

Studies on rice seed quality through analysis of a large-scale T-DNA insertion population

Fang-Fang Fu^{1,*}, Rui Ye^{1,*}, Shu-Ping Xu¹, Hong-Wei Xue¹

¹National Key Laboratory of Plant Molecular Genetics, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Science (SIBS), Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China

A rice (*Oryza sativa*) T-DNA insertion population, which included more than 63 000 independent transgenic lines and 8 840 identified flanking sequence tags (FSTs) that were mapped onto the rice genome, was developed to systematically study the rice seed quality control. Genome-wide analysis of the FST distribution showed that T-DNA insertions were positively correlated with expressed genes, but negatively with transposable elements and small RNAs. In addition, the recovered T-DNAs were preferentially located at the untranslated region of the expressed genes. More than 11 000 putative homozygous lines were obtained through multi-generations of planting and resistance screening, and measurement of seed quality of around half of them, including the contents of starch, amylose, protein and fat, with a nondestructive near-infrared spectroscopy method, identified 551 mutants with unique or multiple altered parameters of seed quality. Analysis of the corresponding FSTs showed that genes participating in diverse functions, including metabolic processes and transcriptional regulation, were involved, indicating that seed quality is regulated by a complex network.

Keywords: rice, T-DNA insertion population, seed quality, near-infrared spectroscopy

Cell Research (2009) 19:380-391. doi: 10.1038/cr.2009.15; published online 17 February 2009

Introduction

Rice (*Oryza sativa*), as a wholesome and nutritious cereal grain, provides the staple food for more than half of the world's population [1]. Along with the growth of economies and population throughout the world, there is an increasing demand for high-yield and high-quality rice. The grain quality of rice, including that of milling, appearance, cooking and eating, and nutrition, is determined by multiple physicochemical properties [2], and studies on the involved key genes and relevant regulatory mechanisms will help to illustrate the regulatory networks of rice quality control and benefit for the breeding efforts.

Starch, which is composed of amylose and amylopectin, comprises ~90% of dry endosperm of rice seed.

The apparent amylose content (AAC) is recognized as an important determinant of the appearance and texture of grain [3]. Thai jasmine rice (KDML 105), one of the best-quality rice with good cooking and eating qualities, has low amylose content (AC) and medium gel consistency (GC) [4].

The nutritional quality includes the proportions of protein, fat, mineral and other substances [5]. The rice protein is rich in methionine and cysteine, and compares favorably with proteins of other cereals, but is poor in lysine and threonine, making it an incomplete protein source for human infants [6]. Storage triacylglycerols are stored in oil bodies containing phospholipids and oleosins at the surface [7], and oil bodies are abundant in embryo and aleurone layer of rice seed. However, the aleurone layer is usually removed by milling because it turns rancid upon storage [8]. In addition, oil also influences other aspects of seed quality, and a recent report showed that the quality and viability of *Arabidopsis* seeds can be enhanced by suppressing phospholipase D [9].

The near-infrared spectroscopy (NIRS) technology, which is based on the fact that several natural products can readily absorb NIR radiation at a specific region or

*These two authors contributed equally to this work.

Correspondence: Hong-Wei Xue

Tel: +86-21-54924059; Fax: +86-21-54924060

E-mail: hwxue@sibs.ac.cn

Received 16 March 2008; revised 9 June 2008; accepted 18 August 2008; published online 17 February 2009

wavelength, has been used for rapid and accurate analysis of chemical and physical properties without sample destruction [10]. NIRS is time-saving and low-cost, and previous studies have used NIRS to determine quality characteristics such as moisture, protein, starch, amylose and lipids in buckwheat, wheat, barley, maize, soybean and rice [11-17]; the practice was widely accepted in the trade of these products.

To date, many genes related to important agronomic traits and seed qualities have been identified through breeding and genetic transformation protocols [1, 5, 18], and quantitative trait locus (QTL) analysis and map-based cloning have been applied to identify loci associated with seed quality control [19, 20]. Chinese geneticists and breeders have made great progress in the hybrid rice technology [21], and marker-assisted selection has been widely used to pyramid functional genes into popular hybrid rice cultivars to improve the seed quality [22, 23]. As a multifactorial trait, modulations of genes involved in photosynthesis, plant architecture and transcriptional networks controlling plant development were proved to be effective in improving seed quality [24]. However, systemic studies on rice seed quality control and the corresponding rice resources are still lacking.

Generation of T-DNA insertion populations is an important strategy to study gene functions on a large-scale basis [25]. Currently, several databases for *Arabidopsis* have been established to collect information on T-DNA insertion populations, including SALK (<http://signal.salk.edu/cgi-bin/tdnaexpress>), GABI-Kat (<http://www.gabi-kat.de/>) and SAIL (http://www.tmri.org/en/partnership/sail_collection.aspx) [26-28]. For rice, several resources with large collections of insertion mutants were also generated, including those in Australia (<http://www.pi.csiro.au/fgtrtpub/>) [29], China (<http://rmd.ncpgr.cn/>; <http://trim.sinica.edu.tw/>) [30-32], Europe (<http://orygenesdb.cirad.fr/>) [33], France (<http://urgi.versailles.inra.fr/OryzaTagLine/>) [34], Japan (<http://pfg101.nias.affrc.go.jp/~miyao/pub/tos17/index.html.en>) [35, 36], Korea (<http://141.223.132.44/pfg/>) [37, 38] and USA (<http://sundarlab.ucdavis.edu/>) [39]. These resources helped to identify and study genes involved in rice development [40, 41]. Some mutants with marked changes of seed quality were also identified [42, 43].

To facilitate the coordination of rice functional genomic studies, especially that on seed quality regulation, we have developed the Shanghai T-DNA Insertion Population (SHIP, <http://ship.plantsignal.cn>) of rice by large-scale *Agrobacterium*-mediated transformation. The flanking sequences of T-DNAs were identified by thermal asymmetric interlaced-PCR (TAIL-PCR) [44, 45] and mapped onto the rice genome. After multiple genera-

tions of planting and resistance screening, more than 11 000 putative homozygous lines were obtained and analysis on the seed quality of around half of them with NIRS identified a total of 551 mutants with altered seed quality. Further analysis on the corresponding flanking sequence tags (FSTs) revealed that genes of diverse molecular functions are involved in seed quality control.

Results

Generation of a rice mutant population of 63 000 T-DNA insertion lines

A project aiming to generate a large T-DNA insertion population was initiated in 2003. To date, using Zhonghua 11 or Nipponbare, two rice *japonica* varieties, as materials, more than 63 000 independent transgenic lines have been generated through *Agrobacterium*-mediated transformation. In all, 15 000 independent lines were analyzed by TAIL-PCR to identify the T-DNA FSTs. Based on the finding from previous studies that isolation of FSTs from the left border was more efficient than that from the right one, due to a higher frequency of inverted T-DNA repeats involving two right borders [46], a set of nested sequence-specific primers corresponding to the left border of T-DNA and a shorter arbitrary degenerate (AD) primer were utilized in the present study. Overall, 12 948 FSTs were isolated till February 2008. A detailed analysis of these sequences by aligning against the rice pseudo-chromosome version 5 (The Institute for Genome Research (TIGR)) indicated that 8 840 (68.3%) of them were mapped onto the rice genome with an average 96.8% homology (all the FSTs have been submitted to RiceGE, <http://signal.salk.edu/cgi-bin/RiceGE>). After removing the T-DNA footprint, the size of the genomic sequence of these FSTs averaged at 241 bp. The other FSTs included 2 127 (16.4%) T-DNA tandem repeats, 1 564 (12.1%) binary vector sequences and sequences that did not hit at all (3.2%, probably due to non-specific amplification).

