indexed inserts present in the current international collections.

Quality of the association between T2 seed bags and FST (more than 80%), deduced from their use in reverse genetics, sequencing twice the FST, and observation of known phenotypes, indicates high reliability of our collection. A similar rate (76%) of FST reconfirmation has been observed in two important *Arabidopsis* T-DNA libraries (O'Malley and Ecker, 2010).

One can estimate that altogether 8686 different rice genes, representing one-fourth to one-fifth of the estimated gene complement, are interrupted by a sequence-indexed insert. Nevertheless, it is likely that not all disruptions result in gene K.O.s, because of inefficiency for some insertions, notably in intron or 3'UTR regions, at preventing translation. However, information deduced from Southern blot analysis suggests that the library actually contains 100 000 and 45 000 *Tos*17 and T-DNA insertion loci, respectively. Even though number of lines to reach genome saturation by inserts relies on an exponential and not a linear relation (Krishnan *et al.*, 2009) and mutagens may exhibit strong insertion bias to certain genes limiting their coverage, this indicates that potentially a much larger set of rice genes can have an insert in an insertion library of rather limited size like OTL. Therefore, upon observation of a phenotype in a line, a subsequent step to the search for presence of sequence-indexed inserts in predicted genes is to survey the cosegregation of other T-DNA and Tos17 insertions residing in the line with the altered trait through DNA blot analysis.

To harness the full potential of the library, one can use PCR surveys in DNA pools (Hirochika, 2001; Lee *et al.*, 2003) using gene- and element-specific primers. However, organization of the T-DNA, integration of backbone sequences, and GC content of the target genomic region may represent important limitations. Nowadays, mutant genome sequencing represent a hopeful alternative and has moreover the potential of revealing hidden lesions in the genome such as point mutations, structural rearrangements, and insertions of other elements mobilized during the transformation/regeneration process (Sabot *et al.*, 2011). Resequencing the genome of Arabidopsis and rice regenerants has recently shown that point mutation is the major source of somaclonal variation (Jiang *et al.*, 2012; Miyao *et al.*, 2012).

Range of phenotypic variation

Field evaluation of insertion lines has been conducted for seed propagation and description of the collection by phenotypic

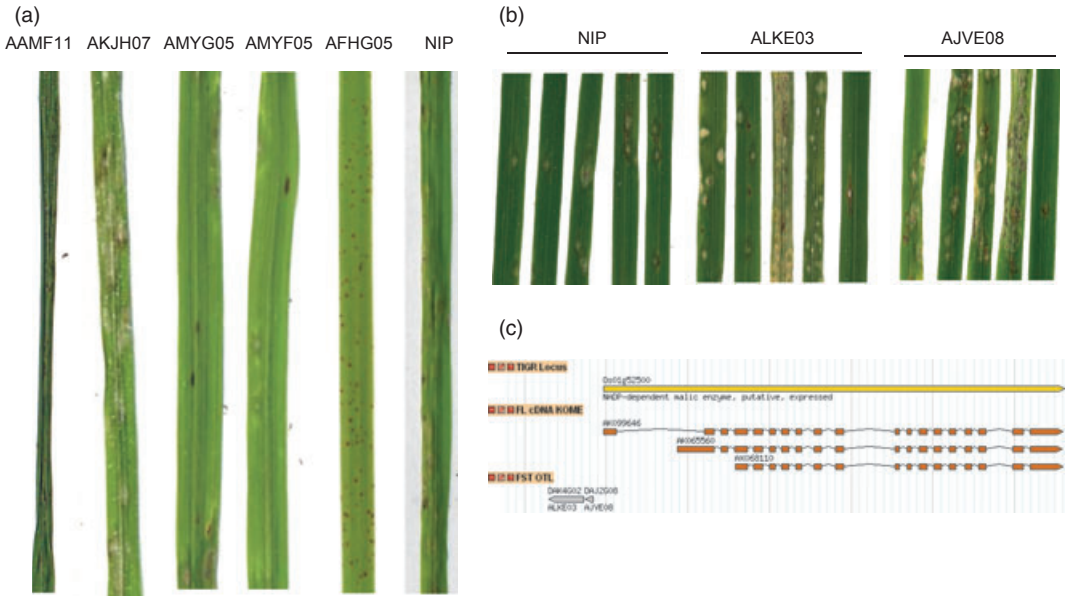

Figure 4 Mutants affected for rice blast resistance in the OTL mutant collection. (a) Representative mutants are shown: increased susceptibility (AAMF11, AKJH07) or increased resistance (AMYG05, AMYF05, AFHG05) as compared to wild-type Nipponbare (NIP). The greyish lesions are typical of susceptibility, while the dark and small lesions are typical of resistance response. (b) The ALKE03 and AJVE08 lines display similar enhanced susceptibility and are both mutated in the same gene as described in (c). (c) Position of the insertion sites in the ALKE03 and AJVE08 lines (grey arrows). All plants were grown 3 weeks and inoculated with the normally virulent isolate FR13 of *Magnaporthe oryzae*.

Table 4 Estimate of tagging frequency of phenotypic alterations by T-DNA and *Tos*17 inserts in the OTL library, based on analysis of mutants exhibiting a lesion-mimic (LSD/NEC) or an altered fungal infection response (EDR/EDS)

	T-DNA			*Tos*17		
	LSD/NEC	EDR/EDS	All phenotypes	LSD/NEC	EDR/EDS	All phenotypes
Total mutants	71	43	114	71	43	114
Mutants analysed	66	43	109	26	26	52
No linkage between insertional element and phenotype	47	33	–	16	13	–
Successful allelism test/total tested	0/4	2/2	2/6	1/5	0/0	1/5
Maximum potentially tagged (linkage and/or allelism)	15 (23%)	12 (28%)	27 (25%)	6 (23%)	13 (50%)	19 (36%)
Strong evidence for tagging	1 (3%)	2 (4.7%)	3 (2.7%)	1 (3.8%)	0 (0%)	1 (1.9%)

EDR, enhanced disease resistance; EDS, enhanced disease susceptibility.

