the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) **85,** 648–659 doi: 10.1111/tpj.13132

Somaclonal variation does not preclude the use of rice transformants for genetic screening

Fu-Jin Wei[1,2], Lin-Yun Kuang[3], Hui-Min Oung[1,†], Sin-Yuan Cheng[1], Hshin-Ping Wu[1], Lin-Tzu Huang[1], Yi-Tzu Tseng[1,4], Wan-Yi Chiou[1,‡], Vicki Hsieh-Feng[2], Cheng-Han Chung[1,§], Su-May Yu[5], Lan-Ying Lee[6], Stanton B. Gelvin[6] and Yue-le C. Hsing[1,*]

[1]*Institute of Plant and Microbial Biology, Academia Sinica, No. 128, Section 2, Yien-Chu-Yuan Road, Taipei 11529, Taiwan,*
[2]*Department of Agronomy, National Taiwan University, No. 1, Section 4, Roosevelt Rd, Taipei 10617, Taiwan,*
[3]*Transgenic Plant Core Facility, Academia Sinica, No. 128, Section 2, Yien-Chu-Yuan Road, Taipei 11529, Taiwan,*
[4]*Institute of Plant Biology, National Taiwan University, No. 1, Section 4, Roosevelt Rd, Taipei 10617, Taiwan,*
[5]*Institute of Molecular Biology, Academia Sinica, No. 128, Section 2, Yien-Chu-Yuan Road, Taipei 11529, Taiwan,*
[6]*Department of Biological Sciences, Purdue University, 201 South University St., West Lafayette, IN, 47907-1392, USA*

Received 21 October 2015; revised 21 December 2015; accepted 20 January 2016; published online 2 February 2016.
*For correspondence (e-mail bohsing@gate.sinica.edu.tw).
† Present address: Institute of Biological Chemistry, Washington State University, PO Box 646430,Pullman, WA, 99164-6430, USA.
‡ Present address: Graduate School of Environmental and Life Science, Okayama University, 3-1-1 Tsushima-naka kita-ku,Okayama, 700-8530, Japan.
§ Present address: Department of Microbiology and Immunology, Drexel University, Philadelphia, PA, 19129, USA.

SUMMARY

Rice (*Oryza sativa*) is one of the world's most important crops. Rice researchers make extensive use of insertional mutants for the study of gene function. Approximately half a million flanking sequence tags from rice insertional mutant libraries are publicly available. However, the relationship between genotype and phenotype is very weak. Transgenic plant assays have been used frequently for complementation, overexpression or antisense analysis, but sequence changes caused by callus growth, *Agrobacterium* incubation medium, virulence genes, transformation and selection conditions are unknown. We used high-throughput sequencing of DNA from rice lines derived from Tainung 67 to analyze non-transformed and transgenic rice plants for mutations caused by these parameters. For comparison, we also analyzed sequence changes for two additional rice varieties and four T-DNA tagged transformants from the Taiwan Rice Insertional Mutant resource. We identified single-nucleotide polymorphisms, small indels, large deletions, chromosome doubling and chromosome translocations in these lines. Using standard rice regeneration/transformation procedures, the mutation rates of regenerants and transformants were relatively low, with no significant differences among eight tested treatments in the Tainung 67 background and in the cultivars Taikeng 9 and IR64. Thus, we could not conclusively detect sequence changes resulting from *Agrobacterium*-mediated transformation in addition to those caused by tissue culture-induced somaclonal variation. However, the mutation frequencies within the two publically available tagged mutant populations, including TRIM transformants or *Tos17* lines, were about 10-fold higher than the frequency of standard transformants, probably because mass production of embryogenic calli and longer callus growth periods were required to generate these large libraries.

Keywords: *Agrobacterium*-mediated transformation, indels, regeneration, rice, single-nucleotide polymorphisms, somaclonal variation.

INTRODUCTION

Rice (*Oryza sativa*) is one of the world's most important crops and is the principal food of nearly 50% of the world's population. Because cereal crops share a large degree of synteny, rice is an excellent model crop for cereal genomics research (Gale and Devos, 1998). With the rising world population, feeding people in a more sustainable and environmentally friendly way is becoming increasingly important. To fulfill this important responsibility, rice researchers need to share resources to better understand the functions of rice genes, especially those encoding important agronomic traits.

With the complete genomic sequencing of rice (IRGSP, 2005), the challenge of the post-genomic era is to analyze systematically the functions of all rice genes. An important approach to defining the function of a novel gene is to abolish or activate its expression. Insertional mutagenesis,

with T-DNA (Jeong *et al.*, 2002; Sallaud *et al.*, 2004; Johnson *et al.*, 2005; Hsing *et al.*, 2007; Wan *et al.*, 2009) or transposable elements such as *Tos17* (Miyao *et al.*, 2003), *Ac*/*Ds* (Kim *et al.*, 2004; Upadhyaya *et al.*, 2006; He *et al.*, 2007; Jiang *et al.*, 2007) or *Spm* (Kumar *et al.*, 2005), is one method to determine gene function and to isolate the target gene responsible for a specific phenotype. Many research groups have established rice insertional mutant resources and provided flanking sequence tag (FST) information for these lines. In mid-2015, about 450 000 integration sites were available in public databases – RiceGE (http://signal.salk.edu/RiceGE/) and OrygenesDB (http://orygenesdb.cirad.fr). Several recent papers (Droc *et al.*, 2013; Wang *et al.*, 2013; Yang *et al.*, 2013) have reviewed these rice mutant resources and their applications.

Unfortunately, these rice insertional mutant libraries have a low tagging efficiency; that is, the relationship between genotype (disrupted or activated gene) and phenotype is very weak. For forward genetic screens, the estimated tagging efficiency is 5–10% for the *Tos17*- and T-DNA-tagged populations (e.g. Droc *et al.*, 2013; Wei *et al.*, 2013). Because tissue culture-free transformation protocols were used to generate Arabidopsis T-DNA–tagged insertion populations, forward genetic screens have a higher tagging efficiency, estimated at about 35% (Budziszewski *et al.*, 2001; McElver *et al.*, 2001). In both of these species, many observed phenotypes are not caused by the integration of T-DNA or the transposon, so other sequence or epigenetic changes must have caused the phenotypes.

Culturing tissues *in vitro* can induce mutations (somaclonal variation; Larkin and Scowcroft, 1981). The *in vitro* growth of rice calli may induce movement of endogenous transposons (Hirochika *et al.*, 1996; Sabot *et al.*, 2011; Miyao *et al.*, 2012; Stroud *et al.*, 2013). Somaclonal variation, resulting from a sum of genetic and epigenetic changes, might occur during callus induction, growth, regeneration or (as appropriate) *Agrobacterium*-mediated transformation.

Insertional mutant resources for rice have been used extensively, so we used next-generation sequencing (NGS) to examine sequence changes in regenerants of the non-transformed rice variety Tainung 67 (TNG67; hereafter termed 'regenerants') and transformants of three rice varieties, as well as several transformed TNG67 mutant lines from the Taiwan Rice Insertional Mutant (TRIM) collection (Hsing *et al.*, 2007). We also monitored the ploidy and sequence changes of several other TRIM lines and compared the results of regeneration and transformation. We found that mutations, especially large deletions, were more prevalent among the T-DNA insertion mutation lines in the TRIM collection than among more recently made regenerants and transformants. Thus, callus growth and *Agrobacterium*-mediated transformation with a 'standard' transformation protocol (i.e. a protocol, such as the one described in Experimental Procedures in Supporting Information, that allows regeneration soon after callus induction) can generate some sequence changes but at a much reduced frequency from that seen in tagged mutant populations such as TRIM or *Tos17* lines.