To test the accuracy of FSTs (whether the T-DNA insertion was indeed at the predicted position), PCR amplification was performed using a primer pair, of which one is the T-DNA left border primer and the other is a reverse primer corresponding to the isolated FSTs. The results showed that, among the 118 lines tested, T-DNA insertions of 69 lines (58.5%) were confirmed to be at the predicted position, which is relatively lower than that in *Arabidopsis* [28].

Further, to analyze the locus numbers of T-DNAs, T1 seeds of 2 097 independent lines were germinated in sterile water containing hygromycin and those with a segregation ratio of 3 resistant:1 sensitive were regarded

as containing a single locus of T-DNA. This preliminary analysis showed that around 64.7% (1 357 lines) of the transgenic lines contain a single locus of T-DNA.

Distribution of T-DNA insertions on rice genome

Assignment of 8 840 FSTs on chromosomes showed that FSTs are distributed all over the chromosomes; however, the distribution was not even and the number of insertions was related to the chromosome size. Calculation of the T-DNA density (per Mb) showed the range from 15.60 (Chr. 12) to 32.27 (Chr. 3), with an average of 23.76, indicating a positive correlation with the chromosome size. The relatively longer chromosomes, 1 and 3, had an apparently higher density of T-DNA insertions (i.e., contained 28.03% of the total number of insertions), while chromosomes 6, 7 and 10-12 had a relatively lower density (Table 1). Linear regression analysis indicated that the chromosome size was significantly positively correlated with the insertion numbers (correlation coefficient = 0.924, $P < 0.001$), as well as with the insertion density (correlation coefficient = 0.729, $P < 0.01$) (Supplementary information, Figure S1).

Further analysis on the distributional relationship between T-DNA insertions and several other features of the rice genome, including expressed genes and transposable elements (annotated by TIGR), was performed. In addition, as small RNAs were proposed to regulate transposon mobility and assembly of heterochromatin [47], the distribution of small RNAs [publicly available data, 48] was also analyzed. The whole genome was divided into non-overlapping windows with the size of 500 kb,

and the numbers of T-DNA insertions, expressed genes, transposable elements and small RNAs mapping to each window were determined, which were then used for analysis of Pearson's correlation coefficient (PCC). The results revealed that the distribution of T-DNA insertions on the whole genome was positively correlated with expressed genes ($PCC = 0.39$, $P < 0.001$), but negatively correlated with transposable elements ($PCC = -0.25$, $P < 0.001$) and small RNAs ($PCC = -0.37$, $P < 0.001$). These correlations show the same trend, but different degrees for each individual chromosome (relatively higher on chromosomes 6 and 7, while lower on chromosomes 2 and 11, Figure 1).

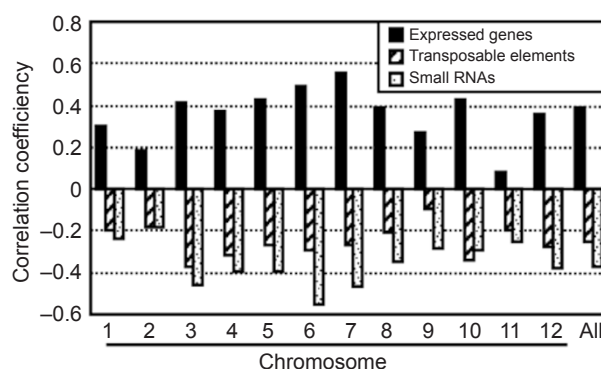


Figure 1 The distribution correlation between T-DNA insertions and expressed genes (black column), transposable elements (striped column) and small RNAs (dot column). "All" indicates the whole genome.

Table 1 Distribution of T-DNA insertions on rice chromosomes

Chr. no.	Chr. size (Mb) ¹	Insertions	Insertion density ²	Frequency (%) ³
1	43.60	1305	29.93	14.76
2	35.93	907	25.24	10.26
3	36.35	1173	32.27	13.27
4	35.24	916	25.99	10.36
5	29.87	797	26.68	9.02
6	31.25	640	20.48	7.24
7	29.69	566	19.06	6.40
8	28.31	682	24.09	7.72
9	23.01	517	22.47	5.85
10	22.88	394	17.22	4.46
11	28.46	514	18.06	5.81
12	27.50	429	15.60	4.85
Total	372.09	8840	23.76	100

¹TIGR rice genome data in release 5.

²Number of T-DNA insertions per Mb chromosome size.

³T-DNA insertions/total insertions.

As T-DNA insertions were positively correlated with expressed genes, the distribution of T-DNA insertions within genes and intergenic regions was further analyzed. To facilitate the analysis, the DNA sequence from ATG to the stop codon, together with the sequence upstream (2 500 bp) from ATG and downstream (750 bp) from the stop codon, was defined as a gene, and the segments which were not included in genes were defined as intergenic regions. Overall, 6 586 (74.50%) T-DNAs located in genes and 2 254 (25.50%) in intergenic regions. Calculation of the insertion density revealed that recovered T-DNAs were preferentially located at the untranslated regions of genes, and that the density of T-DNA insertions near the gene-coding region was higher than the densities of those far from the coding region (Table 2).

Functional annotation (Gene Ontology, “molecular function”) showed the tagged genes fall into multiple functions, and interestingly, some categories were significantly over-represented, including those for hydrolase, nucleic acid binding and motor functions, while some, such as genes in the categories of transcription factor, nuclease, transferase and oxygen binding, were under-represented (Table 3).

Identification of putative homozygous mutant lines and characterization of mutants with altered seed qualities by NIRS

Considering that T-DNA insertion mutagenesis was recessive, and that some phenotypes could not be exhibited in the heterozygous T1 plants, we tried to identify putative homozygous lines by resistance selection and

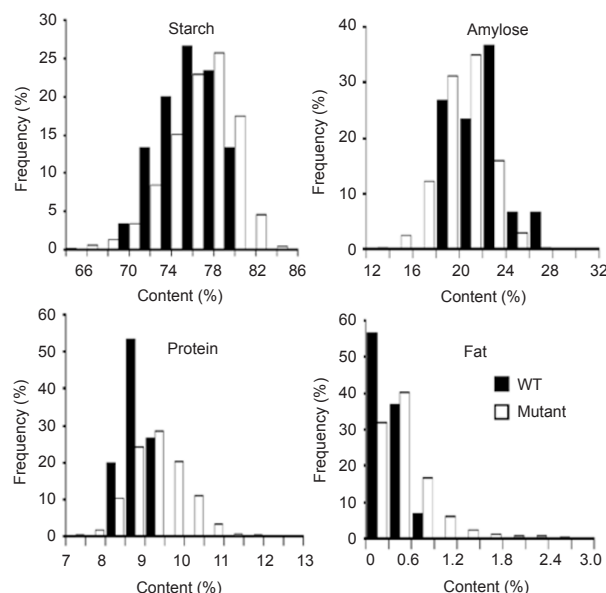


Figure 2 Distribution plots of wild-type (black column) and homozygote population (white column) for starch, amylose, protein and fat contents (for wild type, *n* = 50; for homozygote population, *n* = 5 868).

