

Development of 1047 insertion-deletion markers for rice genetic studies and breeding

Y.X. Zeng*, Z.H. Wen*, L.Y. Ma, Z.J. Ji, X.M. Li and C.D. Yang

State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, China

*These authors contributed equally to this study.

Corresponding author: C.D. Yang

E-mail: yangchangdeng@126.com

Genet. Mol. Res. 12 (4): 5226-5235 (2013)

Received March 14, 2013

Accepted July 18, 2013

Published October 30, 2013

DOI <http://dx.doi.org/10.4238/2013.October.30.7>

ABSTRACT. In this study, a total of 1047 insertion-deletion (InDel) primer pairs distributed across the rice genome were developed and experimentally validated. The primer pairs were designed based on the InDel length polymorphisms between 93-11 (*Oryza sativa* ssp *indica* cv.) and Nipponbare (*Oryza sativa* ssp *japonica* cv.), aiming for utilization between *indica* and *japonica* rice, or between other inter-subspecific rice cultivars. The 1047 primer pairs were dispersed across all 12 of the rice chromosomes, with one InDel marker found every 371.3 kb on average. The InDel length of the markers varied from 3 to 39 bp: 88.2% of the markers contained 6 to 25 bp, only 6.2% of markers were ≤ 5 bp, and 5.6% were ≥ 26 bp. Six hundred and twenty-three (59.5%) of the 1047 InDel markers were shown to amplify well and were polymorphic between Taichung65 and IR8, and 476 (45.5%) markers were polymorphic between Lemont and Yangdao4, while 398 (38.0%) were polymorphic in both combinations. These results demonstrated that the polymerase chain reaction-based InDel markers developed in this study could be of immediate use for rice genetic studies and breeding programs.

Key words: Rice (*Oryza sativa* L.); Molecular marker; Insertion-deletion length polymorphism; InDel

INTRODUCTION

Molecular markers play important roles in both basic and applied research, such as fingerprinting genotypes, analyzing genetic diversity, determining variety identity, marker-assisted breeding, phylogenetic analysis, and map-based cloning of genes (Shen et al., 2004; Liu et al., 2013). Rice is an important food crop for more than half the world's population, and is a model plant for the grasses. A large number of molecular markers have been developed in rice, including expressed sequence tags, restricted fragment length polymorphisms, simple sequence repeat (SSR) markers, single nucleotide polymorphism (SNP) markers, insertion-deletion (InDel) length polymorphism markers, and cleaved amplified polymorphic sequence markers among others (Akagi et al., 1996; Panaud et al., 1996; Temnykh et al., 2001; McCouch et al., 2002; Nasu et al., 2002; International Rice Genome Sequencing Project, 2005; Ren et al., 2005; Chen et al., 2011). The emergence of genomic sequences in rice has increased the development of several locus-specific markers for high-resolution genetic analysis.

SSR, InDel, and SNP markers are frequently used in rice studies. A total of 18,828 SSR markers have been developed in rice, with approximately one SSR marker found every 19.69 kb in the rice genome (International Rice Genome Sequencing Project, 2005). The fact that SSR markers are co-dominant, multi-allelic, high-density, easy to use, and inexpensive makes them the most widely used markers in rice to date (McCouch et al., 2002). The release of the genome sequence of two inbred rice cultivars, *japonica* cv. Nipponbare and *indica* cv. 93-11, has greatly accelerated the identification of SNP and InDel polymorphisms in rice. A total of 1,703,176 SNPs and 479,406 InDel polymorphisms have been identified to date, and the frequencies of SNP and InDel polymorphisms between Nipponbare and 93-11 are 0.71 and 0.20%, respectively, or approximately one SNP every 268 bp and one InDel every 953 bp along the rice genome (Shen et al., 2004). Using whole-genome sequencing of a *japonica* landrace rice line, Omachi, SNP and InDel polymorphisms between Omachi and Nipponbare were reported (Arai-Kichise et al., 2011). The high frequency of SNP and InDel polymorphisms makes the development of high-density SNP and InDel markers possible. The SNP markers have been used more and more frequently over recent years. A total of 213 SNP markers have been established (Nasu et al., 2002). A set of 372 SNP markers were used to genotype 300 inbred rice varieties from 22 rice-growing countries (Chen et al., 2011). A total of 280 polymorphic SNP markers were used in mapping quantitative trait loci for heat tolerance in rice (Ye et al., 2012). A whole-genome resequencing technique was used to examine genome-wide SNPs to construct a genetic map with high resolution using 150 rice recombinant inbred lines (Huang et al., 2009). The InDel polymorphisms have received relatively little attention compared with SNPs. However, InDel markers have more practical value for laboratories without infrastructure to perform SNP genotyping (Liu et al., 2013). InDel polymorphisms occur more frequently than SSR polymorphisms in the rice genome, which has made them more preferable in fine mapping genes. Since InDel markers use the same experimental approaches routinely used for SSR markers, the InDel markers have all the advantages of the SSRs (Liu et al., 2013). Fifty InDel markers were developed for rice by Shen et al. (2004). Although a large number of InDel polymorphisms have been reported in the rice genome, InDel markers with good coverage of the 12 rice chromosomes have not yet been developed.

