RMD: a rice mutant database for functional analysis of the rice genome

Jianwei Zhang, Caishun Li, Changyin Wu, Lizhong Xiong, Guoxing Chen, Qifa Zhang and Shiping Wang*

National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China

ABSTRACT

Rice Mutant Database (RMD, http://rmd.ncpgr.cn) is an archive for collecting, managing and searching information of the T-DNA insertion mutants generated by an enhancer trap system. We have generated ∼129 000 rice mutant (enhancer trap) lines that are now being gathered in the database. Information collected in RMD includes mutant phenotypes, reporter-gene expression patterns, flanking sequences of T-DNA insertional sites, seed availability and others, and can be searched by respective ID, keyword, nucleotide sequence or protein sequence on the website. This database is both a mutant collection for identifying novel genes and regulatory elements and a pattern line collection for ectopic expression of target gene in specific tissue or at specific growth stage.

INTRODUCTION

Rice (*Oryza sativa* L.) has now become a model plant system for genome research of the monocots, especially for cereal species. Although the draft and finished genome sequences of both *indica* and *japonica* subspecies of rice have been released, the function for majority of the genes in the rice genome have not been experimentally studied (1,2). A common strategy for large-scale discovery of gene function is to generate populations of DNA insertion mutants and identify the genes by screening and characterizing the mutations. The enhancer trap system, based on the expression pattern of a reporter gene randomly integrated into the genome, is regarded as a useful tool for identifying novel genes and regulatory elements. This technology was first applied in the bacterium (3) and now is widely used in animal research, especially in *Drosophila* (4). There were also reports on the establishment of such system in *Arabidopsis* (5) and

moss (6). In rice, we have generated more than 129 thousand independent transgenic lines with the enhancer trap construct by *Agrobacterium*-mediated T-DNA insertion (7). The system has three built-in strategies for functional analysis of the rice genome. First, T-DNA insertions cause gene mutations, providing an efficient approach for gene identification and isolation. Second, expression of the reporter gene indicates the presence of an enhancer element in the neighboring genomic region, which can be used for isolation and characterization of the enhancer. Third, the lines showing spatial- or temporal-specific expression of the reporter gene can be used to drive ectopic expression of a transgene, thus useful for unveiling latent functions of unknown or known genes. Large-scale screening and characterization of these rice mutant lines are now in progress. Comprehensive information, including the phenotypes, reporter-gene expression patterns and flanking sequences of T-DNA insertional sites of these enhancer trap lines has been generated. As there is an urgent need for publicly available rice mutant resources in functional genomic studies by the rice research community (8), we undertook the effort to establish the Rice Mutant Database (RMD, http://rmd.ncpgr.cn) for public access. This database together with other insertional rice mutant collections (8,9) should serve as informative resources for accelerating functional genomic studies of rice, as well as other monocots.

ARCHITECTURE AND CONTENT OF THE DATABASE

RMD was constructed using a Redhat Linux system and maintained using MySQL, a relational database management system. The data collected in RMD are placed into three tables (mutant lines, phenotypes and flanking sequences), that are linked via entry identifier (ID) (Figure 1). The general information of current release 20050901 of RMD is summarized in Table 1.

The enhancer trap lines were generated using three *japonica* rice varieties (*O.sativa* ssp. *japonica*), Zhonghua 11 (Z11),

*To whom correspondence should be addressed. Tel: +86 27 87283009; Fax: +86 27 87287092; Email: swang@mail.hzau.edu.cn
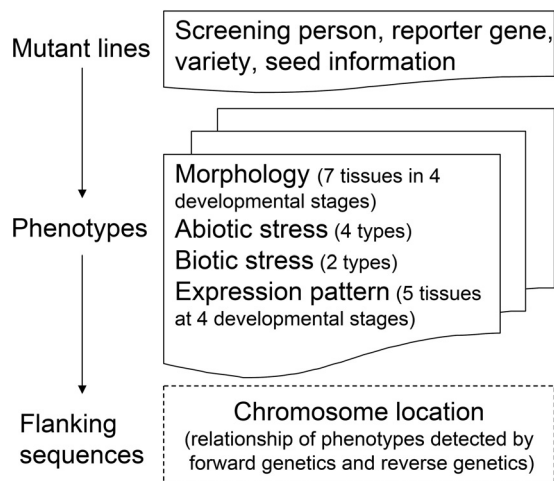
Figure 1. The composition of Rice Mutant Database. The component in dotted frame will be integrated into the database.

Table 1. General information of the Rice Mutant Database, release 20050901

Data type	Number of entries
Total enhancer trap line	128 560
From Zhonghua 11 (Z11)	113 262
From Zhonghua 15 (Z15)	14 197
From Nipponbare (NPB)	1101
Mutant phenotype[a]	23 844
Expression pattern	5511
Flanking sequence[b]	13 738
Mutant photograph	733
Seed available	53 432

[a]The information indicates the lines that had been screened for at least one type of phenotypes (including abiotic and biotic stress screening) listed in Table 2.