RESULTS

Rice genome resequencing and detection of sequence changes

Using the Illumina HiSeq platform we resequenced the genome of the rice variety TNG67 and the genomes of three to four lines each of plants regenerated from rice calli, in the R_0 or T_0 generation, treated in one of eight ways (Table 1). In treatments A and C, the sequence changes revealed somaclonal variation induced by callus tissue culture and regeneration regimes. In treatment B, calli were co-cultivated with *Agrobacterium tumefaciens* EHA105 (Hood *et al.*, 1993), which lacks T-DNA but can transfer virulence effector proteins to plants. In treatments D and E, calli were co-cultivated with the avirulent strain *A. tumefaciens* A136 (Montoya *et al.*, 1978) and thus revealed genome changes caused by incubation of calli with *Agrobacterium* that could not transfer T-DNA or virulence effector proteins. However, these strains could induce a pathogen-associated molecular pattern (PAMP) response (Zipfel *et al.*, 2006). In treatments F and G, rice calli were co-cultivated with *A. tumefaciens* EHA105 containing the T-DNA binary vector pCambia 1305.1 (http://www.cambia.org, hygromycin selection) or pE3730 (bialaphos selection), and thus the plants recovered were transformants. In condition H, calli were bombarded with the plasmid pCambia 1305.1. Although the treatments were different, we maintained the cultures for the same relatively short time in the callus state. We also prepared transformants using Taikeng 9 (TK9), a *japonica* variety, and IR64, an *indica* variety. We additionally sequenced the genomes of four mutant TRIM line transformants (M0048349, M0053677, M0079651 and M0084311) at the T_2 generation. These four TRIM lines are tillering dwarf mutants, with a plant height of less than half that of control plants and tiller numbers at least three times higher than that of control plants (Figure S1 in the Supporting Information). We obtained paired-end sequence data averaging 13- to 26-fold coverage depth (Table S1). Coverage for each genome ranged from 94% to 98% for *japonica* rice and 90% to 92% for *indica* rice (Table S1). For the wild-type TNG67 we also sequenced the genome of plants from two subsequent generations, TNG67_1 and TNG67_2, and checked for nucleotide changes between them.

For the non-transformed regenerants in treatments A–E, we used the R_0 generation and thus detected only heterozygous mutations relative to their wild-type TNG67 progenitor. For the transformants in treatments F–H we

Table 1 Details of the treatments. Treatments A–H as well as IR64 and Taikeng 9 (TK9), the *Agrobacterium* strain, Ti plasmid, binary vector and selection marker used, and the single nucleotide polymorphism (SNP) rate

Treatment	Plants[a]	*Agrobacterium*[b]	Antibiotics[c]	Ti plasmid derivative	Binary vector	Selection	Reporter	SNPs[d]
A	3		–					107.3a[e]
B	4	EHA105	+	pTiBo542				110.2a
C	4		+					109.5a
D	3	At2	+					120.7a
E	3	At2	+		pCAMBIA1305.1			206.0a
F	3	EHA105	+	pTiBo542	pCAMBIA1301	Hygromycin	GUS	188.0a
G	3	EHA105	+	pTiBo542	pE3730	Bialaphos	Venus	139.3a
H	3		+		pCAMBIA1305.1 + bombardment	Hygromycin	GUS	180.0a
IR64	3	EHA105	+	pTiBo542	pCAMBIA1380	Hygromycin	GUS	182.3a
TK9	2	EHA105	+	pTiBo542	pCAMBIA1301	Hygromycin	GUS	126.0a

[a]Number of regenerants or transformants sequenced.
[b]Agrobacterium strain used.
[c]Antibiotic (cefotaxime) used after *Agrobacterium* co-cultivation.
[d]Number indicates average number of SNPs per treatment.
[e]Significance level by Duncan's new multiple range test (Duncan 1955).

used the T_0 generation as well as TK9 and IR64, and thus also detected heterozygous mutations. In this generation, the sequence changes [i.e. single-nucleotide polymorphisms (SNPs) or indels] occurring during callus growth, co-cultivation or regeneration stages should be present in about 50% of the reads in the R_0/T_0 plant. Any homozygous sequence changes differing from that of their progenitor at this generation should have been present before calli were induced. For the T_2 generation TRIM lines, we detected homozygous mutations relative to their wild-type progenitor. The expected homozygosity frequency is $1/4 + (1/2 \times 1/4) = 3/8$, so we multiplied the number of homozygous changes by 8/3 to obtain the initial mutation rates in the T_0 generation.

We analyzed sequence coverage and performed manual curation to detect sequence changes in the NGS data. Because we used the variety TNG67 to generate the mutant lines, we first sequenced TNG67 plants at two subsequent generations. Their FASTQ files are available in the NCBI Archives [accession numbers SRR1525794 (TNG67_1) and SRR1531575 (TNG67_2)]. Although the sequencing depth differed between the datasets, we could still identify all possible SNPs or indels in the overlapping non-highly repetitive regions of the genome (approximately 85% of the total genome). We could not identify any DNA polymorphisms in this region.

To identify heterozygous sequence changes in the genomes, the mismatch frequency of the reads covering the variant site should be between 0.3 and 0.7 to avoid false-negative results caused by sequencing errors. To identify homozygous sequence changes in the genomes, the mismatch frequency should be >0.8 to avoid alterations caused by sequencing errors. We performed filtering of SNPs and indels (Data S1) with the code provided (Data S2), followed by confirmation (Data S3). Figure 1 illustrates the Integrative Genomics Viewer (IGV) histogram at the 6576-kbp region of chromosome 4 of TNG67 (a), M0048349-T2 (b) and M0079651-T2 (c). The Nipponbare RefSeq is used as the coordinate, with the NGS readings of TNG67 and mutants aligned. The sequence depth of TNG67 is 2- to 2.5-fold that of the mutants. In addition, one of the mutants (M0048349-T2, b) may also serve as one of the internal standards to show that a homozygous SNP is present in the mutant M0079651-T2 (c). Similarly, indels that occurred in the NGS data were detected. The SNPs and indels were then confirmed by derived cleaved amplified polymorphic sequences (dCAPS) or PCR. Table S2 lists the primers used. Five SNPs and five indels were chosen randomly and 10 primer pairs were tested. All mutations were correctly identified. Table S3 lists the chromosomal location and sequencing depth of these sequence changes, with Figure 1 showing the IGV view of dCAPS_2 and Figure S2 showing that of Indel_2. Figure S3 illustrates the gel analysis for four SNPs and four indels.

The heterozygous sequence changes for the 25 plants with treatments A–H and the transformants from two varieties other than TNG67 are enumerated in Tables 1 and 2. Table 1 shows the average for three to four replicates, and Table 2 shows the detailed information for each plant, including the homozygous and heterozygous SNPs and indels. With multiple samples for each treatment, we used Duncan's new multiple range test to reveal any effects with different treatments. Although the trend shows relatively higher SNP rates for the three types of transformant (treatments F–H) and TK9 and IR64 transformants versus the five types of regenerants (treatments A–E), statistical analysis revealed no significant difference among any of the treatment conditions. Thus, the estimated mutation rate in these regenerants and transformants, including SNPs and

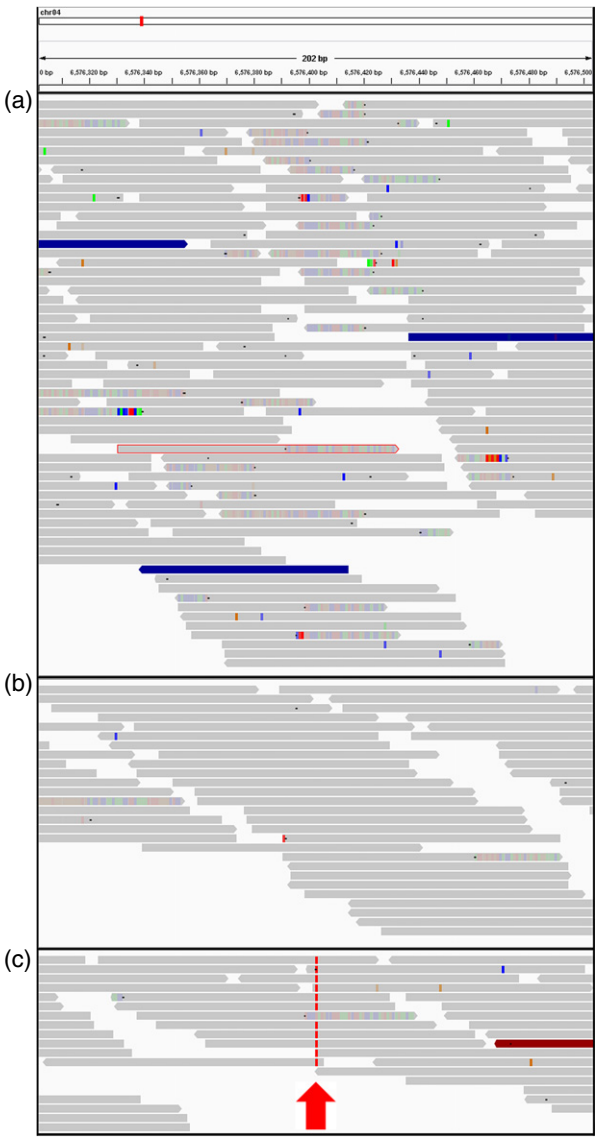

Figure 1. Examples of single nucleotide polymorphism identification revealed by an Integrative Genomics Viewer (IGV) image.
The derived cleaved amplified polymorphic sequence (dCAPS_2), i.e. the 6576-kbp region of chromosome 4, is shown: (a) TNG67, (b) M48349-T2, (c) M79651-T2. Alignments toward the Nipponbare IRGSP 1.0 (top view) are represented by the gray polygons and the mismatched reads by red or blue polygons. Red bar, T; green bar, A; orange bar, G; blue bar, C.

indels, was 107 to 206/373 128 865 × 83% × 2 = (1.7–3.3) × 10^{-7} per site per diploid genome per generation, with 373 128 865 being the size of the reference genome in base pairs for Nipponbare IRGSP 1.0 (Kawahara *et al.*, 2013) and 83% the average genome coverage for these lines.