In this study, we analyzed the InDel sequences between Nipponbare and 93-11, and developed a set of 1047 InDel markers along the 12 rice chromosomes, corresponding to ap-

proximately one marker per 371.3 kb in the rice genome. We experimentally validated all of the InDel markers, and analyzed their applicability in two other *indica/japonica* combinations. The InDel markers developed in the present study are valuable for rice genetic research and marker-assisted breeding.

MATERIAL AND METHODS

Searching InDel polymorphisms between 93-11 and Nipponbare sequences

We searched and collected InDel polymorphisms along the rice genome using 93-11 and Nipponbare as reference sequences. The 93-11 genomic sequences were downloaded from the Beijing Genomics Institute-Rice Information System (BGI-RIS) (<ftp://ftp.genomics.org.cn/pub/ricedb/SynVs9311/9311/Sequence/SupScaffold/>). The downloaded files with the filename extension '.gz' were unzipped using the WinRAR software. The unzipped files were opened in the DNASTar EditSeq program (DNASTAR Inc.). The contig sequences of the 93-11 were used as queries to BLASTN Nipponbare genomic sequences (<http://blast.ncbi.nlm.nih.gov/>). The alignments of 93-11 and Nipponbare sequences were shown as the BLASTN results. The InDel site can be found from the alignment results denoted as '-'

Candidate InDel sequences for primer design

The alignment results of 93-11 and Nipponbare from the BLASTN search in the National Center for Biotechnology Information (NCBI) website were analyzed, and the candidate sequences for primer design were determined to require the following characteristics: 1) the sequence length of the PCR product should usually be 120-480 bp, the InDel length between 93-11 and Nipponbare should usually be >4 bp, and 2) the InDels between 93-11 and Nipponbare should be located relatively in the middle of the candidate sequences, and there should be identical sequences for 93-11 and Nipponbare in places outside of the candidate sequence. This is because SNPs or InDels occurring at places other than the candidate sequence restrict the selection of forward or reverse primers.

Primer design

The candidate sequences (Nipponbare or 93-11) were copied and pasted into the PRIMER PREMIER software (version 5.0, PREMIER Biosoft International, Palo Alto, CA, USA). Primer pairs were automatically searched using PRIMER PREMIER. Primers were limited to 16-26 nucleotides long. The PRIMER PREMIER software will automatically detect 'hairpin', 'dimer', 'cross dimer', and 'false priming' of the forward and reverse primers, and provide a final rating (from 0 to 100) for a specific primer pair. The 'hairpin', 'dimer', and 'cross dimer' indicate the secondary structure formed by primers. The 'false priming' indicates primer binding at the incorrect site of the template. Sequences that provided primer pairs with ratings below 86 were eliminated from consideration. If 'false priming' was detected, the primer pairs were also eliminated from further consideration. The names and the forward and reverse sequences of the 1047 InDel markers developed in this study are listed in the supplementary material ([Tables S1](#) to [S12](#)).