use them for further analysis. Overall, more than 11 000 putative homozygous lines were obtained, and by using primers located at the T-DNA left border and two sides of the predicted insertion sites of FSTs, 49 (71.0%) of 69 putative homozygous lines were confirmed as homozygous at the insertion positions. This provides a good

Table 2 Distribution of T-DNA insertions in the genes and intergenic regions of rice genome

Annotation	Total size (Mb)	Insertions	Insertion density (per Mb)	Frequency (%)
Coding region	146.82	2128	14.49	24.07
Exon	76.73	916	11.94	10.36
Intron	70.09	1212	17.29	13.71
Downstream (~250 bp)	13.90	490	35.25	5.54
Downstream (250-500 bp)	12.68	420	33.12	4.75
Downstream (500-750 bp)	11.28	380	33.69	4.30
Upstream (~500 bp)	27.46	854	31.10	9.66
Upstream (-500-1500 bp)	45.20	1384	30.62	15.66
Upstream (-1500-2500 bp)	31.51	930	29.51	10.52
Total in gene	281.52	6586	23.39	74.50
Intergenic	90.56	2254	24.89	25.50
Total	372.08	8840	23.76	100

“Coding region” is defined as the sequence from ATG to stop codon; “downstream and upstream” represents the 3’ direction relative to the stop codon and 5’ direction relative to ATG, respectively; “gene” is defined as a sequence between 2 500 bp upstream from ATG and 750 bp downstream from the stop codon; “Intergenic” is defined as a sequence in which no predicted gene has been found within 2 500 bp upstream from ATG or within 750 bp downstream from the stop codon.

Table 3 Gene Ontology annotation of FSTs of the insertion population that mapped to the genic region (The GO terms were based on the TIGR product annotations.)

GO terms	FSTs		TIGR ¹		P-value ²
	No.	(%)	No.	(%)	
Hydrolase	657	12.3	2816	10.49	0
Kinase	399	7.47	2197	8.19	0.06
Nuclease	7	0.13	116	0.43	0
Transferase	294	5.51	1850	6.89	0
Other catalytic	510	9.55	2620	9.76	0.6
Carbohydrate binding	175	3.28	864	3.22	0.81
Chromatin binding	7	0.13	25	0.09	0.36
Lipid binding	44	0.82	269	1.00	0.19
Nucleic acid binding	569	10.66	2457	9.15	0
Nucleotide binding	471	8.82	2323	8.66	0.67
Oxygen binding	65	1.22	444	1.65	0.01
Protein binding	557	10.43	2586	9.64	0.05
Other binding	330	6.18	1623	6.05	0.68
Enzyme regulator	34	0.64	169	0.63	0.95
Motor	36	0.67	109	0.41	0
Receptor	82	1.54	471	1.76	0.22
Signal transducer	74	1.39	375	1.40	0.94
Structural molecule	120	2.25	615	2.29	0.83
Transcription factor	329	6.16	1854	6.91	0.03
Transcription regulator	60	1.12	375	1.40	0.09
Translation factor activity	29	0.54	130	0.48	0.54
Transporter	254	4.76	1179	4.39	0.19
Molecular function unknown	237	4.44	1372	5.11	0.03
Total	5340	100	26839	100	—

¹The number and frequency of each GO term in the entire genome of TIGR annotation.

²P-value of the chi-square test. P-value < 0.05 indicates the over-represented or under-represented functional categories in this population.

Table 4 The seed quality values through near-infrared spectroscopy measurement

		Starch (%)			Amylose (%)			Protein (%)			Fat (%)		
		Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD
Value	WT	69.06-79.38	75.02	2.62	18.08-26.56	21.83	2.17	8.35-9.34	8.79	0.31	0.12-0.67	0.32	0.15
	Mu	64.28-85.79	77.51	3.15	12.87-30.45	20.18	2.11	7.28-13.07	9.28	0.69	0.00-2.92	0.51	0.37
Criteria		<70 or >82			<18 or >24			<7.5 or >11			>1.1		
No. of	1st	287 (4.9%)			331 (5.6%)			57 (0.1%)			262 (4.5%)		
candidate Mu	2nd	252 (4.3%)			162 (2.8%)			50 (0.1%)			188 (3.2%)		
Repetition ratio		87.8%			48.9%			87.7%			71.8%		

The average data of WT seeds were obtained from 50 individual samples and the criterion was set as the 99% confidence interval limits of the data (mean ± 2.58 × SD) [51]. In total 5 868 homozygous mutant lines were analyzed.

resource for functional genomic studies through forward genetics (<http://www.plantsignal.cn>).

The obtained putative homozygous lines were further

used for seed quality analysis by measuring the contents of starch, amylose, protein and fat through NIRS. The calibration equations were developed and validated by

Foss Company, and the statistical indices of the calibration and cross-validation set are listed in Supplementary information, Table S1. Overall, 5 868 putative homozygous lines were analyzed (the data distributions are shown in Figure 2). For each of four constituents, the mean values showed no significant difference between the homozygote population and WT (Zhonghua 11); however, the whole data set of the homozygote population showed wider ranges. The mean content of starch in Zhonghua 11 was 75.02%, ranging from 69.06% to 79.38%, which was consistent with the known content of starch in *japonica* varieties (70-80%), while that of the homozygote population was 77.51%, ranging from 64.28% to 85.79%. Contents of amylose, protein and fat also showed similar mean values and wider ranges in the homozygote population (Table 4).

Shapiro-Wilk test revealed that WT followed the normal distribution of quality ($P > 0.05$). Based on this, the 99% confidence interval limit was used as the threshold of mutant selection. In total, 287 lines with altered starch content (4.9% of the total tested lines) were selected in the first-round measurement; among them, 252 lines showed the same trend in the second-round measurement, with a repeat rate of 87.8%. Similarly, 162 lines (2.8%) were identified related to amylose changes. The reported protein content in WT seeds is $8.50 \pm 0.35\%$ ([49], similar to that in our analysis, $8.79 \pm 0.31\%$), and finally 50 lines with altered protein contents were identified. As the fat content in milled rice is very low, the putative mutants with higher fat content ($> 1.1\%$) were identified (188 lines, 3.2%) (Table 4).

Among the identified lines, some showed changes in a single constituent, whereas some showed changes in two or three constituents of seeds (Figure 3), suggesting that the corresponding genes may be involved in some

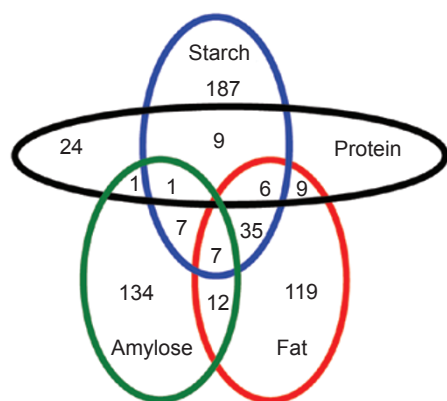


Figure 3 Venn diagrams show overlapping mutant lines with changes in the four constituents of seed quality.