records. The latter could be extended to the behaviour of the insertion lines under various environmental constraints (drought, salinity, high and low temperatures, etc.). 18.2% of the lines exhibited alteration in at least one of the observed traits. Some of the lines cumulated up to 12 variant traits, some of them likely resulting from the pleiotropic effect of some mutations. Comparison of previous field evaluations of the 50 000 NIAS *Tos*17 Nipponbare library (Miyao *et al.*, 2007), 22 000 TRIM T-DNA activation tagging Tainung67 library (Chern *et al.*, 2007), and >100 000 lines of the RMD T-DNA ZhongHua11 library (Zhang *et al.*, 2006) indicates that variation is observed in the same categories and subcategories of traits, although with variable frequency. For some traits such as EDS/EDR (0.4%) and LSD/NEC (0.7%) mutants, we found the same frequency of variation as in other mutant populations. For example, Wu *et al.* (2005) found 0.18% mutants affected for blast resistance and 0.74% lesion-mimic phenotypes in an IR64 deletion mutant collection. The altered traits were also found those exhibiting a wide range of variation in plants regenerated from germinal and somatic, tissue, cell, and protoplast cultures that have been extensively evaluated in the 1970–1990s to harness the potential of somaclonal variation in rice breeding. Frequency, range, and favourable vs. unfavourable feature of the variation were found to be genotype dependent (Sukekiyo and Kimura, 1991). Duration and procedure used in tissue culture are also important factors influencing the frequency of variation, as illustrated by the known accumulation of *Tos*17 copies in cells over time in culture (Hirochika *et al.*, 1996). Differences in frequencies observed between the insertion line libraries may therefore be genotype and tissue culture procedure dependent but may also result from the use of T-DNA that create additional lesions (abortive and nonabortive insertions) and may carry an activation tag (creating dominant mutations and

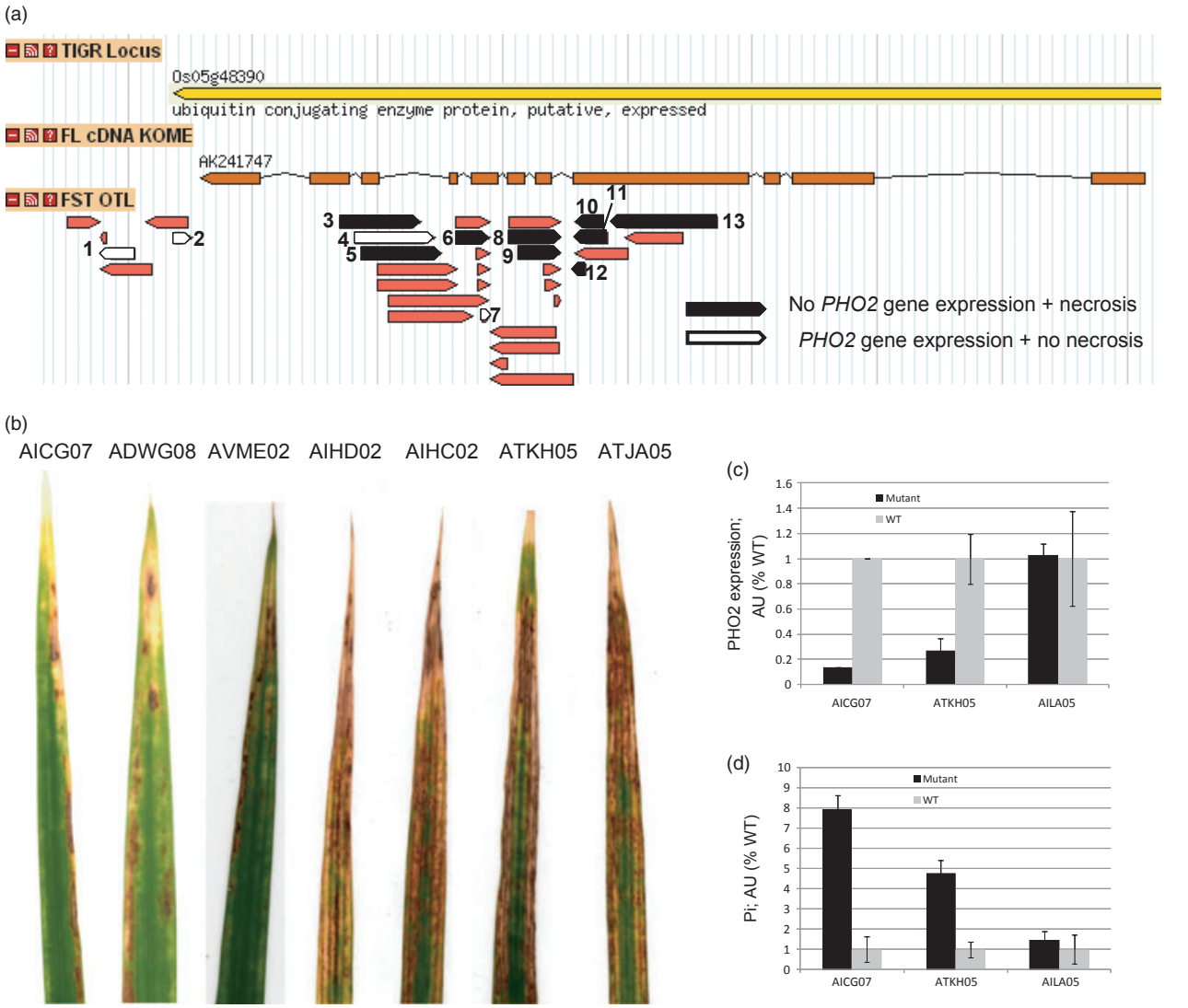

Figure 5 The PHO2 gene is a *Tos*17 integration hot spot. (a) Lines harbouring the *Tos*17 insert (mutant) or sister lines not containing *Tos*17 insertion (Wild-type = WT) were grown in the greenhouse under normal growing conditions. The position of the insertion is depicted in the 13 lines tested (1 = ALWB09, 2 = ATEF09, 3 = ATJA05, 4 = AILA05, 5 = ATKH05, 6 = AEVH04, 7 = AQAG12, 8 = AIHD02, 9 = AIHC02, 10 = ADWG08, 11 = AICG07, 12 = AVME02, and 13 = ADUC03). The black arrows represent lines that showed necrosis, the white arrows the lines that did not display necrosis and the red arrows lines that were not tested. (b) Example of necrotic phenotypes on 8-week-old plants. The relative severity of necrosis is not representative but merely reflects independent experiments. (c) PHO2 gene expression measurements in mutant and their corresponding WT sister lines. RNA were extracted and analysed as described in the study of Delteil *et al.* (2012) using PHO2-specific primers (data not shown). (d) Inorganic phosphate was measured as described in the study of Itaya and Ui (1966).

having the potential of producing a phenotype in inserting in genes accomplishing a redundant function).