When looking for homozygous sequence changes in the genomes of the four TRIM T_2-generation transgenic plants, the mismatch frequency should be greater than 0.8 to avoid alterations caused by sequencing errors. By filtering SNPs and indels, followed by confirmation, the final SNP numbers were 246, 433, 314 and 206 and the indels were

Table 2 Sequence changes in the lines tested. The heterozygous and homozygous single nucleotide polymorphisms (SNPs) and indels, as well as T-DNA integration site for plants with treatments A–H and two other rice varieties

Plant ID	Heterozygous			Homozygous SNP	Integration site
	SNP	Indel	SNP + indel		
A1	86	12	98	51	0
A2	67	17	84	51	0
A3	128	12	140	41	0
B1	74	11	85	55	0
B2	69	18	87	55	0
B3	103	14	117	40	0
B4	139	9	148	66	0
C1	102	22	124	51	0
C2	109	14	123	61	0
C3	103	26	129	37	0
C4	53	7	60	50	0
D1	138	34	172	48	0
D2	36	18	54	46	0
D3	114	21	135	40	0
E1	90	32	122	63	0
E2	176	33	209	49	0
E3	236	50	286	51	0
F1	146	32	178	53	1
F2	134	27	161	53	1
F3	191	34	225	45	1
G1	80	22	102	41	2
G2	98	24	122	42	7
G3	158	35	193	52	1
H1	223	41	264	45	4
H2	131	36	167	47	7
H3	95	14	109	47	8
TK9 1	101	25	126	–	2
TK9 2	142	69	211	–	2
IR64 1	160	22	182	–	2
IR64 2	134	14	148	–	1
IR64 3	195	22	217	–	1

42, 53, 71 and 55 for the TRIM lines M0048349, M0053677, M0079651 and M0084311, respectively (Table S4).

We then checked the intersections of the SNPs and indels among these four TRIM lines. We found 36 SNPs and 12 indels shared by M0079651 and M0084311 and one SNP shared by M0053677, M0079651 and M0084311. We used seeds harvested from the Taiwan Agriculture Research Institute (TARI) field for callus induction. The sub-cultured callus pieces were then randomly used to generate TRIM lines over the last 10 years, and some may share the same genetic background. These 48 (36 + 12) sequence changes must have existed in the seeds used for transformation, and thus they were not used for estimating sequence changes in these transformants. Taken together, the T_2 generations of M0048349, M0053677, M0079651 and M0084311 feature 288 (244 + 42), 485 (432 + 53), 336 (277 + 59) and 212 (169 + 43) homozygous sequence changes, respectively, for a mean of 330.3 ± 57.6 sequence changes per line. To estimate the sequence changes

Table 3 Features of vector-DNA/plant DNA junctions

	Simple ligation	Microhomology	Filler sequence
T-DNA end[a]			
Left T-DNA border	3	7	3
Right T-DNA border	3	2	0
Intact T-DNA ends	4	5	1
Processed T-DNA ends	4	2	4
Subtotal	14	16	8
Vector DNA[b]			
Processed vector DNA	13	10	4

[a]The T-DNA ends of transgenic plants from treatments F and G and varieties TK9 and IR64.
[b]The vector DNA ends of transgenic plants from treatment H.

occurring in the T_0 generation, these numbers were then multiplied by the corresponding factor (8/3), yielding about 550–1300, with a mutation frequency of (8.9–21.0) $\times$ 10^{-7} per site per diploid genome. Thus, the mutation frequency in the four TRIM transformants was 3- to 10-fold higher than that of the regenerants (treatments A–E) and transformants (treatments F–H, and the two other varieties).

Integration of T-DNA into the rice genome

We sequenced *Agrobacterium*-mediated transformants from treatments F and G, the TRIM population (all were in the TNG67 background) and two other rice varieties (TK9 and IR64) for a total of 15 plants. The average copy number of integrated T-DNAs was 1.67 copies per transformed plant (range one to seven copies; Table 2). Although one plant from *Agrobacterium* treatment G contained seven T-DNA copies, particle bombardment (treatment H) generally yielded the highest copy number of transgenes. All other *Agrobacterium*-derived transformants had one or two T-DNA insertions.

Detailed analyses of T-DNA/T-DNA and T-DNA/genomic DNA junctions in transgenic plants have been conducted with dicotyledonous plants such as tobacco (Gheysen *et al.*, 1987), aspen (Kumar and Fladung, 2002), cotton (Zhang *et al.*, 2008), tomato (Thomas and Jones, 2007), petunia (Cluster *et al.*, 1996) and Arabidopsis (Brunaud *et al.*, 2002; Forsbach *et al.*, 2003; Windels *et al.*, 2003; Singer *et al.*, 2012) and monocotyledonous plants such as rice (Kim *et al.*, 2003). These results illustrated simple ligation, microhomology and/or filler DNA at the left or right border junctions. Using the whole-genome sequencing data we also detected similar features in the transgenic plants we produced. For instance, Table 3 indicates that at the left border (LB), seven events had homologous sequences, three had simple ligation and three had filler DNA at the junction position. In addition, the three transgenic plants bombarded with the plasmid pCAMBIA1305.1 (treatment H) also contained simple ligation, microhomology and filler sequences, with a ratio similar to that of *Agrobacterium*-mediated transformants.

In addition, we identified translocations at or near the T-DNA integration site for some T-DNA transformants. Figure S4 illustrates a reciprocal translocation in IR64 transformant 1. Part of the T-DNA integrated into a 28.5-Mbp region of chromosome 1 then linked to a 20-Mbp region of chromosome 8. Only three-quarters of the T-DNA backbone, missing both the right border (RB) and LB, was inserted into the rice genome. In addition, 7- and 17-bp deletions occurred at the two translocation junction sites. The translocation event was confirmed by PCR (Figure S4b) followed by capillary sequencing. The translocation event is heterozygous in this R_0 plant and thus has less PCR product (compare lanes 1 and 2 and lanes 3 and 4). The translocations from chromosome 1 to 8 (lane 6) and then from chromosome 8 to T-DNA (lane 8), followed by from T-DNA to chromosome 1 (lane 10), are present in the transformant only but not in wild-type IR64 plants (lanes 5, 7 and 9). We found two translocation events in the materials we sequenced, one in an IR64 transformant and one in a plant subject to treatment G. The former had two integrated T-DNA loci and the latter had seven.

For treatment H, pCAMBIA1305 was bombarded into embryogenic callus in the absence of *Agrobacterium*. The entire 11-kbp plasmid DNA was integrated into the H2 plant, as shown by the IGV histogram with the 'view as pair' mode (Figure 2). The two dark red polygons (each polygon represents an Illumina read) linked with a dark red line indicate paired reads with a larger insertion size than threshold length. In the genome of H2, the polygon at the very right end linked with the very left end is correct because we drew a linear genome to represent a circular vector (Figure 2b, top lines). In addition, none of the gray polygons are interrupted, so the whole plasmid was integrated into the plant genome. However, the genome of H1 and H3 plants shows truncated plasmid DNA with different lengths of deletions or inverted repeat portions of the bombarded DNA (Figure 2a,c). The truncated regions are random and do not relate to RB or LB sequences. In addition, the H plants contained multiple copies of integrated transgene DNA (four for plant H1, seven for plant H2 and eight for plant H3). One event in plant H2 and two in plant H3 indicate chromosome translocations.