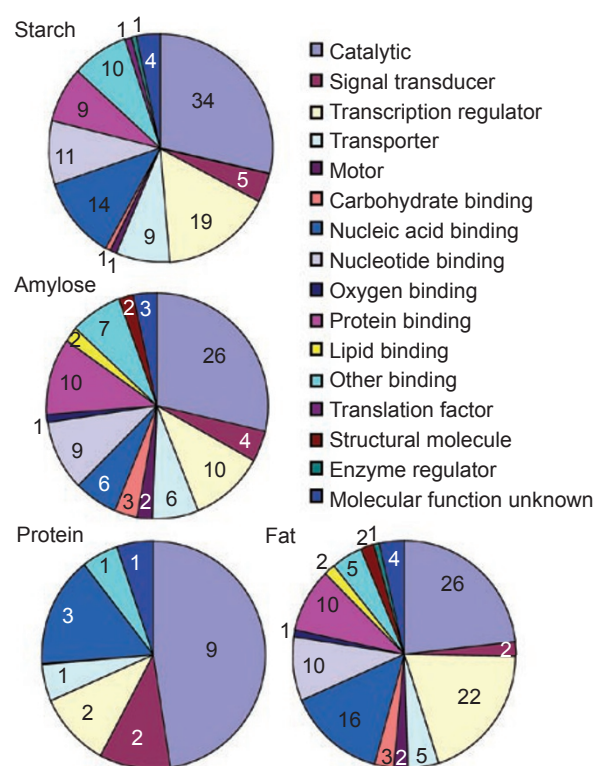


Figure 4 Function annotation of mutant genes associated with seed quality changes by "molecular function" using the Gene Ontology according to the TIGR model.

form of common regulation during the synthesis and metabolism of these constituents, or that perturbation in biosynthesis/metabolism of one constituent may result in the altered content of another. Among 551 lines, 58 lines showed changes in starch and fat contents, agreeing well with the fact that starch metabolism affects fat content [50]. Few mutants had changes in both protein and amylose, implying that the regulation of synthesis and metabolism between protein and amylose was more independent. The number of mutants with altered starch and ACs was 15, similar to the number of mutants with altered protein and fat contents (Figure 3).

Functional categories of genes related to seed quality control

The corresponding mutated genes of the 551 lines were analyzed using the Gene Ontology annotation, and the chi-square test was performed to determine whether the genes have functional preference. The results showed that genes with functions of signal transducer, catalytic, transcription regulator and transporter were present in all kinds of mutants (Figure 4), suggesting that the alteration of seed quality may be due to the modification

of some common genetic factors. In addition, some GO categories were present in one or two kinds of mutants exclusively, and were functionally closely related. For example, many encoded proteins with carbohydrate-binding activity were identified in mutants with altered starch and ACs, while those with lipid-binding activity were enriched in mutants with altered fat contents, indicating that such mutants may be attributable to genes directly involved in the cognate synthesis or metabolism pathways. Other genes have diverse molecular functions, including protein-binding and enzyme regulator activities.

As expected, some of the mutant genes have been reported to be involved in seed quality regulation, and their defective expression resulted in altered seed quality (Table 5). Most of these genes encode enzymes involved in the metabolism pathway of nutrients, and PCR analysis confirmed the T-DNA insertion in these genes (Figure 5A). Further, considering that seed quality can be influenced by many factors, such as the environment or somaclonal variation, mutant line SHIP_ZSF0781 was selected to test the genetic linkage between the insertion and phenotype. This line had an insertion at the upstream of gene Os08g37800, which encodes a carbon catabolite derepressing protein kinase. The homozygote lines contained higher content of starch, and comparison between the homozygote of T2 generation and several independent progeny lines from one heterozygous T1 plant confirmed the genetic linkage between insertion and the increased starch content. Among eight progeny, four (Figure 5B, 4-7) were homozygous (confirmed by PCR analysis, lower panel) and had similarly increased starch content as the homozygote (Figure 5B, lane 2),

and three heterozygous lines (Figure 5B, 8-10) showed slightly increased starch contents.

Data collection of SHIP

All information collected has been integrated into the database SHIP (<http://ship.plantsignal.cn>), which includes 8 840 FSTs mapping on the rice genome, 11 000 putative homozygous lines, seed quality data of 5 868 putative homozygous lines analyzed with NIRS and some observable phenotypes during growth under agronomical conditions. The SHIP database can be accessed freely and searched using keywords or through the BLAST program, and will be updated regularly.

Discussion

Distribution of T-DNA insertions on chromosomes

Our study showed that the distribution of T-DNA insertions was correlated with the chromosome size, which is consistent with previous reports that T-DNA insertions are not randomly distributed on the chromosomes [57, 58]. The larger chromosomes contain more T-DNA insertions and also show higher insertion density (Table 1; Supplementary information, Figure S1). This may be due to the fact that large chromosomes have higher euchromatin, into which the T-DNAs preferentially integrate [59]. In addition, the T-DNA distribution was positively correlated with expressed genes but negatively with transposable elements and small RNAs. Indeed, it has been suggested that the positive correlation between expressed genes and T-DNA insertions may be due to the “open” state of actively transcribed regions, rendering the DNA in these regions more accessible [58, 60, 61].

Table 5 Reported genes that are involved in the seed quality (s) control in the characterized mutants

Quality	Line ID	Starch (%)	Amylose (%)	Protein (%)	Fat (%)	Gene locus	Annotated genes	Ref.
Starch	SHIP_ZSF0092	82.02 (75.02)	23.53 (21.83)	8.02 (8.79)	0.09 (0.32)	Os11g29400	6-Phosphogluconate dehydrogenase	[51]
	SHIP_ZSF0235	83.12 (75.02)	17.25 (21.83)	8.75 (8.79)	0.32 (0.32)	Os03g40270	α -1,4-glucan-protein synthase	[52]
	SHIP_ZSF0781	82.34 (75.02)	22.35 (21.83)	8.68 (8.79)	0.08 (0.32)	Os08g37800	Carbon catabolite derepressing protein kinase	[53]
	SHIP_ZSF0676	69.36 (75.02)	20.72 (21.83)	11.00 (8.79)	2.23 (0.32)	Os01g53930	Hexokinase 1	[54]
	SHIP_ZD0819	83.01 (75.02)	15.42 (21.83)	8.88 (8.79)	0.08 (0.32)	Os10g39440	Hexose transporter	[55]
Amylose	SHIP_ZD0819	83.01 (75.02)	15.42 (21.83)	8.88 (8.79)	0.08 (0.32)	Os10g39440	Hexose transporter	[55]
Fat	SHIP_ZSF0676	69.36 (75.02)	20.72 (21.83)	11.00 (8.79)	2.23 (0.32)	Os01g53930	Hexokinase 1	[54]
	SHIP_ZS3493	69.05 (75.02)	14.50 (21.83)	10.12 (8.79)	1.85 (0.32)	Os12g02340	Nonspecific lipid-transfer protein 3 precursor	[56]

The quality was measured through NIRS. The corresponding amounts of starch, amylose, protein or fat of WT are indicated in parentheses.

