

A rice phenomics study—phenotype scoring and seed propagation of a T-DNA insertion-induced rice mutant population

Chyr-Guan Chern · Ming-Jen Fan · Su-May Yu · Ai-Ling Hour · Po-Chang Lu · Yao-Cheng Lin · Fu-Jin Wei · Sheng-Chung Huang · Shu Chen · Ming-Hsing Lai · Ching-Shan Tseng · Hsing-Mu Yen · Woei-Shyuan Jwo · Chen-Chia Wu · Tung-Lung Yang · Lung-Sheng Li · Yih-Cheng Kuo · Su-Mien Li · Charng-Pei Li · Chiu-Kai Wey · Arunee Trisiriroj · Hsing-Fang Lee · Yue-Ie C. Hsing

Received: 12 February 2007 / Accepted: 23 July 2007 / Published online: 14 August 2007

© Retained by Author 2007

Abstract With the completion of the rice genome sequencing project, the next major challenge is the large-scale determination of gene function. As an important crop and a model organism, rice provides major insights into gene functions important for crop growth or production. Phenomics with detailed information about tagged populations provides a good tool for functional genomics analysis. By a T-DNA insertional mutagenesis approach, we have generated a rice mutant population containing 55,000 promoter trap and gene activation or knockout lines. Approximately 20,000 of these lines have known integration sites. The T0 and T1 plants were grown in net “houses”

for two cropping seasons each year since 2003, with the mutant phenotypes recorded. Detailed data describing growth and development of these plants, in 11 categories and 65 subcategories, over the entire four-month growing season are available in a searchable database, along with the genetic segregation information and flanking sequence data. With the detailed data from more than 20,000 T1 lines and 12 plants per line, we estimated the mutation rates of the T1 population, as well the frequency of the dominant T0 mutants. The correlations among different mutation phenotypes are also calculated. Together, the information about mutant lines, their integration sites, and the phenotypes make this collection, the Taiwan Rice Insertion Mutants (TRIM), a good resource for rice phenomics study. Ten T2 seeds per line can be distributed to researchers upon request.

Chyr-Guan Chern, Ming-Jen Fan, and Su-May Yu have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-007-9218-z) contains supplementary material, which is available to authorized users.

C.-G. Chern · M.-J. Fan · S.-C. Huang · S. Chen · M.-H. Lai · C.-S. Tseng · H.-M. Yen · W.-S. Jwo · C.-C. Wu · T.-L. Yang · L.-S. Li · Y.-C. Kuo · S.-M. Li · C.-P. Li · C.-K. Wey
Taiwan Agricultural Research Institute, Wufeng, Taichung 41301, Taiwan

Present Address:

M.-J. Fan
Department of Biotechnology and Bioinformatics, Asia University, Wufeng, Taichung 41354, Taiwan

S.-M. Yu
Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 11529, Taiwan

A.-L. Hour · P.-C. Lu · Y.-C. Lin · F.-J. Wei · A. Trisiriroj · H.-F. Lee · Y.-I. C. Hsing (✉)
Institute of Plant and Microbial Biology, Academia Sinica, Nankang, Taipei 11529, Taiwan
e-mail: bohling@gate.sinica.edu.tw

Keywords Activation · Knockout · Phenotype · Rice · Segregation · T-DNA

Abbreviations

GM	Genetically modified
MPSS	Massively parallel signature sequence
NPGRC	National Plant Genetic Resource Center
RH	Relative humidity
RAP	Rice Annotation Project
SAGE	Serial analysis of gene expression
TARI	Taiwan Agriculture Research Institute
TNG67	Tainung 67
TRIM	Taiwan Rice Insertion Mutants

Introduction

Rice (*Oryza sativa*) is one of the most important crops in the world. Rice, wheat, and maize together account for

60% of the world's food production, and rice is the principal food of nearly 50% of the world's population. These cereal crops share a large degree of synteny, making rice an excellent model cereal crop for genomics research (Gale and Devos 1998). Rice was the first crop plant chosen for complete genome sequencing because (1) it has the smallest genome size (390 Mb) among the major cereal crops, (2) it can undergo large-scale transformation on a routine basis, (3) the isolation of genes from rice could facilitate isolation of homologous genes from other cereal crops, and (4) much molecular and genetic information (ESTs, markers, genetic, and physical maps, etc.) about rice is available.

With the completion of the genomic sequencing of rice (IRGSP 2005), the challenge of the post-genomic era is to systematically analyze the functions of all genes in the genome. Computation- and curation-based annotation of the genome was initiated to predict the locations of the genes, including exons, introns, and their putative functions (Itoh et al. 2007; Ouyang et al. 2007). In the Rice Annotation Project (RAP1), through manual curation, Itoh et al. suggested that the gene number of rice is approximately 32,000. Researchers at The Institute for Genomic Research Rice Genome Annotation project, by ab initio annotation, identified 42,653 nontransposable element-related genes. The large differences in the annotated gene numbers suggest the necessity for further analysis. Furthermore, since no experiments were carried out in these projects, the annotation is subjected to another layer of uncertainty. In the case of Arabidopsis genome annotation, for instance, the sequencing was completed in 2000 (AGI 2000), but at least 40% of the gene predictions were subsequently found to be erroneous (Hass et al. 2005). Thus, further validation of the present rice gene model and identification of additional genes must be achieved by other computational and experimental approaches.

A variety of investigations have explored rice gene structures and functions, such as full-length cDNA sequences (Kikuchi et al. 2003), whole genome tiling microarrays (Li et al. 2006), gene expression arrays (Wasaki et al. 2006; Li et al. 2006), serial analysis of gene expression (SAGE; Gibbins et al. 2003; Su et al. 2005), massively parallel signature sequence (MPSS; Nakano et al. 2006), proteomics (Komatsu and Tanaka 2005), and generation of large-scale chemical and irradiation-induced mutants (Wu et al. 2005). Among these techniques, an important and direct approach of defining the function of a novel gene is to eliminate or activate its function by insertional mutagenesis. Insertional mutagenesis, with T-DNA or a transposable element, provides opportunities to assign a function to a particular DNA sequence and to isolate the target gene causing a specific phenotype.