PCR amplification and experimental validation of the InDel markers

All of the 1047 primer pairs developed in this study were tested in PCR and polyacrylamide gel analysis. PCR was performed in a 15- μ L reaction volume containing 50-100 ng template DNA, 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl₂, 200 μ M dNTP, 0.2 μ M primer pairs, and 0.7 U *Taq* DNA polymerase (Dingguo Biotech Ltd., Beijing, China). The DNA amplification protocol included an initial 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 55°C, and 1 min at 72°C, with a final extension for 7 min at 72°C. The reactions were performed in a PTC-200 thermal cycler (MJ Research Inc., Waltham, MA, USA). The PCR products were separated on 8% denaturing polyacrylamide gels and visualized using silver staining (Bassam et al., 1991).

Two *indica-japonica* combinations (Taichung65/IR8 and Lemont/Yangdao4) were used for polymorphism validation of the InDel markers developed in this study. Taichung65 (from Taiwan) and Lemont (from the USA) are *japonica* cultivars, whereas Yangdao4 (from Jiangsu Province, China) and IR8 (from IRRI) are *indica* cultivars.

Construction of the physical map and the genetic linkage map

The positions of the 1047 InDel markers in the physical map of the rice chromosomes were constructed according to the alignment of the InDel primers in relation to the Nipponbare reference sequences in GenBank.

In order to assess the value of the InDel markers for genetic research, 12 polymorphic InDel markers between Lemont and Yangdao4 on chromosome 12 were used to construct a genetic linkage map: D1202, D1211, D1220, D1225, D1228, D1239, D1246, D1252, D1260, D1264, D1270, and D1274. The genetic linkage map was constructed using 190 individuals in an F₂ population derived from crossing Lemont with Yangdao4. The linkage map was constructed using MAPMAKER/EXP 3.0 (Lander et al., 1987). The Kosambi mapping function was used to transform the recombination frequency into genetic distances (cM).

RESULTS

Development of InDel markers distributed over the rice genome

There are several InDel polymorphisms between 93-11 and Nipponbare in the rice genome. We selected more than 3500 InDel polymorphic sites dispersed along the 12 rice chromosomes, and tried to transfer them into InDel markers. A majority of candidate DNA templates were found to be unsuitable for primer design because of high (or low) GC content, secondary structure formed by primers, or other reasons. Ultimately, a total of 1047 InDel markers dispersed along the 12 rice chromosomes were developed, approximately one InDel marker every 371.3 kb (the rice genome contains 388.82 Mb according to the International Rice Genome Sequencing Project, IRGSP). The forward and reverse sequences of the 1047 InDel markers developed in this study are listed in the supplementary material ([Tables S1 to S12](#)). The physical positions of the InDel markers across the rice genome are shown in Figure 1. The average interval between adjacent InDel markers varied among the chromosomes, from the shortest (274 kb) on chromosome 1 to the longest (492 kb) on chromosome 11, with an average size of 392 kb across all 12 chromosomes (Table 1).