Tagging frequency

Interest of the use of insertion mutant libraries in forward genetics screens for gene discovery depends on the answer to an important reiterated question: What is the frequency of phenotypic variations tagged by the mutagens? We have mentioned above that phenotypic variation may have many causes such as DNA/histone methylation, point mutations, deletions, larger structural rearrangements, or mobilization of other transposable elements. The only published information so far for rice insertion mutagenesis has long been that of the *Tos*17 NIAS library, which evaluated this frequency to 5% (Nonomura *et al.*, 2003). Using specific screens for three different traits (grain development, apparition of spontaneous lesions, and response to inoculation by a fungal pathogen), we show here that efficiency of tagging is dependent on the trait, likely on its amenability to somaclonal variation, and probably on the robustness of the undertaken phenotypic screen. As the segregation analyses between inserts and the altered trait were carried out through DNA blot analysis, all the inserts containing the probe sequence were simultaneously surveyed. Whereas establishment of linkage was unfruitful in grain-associated phenotypes, it was more successful with the two other traits. For the EDS/EDR/LSD/NEC phenotypes, we found in a first assessment that 19 phenotypes of 52 and 27 phenotypes of 109 could be due to *Tos*17 insertions and T-DNA, respectively (Table 4). These numbers likely overestimated the tagging

frequency as in some cases, linkage analysis was only performed on a small number of plants. When considering only the cases where linkage analysis relies on more than five plants, the maximum tagging efficiency is 12 of 109 (11%) for the T-DNA and 11 of 52 (21%) for *Tos*17. Conversely, a minimum tagging efficiency can be estimated when considering that of 43 EDR/EDS and 71 LSD/NEC phenotypes analysed, two EDS phenotypes (original lines AKJH07 and ALKE03) and two LSD/NEC (APIE05 and AICG07) are likely due to T-DNA or *Tos*17 insertions. Quite similar tagging efficiencies were found for the T-DNA (2.7%) and for *Tos*17 (1.9%) (Table 4). Overall, our data suggest that at least four phenotypes of 114 tested (3.5%) are due to an insertion element in these specific screens. The AKJH07/AEWH07 allelic mutations likely identify a NAC transcription factor. Such genes have already been shown to be required for disease resistance (Wang *et al.*, 2009) or abiotic stress tolerance in rice (Hu *et al.*, 2006). In our case, dwarfism associated with these mutations (data not shown) could also be responsible for the increased susceptibility, as blast susceptibility is highly dependent on the development stage (Ribot *et al.*, 2008). In contrast, the ALKE03/AJVE08 allelic pair in the putative NADP-malic enzyme do not display obvious morphological change. The involvement of this gene as a positive regulator of disease resistance is consistent with previous report (Parker *et al.*, 2009) that the NADP-malic enzyme activity is increasing upon rice blast infection and that this activity seems to be suppressed during compatible interaction. The two other phenotypes potentially tagged are due to mutations in known genes (SPL7: Yamanouchi *et al.* (2002) and OsPHO2: Bari *et al.* (2006)).

Convergence of hints increase the chance of wet validation of tagging

To increase the chance of establishing linkage between phenotypic variation and insertional mutagens, it might be important to rely on a convergence of hints. Such convergence can be the coincidence of enhancer trapping–mediated reporter gene detection and phenotype in a given organ (such as the seed) or coincidence of phenotypes in independent lines which proved to contain allelic sequence-indexed insertions in a given gene. The latter has been illustrated by the lesion-mimic phenotype observed in 9 lines which was found to result from the KO allelic insertions in the *PHO2* gene, which turned to be a hot spot for *Tos*17 inserts.

This example stresses the need to link phenotypic databases through a common vocabulary to fully harness the potential of that information for focusing molecular validation in forward genetics gene discovery studies. Phenotypic information resulting from field observations or specific screens is so far gathered in distinct web-accessible databases (http://tos.nias.affrc.go.jp/miyao/pub/tos17/index.html.en, http://rmd.ncpgr.cn/, http://trim.sinica.edu.tw, http://oryzatagline.cirad.fr). An important step remains to be accomplished to link these databases through the use of a common vocabulary to describe the altered traits. In that aim, the Plant Ontology Consortium (Jaiswal *et al.*, 2005)-controlled vocabularies to describe mutant phenotypes appear the most suitable to ensure the possibility of future crossreferencing between different databases. Using the hierarchical architecture of TO would allow starting a cross-database search for a broad term (e.g. plant morphology) to establish a first list of lines from different insertion libraries that exhibit a convergent phenotype. Refining the search using narrower TO terms (e.g. plant height) and further examining the function of the genes interrupted in this set of lines (e.g. GA-related) may provide precious indications on the pathways involved in the elaboration of the altered trait.

Nowadays, rice insertion libraries represent more than 500 000 lines and 240 000 sequenced inserts (Krishnan *et al.*, 2009). Although allowing the identification of an insertion in a majority of rice genes owing to the insertional bias of the mutagens in gene-rich regions, the number of sequence-indexed inserts remains largely insufficient and lags far behind *Arabidopsis* (385 000 insertion sites), the genome of which representing one-third of rice. Even with such extensive coverage, it has to be kept in mind that 12.2% of the *Arabidopsis* genes yet remain devoid of insertion (O'Malley and Ecker, 2010).

The 30 000 lines of the OTL library may have potential to contain a larger number of lesions in the rice genome covering many more genes, but their full molecular characterization is hardly achievable using conventional PCR-based methods. A promising perspective is the full-genome sequencing of thousands of mutants, which will soon become an affordable effort and would not only reveal virtually all the insertion sites of the known mutagens but also characterize additional hidden lesions (Zuryn *et al.*, 2010). Such an effort would also largely unravel the causes of somaclonal variation residing at the nucleotide level.