Since the late 1990s, much effort worldwide has been invested in generating a tagged rice mutant population. Researchers in Taiwan (Hsing et al. 2007), Korea (Jeong et al. 2002, 2006), China (Chen et al. 2003; Wu et al. 2003), and France (Sallaud et al. 2004) use T-DNA as the vector; those in Japan (Hirochika 2001; Miyao et al. 2003) use the rice endogenous retrotransposon *Tos17*; and those in Australia (Upadhyaya et al. 2006), Europe (van Enchevort et al. 2005), and the United States (Kolesnik et al. 2004) use the maize *Ac/Ds* and *En/Spm* transposable elements. These independent national initiatives have led to the accumulation of mutant lines, characterization of the flanking sequences at insertional points, and presentation of the information in web-accessible databases (Hirochika et al. 2004; An et al. 2005; Guiderdoni et al. 2007; Hsing et al. 2007). Among them, those populations using *Tos17* and *Ac/Ds* only produce gene knock-outs. For those using T-DNA as the vector, some of them provide two functions, including enhancer trap and knockout, and only two groups provide three functions, i.e. gene trap, gene knockout, and activation tagging.

However, difficulties often arise in the identification of a phenotype that would be associated with the mutation. One would be the presence of multigene families that show redundant or highly specialized functions. In addition, since generation of tagged rice population has to go through callus culture which may produce many background mutations, low tagging efficiency existed in *Tos17* (Miyao et al. 2003) or T-DNA-insertion mutant lines (An et al. 2005).

Previously, we constructed ~55,000 T-DNA insertional lines of a *japonica* rice cultivar, Tainung 67 (TNG67), and determined the sequence flanking the T-DNA in ~20,000 of these lines (Hsing et al. 2007). This work is abbreviated as TRIM (Taiwan Rice Insertion Mutants), available at the website <http://www.trim.sinica.edu.tw>. In this TRIM population, the average T-DNA copy number is 1.7, and *Tos17* is not stimulated in the tissue culture stage and won't cause the background mutation, making it an ideal material for use in a tagged population. We also discovered that T-DNA was preferentially integrated into the genic region, with no obvious insertion "hot spot," thus generating high tagging efficiency. The TRIM population, which contains the tri-functional T-DNA, including gene trap, gene knockout, and activation tagging, offers a highly valuable resource for high-throughput rice functional analyses with both forward and reverse genetic approaches. For instance, in promoter-trapping analysis, GUS analysis results indicated integration of genes especially expressed in primary leaves of the germinating seeds, root tips, vascular bundles, and calli. In knockout gene analysis, four genes were used to demonstrate the effect of T-DNA integration. Analysis of a

dominant dwarf plant revealed that the T-DNA integration activated a nearby GA2 oxidase gene.

We grew T0 and T1 plants of the TRIM population in a field meant for genetically modified (GM) crops and examined visible phenotypes systematically over two growing seasons during the last 3 years. Skilled rice breeders scored the phenotypes, and all data were analyzed and stored, along with the flanking sequence information, in a searchable database. In this report, we describe the methodology of propagation. We analyze visible phenotypes in tens of thousands of genes knocked out or activated in the TRIM population, a large-scale resource for rice phenomics study.

Materials and methods

Field management for the T-DNA mutant population

An isolated net field, located at the Taiwan Agriculture Research Institute (TARI), Wufeng, Taiwan, specific for GM crops, was used for growth and propagation of rice T-DNA mutants. The rice growing season in Taiwan each year is about 10 months, with two cropping seasons of about 4–5 months each. Since the rice variety used in this study is not sensitive to day length or temperature, T0 and T1 plants can be grown continuously.

The field has 1.0 ha divided into two regions, or “houses.” Each region is surrounded by two layers of net: a 32-mesh net to 2 m and a 24-mesh net to 5 m from ground level to reduce pollen spread from the field (Fig. S1). A bird net at the top with mesh of $2 \times 2 \text{ cm}^2$ covers the whole area. Each region is divided by a rib, and the wild-type rice variety TNG67 was planted as border and central lines and was used as the control plant for qualitative and quantitative traits, as well as ANOVA. The main rib is broad to allow for the mechanical tractor working in the field. The gate for the net house was always locked to fulfill the requirements of the GM field. Field management, including the application of fertilizers and pesticides, was the same as for a regular paddy rice field in central Taiwan (Lai et al. 1996).

The planting density was $25 \times 25 \text{ cm}^2$, and all plants were of single-seed descent. Each of the two regions contained approximately 4,000 T0 plants and 2,500 T1 lines. Approximately 0.2 ha was used for T0 plants and 0.7 ha for T1 plants each season. T0 plants were grown in the region closest to the entrance and T1 lines in the rest of the area. The area for T0 plants was large, and some empty space was prepared for new seedlings.

Approximately 1,000 T0 seedlings with 4–5 leaves were sent from the tissue culture lab each month. During this time, leaf samples were collected for the subsequent

flanking sequence analysis. All seedlings were tagged with a barcode and then transplanted to the field. A total of ~8,000 T0 plants each year were transplanted continuously from February to August with a growing season of approximately 4 months. T1 lines were transplanted twice a year, in mid-February and late July. Thirty seeds per line were used for germination. The seedling phenotypes were recorded at three-leaf stage, and 12 seedlings were then transplanted into the field in blocks of 3×4 plants. All seeds from each plant were harvested, cleaned, and stored in the National Plant Genetic Resource Center (NPGRC) at TARI.

Phenotype scoring

We used 11 categories divided into 65 subcategories to classify phenotypes in the T1 population (Table 1). Five skilled rice breeders led by two senior breeders took care of the daily screening, recording, and field management. Breeders recorded the plant phenotypes according to the subcategory code number, along with the plant number belonging to each mutant group. This information was then transformed into phenotype traits and segregation ratios. Three important agronomic quantitative traits—heading date, plant height, and panicle number—were also recorded. To determine the T2 seed parameters, 30 seeds per line were used. Ten seeds were used as one duplicate, for a total of three duplicates. The width, length, and height of each seed were recorded before the ten-seed weight was measured. The germination rate was also monitored.