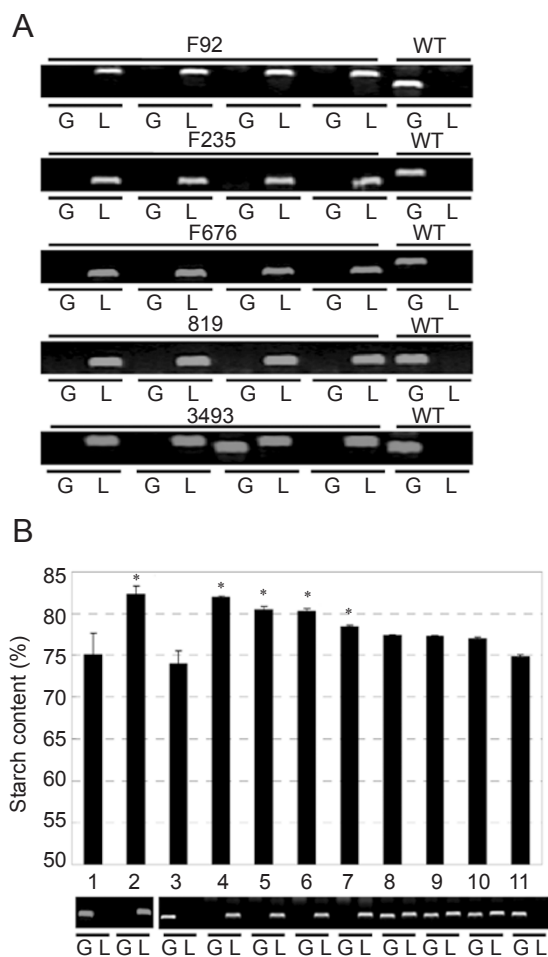


Figure 5 Analysis of the T-DNA insertion and genetic linkage of the seed quality with T-DNA insertion. **(A)** Confirmation of T-DNA insertions. Five homozygotes with mutations corresponding to genes that were previously reported to control seed quality were analyzed by PCR amplification, each using a primer pair matched on the genome (flanking the insertion site, G) or that of a T-DNA left border primer and a reverse primer (L). For each mutant, four independent plants were tested. The corresponding genomic fragment of WT was amplified and used as control. **(B)** Analysis of the genetic linkage in a mutant line SHIP_ZSF0781 with altered starch content. The starch contents of WT (1, 3), homozygote mutant line (2) and different progeny lines from one heterozygous T1 plant (4-11, of which 4-7 are homozygotes, 8-10 are heterozygous lines and 11 is WT) were measured (upper panel), showing good correlation with the gene mutation situation, as analyzed by PCR amplification (lower panel). * indicates the starch content has a significant difference compared with WT ($P < 0.05$).

The negative correlation with small RNAs suggests that small RNAs also play roles in influencing T-DNA integration by maintaining the DNA in a condensed state. Another explanation for the few T-DNA distributions in

transposable elements is that the drug resistance marker gene in T-DNA would be silenced by the small RNAs originating from such regions. As reported, in many eukaryotes transposable elements can generate 21-24-nucleotide siRNAs [48, 62], which are involved in the silencing of flanking genes [63]. T-DNAs were also preferentially recovered from the sites upstream and downstream of genes near the coding region, likely due to the presence of multiple *cis*-elements that bind transcription factors or other regulatory factors, creating a more “open state” than the coding sequence [58].

The homozygous T-DNA insertion population provides a good resource for functional genomic studies

As a staple crop for human food, many efforts have been made to improve the yield and seed quality of rice by breeding [21]. QTL and map-based cloning were used to identify genes associated with agricultural traits. One disadvantage of these strategies is that they are time-consuming, whereas it is much easier to identify target genes based on phenotypes caused by T-DNA insertion mutations. Alternatively, T-DNA insertion populations provide a large platform for studying gene functions using forward or reverse genetics strategies. Nowadays, many T-DNA insertion populations have been developed, but most of them were not screened for homozygotes, and many recessive characters could be missed. The homozygotes obtained in this study are useful for characterizing mutants with altered seed quality. Additionally, the obtained homozygote population can be used for other kinds of functional screenings, such as responses to hormones or stress conditions and so on. As 2 920 unigenes with diverse molecular functions were harbored in the population, it would facilitate functional studies of these genes through the reverse genetics strategy. Systemic studies on seed quality control and the identification of 551 putative mutants, and confirmation of the altered seed qualities caused by T-DNA insertion (Figure 5), demonstrated that the insertion population, especially the homozygote population, would be of great value in mutant screening, particularly for those causing a recessive phenotype.

Genes associated with seed quality control are involved in diverse molecular functions

Mutants with alterations in the four kinds of constituents were identified and some mutants show changes of multiple constituents. Among the identified mutant lines, the most frequent overlap was between starch and fat, whereas the mutants for protein and ACs showed the least frequent overlap (Figure 3). Although amylose is one composition of starch, only 15 mutants have altered

amounts of both starch and amylose. This may be due to the fact that amylose constitutes only up to ~20% of the total starch, and mutants with altered amylose may not necessarily result in significant changes in total starch.

Some mutations are found in genes involved in the individual metabolism pathway, suggesting that seed quality can be regulated by directly modifying the genes involved in the synthesis or metabolism pathways of starch, fat and so on [41-43]. Additionally, other genes fall into many other diverse molecular function categories and some genes are associated with more than one changed constituents of seeds, indicating the close relationship between the regulation of synthesis and metabolism of these constituents. For example, some metabolism pathways, such as glycolysis, are involved in all of the synthesis pathways of these constituents. Also, the genes associated with each individual constituent may be regulated by the same regulatory factors, for example by the same transcription factors.

Among the mutated genes, some have been previously shown to play roles in regulating seed quality (Table 5), such as those coding for enzymes involved in the biosynthesis of starch and fat. A gene for carbon catabolite derepressing protein kinase was identified, which is an ortholog of two conserved *Arabidopsis thaliana* protein kinases KIN10/KIN11. KIN10/KIN11 control convergent reprogramming of transcription in plant energy signaling, the mutation of which impairs starch mobilization at night and growth [64]. In yeast and mammals, the orthologous genes Snf1/AMPK are inactivated by sugars and play central roles in energy signaling [65]. Our study suggests that the function of this family of protein kinase is also conserved in rice. The modification of cognate genes can be used as a strategy to improve energy production, especially starch accumulation in crops.

Previous studies have focused on genes directly in-

involved in the biosynthesis/metabolism of starch, protein and fat [66-73]; our analysis suggests that many other genes may also be involved in seed quality control. Analysis of the underlying complex regulatory network should broaden our understanding of the relevant regulatory mechanisms.

In conclusion, a population of 11 000 putative homozygous lines was generated, and a part of them was used to measure the seed quality using NIRS, resulting in the characterization of mutants with increased or decreased contents of starch, amylose and protein and with higher contents of fat. Further studies will provide new insights into the regulation of seed quality; in addition, these mutants can be used as genetic materials directly for different purposes of the food industry and brewing.