We also observed integration of plasmid DNA into mitochondrial DNA in plant H2 (cases A and B, Figure S4c). The junctions of the integration were confirmed by PCR and capillary sequencing. However, we found one end of the integration, from mitochondrial DNA to plasmid DNA (case A) and from chromosome 10 to mitochondria (case B), but failed to find the other end. Our inability

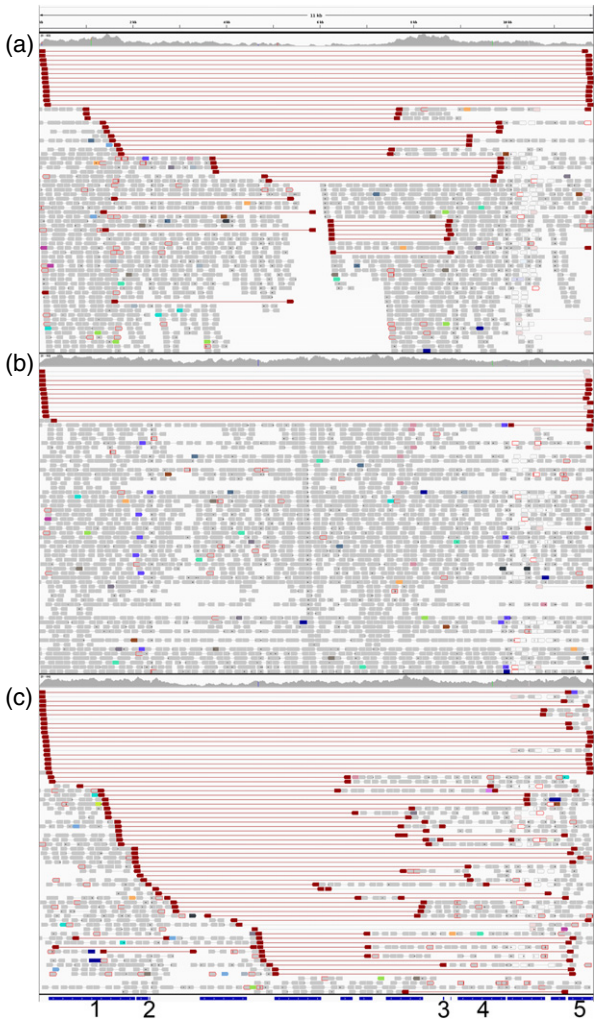

Figure 2. Paired reads of integrated DNA.
The paired reads of the integrated DNA into the rice genome of three plants subject to treatment H: (a) H1 T_0 plant, (b) H2 T_0 plant, (c) H3 T_0 plant. A 12-kbp region of pCAMBIA1305.1 DNA is shown. Alignments toward the pCAMBIA DNA (top row) are represented as gray polygons. The Integrative Genomics Viewer (IGV) histogram in 'view as pair' mode is shown. For the bottom row, the dark blue bars indicate annotated genes in pCAMBIA1305.1. 1, β-glucuronidase (GUS) gene; 2, right border (RB); 3, left border (LB); 4, hygromycin^r gene (Hyg); 5, 35S promoter.

to identify the other junction of the insertion may result from lack of reading depth (22×). Such one-end integration was also found in cases C and D. We found plasmid DNA integrated into chromosome 10 (case C) and 7 (case D). This result was confirmed by PCR and sequencing (Table S2), but we did not find the other end. We did not observe integration into the chloroplast genome in these three H treatment plants.

Analysis of SNPs and indels in the regenerants and transformants

We plotted the chromosomal locations of the 1420 curated SNPs and small indels in four TRIM lines (Figure 3). Similar plots for the 17 regenerants (treatments A–E, Figure S5a), six transformants (treatments F and G, Figure S5b), the three transformants with bombardment (treatment H, Figure S5c) and two other varieties (Figure S5d) are also illustrated. The sequence changes were distributed among all 12 rice chromosomes. Both SNPs (gray bars) and indels (red bars) were randomly distributed and were not related to T-DNA integration sites (green arrows). Of note, the SNPs and indels were present in heterochromatin, such as in peri-centromeric regions. All classes of base changes were detected (Figure S6). The ratio of transitions to transversions was 1.2:1, the most frequent being C/G to T/A and C/G to A/T among transversions. The SNPs or indels preferentially occurred at G or C (Figure S7), whereas the frequency of the G/C ratio for the nearby regions was similar to that of the whole genome (about 43%; IRGSP, 2005). We also checked the frequency of SNP occurrence and the nearby sequences. About 6.5% of the SNPs occurred as two consecutive mutations, and about 84% of these consecutive mutations had identical 5′ and 3′ sides (Table S5). For the SNPs that occurred as singletons, about half were similar to the adjacent base, on the 5′ or 3′ side.

Figure S8 illustrates the distribution of the 197 unique indels in the genomes of the four TRIM transformants. These indels ranged from 40-bp deletions to 16-bp insertions, with the peak at a 1-bp deletion. In total, 32 of the indels occurred as dimer or trimer simple sequence repeats (SSRs); 16 represented deletion of SSRs and 16 insertion of SSRs (Table S6). The SSR copy number changes were 6, 11, 5 and 10 for M0048349, M0053677, M0079651 and M0084311, respectively. The largest change was a deletion of 20 copies of a TA dimer in the genome of M0084311. The locations, types (deletion or addition) and dimer or trimer characteristics of the SSRs are given in Table S6.

We used ANOVA to evaluate the significance of the sequence changes. The distribution of SNP types (transition or transversion) did not differ among the four TRIM lines used as replicates (Table S7). With lines used as the second factor, the SNP distribution did not differ among chromosomes ($P = 0.45$) but did differ among lines ($P \leq 0.01 \times 10^{-6}$; Table S8).

Several large deletions and insertions occurred in the four TRIM transformant lines

Using a new homemade DNA sequence analysis program we identified three large deletions in the TRIM transformants. These deletions were subsequently revealed by the IGV browser (Figures 4 and S9). The 26.2-kbp deleted region in M0048349 chromosome 4 is illustrated by several paired red polygons and the lack of many gray polygons (Figure 4). Another two deletions, of 13.8 and 6.7 kbp, were detected in M0053677 chromosomes 3 and 1, respectively (Figure S9). Again, the deleted region is indicated by

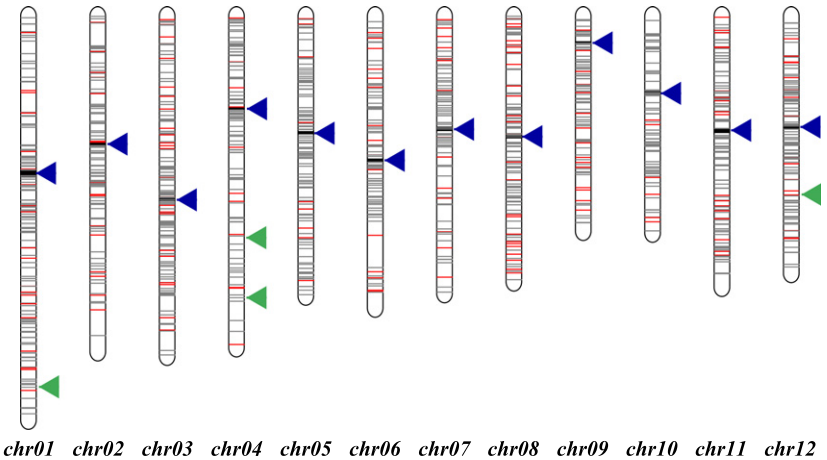

Figure 3. Distribution of homozygous sequence changes in four transformants.
Distribution of homologous single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) in the four Taiwan Rice Insertional Mutagenesis (TRIM) transformants in the 12 chromosomes. In total, 1124 SNPs and 197 indels were used. The positions of centromeres are indicated in blue and the integration sites of T-DNA are in green. Grey bars indicate SNPs and red bars indicate indels.

several paired red polygons and the lack of many gray polygons (Figure 4, panels b and c). Three primer sets that would amplify the region encompassing the deletion were designed (Table S2) and the PCR products were sequenced, thereby confirming these deletions. These deletions did not contain any additional nucleotides between the two flanking sites. Table S8 lists the genes within these deleted sequences in wild-type plants. In total, five, one and one non-transposable element (TE) genes were deleted in the 26.2-, 13.8- and 6.7-kbp regions, respectively. Figure S9(a) also illustrates a heterozygous 0.9-kbp tandem repeat that occurred in chromosome 5 of M0048349 near a hypothetical protein gene. The green polygons indicate that they can be paired, but the directions are wrong; thus, an insertion is present in the region. Analysis of paired-end sequencing data identified another tandem repeat in the genomes of the TRIM line M0079651 (Table S9). One 2.2-kbp repeat was located on chromosome 1 in a hypothetical protein-coding gene. We did not find any large deletions in the genomes of plants with treatments A–H, nor in transformants of IR64 or TK9.