Seed handling and storage

The T1 and T2 seeds harvested from the field were transferred to a quarantine head house at NPGRC. All the materials were examined, registered, and labeled with a barcode. Seeds were cleaned and selected by hand to eliminate unfilled and bad seeds. Photos were taken for seeds with mutated morphology. After cleaning, seed lots were transferred to a seed drying room at $20 \pm 2.5^\circ\text{C}$ and 8–10% relative humidity (RH) to reduce the seed moisture content to 5–8%. The seeds were immediately transferred to a seed packing room at $20 \pm 2.5^\circ\text{C}$, $50 \pm 3\%$ RH. T1 seeds were packed into aluminum cans, which were labeled with barcode, and stored in a long-term storage room at $-12 \pm 2^\circ\text{C}$, $30 \pm 3\%$ RH. For T2 seeds, 30 seeds were packed into aluminum foil bags, for a total of 20 bags. Bags were packed into an aluminum can and stored in a medium-term storage room at $1 \pm 2^\circ\text{C}$, $40 \pm 3\%$ RH, ready for distribution. The remaining seeds were packed in one bag and stored in a long-term storage room. Detailed

Table 1 Summary of phenotype classification and frequency of phenotypes occurring in the TRIM population

Category	Code	Sub-category	Counts ^a	Frequencies ^b (%)
Growth	1	Germination rate	1	0.02
	2	Lethal	139	3.42
	3	Abnormal plants	10	0.25
	4	Weak	105	2.58
Leaf color	11	Albino	13	0.32
	12	Yellow leaf	21	0.52
	13	Dark green leaf	931	22.90
	14	Pale green leaf	286	7.04
	15	Bluish green leaf	69	1.70
	16	Stripe	46	1.13
	17	Zebra	7	0.17
	18	Others	159	3.91
Leaf morphology	21	Wide leaf	186	4.58
	22	Narrow leaf	1,119	27.53
	23	Long leaf	370	9.10
	24	Short leaf	540	13.28
	25	Drooping leaf	66	1.62
	26	Rolled leaf	550	13.53
	27	Spiral leaf	323	7.95
	28	Brittle leaf/culm	14	0.34
	29	Thin lamina joint	5	0.12
	30	Withering	160	3.94
	31	Others	127	3.12
Plant morphology	41	Semidwarf ^c	872	21.45
	42	Dwarf ^c	1,242	30.55
	43	Extremely dwarf ^c	111	2.73
	44	Long culm	24	0.59
	45	Eract	194	4.77
	46	Spread-out	413	10.16
	47	Thin culm	873	21.48
	48	Thick culm	21	0.52
	49	Lazy	1	0.02
Mimic response	51	Lesion mimic	258	6.35
Tiller	61	High tiller position	8	0.20
	62	Low tiller position	0	0.00
	63	Monoculm	74	1.82
	64	Few panicle ^d	352	8.66
	65	Many panicle ^d	1,092	26.86
Heading date	71	Early heading	14	0.34
	72	Late heading	195	4.80
	73	No heading	140	3.44
Flower	75	Abnormal hull	38	0.93
	76	Abnormal floral organ	1	0.02
	77	With awn	36	0.89
	78	Abnormal hull	9	0.22
	79	Abnormal hull color	91	2.24

Table 1 continued

Category	Code	Sub-category	Counts ^a	Frequencies ^b (%)
Panicle morphology	81	Long panicle	35	0.86
	82	Short panicle	939	23.10
	83	Sparse panicle	270	6.64
	84	Dense panicle	189	4.65
	85	Vivipary	0	0.00
	86	Shattering	5	0.12
	87	Neck leaf	36	0.89
	88	Abnormal panicle shape	178	4.38
	89	Others	173	4.26
Seed fertility	91	Sterile	571	14.05
	92	Low fertility	403	9.91
Seed morphology	101	Large grain	29	0.71
	102	Small grain	1,063	26.15
	103	Slender grain	29	0.71
	104	Others	64	1.57

^a 22,665 T1 lines were screened and 4,065 showed obvious traits while growing under normal field conditions. The counts are numbers of mutant lines showing each specific subcategory, with the sum of 15,291

^b The frequencies were obtained by dividing the count by 4,065, the total number of lines with obvious phenotypes

^c Semidwarfism: 70–80% of normal plant height; dwarfism: 40–70% of normal plant height; severe dwarfism: <30 cm

^d Few panicles: <5 panicles per plant; many panicles: >25 panicles per plant

information about seeds and a seed request form can be found at <http://www.npgrc.tari.gov.tw/>

Statistical analysis

Statistical parameters and frequency distributions of seed traits and plant quantitative traits were estimated by using SAS/MEANS, UNIVARIATE, and GLM procedures. To reveal the correlation between phenotype subcategories, the phenotype records were transformed into 1 and 0, where 1 indicates the existence and 0 the absence of the specific mutant trait. The association between pairs of subcategories was then estimated by use of the Pearson correlation coefficient and tested under the null hypothesis of $\rho = 0$, with SAS/CORR. For some mutant codes of interest, we measured association by the chi-square test and SAS/FREQ.

Results

Field management for seed propagation and phenotype scoring

In 2002, we began production of rice T-DNA insertional mutants using seeds of TNG67. Seeds were propagated and selected for phenotypic uniformity under field conditions at

TARI. They were then used for embryogenic callus induction and subsequent transformation with *Agrobacterium*, as indicated previously (Hsing et al. 2007). Those seeds were also used as the wild type in the field for the evaluation of T1 plants.

We produced ~1,000 independent transformants each month. The T1 plants were then planted in two cropping seasons each year, examined for obvious phenotypes, and results were recorded by phenotype code and photography. Quantitative traits were measured before harvest. T2 seeds from the plants belonging to the same phenotype group (see below) were harvested separately.

Phenotypic profiling

Mutation subcategories

Once every 1 or 2 weeks over the growing season, we examined T1 plants for 11 phenotype categories, including overall growth condition, leaf color, leaf morphology, plant morphology, mimic response, tiller, heading date, flower, panicle, seed fertility, and seed morphology. These categories were further divided into 65 subcategories, including “others”. Table 1 illustrates the codes and traits of all phenotypes and the number and frequency of lines exhibiting each trait. During a three-year period, 22,665 mutant lines were propagated and their phenotypes scored. Of

them, 4,065 lines had at least one clearly visible mutant phenotype, that is, 17.9% of the T1 population with mutated traits while growing under normal field conditions. Of the 11 categories, mutations in leaf morphology, plant morphology, and panicle morphology were most frequently observed. The total number of phenotypes was 15,291 (Table 1). Each line with obvious mutated phenotypes contained a mean of three mutant phenotypes (range 1–12). The mutant subcategories with frequencies higher than 20% were dark green leaf, narrow leaf, semidwarfism, dwarfism, thin culm, many panicles, short panicles, and small grains.