Materials and Methods

Plant materials and transformation

Rice (*Oryza sativa*) plants were cultivated in a phytotron with a light (12 h, 28 °C)/dark (12 h, 22 °C) cycle. Two *japonica* varieties, Nipponbare and Zhonghua 11, were used for transformation at the early stage of this project. In the later stage only Zhonghua 11 was used and most of the transformants are with Zhonghua 11 background. The construct pSMR-J18R [74] was used to transform rice through *Agrobacterium*-mediated transformation [75]. For seed quality analysis, only the putative homozygous transformants from Zhonghua 11 were measured.

TAIL-PCR and sequence analysis

For each transformant, genomic DNA was extracted from the leaf sample (25 mg) of a 2-week-old seedling. TAIL-PCR amplifications were performed as reported previously [44, 45, 58]. The nest primers for the left border of T-DNA and AD primers are listed in Supplementary information, Table S2. All the flanking sequences obtained were aligned against the rice pseudochromosomes by BLASTN [76], with a cut-off of $1e^{-5}$ (if several hits were detected, that with the lowest E-value was chosen). The mapped insertion sites were then annotated using TIGR genome

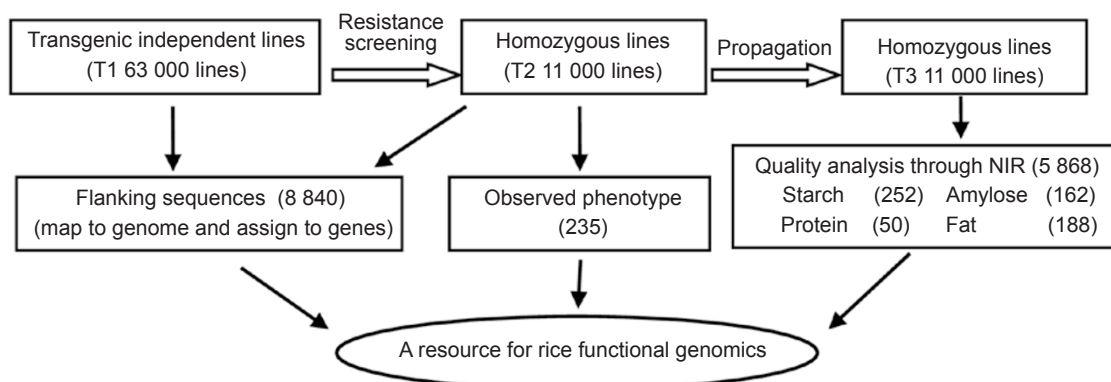


Figure 6 A flow diagram showing mutant analysis and data collection present in SHIP.

annotations release 5 (<http://rice.tigr.org>) and the insertion patterns were indicated as exon, intron, upstream (within 2 500 bp from ATG) or downstream (within 750 bp from the stop codon), or intergenic.

Test of locus number of T-DNA insertion and homozygote screening

To test the locus number of T-DNA insertion, the segregation ratio of hygromycin resistance was calculated. Thirty seeds from each T0 plant were sowed in sterile water containing 30 mg/l hygromycin, and cultivated in a phytotron with a light (18 h, 28 °C)/dark (6 h, 22 °C) cycle for 7 days. The line showing 3 resistant:1 sensitive segregation ratio of hygromycin resistance was considered as containing a single locus of T-DNA. To obtain the putative homozygous lines, T2 seeds were harvested from individual T1 plants and used to analyze the hygromycin resistance. For each line, 30 seeds of 12 individual plants were sowed in sterile water containing hygromycin and those showing complete hygromycin resistance were considered as homozygotes.

Measurement of seed quality by NIRS

Homozygous T2 plants were grown in the field for propagation and the harvested T3 seeds were used for seed quality analysis (Figure 6). For each line, 25 g of paddy sample was dehulled to brown rice, and then milled for 90 s to well-milled rice, which was used for analysis using an NIRSystems model 6500 spectrophotometer (Foss NIRSystems, Inc., Silver Spring, MD, USA) equipped with a transport module, in the reflectance mode. Different calibration equations for starch, amylose, protein and fat were developed and validated by Foss Company. The value of each line was the average of three independent measurements.

To set the criteria for selecting the mutants with altered seed qualities, the distributions of each content in wild-type (Zhonghua 11) plants were analyzed. All data sets were tested for normality using Shapiro-Wilk test (when $P > 0.05$, the data followed the normal distribution) [77], and the confidence limits of the 99% confidence interval were calculated and set as the criteria for mutant selection [78]. Lines with values beyond the criteria were considered to have altered seed quality. After confirmation by second-round measurement, those with the same change trend were selected and analyzed.

Functional annotation of the characterized genes

The annotation information from rice gene ontology (<http://www.tigr.org>) was used to annotate the identified genes. To determine the significance of over- or under-representation of group genes for each GO term, the chi-square test was used.

Confirmation of the mutant lines related to reported genes

Five lines with putative mutation of the previously reported genes were selected and genomic DNA was extracted from four independent plants of each homozygous line. The T-DNA insertion was confirmed through PCR amplification using primers flanking the putative insertion site. The primers specific to each candidate gene were as follows: SHIP_ZS F0092 (F92-1: 5'-AGG GGC TGG AGT GCT AAA TA-3'; F92-2: 5'-GGA AAT AGA ATG ATA AAG GCG T-3'), SHIP_ZSF0235 (F235-1: 5'-TCT GTG ACC ACC TGA GCC T-3'; F235-2: 5'-GCC TTG ACC ATT CGT CCA TA-3'), SHIP_ZSF0781 (F781-1: 5'-GAG ACG AAT CTT TTG

AGC CT-3'; F781-2: 5'-GCT TGG ATT TGG ATG GAC GG-3'), SHIP_ZSF0676 (F676-1: 5'-CAG ACG ACG ACG CTG ACG A-3'; F676-2: 5'-CAA GTT GGC AAA CGA TGG CT-3'), SHIP_ZD0819 (819-1: 5'-GTT GAA CGC GGA GAT AGA TTA-3'; 819-2: 5'-TGA GCA TTA GCA GTA GAG CC-3'), SHIP_ZS3493 (3493-1: 5'-AAC TTG CTC GTT TCA CTT CA-3'; 3493-2: 5'-CAA TGA GAT GGA GGG AGT AG-3'). The T-DNA left border primer NTL2 (5'-ATA GGG TTT CGC TCA TGT GTT GAG C-3') was used.

For detailed analysis of mutant SHIP_ZSF0781, eight independent progenies separated from one heterozygous parent were tested through PCR amplification. The quality of seeds harvested from these progenies independently was measured through NIRS.

Acknowledgments

This study was supported by the National High-Tech Program (863) (Nos. 2002AA2Z1003 and 2006AA10A102) and the State Key Basic Research Program (973) (No. 2005CB120803). We thank Prof Qi-Fa Zhang (Huazhong Agricultural University) for kindly providing the construct pSMR, Liang-Jiao Xue and Qiu-Ming Shi from our laboratory for help with database construction, and all lab members for flanking sequence isolation.