Experimental procedures

Generation of the library

Plasmid constructs

Three T-DNA vectors—p4978, p4984, and p4956ET15—have been used to generate the library (Figure 1). Preparations of the p4978 and p4956ET15 vectors have been described in detail in the studies of Sallaud *et al.* (2004) and Johnson *et al.* (2005), respectively. The p4984 vector has been constructed by inserting an ubi: bar :nos3′ sequence flanked by the terminal inverted repeats of the maize Ac element as a XhoI-SpeI fragment in the EcoICRI site of p4978 (W. Tucker and A. Betzner, ANU, unpublished). The three constructs have been built on the pCAMBIA1300 backbone (R. Jefferson, CAMBIA, Canberra, Australia).

Transformation

First-generation (T0) T-DNA plants have been produced by cocultivation of secondary, seed embryo scutellum–derived calluses of the japonica cultivar Nipponbare of rice (*Oryza sativa* L.) with *Agrobacterium tumefaciens* strain EHA105 or LBA4404 carrying the T-DNA above-mentioned constructs, according to the procedure described by Sallaud *et al.* (2003). Nipponbare seeds serving at tissue culture were first provided by the National Institute of Agrobiological Sciences (NIAS, Tsukuba, Japan) and then seed increased by our own. The transformation efficiency averaged 4 T0 T-DNA plants per cocultivated secondary, seed embryo scutellum–derived callus (4.4, 2.3, and 4.0 for the p4978, p4984, and p4956ET15 constructs, respectively). T1 seeds have been harvested from 19 000 p4978, 3000 p4984 and 9000 p4956ET15 T0 plants.

Molecular characterization of the T-DNA and *Tos*17 inserts

Generation of flanking sequence tags

Flanking regions of inserts were amplified from the left border (LB) of T-DNA inserts and from the right LTR of *Tos*17 inserts

according to the adapter-PCR protocols described in the studies of Sallaud *et al.*, 2004 and Piffanelli *et al.*, 2007, respectively, using genomic DNA isolated from the T0 p4978, p4984, and p4956 ET 15 plants. In addition, walk-PCR was conducted from the right border (RB) of T-DNA inserts in *EcoR*V or *Ssp*I DNA digests (D. Mieulet and P. Piffanelli, unpublished). The sequencing of 18 156 RB PCR products led to the anchoring of 7296 (40%) informative sequences on the rice genome. The frequency of informative sequences was lower than that resulting from the sequencing of regions flanking the T-DNA insert LB (57%). This was attributable to a frequent RB : LB and RB : RB tandem organization of T-DNA copies at the integration loci, representing 37.3% of the sequences. The 7296 anchored sequences represented 5262 discrete insertion points of the T-DNA. Redundancy was mainly due to the use of two independent PCRs from the same DNA sample digested with either *EcoR*V or *Ssp*I. To a lesser extent (5%), it may also result from the analysis of clonal transformation events regenerated from the same cocultivated callus piece.

Southern blot analysis

Southern blot analyses were used to determine the number of copy and integration loci of the mutagens, verify disruption of target genes by a given mutagen, and analyse linkage between a mutation and a mutagen. Genomic DNA was extracted from 250 to 500 mg of T0 regenerated plants or of 6-week-old T1/T2 seedlings with the MATAB method. Eight micrograms of DNA was digested with *Eco*RI, *Xba*I, *Bam*HI, or *Hin*dIII for T-DNA and *Tos*17 insertion lines for 12 h at 37 °C in a final 200 μL. Samples were then precipitated with 500 L of absolute ethanol and 10 μL of NaCl 5 M for 1 h at −20 °C and then washed with 500 μL of ethanol 70%. Samples were centrifuged at 16 000 ***g*** for 15 min, dried for 30 min, and resolubilized in 20 μL of distilled water. Samples were then loaded onto a 0.8% agarose gel for electrophoresis in TAE buffer. DNA was transferred to Hybond-N_+ membrane in alkaline media. Hybridization was first performed with probes labelled with α-[^{32}p]-dCTP and the Mega Prime DNA labelling system according manufacturer (Amersham, UK). Autoradiography was carried out using PhosphorImager. Membranes were then des-hybridized and hybridized again with the gene-specific probe.

Estimation of insert copy and locus number in the library: T-DNA copy and locus number

A precise estimation of the average number of T-DNA copies was provided by the Southern blot analysis of *Sst*I-digested DNA blots of a random set of 384 p4978 T0 plants, hybridized to the *hpt* and *gusA* probes (Figure S1). Additional data on both T-DNA copies and integration loci were generated by the linkage analysis in T1 progenies of 47, 37, and 27 p4978, p4984, and p4956ET15 plants, hybridized at least to the *hpt* probe. *Tos*17 *copy number :* the hybridization of *Xba*I DNA blots of the above-mentioned set of 384 p4978 T0 plants with a reverse transcriptase (RT) probe specific to the Ty1 copia retrotransposon *Tos*17 showed that they contain an average of 3.3 new copies of the element that segregate in an independent manner in T1 progenies (Piffanelli *et al.*, 2007).

FST quality assessment

To assess the quality of the FST data generated from the T-DNA and *Tos*17 inserts in the library, Southern blot analysis was carried out on T2 progeny seedlings resulting from the field seed increase, using a gene-specific probe for 80 and 56 independent T-DNA and *Tos*17 insertions, respectively. Whether the insertion is at the site predicted from FST data, we anticipate the occurrence of a segregating polymorphism because of the integration of the insertion mutagen within the sequence, in the progeny plants. The FSTs were first identified using OryGenesDB (http://orygenesdb.cirad.fr/) using search tool of the Genome Browser. Gene-specific probes primers were designed using the 'primer designer' tool of OryGenesDB with default parameters, except for the restriction enzyme parameter, set to either *Eco*RI, *Xba*I, *Bam*HI, or *Hin*dIII. Gene-specific probes were amplified using Biolabs Taq according manufacturer instructions with the following amplification conditions: 94 °C 1 min, 30 cycles of 94 °C 1 min, 52 °C 1 min, 72 °C 2 min, and a final elongation step of 72 °C 8 min. Probes were gel-purified using Zymoclean Gel DNA recovery kit and eluted in 10 μL Tris–HCl 10 mM. Around 50–100 ng of gel-purified probes were used for labelling and used along *hph* and *Tos*17 specific probes for determining segregation of polymorphisms at the target locus along with those of T-DNA/*Tos*17 inserts.