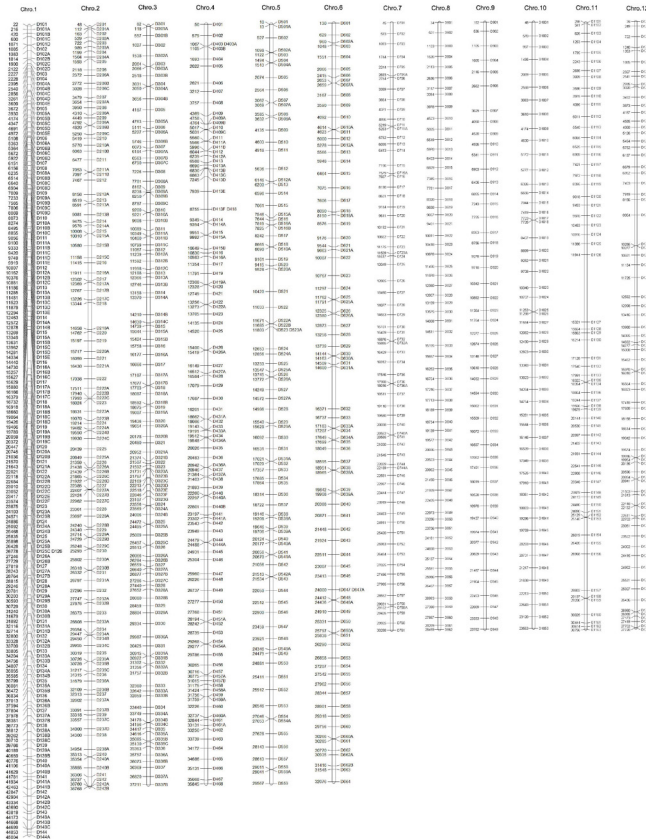


Figure 1. Distribution of the 1047 InDel markers on rice chromosomes. Marker names are listed to the right of the chromosomes, the digit at the left of the corresponding marker name indicates the physical position (kb) of the marker in the rice genome. This physical map was constructed based on the Nipponbare sequences in GenBank.

Table 1. InDel markers developed in this study and their polymorphisms in two *indica-japonica* combinations.

Chromosome	Number of InDel markers developed	Average interval between adjacent markers (kb)*	Number of polymorphic markers between Taichung65 and IR8	Number of polymorphic markers between Lemont and Yangdao4	Number of polymorphic markers at both combinations
1	165	274	107	64	52
2	116	319	74	68	56
3	116	323	78	61	56
4	107	341	57	49	39
5	87	344	47	36	29
6	84	385	55	32	31
7	70	438	47	38	35
8	62	466	29	20	15
9	49	491	24	22	17
10	53	452	26	20	17
11	63	492	39	33	26
12	75	374	40	33	25
Total	1047	392	623	476	398

*The interval between adjacent markers was calculated according to the physical distances of the markers in Nipponbare reference sequences.

The InDel length of the 1047 markers varied from 3 to 39 bp, according to the 93-11 and Nipponbare reference sequences. The majority of the 1047 markers (88.2%) were 6-25 bp InDel in length: 24.4% were 6-10 bp, 28.0% were 11-15 bp, 21.1% were 16-20 bp, and 14.7% were 21-25 bp. Only 6.2% of the markers were ≤ 5 bp in length, and 5.6% of the markers were ≥ 26 bp in length (Figure 2).

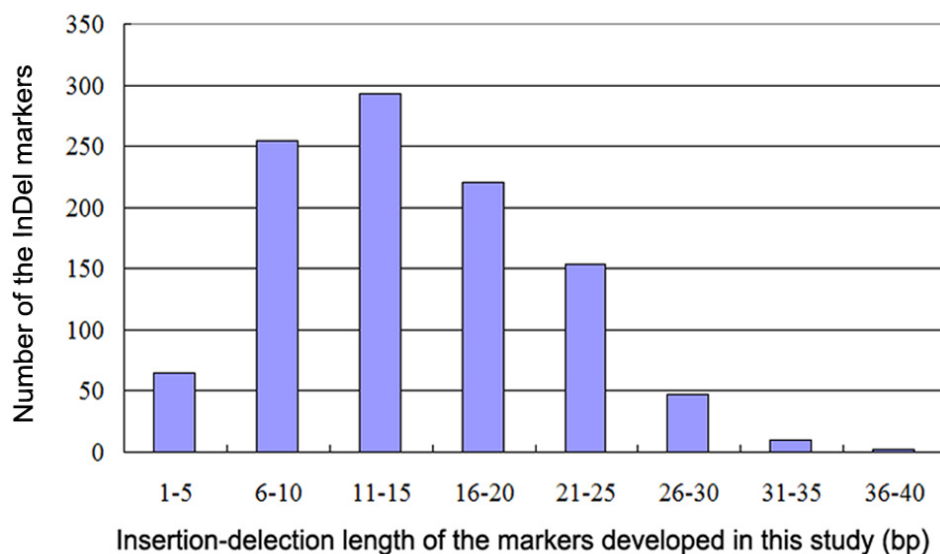


Figure 2. Distribution of the insertion-deletion length of the 1047 InDel markers.