Effect of sequence changes on the regenerants and transformants

Table 4 summarizes the effects of SNPs on the genomes of the regenerants (treatments A–E), transformants (treat-

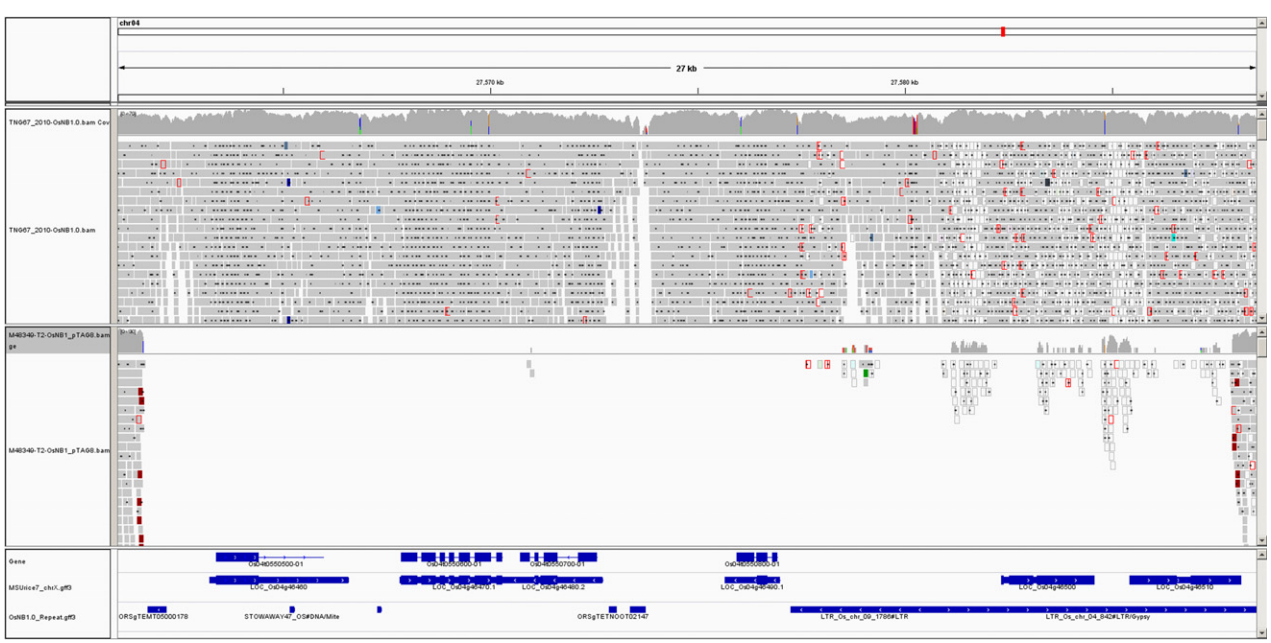

Figure 4. View of a large deletion in one Taiwan Rice Insertional Mutagenesis (TRIM) line.
The Integrative Genomics Viewer view of aligned reads of a 26.2-kbp deletion region in M0048349. A 27-kbp region is shown. The pair read sequencing was performed on an Illumina platform. Alignments toward the Nipponbare IRGSP 1.0 (top row) are represented as gray polygons and mismatched reads as red polygons. The second row is the alignment of TNG67 sequences and the third row the alignment of M0048349. For the bottom row, the top line indicates the annotated rice genes by RAP-DB, and the second line by MSU rice 7.0, and the last line the repeat unit data downloaded from RAP-DB.

ments F–H and the other two varieties) and the four TRIM transformants. Of the many SNPs in these plants, about 80% were in intergenic or promoter regions. Of the remaining mutations, 28% had changes in non-synonymous codons, frame shifts, splice sites or early stop mutations and thus altered amino acid sequences.

Five dwarf tillering mutants have been reported in rice, and genes corresponding to the *d3*, *d10*, *d14*, *d17* and *d27* mutants have been cloned (Ishikawa *et al.*, 2005; Zou *et al.*, 2006; Arite *et al.*, 2007; Lin *et al.*, 2009). All these genes encode proteins responsible for synthesis or regulation of strigolactone. The M0048349 mutant line has a 26.2-kbp deletion. One of the deleted genes is LOC_Os04g46470, the *D17* gene, and this mutant displays the *d17* phenotype. The M0053677 line has a 13.8-kbp deletion. One of the deleted genes is LOC_Os03g10620, the *D14* gene, whose mutant displays the *d14* phenotype. Both M0079651 and M0084311 have 36 common SNPs and 12 common indels. A 1-bp deletion (G), one of these indels, occurred at the splice site of the second intron of LOC_Os11g37650, the *D27* gene, which resulted in the dwarf tillering phenotype. Each of these mutant lines contains sequence changes in genes responsible for strigolactone synthesis. However, in no instance did we detect a T-DNA insertion in these four lines in any gene involved in strigolactone synthesis; that is, the T-DNA insertion did not cause the mutant phenotype.

Table 4 Characteristics of sequence changes for the regenerants and transformants

	Treatments A–E (17 plants)	Treatments G, H (9 plants)	TK9 and IR64 (5 plants)	TRIM-T_2s (4 plants)
Intergenic	1423	991	549	880
Promoter	403	280	182	229
5′ UTR	20	25	14	14
Coding region				
Codon deletion	1	0	2	0
Synonymous	33	14	14	16
Non-synonymous	70	46	39	43
Frame shift	10	7	5	5
Splicing site	10	8	0	4
Translation start site	1	7	2	3
Translation stop site	9	2	1	1
Intron	141	100	57	95
3′ UTR	52	41	19	30

The reference genome was Nipponbare IRGSP-1.0, and RAP-DB non-transposable element gene annotation databases were used. To eliminate the redundant assignment (so each indel may only be annotated to one event) the priorities were: first, non-synonymous coding, splice site donor, start gained, start lost, stop gained, stop lost, frame shift, codon deletion; second, 5′ UTR, upstream; third, synonymous, 3′ UTR, intron, splice site region; fourth, intergenic, downstream. UTR, untranslated region.

We used the sequences of the known dwarf tillering genes *D3*, *D10*, *D14*, *D17* and *D27* to search the TRIM flanking-sequence database but did not find any T-DNA integration events at or near these genes. According to the literature, these mutants should have high tiller numbers and a semi-dwarf-to-dwarf phenotype. We used the phenotypes of high tillering (more than 40 tillers) and dwarfism (plant height less than 60 cm) to search the TRIM phenomics database (Chern *et al.*, 2007) and identified 123 mutant lines that show this phenotype; 57 of these lines had FST data. The genes disrupted by T-DNA integration in these lines encoded proteins such as a peptidase, a protein kinase, a GTPase-activating protein, a transferase, a phytosulfokine receptor and a hypothetical protein; none of these genes is related to genes known to be involved in strigolactone synthesis or regulation. These results correspond to the low tagging efficiency noted in rice insertion mutant populations such as TRIM.

To detect whether *Agrobacterium* genomic DNA is present in the transformants, we performed a similarity search. A few reads (e.g. 73 entries out of 57 463 336 reads from M0053677-T_2 plants) have 100% identity to the 16S RNA sequence of *Alistipes* sp., *Bacteroides* sp., *Blautia* sp., *Clostridium* sp., etc. but none are similar to *Agrobacterium* genomic DNA. In addition, these short reads did not share identity between them. Thus, there was no *Agrobacterium* genomic DNA in the transformants produced in this study and the TRIM lines we sequenced. These few microbial sequences among the sequenced rice genomes indicate that we should pay attention to contamination by microbial DNA during sample preparation.

Sequence changes in some TRIM or Tos17 lines

To reveal additional sequence changes, we examined several TRIM lines for mutations in genes known to be responsible for the corresponding phenotypes. The rice *d1* mutant has an obvious phenotype (round seeds, dense panicle and dwarf stature), and the gene responsible for this phenotype was cloned and studied more than 10 years ago (Ashikari *et al.*, 1999; Ueguchi-Tanaka *et al.*, 2000). We chose four TRIM mutant lines with the *d1* phenotype and designed six primer sets (Table S2) to cover the entire *D1* gene region. Capillary sequencing indicated that SNPs in M0000625, M0005254 and M0033961 created an early translational stop, and a SNP in M0001475 abolished an intron splice donor site. Therefore, SNPs, rather than T-DNA insertions, caused the *d1* phenotype in these four lines.