Phenotypic characterization of the library

T1 progeny screening and seed propagation

T1 seeds harvested on greenhouse-grown T0 plants have been used for specific screenings and seed increase. Observation of phenotypic alterations in grain development and filling has been made first on T1 panicles, and the phenotype was later confirmed on panicles of greenhouse- or field-grown T2 and T3 progeny plants. Five T1 seeds have been used for the detection of GUSA activity and GFP fluorescence on half seeds. Twenty T1 seeds have been used for detecting changes in response to inoculation with *M. oryzae* at the four- to five-leaf stage of plants grown in trays in phytotron. Twenty-five T1 seeds have been shipped to CIAT for field evaluation and seed propagation under agronomic conditions.

Morphophysiological alterations under field conditions

We carried out a multitrait scoring of the entire collection at the CIAT headquarter experimental station (Cali, Colombia). Owing to the very large number of materials to handle, the observations were organized into six successive field trials of about 5000 lines, between 2003 and 2008.

Trait ontology nomenclature

A list of phenotypic traits was established from data mining of several rice phenotypic databases (http://www.gramene.org, http://www.grs.nig.ac.jp/rice/oryzabase, http://www.irri.org/genomics) and was used as a guide for observations (see Table S3). An English-Spanish-French lexical of botanical and agronomic terms was established to facilitate phenotype identification.

Field experimental layout

Two fields of two hectares each were prepared following the requirements of the Colombian Agropecuary Institute, ICA (Instituto Colombiano Agropecuario). Each field was organized in 40 pools of 33 m × 10 m separated by dikes and irrigation channels. The entire surface was covered by nylon nets to avoid any damage or seed dissemination that could be caused by birds. A basic fertilization composed of monoammonium phosphate, iron sulphate, potassium chloride, and microelements was applied. The field was irrigated twice a week.

Screenhouse observations on seedlings

Up to twenty-five T_1 seeds per T_0 plant were sown in the screenhouse. Sowing was carried out in batches of 1250 lines, with about 3-week intervals between the batches. The seeds were heat-treated for 3 days at 50 °C to break dormancy and planted in plastic trays with a mixture of CIAT (67%) and Santander de Quilichao (33 %) soils. The soil characteristics and composition were as follows. CIAT soil: pH 7.39; Organic matter 17.19 g/kg; P-Brayll 71.87 mg/kg; K 0.69 cmol/kg; Ca 13.56 cmol/kg; Mg 7.61 cmol/kg; Na 0.24 cmol/kg; CIC 18.85 cmol/kg; S 64.48 mg/kg; B 0.95 mg/kg; Fe 1.42 mg/kg; Mn 54.61 mg/kg; Cu 0.44 mg/kg; Zn 5.52 mg/kg. Santander soil: pH 6.32; Organic matter 36.81 g/kg; P-Brayll 39.80 mg/kg; K 0.53 cmol/kg; Ca 9.72 cmol/kg; Mg 5.55 cmol/kg; Na 0.14 cmol/kg; CIC 20.40 cmol/kg; S 48.00 mg/kg; B 0.36 mg/kg; Fe 6.76 mg/kg; Mn 49.89 mg/kg; Cu 0.85 mg/kg; Zn 3.68 mg/kg. Twelve lines were planted per tray. Germination rate was determined at 10 days after sowing (DAS). The first phenotypic observations were made at 18–20 DAS.

Field observations

The plantlets were transplanted in the field at 25 DAS. Lines were sown every 50 cm, and plants within a line every 20 cm. Control lines of Nipponbare cv. were planted for each 10 T-DNA lines to facilitate the comparison with wild phenotype. Phenotypic analyses were carried out at different stages of the plant growth, using the list of traits as a guide. A first round of observation was made on 45-day-old plants. A second evaluation was made at flowering stage, and the ultimate observation was made at maturity before harvesting. These three rounds of observation maximized the chances to detect phenotypic variations, as various traits might be observed at only one of these stages. Moreover, this permitted to follow the evolution of a suspected phenotype at early stage and possibly confirm or invalidate it.

Paddy Genes Book database

A local Excel™ database of all data relative to growth conditions, germination, flowering, and phenotypic observations was set up (M. Lorieux, unpublished; http://mapdisto.free.fr). This database, called *Paddy Genes Book*, was mainly used as a working tool to facilitate data entry and compilation. However, it also can be used for data browsing, and several options for searching for lines or traits according different criteria are available. Moreover, the database offers tools for computing basic statistics over traits and lines. Tools for facilitating gene discovery by both forward and reverse genetics approaches were implemented, which allow linking the phenotypic information to the FST positions. Searches using the TO instead of trait codes are allowed. These features permit fast and powerful identification of candidate genes or gene families that can be confirmed later on by line tagging. This database also displays photographs of the mutant phenotypes: more than 40 000 pictures are available.

Specific screen for alterations in grain development and filling

Reporter gene expression: GUS activity was detected on half seeds by the GUS histochemical assay (Jefferson, 1987). GFP fluorescence in living tissue was detected using a Leica MZ FLIII fluorescence stereomicroscope (Leica microsystems GmbH, Wetzlar, Germany) and GFP Plus fluorescence filter set [GFP2, 480-nm excitation filter (bandwidth of 40 nm) and 510-nm barrier filter]. Images were collected using a Leica DC 300F digital camera (Leica microsystems GmbH).

Specific screen for alterations in response to Magnaporthe oryzae inoculation

The inoculation procedure was detailed in the study of Berruyer *et al.*, 2003. We chose a strain that was virulent on the Nipponbare cultivar (*M. oryzae* strain FR13; CIRAD collection—kind gift from D. Tharreau). This strain triggers severe disease characterized by greyish lesions on young plants (<3 weeks old). Approximately 10 000 lines were screened for altered disease response 2 weeks after sowing (three-leaf stage).