Table 2 shows the mutant phenotypes and segregation ratio of the three sublimes in a T1 mutant line M0026806. Three phenotype traits were recorded for M0026806_B and seven for M0026806_C. Although M0026806_B and M0026806_C plants share common phenotypic traits such as semidwarfism, M0026806_C plants showed several different traits such as late heading and sterile seeds. The plant types also differ—spread-out type in the B group and erect type in the C group. Of the 12 plants grown, 8, 3, and 1 plants belonged to the M0026806, M0026806_B, and M0026806_C groups, respectively. The seeds from different groups were harvested separately. The phenotypes of the T1 lines were scored in a similar manner and then recorded systematically. Of the 4,065 lines showing mutated phenotypes, 3,740 have one mutant group, 303 and 22 lines have two and three mutant groups, respectively.

Correlation between the phenotypic traits

While working in the field, we noticed that some of the mutant phenotypes were highly related, so we analyzed the correlation between the subcategories of mutations (Supplementary Table 1). Under a highly stringent condition (i.e., $P < 0.0001$), strong correlations were found between dark green leaf and wide leaf, wide leaf and dwarfism, slender leaf and dwarfism, slender leaf and slender culm, long leaf and

long culm, long leaf and slender culm, short leaf and dwarfism, dwarfism and small seed, dwarfism and short panicle, long culm and big seeds, long culm and long panicle, thin culm and small seed, and thin culm and short panicle.

Variations in quantitative traits in wild type and mutant plants

We recorded three important agronomic quantitative traits—plant height, panicle number, and heading date—for each T1 plant from two cropping seasons in 2004 and 2005. The quantitative trait data were normally distributed for all three traits, with the heading date as the example (Fig. S2). ANOVA of the three parameters in the wild type in two cropping seasons and nine blocks revealed significant variations by season, block, or season by block (Supplementary Table 2). ANOVA of the parameters in the mutant population revealed higher variations by year, season, or year by season (Supplementary Table 3). Since the weather conditions in the first and the second cropping seasons and years differed, the field location and management would change the growth conditions. Information on year and cropping season for mutant lines is available in the searchable database.

Analysis of the three quantitative traits of wild type and mutant lines during the first or second cropping season revealed more variation in the mutant lines than that for the wild type (Table 3). Panicle number in the wild type ranged from 8 to 20 but 1–90 in the mutant lines. Plant height in the wild type ranged from 97 to 122 cm but 18–161 cm in the mutant lines. Heading date in the wild type ranged from 58 to 85 but 23–122 in the mutant lines.

Variations in plant type and seed morphology

Quantitative traits varied greatly in our TRIM population (Table 3). Here we use plant type and seed morphology as examples to illustrate the variations.

Table 2 An example of the phenotype record for the T1 plants of one mutant line, M00026806^a

	M00026806	M00026806_B	M00026806_C
Plant #	8	3	1
Plant height (cm)	101.5	71.7	69.0
Panicle number	28.8	18.3	24.0
Heading date	69.0	66.0	86.0
Phenotype records	Wild type	Semi-dwarf Spread-out tillers Small seeds	Narrow leaf Semi-dwarf Erect tillers Thin culm Sterile seeds Small seeds Late heading

^a This mutant line was propagated during the second cropping season of 2004, and 12 T1 plants were scored for phenotypes. Three sublimes were detected in this T1 line: M00026806 is the wild type, M00026806_B the first mutant group, and M00026806_C the second mutant group. Indicated is the number of plants belonging to each of the three groups

Table 3 Data on panicle number, plant height and heading date for T1 plants grown in the first (1st) and second (2nd) cropping seasons of 2004 and 2005

	Panicle numbers				Plant height (cm)				Heading date (days)			
	1st		2nd		1st		2nd		1st		2nd	
	WT	Mutant	WT	Mutant	WT	Mutant	WT	Mutant	WT	Mutant	WT	Mutant
Av.	12.88	14.88	13.39	16.84	109.84	104.76	106.84	96.00	89.12	85.95	63.97	67.59
SD	2.20	4.62	2.20	6.19	4.34	10.44	3.78	12.27	1.41	4.96	1.70	5.40
CV	17.06	31.07	16.42	36.78	3.95	9.96	3.54	12.78	1.58	5.77	2.66	8.00
Max.	19.83	85.60	20.33	90.50	122.33	161.50	117.17	138.00	95	122.00	71	111.00
Min.	8.50	1.00	8.83	1.00	98.00	21.50	97.83	18.00	86	65.23	58	23.00
95%	16.50	22.11	17.40	25.67	117.00	115.00	113.17	109.67	92	93.00	66	76.00
5%	9.33	9.25	10.33	9.50	102.50	85.00	101.17	71.00	87	79.18	61	61.00

Data are for 250 wild type (WT) and 11,560 mutant plants in the first cropping season, and 250 wild type and 11,493 mutant plants in the second season

Figure 1 shows photos of eight plants transplanted at the same time for the same growth period. The wild-type plant is at early heading stage, contains approximately 20 tillers and is almost 100 cm tall (panel A). Panels B to H illustrate mutants with decreased height (C, D) or tiller number (E), changes in leaf angle (G), tiller angle (B), leaf shape (C, D, F), and plant type (F, H).

The seed of the wild type illustrated in Fig. 2A is awnless and contains short glumes at both sides, and one lemma (wider) and one palea to protect the kernel inside. The length-to-width ratio of the seed is approximately 1.7, and the color is yellowish. Several obvious mutants show changes in seed size (C, E, F), seed shape (C, D), open hull (E, F), and presence of long glume (G, H), or awn (I, J).

Mutant segregation analysis

T1 plants show segregation in the field

Data on segregation rate were obtained from 12 T1 plants per line. Panel A of Fig. 3 shows two albino mutants of the 12 plants in a mutant line, which indicates a putative recessive trait. Panel B shows eight mutants with slender and curled leaves, representing a putative dominant trait. Segregation information in each mutant group is available in the searchable database.