We also examined a seedling lethal mutant of the *Tos17* line NC5956. Map-based screening revealed no PCR product containing the R1963 marker located at 118.1 cM of chromosome 8. To detect the extent of this deletion, a ser-

ies of primer pairs were designed in the 5′ and 3′ directions (Table S2). After walking to the break points, a primer set flanking the deletion was designed and the PCR product sequenced. We detected a 101 151-bp deletion at position 28 Mbp of chromosome 8; no additional nucleotides were inserted into this deleted region. This 101-kbp region contains 20 genes, including four hypothetical proteins and four TE genes.

We also examined the TRIM population for potential polyploid lines. M0045377, a TRIM line with thick culm, increased plant height and longer panicles, was identified, and DNA from leaf tissue of the mutant and wild-type plants was examined by flow cytometry. The mutant plant was tetraploid (Figure S10).

At TARI we propagated thousands of T_1 TRIM lines for two cropping seasons each year for phenomics analysis; 30 seeds per line were imbibed, but only 12 were transplanted onto the paddy field for phenomics screening and propagation. We bulked and extracted DNA from the remaining 15 seedlings per line and subjected it to targeting-induced local lesions in genomes (TILLING) analysis. We designed nine pairs of primer sets (Table S2) covering most of the rice waxy genes (6297 bp) and performed TILLING on 200 bulked DNA samples. Figure S11 illustrates two sets of TILLING gels. Each panel shows the analysis of 10 bulked DNA samples. Lane 2 in panel (a) and lane 10 in panel (b) show different bands, which indicate the presence of SNPs in at least one of the 15 bulked lines. We then performed capillary sequencing of the PCR products and confirmed sequence changes. Of the $15 \times 200 \times 6297$ bp = 18 891 000 bp screened, we identified six SNPs, confirmed by sequencing, for an estimated mutation rate of 3.2×10^{-7} per site per diploid genome. This rate is approximately the same as that for the TRIM mutants identified above.

DISCUSSION

TRIM as a resource for rice functional genomics

We used *Agrobacterium* T-DNA as an insertional mutagen of TNG67, a *japonica* rice variety. The raw data from the VCF output indicated that there were 180 969 homozygous SNPs and 30 823 homozygous indels between Nipponbare and TNG67. These numbers became 162 053 and 28 871, respectively, after stringent filtering. Because the SNP rate is quite low between these two *japonica* varieties, we feel justified in using the Nipponbare sequence (IRGSP v.1.0) and the NGS data of TNG67 to search for any sequence changes in the mutants studied.

A database of the resulting collection of mutagenized lines, TRIM (Hsing *et al.*, 2007), can be accessed at http://trim.sinica.edu.tw. We obtained detailed phenomics information for 12 T_1 plants per line; the data are available in the TRIM database (Chern *et al.*, 2007; Lo *et al.*, 2016). As of mid-2015, 127 217 lines were included in the TRIM collection; 73 248 lines have phenomics data and 58 117 have FST data.

Tos17 did not transpose in the regenerants or transformants produced

From the genome sequencing data, we identified three copies of *Tos17* in each of the three rice varieties used. They are located at 1.0 Mbp of chromosome 2, 26.7 Mbp of chromosome 7 and 15.4 Mbp of chromosome 10 for TNG67 and TK9. Therefore, *Tos17* elements are at the same chromosome locations for these two *japonica* varieties. The locations on chromosomes 7 and 10 are exactly the same as for the variety Nipponbare, with an extra copy on chromosome 2. Although the *indica* rice variety IR64 also contains three copies of *Tos17*, the locations are different. In IR64, *Tos17* is located at 28.7 Mbp of chromosome 1, 26.9 Mbp of chromosome 2 and 30.3 Mbp of chromosome 2. In addition, of all the 17 regenerants and 18 transformants used in our study, none contained more than three copies of *Tos17*. Therefore, *Tos17* does not transpose during callus growth and regeneration periods for the three varieties used.

Pre-existing sequence changes among seed batches

Our sequencing and identification of SNPs and indels among four TRIM lines suggested that 48 sequence changes must have existed in the seeds used for transformation. Table 2 lists the heterozygous and homozygous sequence changes in plants having treatments A–H. These are R_0 or T_0 plants, and thus the homozygous SNPs and indels must have pre-existed in the seed batches used for callus induction before regeneration and transformation. The mean number about 50 sequence changes per plant, and some of the SNPs/indels are identical in different plants, again indicating that they were pre-existing.

We sequenced four TRIM lines with the dwarf tillering phenotype. These lines were independently generated at different times. Regardless of this we identified a common 1-bp deletion sequence change in the *D27* gene of TRIM lines M0079651 and M0084311. Therefore we checked the number of dwarf tillering mutants in our TRIM phenomics database. Since 2002, we have generated about 130 000 TRIM lines, numbered from the beginning of the project. We checked the frequency of the dwarf tillering phenotype every 5000 lines. The mean frequency per 5000 lines was 3.8 ± 3.1. However, 21 such mutants occurred in lines 80 001–85 000. The high numbers of dwarf tillering mutants in the two subgroups indicated that this specific 1-bp deletion may have been present as a recessive gene in a particular seed lot used to generate callus for transformation. We performed sequencing of the *D27* gene region of two lines, M0079745 and M0084921 (primer sequence in Table S2). One line, M0079745, indeed had the identical deletion (G) at the splice site.

Mutation during rice callus propagation, regeneration and transformation

The spontaneous mutation rate of the Arabidopsis genome was estimated at 7×10^{-9} base substitutions per site per generation (Ossowski *et al.*, 2010). Jiang *et al.* (2011) used NGS to estimate the genome-wide sequence variation of five R_1 generation Arabidopsis regenerant lines and identified SNPs and small indels (<2 bp). The mutation rate was $(10.5 \pm 1.0) \times 10^{-7}$ base substitutions per site per generation, around 60- to 350-fold higher than the spontaneous mutation rate. These results suggested that plant regeneration could induce mutations.

Maintenance of callus and the regeneration of rice plants from tissue culture result in transposon movement. Several studies have investigated sequence changes in regenerated rice mutant lines after *Tos17* movement or T-DNA integration. Miyao *et al.* (2012) sequenced the genomes of three *Tos17* mutant lines and concluded that, other than integration of the retrotransposon *Tos17*, SNPs and indels were the major causes of somaclonal variation. The estimated mutation rate was 17.4×10^{-7} per site per regenerant. Using a similar strategy, Sabot *et al.* (2011) studied transposition events in another rice *Tos17* mutant line; these authors detected the transposition of another 11 long-terminal repeat retrotransposons and 12 miniature inverted-repeat transposable elements. Whole-genome sequencing of a transgenic rice line producing an edible vaccine against cedar pollen allergy, at the T_5 generation, and its precursor non-transgenic line revealed a mutation rate of 5.5×10^{-7} per site per generation (Kawakatsu *et al.*, 2013). The mutation rates of Arabidopsis (Jiang *et al.*, 2011), rice *Tos17* mutants (Miyao *et al.*, 2012) and rice TRIM mutants (this study) all indicated mutation frequencies of the same log scale. However, in the present work the SNP and indel rates were $(1.7–3.3) \times 10^{-7}$ and $(8.9–21.0) \times 10^{-7}$ per site per diploid genome, respectively, for regenerants/transformants from treatments A–H, as well as for two other rice varieties and the four TRIM transformants, respectively. Thus, mutation rates of regenerants and transformants using the modified Salaud protocol are 3- to 10-fold lower than those for the TRIM and *Tos17* mutants. All these mutation rates are about 100-fold higher than the spontaneous mutation rate.

Zhang *et al.* (2014) recently sequenced the genome of a tissue culture-originating rice inbred line after selfing for eight generations. Compared with its wild-type progenitor Hitomebore, the line showed 54 268 DNA polymorphisms, including 37 332 SNPs and 16 936 small indels. Their calculated average mutation rate was 5×10^{-5} per site per diploid genome, much higher than that of other studies. With similar sequencing depth in the two studies, the major difference for mutation estimation methods is that Zhang *et al.* used 10–100 coverage as a cut-off value, whereas we used 8–30. Many repetitive sequences will be present in high-coverage genome regions. Thus, the reason for the differences in calculated mutation rate may be the standard of quality control for detecting sequence changes and the mutagens used.