References

- 1 Wang YH, Xue YB, Li JY. Towards molecular breeding and improvement of rice in China. *Trends Plant Sci* 2005; **10**:610-614.
- 2 Li ZF, Wan JM, Xia JF, Zhai HQ. Mapping quantitative trait loci underlying appearance quality of rice grains (*Oryza sativa* L.). *Yi Chuan Xue Bao* 2003; **30**:251-259.
- 3 Juliano BO. Varietal impact on rice quality. *Cereal Food World* 1998; **43**:207-222.
- 4 Lanceras JC, Huang ZL, Naivikul O, Vanavichit A, Ruanjaichon V, Tragoonrun S. Mapping of genes for cooking and eating qualities in Thai jasmine rice (KDML105). *DNA Res* 2000; **7**:93-101.
- 5 Bajaj S, Mohanty A. Recent advances in rice biotechnology-towards genetically superior transgenic rice. *Plant Biotechnol J* 2005; **3**:275-307.
- 6 Koo WW, Lasekan JB. Rice protein-based infant formula: current status and future development. *Minerva Pediatr* 2007; **59**:35-41.
- 7 Tzen JT, Huang AH. Surface structure and properties of plant seed oil bodies. *J Cell Biol* 1992; **117**:327-335.
- 8 Ye X, Al-Babili S, Klott A, *et al.* Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 2000; **287**:303-305.
- 9 Devaiah SP, Pan X, Hong Y, Roth M, Welti R, Wang X. Enhancing seed quality and viability by suppressing phospholipase D in *Arabidopsis*. *Plant J* 2007; **50**:950-957.
- 10 Kim KS, Park SH, Choung MG. Nondestructive determination of lignans and lignan glycosides in sesame seeds by near infrared reflectance spectroscopy. *J Agric Food Chem* 2006; **54**:4544-4550.
- 11 Hong JH, Ikeda K, Kref I, Yasumoto K. Near-infrared diffuse reflectance spectroscopic analysis of the amounts of moisture,

- protein, starch, amylose, and tannin in buckwheat flours. *J Nutr Sci Vitaminol* 1996; **42**:359-366.
- 12 Manley M, van Zyl L, Osborne BG. Using Fourier transform near infrared spectroscopy in determining kernel hardness, protein and moisture content of whole wheat flour. *J Near Infrared Spectrosc* 2002; **10**:71-76.
- 13 Beecher B, Bowman J, Martin JM, *et al.* Hordoindolines are associated with a major endosperm-texture QTL in barley (*Hordeum vulgare*). *Genome* 2002; **45**:584-591.
- 14 Baye MT, Pearson CT, Settles AM. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *J Cereal Sci* 2006; **43**:236-243.
- 15 Choung MG, Kang ST, Han WY, *et al.* Determination of fatty acid composition in soybean seed using near infrared reflectance spectroscopy. *Korean J Breed* 2005; **37**:197-202.
- 16 Wu JG, Shi CH. Prediction of grain weight, brown rice weight and amylase content in single grains using near-infrared reflectance spectroscopy. *Field Crop Res* 2004; **87**:13-21.
- 17 Kim YH, Kang CS, Lee YS. Quantification of tocopherol and tocotrienol content in rice bran by near infrared reflectance spectroscopy. *Korean J Crop Sci* 2004; **49**:211-215.
- 18 Giri CC, Vijaya Laxmi G. Production of transgenic rice with agronomically useful genes: an assessment. *Biotechnol Adv* 2000; **18**:653-683.
- 19 Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard JH. QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* x *O. glaberrima*. *Theor Appl Genet* 2004; **109**:630-639.
- 20 Fan CC, Yu XQ, Xing YZ, Xu CG, Luo LJ, Zhang Q. The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. *Theor Appl Genet* 2005; **110**:1445-1452.
- 21 Virmani SS, Sun ZX, Mou TM, Jauhar Ali A, Mao CX. Two-line Hybrid Rice Breeding Manual. Los Baños (Philippines): International Rice Research Institute, 2003.
- 22 Zhou PH, Tan YF, He YQ, Xu CG, Zhang Q. Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. *Theor Appl Genet* 2003; **106**:326-331.
- 23 Wang J, Wan X, Li H, Pfeiffer WH, Crouch J, Wan J. Application of identified QTL-marker associations in rice quality improvement through a design-breeding approach. *Theor Appl Genet* 2007; **115**:87-100.
- 24 Van Camp W. Yield enhancement genes: seeds for growth. *Curr Opin Biotechnol* 2005; **16**:147-153.
- 25 Bouche N, Bouchez D. *Arabidopsis* gene knockout: phenotypes wanted. *Curr Opin Plant Biol* 2001; **4**:111-117.
- 26 Alonso JM, Stepanova AN. T-DNA mutagenesis in *Arabidopsis*. *Methods Mol Biol* 2003; **236**:177-188.
- 27 Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B. An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol Biol* 2003; **53**:247-259.
- 28 Sessions A, Burke E, Presting G, *et al.* A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* 2002; **14**:2985-2994.
- 29 Padhyaya NM, Zhu QH, Zhou XR, *et al.* Dissociation (Ds) constructs, mapped Ds launch pads and a transiently-expressed transposase system suitable for localized insertional mutagenesis in rice. *Theor Appl Genet* 2006; **112**:1326-1341.
- 30 Chen S, Jin W, Wang M, *et al.* Distribution and characterization of over 1000 T-DNA tags in rice genome. *Plant J* 2003; **36**:105-113.
- 31 Zhang J, Li C, Wu C, *et al.* RMD: a rice mutant database for functional analysis of the rice genome. *Nucleic Acids Res* 2006; **34**:D745-D748.
- 32 Chern CG, Fan MJ, Yu SM, *et al.* A rice phenomics study--phenotype scoring and seed propagation of a T-DNA insertion-induced rice mutant population. *Plant Mol Biol* 2007; **65**:427-438.
- 33 van Enkevort LJ, Droc G, Piffanelli P, *et al.* EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice. *Plant Mol Biol* 2005; **59**:99-110.
- 34 Sallaud C, Gay C, Larmande P, *et al.* High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. *Plant J* 2004; **39**:450-464.
- 35 Hirochika H. Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr Opin Plant Biol* 2001; **4**:118-122.
- 36 Miyao A, Tanaka K, Murata K, *et al.* 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* 2003; **15**:1771-1780.
- 37 Jeong DH, An S, Kang HG, *et al.* T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol* 2002; **130**:1636-1644.
- 38 Jeong DH, An S, Park S, *et al.* Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. *Plant J* 2006; **45**:123-132.
- 39 Kumar CS, Wing RA, Sundaresan V. Efficient insertional mutagenesis in rice using the maize En/Spm elements. *Plant J* 2005; **44**:879-892.
- 40 Lee S, Kim J, Son JS, *et al.* Systematic reverse genetic screening of T-DNA tagged genes in rice for functional genomic analyses: MADS-box genes as a test case. *Plant Cell Physiol* 2003; **44**:1403-1411.
- 41 Jung KH, Han MJ, Lee DY, *et al.* Wax-deficient anther1 is involved in cuticle and wax production in rice anther walls and is required for pollen development. *Plant Cell* 2006; **18**:3015-3032.
- 42 Kang HG, Park S, Matsuoka M, An G. White-core endosperm floury endosperm-4 in rice is generated by knockout mutations in the C-type pyruvate orthophosphate dikinase gene (OsPPDKB). *Plant J* 2005; **42**:901-911.
- 43 Ryoo N, Yu C, Park CS, *et al.* Knockout of a starch synthase gene OsSSIIIa/Flo5 causes white-core floury endosperm in rice (*Oryza sativa* L.). *Plant Cell Rep* 2007; **26**:1083-1095.
- 44 Liu YG, Mitsukawa N, Osumi T, Whittier RF. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J* 1995; **8**:457-463.
- 45 Liu YG, Whittier RF. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* 1995; **25**:674-681.
- 46 De Block M, Debrouwer D. Two T-DNA's co-transformed into *Brassica napus* by a double *Agrobacterium tumefaciens* infection are mainly integrated at the same locus. *Theor Appl Genet*