Some T0 plants show dominant traits

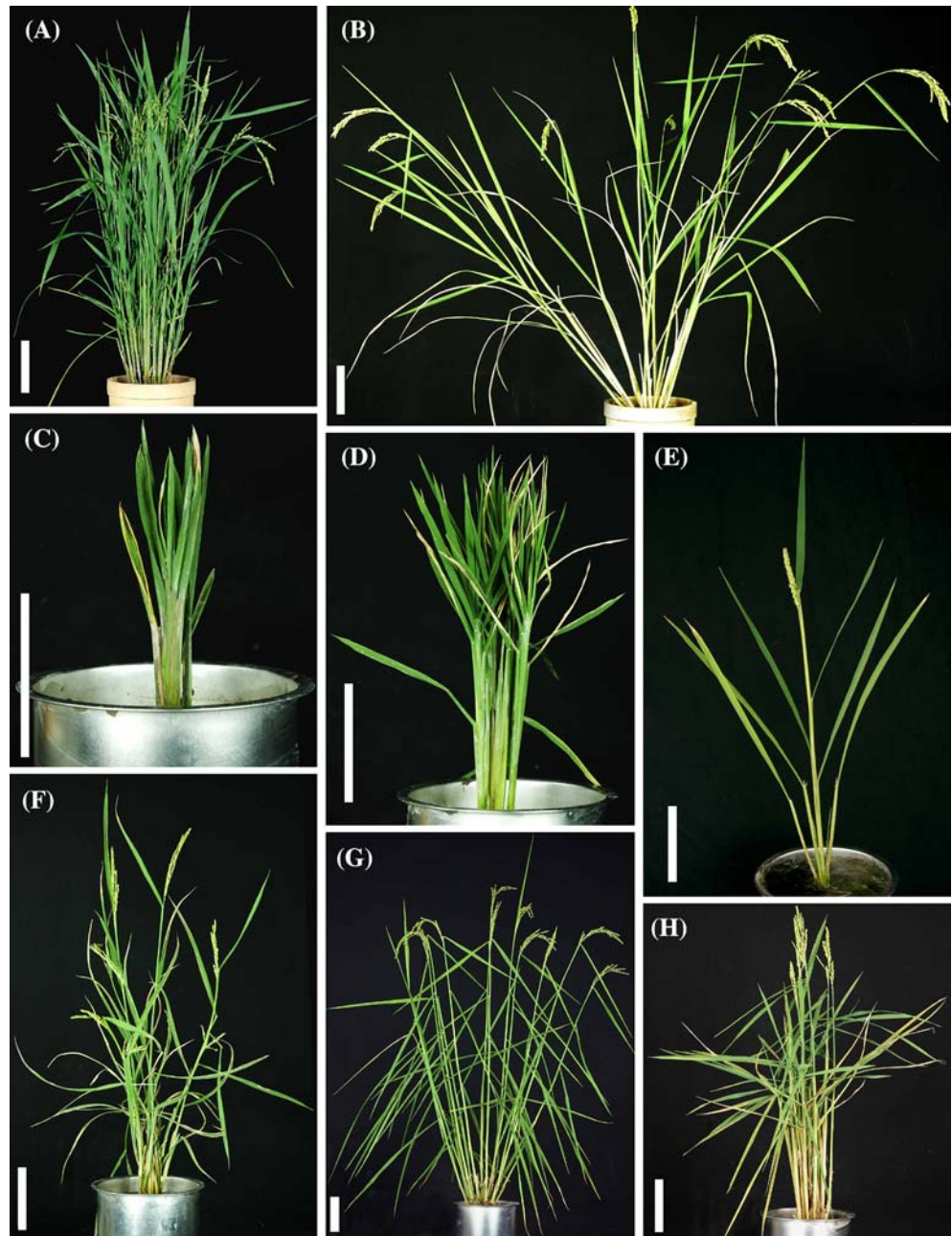
We use two binary vectors in our TRIM system. Tag4, with the function of gene trap and gene knockout, was used in the first ~18,000 lines. Tag8 was used beginning with mutant line number 18,000. It consists of eight copies of

35S enhancers and thus provides an additional function—activation tagging. Among these mutants, some T0 plants showed dominant mutant phenotypes. Figure 4 illustrates three examples. Panel A shows M0066614 (left), a severe dwarf plant, and M0066604 (right), a semidwarf plant, whose phenotype is quite similar to indica rice. Panel B shows M0055554, a dominant severe dwarf mutant with wide, dark-green, and twisted leaves. M0066614 and M0055554 did not produce viable seeds. From those producing seeds, the dominant phenotypes were observed in the following generation. The possibility of finding dominant mutants in the T0 population is about 1% in lines for which the tag8 vector was used.

T-DNA integration and mutant phenotype

Analysis of genotypes and phenotypes reveals co-segregation between a mutant phenotype and the results of T-DNA tagging. For instance, in our previous report (Hsing et al. 2007), we described a dominant mutant line M0047191, a GA-insensitive, severely dwarfed plant, with a T-DNA integrated 2.1 kb upstream of a GA2 oxidase gene. Gene expression analysis and hormone treatment response indicated a correlation between the adult phenotype and results of T-DNA integration. Another example is the mutant line M0055554 (Fig. 4A). Its T0 plant shows a dominant severe dwarfism, with dark-green and twisted leaves. Flanking sequence analysis revealed integration of a T-DNA 4.6 kb upstream of a *CYP72* homolog, a cytochrome P450 protein, which was demonstrated to encode an enzyme in a brassinolide synthesis pathway in Arabidopsis (Nakamura et al. 2005). In the activated line *chi2*, with the *CYP72C1* gene overexpressed, the mature Arabidopsis plant exhibited dominant phenotypic features such

Fig. 1 Examples of variation in T1 plant morphology. All eight plants were transplanted to the field and photos were taken on the same date. Panel **A**, the wild-type plant. Panel **B** illustrates a *lazy* mutant showing the spreading growth. Panel **C** shows an extremely dwarf plant, with short, thick, wide, and dark-green leaves. The leaf shape and ratio of leaf blade to leaf sheath differ from those of the wild type. The plant heads very late, with short panicles and small grains. Panel **D** shows another extremely dwarf plant, with short, narrow, and dark green leaves. The plant does not head. The plant in panel **E** contains only a single tiller (*monoculum*). Panel **F** illustrates a semidwarf plant with an abnormal twisted growth pattern. It has short, spiral leaves, with abnormal hull and small seeds. Panel **G** shows a mutant plant with normal plant height. The leaf collar is thick, with rolled and drooping leaves. Panel **H** illustrates a semidwarf plant with erect shoots. The leaf collar is thick, with a leaf-to-stem angle of 90°. The panicles are relatively short. Bar = 10 cm in each panel



as dwarfism, reduced male fertility, and dark-green, rounded epinastic leaves. Thus, the phenotype of both *Arabidopsis chi2* and rice M005554 mutants, which share the characteristics of brassinolide-deficient mutants, is caused by the enhanced activity of the inserted DNA.