Acknowledgements

This work was supported by several grants of the French plant genomics collaborative programme Génoplante (OsCrR1, OsCrGF, Osmu2, B1, and M1 projects) and of the National Research Agency (ANR) (ANR-GNP-05086G CAGRILL project). The technical assistance of Loic Fontaine, Rémy Michel, Christian Chaine, Frédéric Salles, Murielle Boumbou-Portefaix, Rosie Sevilla, Lucette Gracia, Laurent Rosso, Jérôme Veyret, Carole Maisonneuve, Nicolas Cennes, Anne Laure Latrilhe, Sylvie M'Bello, Vanessa Guerin and Véronique Chalvon is greatly acknowledged.

References

Agrawal, G.K., Yamazaki, M., Kobayashi, M., Hirochika, R., Miyao, A. and Hirochika, H. (2001) Screening of the rice viviparous mutants generated by endogenous retrotransposon Tos17 insertion. Tagging of a zeaxan thin epoxidase gene and a novel OsTATC gene. *Plant Physiol.* **125**, 1248–1257.

Bari, R., Datt Pant, B., Stitt, M. and Scheible, W.-R.D. (2006) PHO2, MicroRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiol.* **141**, 988–999.

Berruyer, R., Adreit, H., Milazzo, J., Gaillard, S., Berger, A., Dioh, W., Lebrun, M.H. and Tharreau, D. (2003) Identification and fine mapping of Pi33 the rice resistance gene corresponding to the *Magnaporthe grisea* avirulence gene ACE1. *Theor. Appl. Genet.* **107**, 1139–1147.

Bruce, M., Hess, A., Bai, J., Mauleon, R., Diaz, M.G., Sugiyama, N., Bordeos, A., Wang, G.-L., Leung, H. and Leach, J. (2009) Detection of genomic deletions in rice using oligonucleotide microarrays. *BMC Genomics*, **10**, 129.

Chern, C.-G., Fan, M.-J., Yu, S.-M., Hour, A.-L., Lu, P.-C., Lin, Y.-C., Wei, F.-J., Huang, S.-C., Chen, S., Lai, M.-H., Tseng, C.-S., Yen, H.-M., Jwo, W.-S., Wu, C.-C., Yang, T.-L., Li, L.-S., Kuo, Y.-C., Li, S.-M., Li, C.-P., Wey, C.-K., Trisiriroj, A., Lee, H.-F. and Hsing, Y.-I. (2007) A rice phenomics study—phenotype scoring and seed propagation of a T-DNA insertion-induced rice mutant population. *Plant Mol. Biol.* **65**, 427–438.

Chin, H.G., Choe, M.S., Lee, S.-H., Park, S.H., Park, S.H., Koo, J.C., Kim, N.Y., Lee, J.J., Oh, B.G., Yi, G.H., Kim, S.C., Choi, H.C., Cho, M.J. and Han, C.-D. (1999) Molecular analysis of rice plants harboring an Ac/Ds transposable element-mediated gene trapping system. *Plant J.* **19**, 615–623.

Delteil, A., Blein, M., Faivre-Rampant, O., Guellim, A., Estevan, J., Hirsch, J., Bevitori, R., Michel, C. and Morel, J.-B. (2012) Building a mutant resource for the study of disease resistance in rice reveals the pivotal role of several genes involved in defence. *Mol. Plant Pathol.* **13**, 72–82.

Droc, G., Ruiz, M., Larmande, P., Pereira, A., Piffanelli, P., Morel, J.B., Dievart, A., Courtois, B., Guiderdoni, E. and Périn, C. (2006) OryGenesDB: a database for rice reverse genetics. *Nucleic Acids Res.* **34**, 736–740.

Droc, G., Périn, C., Fromentin, S.B. and Larmande, P. (2009) OryGenesDB 2008 update: database interoperability for functional genomics of rice. *Nucleic Acids Res.* **37**, 992–995.

Fujiwara, T., Maisonneuve, S., Isshiki, M., Mizutani, M., Chen, L., Wong, H.L., Kawasaki, T. and Shimamoto, K. (2010) Sekiguchi lesion gene encodes a cytochrome P450 monooxygenase that catalyzes conversion of tryptamine to serotonin in rice. *J. Biol. Chem.* **285**, 11308–11313.

Hirochika, H. (2001) Contribution of the *Tos*17 retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* **4**, 118–122.

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.

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q. and Xiong, L. (2006) Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc. Natl Acad. Sci. USA*, **103**, 12987–12992.

Iida, S. and Terada, R. (2005) Modification of endogenous natural genes by gene targeting in rice and other higher plants. *Plant Mol. Biol.* **59**, 205–219.

Ikeda, A., Ueguchi-Tanaka, M., Sonoda, Y., Kitano, H., Koshioka, M., Futsuhara, Y., Matsuoka, M. and Yamaguchi, J. (2001) Slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. *Plant Cell*, **13**, 999–1010.

Itaya, K. and Ui, M. (1966) A new micromethod for the colorimetric determination of inorganic phosphate. *Clin. Chim. Acta.*, **14**, 361–366.

Jaiswal, P., Avraham, S., Ilic, K., Kellogg, E.A., McCouch, S., Pujar, A., Reiser, L., Rhee, S.Y., Sachs, M.M., Schaeffer, M., Stein, L., Stevens, P., Vincent, L., Ware, D. and Zapata, F. (2005) Plant ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comp. Funct. Genomics*, **6**, 388–397.

Jefferson, R.A. (1987) Assaying chimeric genes in plants from gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387–405.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.H., Lee, J., Kim, C., Jang, S., Lee, S., Yang, K., Nam, J., An, K., Han, M.J., Sung, R.J., Choi, H.S., Yu, J.H., Choi, J.H., Cho, S.Y., Cha, S.S., Kim, S.I. and An, G. (2000) T-DNA insertional mutagenesis for functional genomics in rice. *Plant J.* **22**, 561–570.

Jeong, D.H., An, S.Y., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K.S. and An, G.H. (2002) T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol.* **130**, 1636–1644.

Jiang, C., Mithani, A., Gan, X., Belfield, E.J., Klingler, J.P., Zhu, J.-K., Ragoussis, J., Mott, R. and Harberd, N.P. (2012) Regenerant arabidopsis lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. *Curr. Biol.* **21**, 1385–1390.

Johnson, A.A.T., Hibberd, J.M., Gay, C., Essah, P.A., Haseloff, J., Tester, M. and Guiderdoni, E. (2005) Spatial control of transgene expression in rice (*Oryza sativa* L.) using the GAL4 enhancer trapping system. *Plant J.* **41**, 779–789.