[b]In total 11 910 mutant lines have one flanking sequence for each, 753 lines two sequences for each, 67 lines three sequences for each, 16 lines four sequences for each, 9 lines five sequences for each and 2 lines six sequences for each.

Zhonghua 15 (Z15) and Nipponbare (NPB). The data of mutant phenotypes, which refer to novel genes controlling the phenotypes, are collected by screening types falling into the three categories: morphology under normal growth conditions, abiotic stress and biotic stress (Table 2 and Supplementary Table 1). Each of the screening types involves two or more attributes referred to as 'traits'. Phenotypes were scored for each of the traits at different developmental stages. The phenotype data for each trait were routinely collected based on 20 T_1 (in some cases T_2) plants per entry, from which segregation ratios of the traits were also scored. The expression patterns of the reporter gene, which refer to the regulatory elements, are also embodied into the category of 'Screening type' (Table 2 and Supplementary Table 1). The expression patterns were examined for five different tissues of T_0 (in some cases T_1) plants at various developmental stages. Majority of the flanking sequences of T-DNA insertional sites were obtained by thermal asymmetric interlaced-PCR and a small proportion of the flanking sequences were obtained by plasmid rescue. All of these sequences are anchored on rice chromosomes.

SEARCH THE DATABASE

RMD can be freely accessed using a web browser at http://rmd.ncpgr.cn. User can search information from the database in two different ways as described below (Figure 2).

Keyword search

Keyword can be the entry ID of enhancer trap line, single word or phrase. Each ID consists of eight characters, such as 02Z15DJ04 and 03Z11BO64. The first two digits of each ID represent the year in which the line was generated ('02' for year 2002, '03' for year 2003 and so on). The next three characters are the abbreviation of rice variety with which the mutant line was generated (Table 1). The last four characters represent the order of the line produced. For quick searching the database, user can use the ID as query word and the querying result displays all details for the entry, including its morphological mutation, traits screened under stress conditions, expression pattern of the reporter gene in different tissues at different developmental stages, validated flanking sequence(s) and seed availability. If one does not know the ID, any keywords, especially the words or phrases listed in Supplementary Table 1 or the suggested keywords provided on the website, can be used for searching. The querying result will show a list of entries that are linked to the individual subclass pages. Three Boolean operators, 'AND', 'OR' and 'NOT' with all letters entered in UPPERCASE, can be used in combination with keywords for refining the search statement.

A wizard is provided for advanced search (Figure 2). Two advanced search pages for Microsoft Internet Explorer (IE) browser and non-IE browsers are designed, respectively. The system will provide the appropriate page automatically when the user clicks the 'Advanced search' button. First, user can choose one of the four screening types for mutant categories from the selectable menu for IE browser or radio buttons for non-IE browsers. Second, type of traits screened can be selected from a clickable menu for IE browser or radio buttons for non-IE browsers. The last step is to select the phenotype from a clickable list for IE browser or checkboxes for non-IE browsers. For more targeted search, one or more of the four checkboxes for developmental stages can be used for the search. The querying results will display one or a list of entries that are linked to the individual subclass pages.

BLAST search

The BLAST program (10), including BLASTN, TBLASTN and TBLASTX, is integrated into RMD. User can use either nucleotide sequence or protein sequence as query to search the flanking sequences of T-DNA insertional sites. The querying results will display the corresponding ID of mutant line as well as all the alignments of homolog sequences.

FUTURE PLANS

RMD is a new database. The majority of data in this release (20050901) is generated by one of the three groups, Wuhan group, in the National Special Key Program on Rice Functional Genomics of China. The complete screening with screening traits listed in Table 2 or Supplementary Table 1 and collecting flanking sequences from ~129 000 independent

Table 2. Mutation phenotypes in RMD[a]

Screening type	Trait[b]	Number of phenotype in trait	Developmental stage[c]			
			Seedling	Tillering	Heading	Ripening
Morphology under normal growth conditions	Leaf	25	√	√	—	—
	Flag leaf	8	—	—	√	—
	Root	8	√	—	—	—
	Culm	11	√	√	—	—
	Spikelet	6	—	—	√	—
	Panicle	10	—	—	√	—
	Seed	7	—	—	—	√
Abiotic stress	Abscisic acid	3	√	—	—	—
	Drought	3	√	—	—	—
	Low nitrogen	3	√	—	—	—
	Low phosphorus	3	√	—	—	—
Biotic stress	Bacterial blight	2	—	√	—	—
	Blast	2	√	√	—	—
Expression	Leaf	7	√	√	√	—
	Root	7	√	—	—	—
	Culm	2	—	√	√	—
	Flower	11	—	—	√	—
	Seed	4	—	—	—	√