To produce plants for treatments A–H, we followed a modified Salaud protocol. Calli were grown for about 2 weeks before *Agrobacterium* co-cultivation. Calli that were regenerated without transformation were also grown for about 2 weeks. It takes about 3 months from induction of callus until regenerants or transformants are obtained. However, in resource centers such as those generating *Tos17* and TRIM lines people tended to subculture and amplify to produce more calli and thus prolonged the callus growth period for 3–6 months before *Agrobacterium* co-cultivation and regeneration. In addition, TRIM lines were derived from rice suspension cultures instead of calli grown on solidified medium. All these procedures increase the probability of more sequence changes. In this study, although sequence changes do occur in treatments A–H and transformants of the two other varieties examined, the frequencies did not differ from each other significantly.

Castle *et al.* (1993) showed that about 17% of the screened T-DNA-tagged Arabidopsis mutants contained chromosomal rearrangements. In our study, the proportion (about 15%) and characteristics (in plants with more than one T-DNA insertion) of chromosome rearrangement events in rice plants with *Agrobacterium*-mediated transformation were similar to those seen with Arabidopsis transformants. Many studies also illustrated inversions and translocations in *Agrobacterium*-transformed Arabidopsis plants (Nacry *et al.*, 1998; Tax and Vernon, 2001; Lafleuriel *et al.*, 2004; Curtis *et al.*, 2009) and rice plants (Majhi *et al.*, 2014). The previous studies used DNA blots or molecular markers to demonstrate translocations; here, we used whole-genome sequencing to reveal detailed changes.

CONCLUSIONS

In this study we developed a protocol for stringent filtering of SNP or indel raw data. We suggest that attention is paid to the type or generation of plant that is sequenced (T_0, T_1 or T_2) and normalizing the mutation numbers to the original generation (R_0 or T_0). In our studies we could not detect a significant number of sequence changes in the rice genome caused by *Agrobacterium*-mediated transformation above that of tissue culture-induced somaclonal variation. To prepare transgenic plants we suggest: (i) preparing calli from different seed batches, (ii) shortening the callus growth period during the transformation process, (iii) producing at least 10 independent transformants, and (iv) performing genotype and phenotype analyses of each transformant. Despite the low tagging efficiency for the T-DNA and *Tos17*

insertion populations, these transgenic plants can still be used for complementation, overexpression or antisense analysis.

EXPERIMENTAL PROCEDURES

For materials and methods, see Data S1–S3 and Figures S12 and S13.

Sequence data access

The raw read data for this project have been submitted to the National Center for Biotechnology Information Sequence Read Archive. The Sequence Read Archive experiment numbers are TNG67, SRR1525794 and SRR1531575; treatment A1-R_0, SRR2767684; treatment A2-R_0, SRR2767685; treatment A3-R_0, SRR2767686; treatment B1-R_0, SRR2767687; treatment B2-R_0, SRR2767688; treatment B3-R_0, SRR2767689; treatment B4-R_0, SRR2767690; treatment C1-R_0, SRR2767691; treatment C2-R_0, SRR2767692; treatment C3-R_0, SRR2767693; treatment C4-R_0, SRR2767694; treatment D1-R_0, SRR2767695; treatment D2-R_0, SRR2767696; treatment D3-R_0, SRR2767697; treatment E1-R_0, SRR2767698; treatment E2-R_0, SRR2767699; treatment E3-R_0, SRR2767700; treatment F1-R_0, SRR2767701; treatment F2-R_0, SRR2767702; treatment F3-R_0, SRR2767703; treatment G1-R_0, SRR2767704; treatment G2-R_0, SRR2767705; treatment G3-R_0, SRR2767706; treatment H1-R_0, SRR2767707; treatment H2-R_0, SRR2767708; treatment H3-R_0, SRR2767709; M0048349-T_2, SRR2767710; M0053677-T_2, SRR2767711; M0079651-T_2, SRR2767712; M0084311-T_2, SRR2767713; TK9, SRR2079340; TK9 1-R_0, SRR2767714; TK9 2-R_0, SRR2767715; IR64 1-R_0, SRR2767778; IR64 2-R_0, SRR2767800; IR64 3-R_0 SRR2767818.

ACKNOWLEDGEMENTS

We thank Ms L. H. Wu for technical support and Ms Laura Smales (BioMedEditing, Toronto, Canada) for English editing. This work was supported by an Academia Sinica (Taiwan) Thematic Grant AS022342 to LYK, LYL, SBG and YICH.

SUPPORTING INFORMATION

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

Figure S1. The morphology of wild-type plants and four dwarf tillering transformants.

Figure S2. Examples of indel identification revealed by the Integrative Genomics Viewer image.

Figure S3. Validation of single-nucleotide polymorphisms and small deletions.

Figure S4. Analysis of some unusual plant DNA/transgene junctions.

Figure S5. Distribution of heterozygous sequence changes.

Figure S6. Mutation rates of homozygous single-nucleotide polymorphisms in the four transformants.

Figure S7. Homozygous single-nucleotide polymorphisms and their nearby regions.

Figure S8. The distribution of indel sizes in the four transformants.

Figure S9. Integrative Genomics Viewer views of the aligned reads of transformants.

Figure S10. Ploidy analysis of TNG67 and one Taiwan Rice Insertional Mutant (TRIM) line.

Figure S11. Analysis of Taiwan Rice Insertional Mutant (TRIM) mutants in the waxy gene.

Figure S12. Examples of single-nucleotide polymorphisms and indels validated by Integrative Genomics Viewer images.

Figure S13. Distribution of sequence depth for various lines.

Table S1. Summary of sequencing data.

Table S2. Primer sets used in the studies.

Table S3. Information of single-nucleotide polymorphisms and indels used for confirmation.

Table S4. Detected sequence variations in the four Taiwan Rice Insertional Mutant (TRIM) transformants.

Table S5. Single-nucleotide polymorphisms and nearby sequences in the four Taiwan Rice Insertional Mutant (TRIM) transformants.

Table S6. Simple sequence repeat (SSR) copy changes in the four Taiwan Rice Insertional Mutant (TRIM) transformants.

Table S7. ANOVA analysis I.

Table S8. ANOVA analysis II.

Table S9. Large indels in Taiwan Rice Insertional Mutant (TRIM) transformants and the affected genes.

Table S10. Detected sequence variations in the three treatment A regenerants.

Table S11. Sequencing information for the materials used.

Data S1. Experimental procedures.

Data S2. Explanation and home-made code.

Data S3. Analysis and validation of the sequence changes.

REFERENCES

Arite, T., Iwata, H., Ohshima, K., Maekawa, M., Nakajima, M., Kojima, M., Sakakibara, H. and Kyozuka, J. (2007) DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud outgrowth in rice. *Plant J.* **51**, 1019–1029.

Ashikari, M., Wu, J., Yano, M., Sasaki, T. and Yoshimura, A. (1999) Rice gibberellin-insensitive dwarf mutant gene *Dwarf 1* encodes the a-subunit of GTP-binding protein. *Proc. Natl Acad. Sci. USA*, **96**, 10284–10289.

Brunaud, V., Balzergue, S., Dubreucq, B. *et al.* (2002) T-DNA integration into the *Arabidopsis* genome depends on sequences of pre-insertion sites. *EMBO Rep.* **3**, 1152–1157.

Budziszewski, G.J., Lewis, S.P., Glover, L.W. *et al.* (2001) *Arabidopsis* genes essential for seedling viability: isolation of insertional mutants and molecular cloning. *Genetics*, **159**, 1765–1778.

Castle, L.A., Errampalli, D., Atherton, T.L., Franzmann, L.H., Yoon, E.S. and Meinke, D.W. (1993) Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis*. *Mol. Gen. Genet.* **241**, 504–514.

Chern, C.G., Fan, M.J., Yu, S.M. *et al.* (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.

Cluster, P.D., O'Dell, M., Metzlaff, M. and Flavell, R.B. (1996) Details of T-DNA structural organization from a transgenic *Petunia* population exhibiting co-suppression. *Plant Mol. Biol.* **32**, 1197–1203.

Curtis, M., Belcram, K., Bollmann, S., Tominey, C., Hoffman, P., Mercier, R. and Hays, J. (2009) Reciprocal chromosome translocation associated with TDNA-insertion mutation in *Arabidopsis*: genetic and cytological analyses of consequences for gametophyte development and for construction of doubly mutant lines. *Planta*, **229**, 731–745.