Jung, K.H., Hur, J., Ryu, C.H., Choi, Y., Chung, Y.Y., Miyao, A., Hirochika, H. and An, G.H. (2003) Characterization of a rice chlorophyll-deficient mutant using the T-DNA gene-trap system. *Plant Cell Physiol.* **44**, 463–472.

Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C.S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z.G., Sundaresan, V. and Ramachandran, S. (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *Plant J.* **37**, 301–314.

Krishnan, A., Guiderdoni, E., An, G., Hsing, Y.-I.C., Han, C.-D., Lee, M.C., Yu, S.-M., Upadhyaya, N., Ramachandran, S., Zhang, Q., Sundaresan, V., Hirochika, H., Leung, H. and Pereira, A. (2009) Mutant resources in rice for functional genomics of the grasses. *Plant Physiol.* **149**, 165–170.

Kumar, C.S., Wing, R.A. and Sundaresan, V. (2005) Efficient insertional mutagenesis in rice using the maize En/Spm elements. *Plant J.* **44**, 879–892.

Larmande, P., Gay, C., Lorieux, M., Perin, C., Bouniol, M., Droc, G., Sallaud, C., Perez, P., Barnola, I., Biderre-Petit, C., Martin, J., Morel, J.B., Johnson, A.A.T., Bourgis, F., Ghesquiere, A., Ruiz, M., Courtois, B. and Guiderdoni, E. (2008) Oryza tag line, a phenotypic mutant database for the genoplante rice insertion line library. *Nucleic Acids Res.* **36**, D1022–D1027.

Lee, S., Kim, J., Son, J.S., Nam, J., Jeong, D.H., Lee, K., Jang, S., Yoo, J., Lee, J., Lee, D.Y., Kang, H.G. and An, G.H. (2003) Systematic reverse genetic screening of T-DNA tagged genes in rice for functional genomic analyses: MADS-box genes as a test case. *Plant Cell Physiol.* **44**, 1403–1411.

Lee, S., Jung, K.H., An, G.H. and Chung, Y.Y. (2004a) Isolation and characterization of a rice cysteine protease gene, OSCP1, using T-DNA gene-trap system. *Plant Mol. Biol.* **54**, 755–765.

Lee, S.C., Huh, K.W., An, K., An, G. and Kim, S.R. (2004b) Ectopic expression of a cold-inducible transcription factor, CBF1/DREB1b, in transgenic rice (*Oryza sativa* L.). *Mol. Cells*, **18**, 107–114.

Li, J., Jiang, D., Zhou, H., Li, F., Yang, J., Hong, L., Fu, X., Li, Z., Liu, Z., Li, J. and Zhuang, C. (2011) Expression of RNA-interference/antisense transgenes by the cognate promoters of target genes is a better gene-silencing strategy to study gene functions in rice. *PLoS One*, **6**, e17444.

Lorrain, S., Vailleau, F., Balagué, C. and Roby, D. (2003) Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants? *Trends Plant Sci.* **8**, 263–271.

Matsumoto, T., Wu, J.Z., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., Mizuno, H., Yamamoto, K., Antonio, B.A., Baba, T., Sakata, K., Nagamura, Y., Aoki, H., Arikawa, K., Arita, K., Bito, T., Chiden, Y., Fujitsuka, 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, J.Y., 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, H.S., Yano, M., Sasaki, T., Yuan, Q.P., Shu, O.T., Liu, J., Jones, K.M., Gansberger, K., Moffat, K., Hill, J., Bera, J., Fadrosh, D., Jin, S.H., Johri, S., Kim, M., Overton, L., Reardon, M., Tsitrin, T., Vuong, H., Weaver, B., Ciecko, A., Tallon, L., Jackson, J., Pai, G., Van Aken, S., Utterback, T., Reidmuller, S., Feldblyum, T., Hsiao, J., Zismann, V., Iobst, S., de Vazeille, A.R., Buell, C.R., Ying, K., Li, Y., Lu, T.T., Huang, Y.C., Zhao, Q., Feng, Q., Zhang, L., Zhu, J.J., Weng, Q.J., Mu, J., Lu, Y.Q., Fan, D.L., Liu, Y.L., Guan, J.P., Zhang, Y.J., Yu, S.L., Liu, X.H., Zhang, Y., Hong, G.F., Han, B., Choisne, 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, H.R., Rambo, T., Currie, J., Collura, K., Luo, M.Z., Yang, T.J., Ammiraju, J.S.S., Engler, F., Soderlund, C., Wing, R.A., Palmer, L.E., de la Bastide, M., Spiegel, L., Nascimento, L., Zutavern, T., O'Shaughnessy, A., Dike, S., Dedhia, N., Preston, R., Balija, V., McCombie, W.R., Chow, T.Y., Chen, H.H., Chung, M.C., Chen, C.S., Shaw, J.F., Wu, H.P., Hsiao, K.J., Chao, Y.T., Chu, M.K., Cheng, C.H., Hour, A.L., Lee, P.F., Lin, S.J., Lin, Y.C., Liou, J.Y., Liu, S.M., Hsing, Y.I., Raghuvanshi, S., Mohanty, A., Bharti, A.K., Gaur, A., Gupta, V., Kumar, D., Ravi, V., Vij, S., Kapur, A., Khurana, P., Khurana, P., Khurana, J.P., Tyagi, A.K., Gaikwad, K., Singh, A., Dalal, V., Srivastava, S., Dixit, A., Pal, A.K., Ghazi, I.A., Yadav, M., Pandit, A., Bhargava, A., Sureshbabu, K., Batra, K., Sharma, T.R., Mohapatra, T., Singh, N.K., Messing, J., Nelson, A.B., Fuks, G., Kavchok, S., Keizer, G., Llaca, E.L.V., Song, R.T., Tanyolac, B., Young, S., Il, K.H., Hahn, J.H., Sangsakoo, G., Vanavichit, A., de Mattos, L.A.T., Zimmer, P.D., Malone, G., Dellagostin, O., de Oliveira, A.C., Bevan, M., Bancroft, I., Minx, P., Cordum, H., Wilson, R., Cheng, Z.K., Jin, W.W., Jiang, J.M., Leong, S.A., 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., Bruskiewich, R., Bureau, T., Miyao, A., Hirochika, H., Nishikawa, T., Kadowaki, K., Sugiura, M. and Project I.R.G.S. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K. and 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.