[a]Supplementary Table 1 for detail phenotypes of this table.
[b]Abscisic acid: plants treated with 30-80 μM of abscisic acid; Drought stress: stopping watering at 3-4 leaf stage and counting the number of days to leaf rolling; Low nitrogen: plants cultured in solution containing 1/10 of normal (1.428 mM NH_4NO_3) nitrogen; Low phosphorus: plants cultured in solution containing 1/30 of normal (0.323 mM NaH_2PO_4) phosphorus; Bacterial blight: inoculation with *Xanthomonas oryzae* pv. *oryzae*; Blast: inoculation with *Pyricularia grisea* Sacc.
[c]The symbol '√' indicates the stage at which screening was conducted.
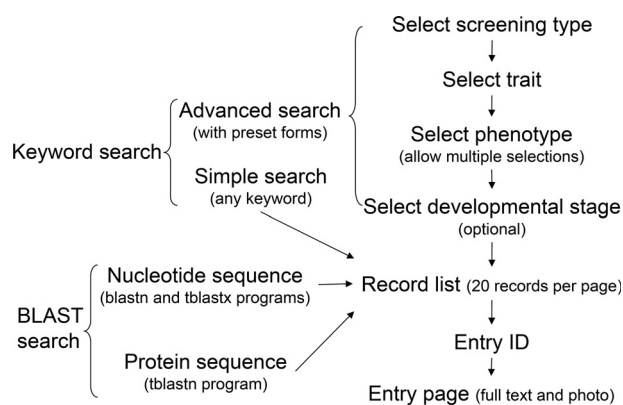
Figure 2. Query of data in Rice Mutant Database.

enhancer trap lines generated by Wuhan group are expected to be finished within next few years. In addition, data collected by other two (Beijing and Shanghai) groups of the same program will give additional information for RMD and enhance its utility. It is expected that RMD will finally contain information from 200 000 independent transgenic lines. Furthermore, the flanking sequences of T-DNA insertional sites will be co-mapped with >15 000 expressed sequence tag loci, which have been co-mapped with large numbers of rice quantitative trait loci for agronomic traits (11), to establish the relationship from genes to phenotypes detected by forward and reverse genetics strategies.

AVAILABILITY

RMD can be freely accessed at http://rmd.ncpgr.cn. Seeds of the enhancer trap lines or mutants in this database can be requested for research purpose by downloading the request form at RMD page of 'Contact us' and sending an e-mail request to Dr Changyin Wu (cywu@mail.hzau.edu.cn).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Yu,J., Hu,S., Wang,J., Wong,G.K., Li,S., Liu,B., Deng,Y., Dai,L., Zhou,Y., Zhang,X. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, **296**, 79–92.
2. International Rice Genome Sequencing Project (2005), The map-based sequence of the rice genome. *Nature*, **436**, 793–800.
3. Casadaban,M.J. and Cohen,S.N. (1979) Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: *in vivo* probe for transcriptional control sequences. *Proc. Natl Acad. Sci. USA*, **76**, 4530–4533.
4. Bellen,H.J. (1999) Ten years of enhancer detection: lessons from the fly. *Plant Cell*, **11**, 2271–2281.
5. Campisi,L., Yang,Y., Yi,Y., Heilig,E., Herman,B., Cassista,A.J., Allen,D.W., Xiang,H. and Jack,T. (1999) Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. *Plant J.*, **17**, 699–707.
6. Hiwatashi,Y., Nishiyama,T., Fujita,T. and Hasebe,M. (2001) Establishment of gene-trap and enhancer-trap systems in the moss *Physcomitrella patens. Plant J.*, **28**, 105–116.

7. Wu,C., Li,X., Yuan,W., Chen,G., Kilian,A., Li,J., Xu,C., Li,X., Zhou,D.X., Wang,S. *et al.* (2003) Development of enhancer trap lines for functional analysis of the rice genome. *Plant J.*, **35**, 418–427.

8. Hirochika,H., Guiderdoni,E., An,G., Hsing,Y.I., Eun,M.Y., Han,C.D., Upadhyaya,N., Ramachandran,S., Zhang,Q., Pereira,A. *et al.* (2004) Rice mutant resources for gene discovery. *Plant Mol. Biol.*, **54**, 325–334.

9. Sallaud,C., Gay,C., Larmande,P., Bes,M., Piffanelli,P., Piegu,B., Droc,G., Regad,F., Bourgeois,E., Meynard,D. *et al.* (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. *Plant J.*, **39**, 450–464.

10. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

11. Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D., Han,B., Zhang,Q. and Wang,S. (2005) Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite *indica* rice cultivar Minghui 63. *Plant J.*, **42**, 772–780.