Available website

We constructed a database containing information about the TRIM population. For each individual mutant line, three folders of information are available on flanking sequence, phenotype description, and T2 seed morphology.

This website contains a user-friendly search engine linked to functional genomics research. Users may search the flanking sequence database by BLAST or search putative knockout or activated genes by keyword. They also may search by use of the phenotype subcategory or quantitative traits mentioned above.

Distribution of the T2 seeds

Table 4 lists the current seed numbers stored in the seed stock center, with the mean, median, mode, maximum, and minimum numbers of seeds. The mean number of T1 seeds

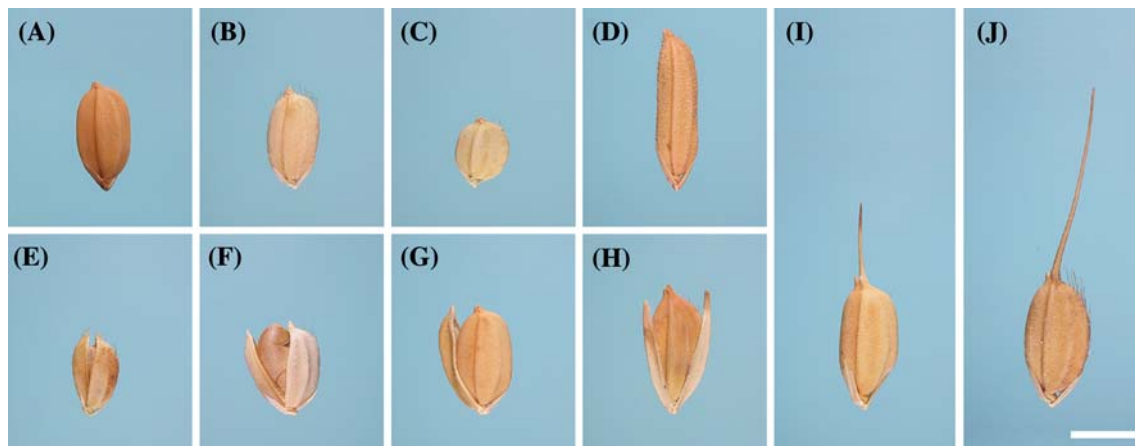


Fig. 2 Examples of variations in T2 seed morphology. Panel **A**, the seed of the wild type. Panel **B**, the seed color is lighter than that of the wild type. Panel **C** illustrates a small seed with light color and round shape, with length-to-width ratio of approximately 1.1. Panel **D** shows a long seed, with a length-to-width ratio of 3.6. Panel **E** shows a small

seed with an opened hull. Panel **F** shows another small opened seed. The seed in panel **G** has a single extra-long glume, whereas both of the glumes of the seed in panel **H** are extra long. The seed in panel **I** has a short awn, whereas the one in panel **J** has a very long awn. Bar = 5 mm in each panel

Fig. 3 Examples of phenotype segregation in T1 lines in the field. Panel **A**, 2 of the 12 T1 plants are albino (*arrow*), illustrating a recessive trait. Panel **B**, 8 of the 12 T1 plants have slender and curled leaves (*arrow*), illustrating a dominant trait



is 380 and T2 seeds approximately 10,000. Since the T1 seed number is limited, they will not be distributed. We will distribute 10 T2 seeds per line upon request. Detailed information is available on the NPGRC website.

Discussion

We propagated and collected phenotypic data by visual observation of 22,665 T-DNA-insertion rice T1 lines and thus provide a tool, the TRIM population, for phenome analysis of rice plants. Among these lines, 4,065 mutants had clearly visible phenotypes. We analyzed three quantitative traits. With the mutation phenotype information from this large-size mutant population, detailed analysis including mutation rates, high-frequency phenotypes, and correlated mutant traits were performed. All the phenotype data and segregation information, together with the searchable flanking sequence data, are available on the TRIM Web site. Thirty T2 seeds of each line are available for distribution.

TNG67 as an ideal material for rice insertional mutant study

TNG67 has been a popular *japonica* variety in Taiwan since the late 1970s. Because TNG67 is insensitive to temperature and photoperiod, it usually sets seed in a reasonable time (4 months). We used this variety to generate T-DNA insertional lines almost year round. Like most of the rice varieties in Taiwan, TNG67 is used for two cropping seasons each year. In the first cropping season, temperature and day length increase during the growth period. In the second growing season, temperature and day length decrease. Rice is a short-day plant, with a critical day length of approximately 15 h. Nipponbare, the variety used for the international genome sequencing work, is sensitive to both temperature and day length, and the growing condition has to be carefully controlled. Use of TNG67, however, would double the efficiency of field utilization and would not require additional artificial light for the promotion or prevention of heading. In addition, TNG67 was found to

Fig. 4 Example of dominant mutant phenotypes in T0 plants. Panel A. About four months after being transferred to the field, the left front mutant, M0066614 (*arrow*), showed a severe dwarf and the right front a semidwarf mutant, M0066604 (*arrowhead*). Panel B. About 45 days after transplantation, mutant line M0055554 showed dominant traits of severe dwarfism, with broad and curled leaves