Miyao, A., Iwasaki, Y., Kitano, H., Itoh, J., 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.

Miyao, A., Nakagome, M., Ohnuma, T., Yamagata, H., Kanamori, H., Katayose, Y., Takahashi, A., Matsumoto, T. and Hirochika, H. (2012) Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant Cell Physiol.* **53**, 256–264.

Mori, M., Tomita, C., Sugimoto, K., Hasegawa, M., Hayashi, N., Dubouzet, J., Ochiai, H., Sekimoto, H., Hirochika, H. and Kikuchi, S. (2007) Isolation and molecular characterization of a *Spotted leaf 18* mutant by modified activation-tagging in rice. *Plant Mol. Biol.* **63**, 847–860.

Nonomura, K.I., Miyoshi, K., Eiguchi, M., Suzuki, T., Miyao, A., Hirochika, H. and Kurata, N. (2003) The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *Plant Cell*, **15**, 1728–1739.

Nonomura, K.L., Nakano, M., Fukuda, T., Eiguchi, M., Miyao, A., Hirochika, H. and Kurata, N. (2004) The novel gene *HOMOLOGOUS PAIRING ABERRATION IN RICE MEIOSIS1* of rice encodes a putative coiled-coil protein required for homologous chromosome pairing in meiosis. *Plant Cell*, **16**, 1008–1020.

O'Malley, R.C. and Ecker, J.R. (2010) Linking genotype to phenotype using the Arabidopsis unimutant collection. *Plant J.* **61**, 928–940.

Parker, D., Beckmann, M., Zubair, H., Enot, D.P., Caracuel-Rios, Z., Overy, D.P., Snowdon, S., Talbot, N.J. and Draper, J. (2009) Metabolomic analysis reveals a common pattern of metabolic re-programming during invasion of three host plant species by *Magnaporthe grisea*. *Plant J.* **59**, 723–737.

Piffanelli, P., Droc, G., Mieulet, D., Lanau, N., Bès, M., Bourgeois, E., Rouvière, C., Gavory, F., Cruaud, C., Ghesquière, A. and Guiderdoni, E. (2007) Large-scale characterization of *Tos*17 insertion sites in a rice T-DNA mutant library. *Plant Mol. Biol.* **65**, 587–601.

Ribot, C., Hirsch, J., Balzergue, S., Tharreau, D., Nottéghem, J.-L., Lebrun, M.-H. and Morel, J.-B. (2008) Susceptibility of rice to the blast fungus, *Magnaporthe grisea*. *J. Plant Physiol.* **165**, 114–124.

Sabot, F., Picault, N., El-Baidouri, M., Llauro, C., Chaparro, C., Piegu, B., Roulin, A., Guiderdoni, E., Delabastide, M., McCombie, R. and Panaud, O. (2011) Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. *Plant J.* **66**, 241–246.

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

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, J.C., Delseny, M. and Guiderdoni, E. (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. *Plant J.* **39**, 450–464.

Sukekiyo, Y. and Kimura, Y. (1991) Somaclonal variation in protoplast-derived rice plants. *Gamma Field Symposia*, **30**, 43–58.

Till, B., Cooper, J., Tai, T., Colowit, P., Greene, E., Henikoff, S. and Comai, L. (2007) Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol.* **7**, 19.

Wang, X.E., Basnayake, B.M.V.S., Zhang, H., Li, G., Li, W., Virk, N., Mengiste, T. and Song, F. (2009) The arabidopsis ATAF1, a NAC transcription factor, is a negative regulator of defense responses against necrotrophic fungal and bacterial pathogens. *Mol. Plant Microbe Interact.* **22**, 1227–1238.

Warthmann, N., Chen, H., Ossowski, S., Weigel, D. and Hervé, P. (2008) Highly specific gene silencing by artificial miRNAs in rice. *PLoS One*, **3**, e1829.

Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M.B., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D. and Stoutjesdijk, P.A. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* **27**, 581–590.

Wu, J.L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M.R., Ramos-Pamplona, M., Mauleon, R., Portugal, A. and Ulat, V.J. (2005) Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol. Biol.* **59**, 85–97.

Yamanouchi, U., Yano, M., Lin, H., Ashikari, M. and Yamada, K. (2002) A rice spotted leaf gene, Spl7, encodes a heat stress transcription factor protein. *Proc. Natl Acad. Sci. USA*, **99**, 7530–7535.

Zhang, J., Li, C., Wu, C., Xiong, L., Chen, G., Zhang, Q. and Wang, S. (2006) RMD: a rice mutant database for functional analysis of the rice genome. *Nucleic Acids Res.* **34**, D745–D748.

Zhang, Q., Li, J., Xue, Y., Han, B. and Deng, X.W. (2008) Rice 2020: a call for an international coordinated effort in rice functional genomics. *Mol. Plant*, **1**, 715–719.

Zuryn, S., Le Gras, S.P., Jamet, K. and Jarriault, S. (2010) A strategy for direct mapping and identification of mutations by whole-genome sequencing. *Genetics*, **186**, 427–430.

Supporting information

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

Figure S1 T-DNA organization in p4978 lines.
Figure S2 Assessment of the quality of the library by Southern blot analysis: example of MSP1.
Figure S3 Distribution of the number of altered traits in variant lines field evaluated in CIAT, Colombia.
Figure S4 Examples of lesion mimic phenotypes in the OTL mutant collection.
Figure S5 The AGGC08 line triggers lesions upon *Magnaporthe oryzae* infection.
Table S1 Distribution of the distinct T-DNA and *Tos*17 insertion loci of the OTL library among the 12 rice chromosomes and number and % of inserts in the coding sequence compartment of predicted genes residing on these chromosomes.
Table S2 Correspondence between categories and sub categories of traits observed in previous field evaluations of Nipponbare NIAS *Tos*17 lines (Miyao *et al.*, 2007) and Tainung67 TRIM T-DNA activation tagging lines (Chern *et al.*, 2007) with those of this study.
Table S3 Frequency of occurrence of the altered traits in the OTL library according to the trait ontology categories.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.