- 1991; **82**:257-263.
- 47 Aravin AA, Lagos-Quintana M, Yalcin A, *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* 2003; **5**:337-350.
- 48 Nobuta K, Venu RC, Lu C, *et al.* An expression atlas of rice mRNAs and small RNAs. *Nat Biotechnol* 2007; **25**:473-477.
- 49 Shen XH, Cui HR, Zhang HL, Cheng SH. Comparative studies on physical-chemical properties and major nutritional components of rice grain in herbicide-resistant transgenic rice and its parent. *Chin J Rice Sci* 2004; **18**:181-183.
- 50 Vigeolas H, Möhlmann T, Martini N, Neuhaus HE, Geigenberger P. Embryo-specific reduction of ADP-Glc pyrophosphorylase leads to an inhibition of starch synthesis and a delay in oil accumulation in developing seeds of oilseed rape. *Plant Physiol* 2004; **136**:2676-2686.
- 51 Panneerselvam R, Abdul Jaleel C, Somasundaram R, *et al.* Carbohydrate metabolism in *Dioscorea esculenta* (Lour.) Burk. tubers and *Curcuma longa* L. rhizomes during two phases of dormancy. *Colloids Surf B Biointerfaces* 2007; **59**:59-66.
- 52 Rothschild A, Wald FA, Bocca SN, Tandecarz JS. Inhibition of UDP-glucose: protein transglucosylase by a maize endosperm protein factor. *Cell Mol Biol* 1996; **42**:645-651.
- 53 Takano M, Kajiya-Kanegae H, Funatsuki H, Kikuchi S. Rice has two distinct classes of protein kinase genes related to SNF1 of *Saccharomyces cerevisiae*, which are differently regulated in early seed development. *Mol Gen Genet* 1998; **260**:388-394.
- 54 Focks N, Benning C. wrinkled1: A novel, low-seed-oil mutant of *Arabidopsis* with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol* 1998; **118**:91-101.
- 55 Toyota K, Tamura M, Ohdan T, Nakamura Y. Expression profiling of starch metabolism-related plastidic translocator genes in rice. *Planta* 2006; **223**:248-257.
- 56 Bun-ya M, Muro Y, Niki T, Kondo J, Kamiryo T. New aspects of sterol carrier protein 2 (nonspecific lipid-transfer protein) in fusion proteins and in peroxisomes. *Cell Biochem Biophys* 2000; **32**:107-116.
- 57 Schneeberger RG, Zhang K, Tatarinova T, *et al.* Agrobacterium T-DNA integration in *Arabidopsis* is correlated with DNA sequence compositions that occur frequently in gene promoter regions. *Funct Integr Genomics* 2005; **5**:240-253.
- 58 Zhang J, Guo D, Chang Y, *et al.* Non-random distribution of T-DNA insertions at various levels of the genome hierarchy as revealed by analyzing 13 804 T-DNA flanking sequences from an enhancer-trap mutant library. *Plant J* 2007; **49**:947-959.
- 59 Wu J, Maehara T, Shimokawa T, *et al.* A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell* 2002; **14**:525-535.
- 60 Barakat A, Gallois P, Raynal M, *et al.* The distribution of T-DNA in the genomes of transgenic *Arabidopsis* and rice. *FEBS Lett* 2000; **471**:161-164.
- 61 Sha Y, Li S, Pei Z, Luo L, Tian Y, He C. Generation and flanking sequence analysis of a rice T-DNA tagged population. *Theor Appl Genet* 2004; **108**:306-314.
- 62 Lippman Z, Gendrel AV, Black M, *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* 2004; **430**:471-476.
- 63 Saze H, Kakutani T. Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J* 2007; **26**:3641-3652.
- 64 Baena-Gonzalez E, Rolland F, Thevelein JM, Sheen J. A central integrator of transcription networks in plant stress and energy signalling. *Nature* 2007; **448**:938-942.
- 65 Hardie DG, Carling D, Carlson M. The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell? *Annu Rev Biochem* 1998; **67**:821-855.
- 66 Mizuno K, Kawasaki T, Shimada H, *et al.* Alteration of the structural properties of starch components by the lack of an isoform of starch branching enzyme in rice seeds. *J Biol Chem* 1993; **268**:19084-19091.
- 67 Itoh K, Ozaki H, Okada K, Hori H, Takeda Y, Mitsui T. Introduction of Wx transgene into rice wx mutants leads to both high- and low-amylose rice. *Plant Cell Physiol* 2003; **44**:473-480.
- 68 Satoh H, Nishi A, Yamashita K, *et al.* Starch-branching enzyme I-deficient mutation specifically affects the structure and properties of starch in rice endosperm. *Plant Physiol* 2003; **133**:1111-1121.
- 69 Fujita N, Yoshida M, Asakura N, *et al.* Function and characterization of starch synthase I using mutants in rice. *Plant Physiol* 2006; **140**:1070-1084.
- 70 Fujita N, Yoshida M, Kondo T, *et al.* Characterization of SSIIIa-deficient mutants of rice: the function of SSIIIa and pleiotropic effects by SSIIIa deficiency in the rice endosperm. *Plant Physiol* 2007; **144**:2009-2023.
- 71 Takemoto Y, Coughlan SJ, Okita TW, Satoh H, Ogawa M, Kumamaru T. The rice mutant esp2 greatly accumulates the glutelin precursor and deletes the protein disulfide isomerase. *Plant Physiol* 2002; **128**:1212-1222.
- 72 Jiang SM, Zhu SS, Liu SJ, Jiang L, Xu LL, Wan JM. Screening and genetic analysis of rice glutelin mutant. *J Genet Genom Bao* 2003; **30**:641-645.
- 73 Wu GT, Lang CX, Hu ZH, *et al.* Raising fat content in transgenic rice by anti-PEP gene transformation. *J Plant Physiol Plant Mol Biol* 2006; **32**:339-344.
- 74 Wu C, Li X, Yuan W, *et al.* Development of enhancer trap lines for functional analysis of the rice genome. *Plant J* 2003; **35**:418-427.
- 75 Hiei Y, Ohta S, Komari T, Kumashiro T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J* 1994; **6**:271-282.
- 76 Altschul SF, Madden TL, Schaffer AA, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**:3389-3402.
- 77 Sürücü B, Koç E. Assessing the validity of a statistical distribution: some illustrative examples from dermatological research. *Clin Exp Dermatol* 2008; **33**:239-242.
- 78 Tate RF, Klett GW. Optimal confidence intervals for the variance of a normal distribution. *J Am Stat Assoc* 1959; **54**:674-682.

(Supplementary information is linked to the online version of the paper on the Cell Research website.)