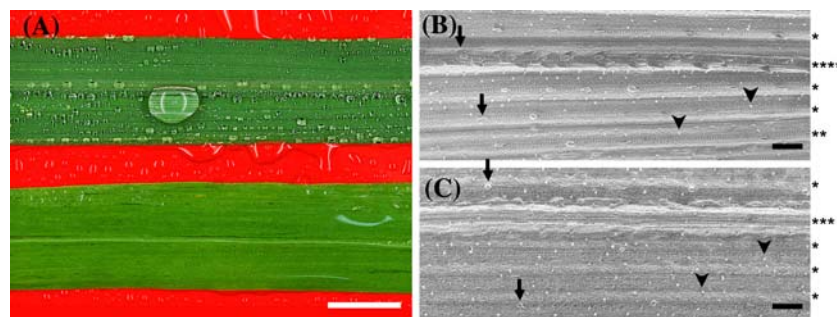
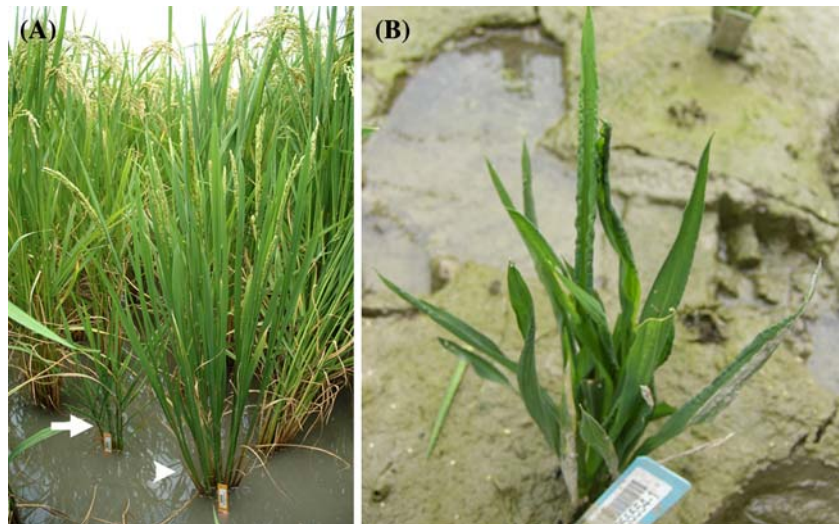


Fig. 5 A novel *trichome-less* mutant. Panel A. The upper leaf surface of the wild type (*upper panel*) and mutant (*lower panel*) after being sprayed with water. Bar = 1 cm. Panel B. Scanning electronic micrograph of the upper leaf surface of the wild type. Bar = 200 μ m. Panel C. Scanning electronic micrograph of the upper leaf surface of

the *trichome-less* mutant. Bar = 200 μ m. Arrows indicate the type I trichomes, and arrowheads indicate the type II trichomes. Stars at the right side indicate files of silica cells; the numbers of the stars indicate the line number in each file

Table 4 The T1 and T2 seed numbers harvested per mutant line

	Mean	Median	Mode	Maximum	Minimum
T1 seeds	380	245	80	4,470	0
T2 seeds	10,080	10,600	13,000	31,000	0

The data are calculated from 41,840 T1 lines and 14,709 T2 lines

harbor a mean of 0.12 newly transposed *Tos17* copies (Hsing et al. 2007), whereas mutant lines of Nipponbare (Guiderdoni, E., CIRAD) and Dongjin (An, G., POS-TECH) harbored a mean of 3–4 new copies, respectively (Guiderdoni et al. 2007). TNG67 has a reasonable rate of single-nucleotide polymorphism to Nipponbare (Hour et al. 2007), and locating flanking sequence tags to the exact integration site in the genome is not difficult (Hsing et al. 2007). All these factors make TNG67 ideal material for rice functional genomics study.

Experience from producing and phenotyping the TRIM population

Large field and intensive labor are required for propagation and screening

Our rice field underwent two cropping seasons each year. We have two senior and five regular rice breeders handling the field screening work. During transplanting/harvesting, more than 20 people are required to work in the field. With this huge manual effort, however, the speed of propagation

for T1 plants still cannot match the generation of T0 plants from the tissue culture lab. Currently, the T0 plant number is higher than 70,000, but we only propagate about 30,000 T1 lines. Furthermore, since the paddy field is a net house, the weather affects the growth conditions. An expanded field may help to speed up the propagation.

Several well-trained rice breeders are required for the phenotype-screening work

Five well-trained rice breeders worked each day in the field. Since the area of the field is limited, we used the same field continuously. Unwanted “volunteer” rice may contaminate the field if not removed. We planted all the rice plants in a 25 × 25 cm array; thus, plants growing at unexpected location could be recognized and discarded. The breeders removed plants at least once a week before the heavy tiller stage (i.e., the close of canopy). As well, breeders had to be familiar with the phenotypic traits, because they were grading the phenotypes in the large field.

Careful investigation leads to discovery of novel mutation phenotype

During the screening process after a rainfall, we noticed that for one particular T1 line, 3 of the 12 plants showed interesting phenotypes. After a rain, the surface of normal rice leaves showed many small dew drops (Fig. 5A). However, the mutant leaf surface contained only a thin film of water. Scanning electron microscopy of the upper side of a normal rice leaf as shown in Fig. 5B revealed two kinds of trichomes: type I—large, hooked, unidirectional, located on the surface of files of silica cells, 1–4 lines per file, over a thin vascular bundle; and type II—small, random-directional, located randomly. The density of type I trichomes is very high in wild-type leaves but much less in mutant leaves (Fig. 5C), with a ratio of type I trichomes in the wild type to mutant of about 10:1. The same is true for the lower leaf side (data not shown). With few type I trichomes, the surface tension is reduced, so the water on the leaf surface has a different shape in this *trichome-less* mutant.

Good seed stock facilities required

The T1 and T2 seeds prepared in the current work are treasures for rice functional genomic research. They are stored in National Plant Genetic Resources Center, where both temperature and relative humidity are kept low for

long-term storage. The seed viability may be at least 50 years under this condition. As well, seed packages are marked and tagged with a barcode. All information on seed registration, storage, and distribution record is available on NPGRC website.

Comparison and cross-talk of the phenotypes with other rice mutant populations

Several rice mutant phenotype databases are now available, such as mutant populations of *Tos17*-tagged Nipponbare (Miyao et al. 2004), T-DNA tagged Zhonghua 11 (Zhang et al. 2006), and chemical- and irradiation-induced IR64 (Wu et al. 2005). In addition, a unified vocabulary for plant structure ontology was recently suggested (Yamazaki and Jaiswal 2005; Ilic et al. 2007). Each group uses different descriptions and codes for mutant traits. Supplementary Table 4 shows the link/comparison among these groups. A unified vocabulary for the mutant collections should be promoted so that the mutant traits from different groups may be compared.