Droc, G., An, G., Wu, C. *et al.* (2013) Mutant resources for functional analysis of the rice genome. In *Genetics and Genomics of Rice* (Zhang, Q. and Wing, R.A., eds). New York: Springer, pp. 81–115.

Duncan, D.B. (1955) Multiple range and multiple F tests. *Biometrics*, **11**, 1–42.

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

Gale, M.D. and Devos, K.M. (1998) Comparative genetics in the grasses. *Proc. Natl Acad. Sci. USA*, **95**, 1971–1974.

Gheysen, G., Montagu, M.V. and Zambryski, P. (1987) Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangements of target plant DNA sequences. *Proc. Natl Acad. Sci. USA*, **84**, 6169–6173.

He, C., Dey, M., Lin, Z., Duan, F., Li, F. and Wu, R. (2007) An efficient method for producing an indexed, insertional-mutant library in rice. *Genomics*, **89**, 532–540.

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.

Hood, E.E., Gelvin, S.B., Melchers, L. and Hoekema, A. (1993) New *Agrobacterium* helper plasmids for gene transfer to plants. *Transgenic Res.* **2**, 208–218.

Hsing, Y.I., Chern, C.G., Fan, M.J. *et al.* (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol. Biol.* **63**, 351–364.

IRGSP. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Ishikawa, S., Maekawa, M., Arite, T., Onishi, K., Takamure, I. and Kyozuka, J. (2005) Suppression of tiller bud activity in tillering dwarf mutants of rice. *Plant Cell Physiol.* **46**, 79–86.

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

Jiang, S.Y., Bachmann, D., La, H., Ma, Z., Venkatesh, P.N., Ramamoorthy, R. and Ramachandran, S. (2007) *Ds* insertion mutagenesis as an efficient tool to produce diverse variations for rice breeding. *Plant Mol. Biol.* **65**, 385–402.

Jiang, C., Mithani, A., Gan, X., Belfield, E.J., Klingler, J.P., Zhu, J.K., Ragoussis, J., Mott, R. and Harberd, N.P. (2011) Regenerant *Arabidopsis* lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. *Curr. Biol.* **21**, 1385–1390.

Johnson, A.A., 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.

Kawahara, Y., de la Bastide, M., Hamilton, J.P. *et al.* (2013) Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)*, **6**, 4.

Kawakatsu, T., Kawahara, Y., Itoh, T. and Takaiwa, F. (2013) A whole-genome analysis of a transgenic rice seed-based edible vaccine against cedar pollen allergy. *DNA Res.* **20**, 623–631.

Kim, S.R., Lee, J., Jun, S.H., Park, S., Kang, H.G., Kwon, S. and An, G. (2003) Transgene structures in T-DNA-inserted rice plants. *Plant Mol. Biol.* **52**, 761–773.

Kim, C.M., Piao, H.L., Park, S.J. *et al.* (2004) Rapid, large-scale generation of *Ds* transposant lines and analysis of the *Ds* insertion sites in rice. *Plant J.* **39**, 252–263.

Kumar, S. and Fladung, M. (2002) Transgene integration in aspen: structures of integration sites and mechanism of T-DNA integration. *Plant J.* **31**, 543–551.

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.

Lafleuriel, J., Degroote, F., Depeiges, A. and Picard, G. (2004) A reciprocal translocation, induced by a canonical integration of a single T-DNA, interrupts the *HMG-I/Y Arabidopsis thaliana* gene. *Plant Physiol. Biochem.* **42**, 171–179.

Larkin, P.J. and Scowcroft, W.R. (1981) Somaclonal variation - a novel source of variability from cell cultures for plant improvement. *Theor. Appl. Genet.* **60**, 214.

Lin, H., Wang, R., Qian, Q. *et al.* (2009) DWARF27, an iron-containing protein required for the biosynthesis of strigolactones, regulates rice tiller bud outgrowth. *Plant Cell*, **21**, 1512–1525.

Lo, S.F., Fan, M.J., Hsing, Y.I. *et al.* (2016) Genetic resources offer efficient tools for rice functional genomics research. *Plant, Cell Environ.* doi: 10.1111/pce.12632.

Majhi, B.B., Shah, J.M. and Veluthambi, K. (2014) A novel T-DNA integration in rice involving two interchromosomal translocations. *Plant Cell Rep.* **33**, 929–944.

McElver, J., Tzafrir, I., Aux, G. *et al.* (2001) Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics*, **159**, 1751–1763.

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., 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.

Montoya, A.L., Moore, L.W., Gordon, M.P. and Nester, E.W. (1978) Multiple genes coding for octopine-degrading enzymes in *Agrobacterium*. *J. Bacteriol.* **136**, 909–915.

Nacry, P., Camilleri, C., Courtial, B., Caboche, M. and Bouchez, D. (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics*, **149**, 641–650.

Ossowski, S., Schneeberger, K., Lucas-Lledo, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D. and Lynch, M. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, **327**, 92–94.

Sabot, F., Picault, N., El-Baidouri, M. *et al.* (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., Gay, C., Larmande, P. *et al.* (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. *Plant J.* **39**, 450–464.

Singer, K., Shiboleth, Y.M., Li, J. and Tzfira, T. (2012) Formation of complex extrachromosomal T-DNA structures in *Agrobacterium tumefaciens*-infected plants. *Plant Physiol.* **160**, 511–522.

Stroud, H., Ding, B., Simon, S.A., Feng, S., Bellizzi, M., Pellegrini, M., Wang, G.L., Meyers, B.C. and Jacobsen, S.E. (2013) Plants regenerated from tissue culture contain stable epigenome changes in rice. *ELife*, **2**, e00354.

Tax, F.E. and Vernon, D.M. (2001) T-DNA-associated duplication/translocations in *Arabidopsis*. Implications for mutant analysis and functional genomics. *Plant Physiol.* **126**, 1527–1538.

Thomas, C.M. and Jones, J.D. (2007) Molecular analysis of *Agrobacterium* T-DNA integration in tomato reveals a role for left border sequence homology in most integration events. *Mol. Genet. Genomics*, **278**, 411–420.

Ueguchi-Tanaka, M., Fujisawa, H., Kobayashi, M., Ashikari, M., Iwasaki, Y., Kitano, H. and Matsuoka, M. (2000) Rice dwarf mutant *d1*, which is defective in the a subunit of the heterotrimeric G protein, affects gibberellin signal transduction. *Proc. Natl Acad. Sci. USA*, **97**, 11638–11643.

Upadhyaya, N.M., Zhu, Q.H., Zhou, X.R. *et al.* (2006) Dissociation *Ds* constructs, mapped *Ds* launch pads and a transiently-expressed transposase system suitable for localized insertional mutagenesis in rice. *Theor. Appl. Genet.* **112**, 1326–1341.

Wan, S., Wu, J., Zhang, Z. *et al.* (2009) Activation tagging, an efficient tool for functional analysis of the rice genome. *Plant Mol. Biol.* **69**, 69–80.

Wang, N., Long, T., Yao, W., Xiong, L., Zhang, Q. and Wu, C. (2013) Mutant resources for the functional analysis of the rice genome. *Mol. Plant*, **6**, 596–604.

Wei, F.J., Droc, G., Guiderdoni, E. and Hsing, Y.I. (2013) International Consortium of Rice Mutagenesis: resources and beyond. *Rice*, **6**, 39.

Windels, P., De Buck, S., Van Bockstaele, E., De Loose, M. and Depicker, A. (2003) T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences. *Plant Physiol.* **133**, 2061–2068.

Yang, Y., Li, Y. and Wu, C. (2013) Genomic resources for functional analyses of the rice genome. *Curr. Opin. Plant Biol.* **16**, 157–163.

Zhang, J., Cai, L., Cheng, J. *et al.* (2008) Transgene integration and organization in cotton (*Gossypium hirsutum* L.) genome. *Transgenic Res.* **17**, 293–306.

Zhang, D., Wang, Z., Wang, N. *et al.* (2014) Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. *PLoS ONE*, **9**, e96879.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D., Boller, T. and Felix, G. (2006) Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell*, **125**, 749–760.

Zou, J., Zhang, S., Zhang, W., Li, G., Chen, Z., Zhai, W., Zhao, X., Pan, X., Xie, Q. and Zhu, L. (2006) The rice *HIGH-TILLERING DWARF1* encoding an ortholog of *Arabidopsis MAX3* is required for negative regulation of the outgrowth of axillary buds. *Plant J.* **48**, 687–698.