Conclusion

In the present work, we generated a large T-DNA tagged rice mutant population, the TRIM population. With intensive phenomics study, the population is ready for large-scale functional genomics analysis. The TRIM population has many advantages: (1) use of a tri-function vector for promoter trapping, gene knockout and activation may have produced novel mutations not obtainable with traditional mutagenesis or *Tos17/Ac/Ds/T-DNA* knockout tagging strategies; (2) users may search the database by several different approaches, including sequence search using BLAST, putative knockout or activated gene search with keywords, or plant/seed phenotype search; (3) the database provides genetic segregation data; (4) many mutant phenotypes are dominant; (5) TNG67, a local *japonica* variety, used as the material for generating the mutants, is not sensitive to photoperiod or temperature, so the mutant lines have wide application; (6) all the T1 seeds and some of the T2 seeds are stored under standard long-term storage conditions; and finally, (7) the database resource can support large-scale phenomics study.

There are some drawbacks for use of the TRIM population: (1) the propagation of T1 plants and thus T2 seeds is slow, since we have to grow them in a field specific for GM organisms; (2) production of T0 and T1 plants in the field may be negatively influenced by environmental factors; (3) DNA for FST determination was extracted from T0 plants just before they were transplanted to the field, and some

plants were lost by field work or environmental disorders; (4) a TAIL-PCR method was used to retrieve the FST information because of its ability to be performed on a large scale; however, the success rate is about 60%, and the information on the integration sites of many mutant lines is still missing. Currently, we are applying adaptor ligation PCR methods to the lines showing mutant traits, but the progress is relatively slow; and (5) the tagging rate (<10%) is low, as in the other tagged rice mutant populations.

Acknowledgements We greatly acknowledge the skilled technical support of Dr. Hua-Mei Chang and Mr. Cheng-Tung Lo from Vita Genomics in FST generation. We also thank Drs. Anthony H.C. Huang and Laura Heraty for discussion and critical review of this manuscript. This work was supported by grants from Academia Sinica (94S-1502 and 94F005-1 to SM Yu and 94S-1501 and 94F005-3 to YI Hsing) and the National Science Council (NSC94-2317-B-055-003, NSC95-2317-B-055-002, NSC96-2317-B-055-001 to CG Chern, NSC95-2317-B-055-001, NSC95-2317-B-055-002, NSC95-2317-B-055-003, and NSC5-2317-B-468-002 to MJ Fan; NSC94-2321-001-023 and NSC95-2321-B-001-015 to SM Yu) of Taiwan.

References

- AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- An G, Lee S, Kim SH et al (2005) Molecular genetics using T-DNA in rice. *Plant Cell Physiol* 46:14–22
- Chen S, Jin W, Wang M et al (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. *Plant J* 36:105–113
- Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. *Science* 282:656–659
- Gibbings JG, Cook BP, Dufault MR et al (2003) Global transcript analysis of rice leaf and seed using SAGE technology. *Plant Biotechnol J* 1:271–285
- Guideroni E, An G, Yu S-M et al (2007) T-DNA insertion mutants as a resource for rice functional genomics. In: Upadhyaya NM (ed) *Rice functional genomics – challenges, progress and prospects*. Springer, pp 181–221
- Haas BJ, Wortman JR, Ronning CM et al (2005) Complete reannotation of the *Arabidopsis* genome: methods, tools, protocols and the final release. *BMC Biol* 3:7
- Hirochika H (2001) Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr Opin Plant Biol* 4:118–122
- Hirochika H, Guideroni E, An G et al (2004) Rice mutant resources for gene discovery. *Plant Mol Biol* 54:325–334
- Hour AL, Lin YC, Lu WF et al (2007) Detection of SNPs between Tainung 67 and Nipponbare rice cultivars. *Bot Stud* 48:243–253
- Hsing YI, Chern CG, Fan MJ et al (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol Biol* 63:351–364
- Ilic K, Kellogg EA, Jaiswal P et al (2007) The plant structure ontology, a unified vocabulary of anatomy and morphology of a flowering plant. *Plant Physiol* 143:587–599
- IRGSP (2005) The map-based sequence of the rice genome. *Nature* 436:793–800
- Itoh T, Tanaka T, Barrero RA et al (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res* 17:175–183
- Jeong DH, An S, Kang HG et al (2002) T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol* 130:1636–1644
- Jeong DH, An S, Park S et al (2006) Generation of a flanking sequence-tag database for activation-tagging lines in *japonica* rice. *Plant J* 45:123–132
- Kikuchi S, Satoh K, Nagata T et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science* 301:376–379
- Kolesnik T, Szeverenyi I, Bachmann D et al (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *Plant J* 37:301–314
- Komatsu S, Tanaka N (2005) Rice proteome analysis: a step toward functional analysis of the rice genome. *Proteomics* 5:938–949
- Lai MH, Chen CC, Kuo YC et al (1996) The relationship between grain productivity and nitrogen-fertilizer rate of currently cultivated rice cultivars. I. The effects of different nitrogen rates on grain yield and yield components in rice. *J Agric Res China* 45:203–207
- Li L, Wang X, Stolc V et al (2006) Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet* 38:124–129
- Miyao A, Tanaka K, Murata K et al (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
- Nakamura M, Satoh T, Tanaka S et al (2005) Activation of the cytochrome P450 gene, CYP72C1, reduces the levels of active brassinosteroids in vivo. *J Exp Bot* 56:833–840
- Nakano M, Nobuta K, Vemaraju K et al (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Res* 34:D731–735
- Ouyang S, Zhu W, Hamilton J et al (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res* 35:D883–887
- 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
- Su C-L, Chung C-I, Lin Y-C et al (2005) Statistical analysis of rice SAGE data. *J Genet Mol Biol* 16:248–260
- Upadhyaya NM, Zhu QH, Zhou XR 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
- van Enckevort LJ, Droc G, Piffanelli P et al (2005) EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice. *Plant Mol Biol* 59:99–110
- Wasaki J, Shinano T, Onishi K et al (2006) Transcriptomic analysis indicates putative metabolic changes caused by manipulation of phosphorus availability in rice leaves. *J Exp Bot* 57:2049–2059
- Wu C, Li X, Yuan W et al (2003) Development of enhancer trap lines for functional analysis of the rice genome. *Plant J* 35:418–427
- Wu JL, Wu C, Lei C et al (2005) Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol Biol* 59:85–97
- Yamazaki Y, Jaiswal P (2005) Biological ontologies in rice databases. An introduction to the activities in Gramene and Oryzabase. *Plant Cell Physiol* 46:63–68
- Zhang J, Li C, Wu C et al (2006) RMD: a rice mutant database for functional analysis of the rice genome. *Nucleic Acids Res* 34:D745–748