Potential of the InDel markers for genetic research

The 1047 InDel markers were designed based on the InDel polymorphisms between 93-11 and Nipponbare reference sequences. Therefore, these markers should theoretically be polymorphic between 93-11 and Nipponbare. To test their polymorphisms in other *indica-japonica* varieties, we analyzed them in two *indica-japonica* combinations (Taichung65/IR8 and Lemont/Yangdao4). Of all 1047 InDel markers, 623 markers (59.5%) showed polymorphism between Taichung65 and IR8, 476 markers (45.5%) showed polymorphism between Lemont and Yangdao4, and 398 markers (38.0%) were polymorphic in both of the combinations (Table 1). The polymorphic markers are indicated in [Tables S1](#) to [S12](#). These results demonstrated that the InDel markers developed based on 93-11 and Nipponbare sequences can be used in other *indica-japonica* combinations, although the polymorphic rate of the markers varied with different combinations.

We further selected 12 polymorphic InDel markers on chromosome 12 to genotype an F_2 population consisting of 190 individuals derived by crossing Lemont and Yangdao4. A linkage map consisting of 11 InDel markers was constructed (Figure 3), and the marker D1202 did not link with the 11 InDel markers in this map. The positions of the 11 InDel markers in the genetic linkage map agreed with those of the physical map (Figures 1 and 3).

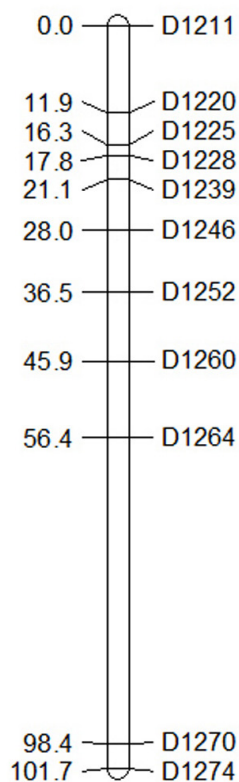


Figure 3. Linkage map of rice chromosome 12 constructed using 190 F_2 individuals derived from a cross between Lemont and Yangdao4.

These results indicated that the developed InDel markers are useful for identifying genetic composition and provide a valuable platform for the molecular dissection of traits.

Relationship between InDel length and polymorphic marker rate

The InDel lengths of the 1047 InDel markers ranged from 3 to 39 bp. It was not clear whether the polymorphic marker rate in a specific *indica-japonica* cross increased with increasing length of the InDel markers. We classified the InDel markers according to their InDel lengths, and calculated the polymorphic marker rate in two *indica-japonica* combinations (Table 2). In the Taichung65/IR8 combination, the correlation coefficient between polymorphic marker rate and InDel length was -0.59888 ($P = 0.1167$), and the correlation coefficient between polymorphic marker rate and InDel length in the Lemont/Yangdao4 combination was -0.19822 ($P = 0.6380$). Therefore, there was no clear relationship between the polymorphic marker rate in a specific *indica-japonica* combination and the InDel length of the developed markers. Therefore, we could not increase the polymorphic marker rate by developing markers with longer InDel sequences.

Table 2. Distribution of the polymorphic marker number at two *indica-japonica* combinations listed according to the InDel length of the markers.

InDel length (bp)	Number (%) of polymorphic markers between Taichung65 and IR8	Number (%) of polymorphic markers between Lemont and Yangdao4
1-5	40 (61.5%)	25 (38.5%)
6-10	146 (57.3%)	112 (43.9%)
11-15	186 (63.5%)	148 (50.5%)
16-20	126 (57.0%)	95 (43.0%)
21-25	97 (63.0%)	71 (46.1%)
26-30	25 (53.2%)	23 (48.9%)
31-35	2 (20%)	1 (10%)
36-40	1 (50%)	1 (50%)
Total	623	476

DISCUSSION

SSR markers have several advantages such as co-dominance, multi-allelic, and technically easy and inexpensive to use, which has made them the most widely used molecular markers in rice since their release (Panaud et al., 1996; Temnykh et al., 2001; McCouch et al., 2002; International Rice Genome Sequencing Project, 2005). SNP and InDel polymorphisms, the first and the second largest polymorphic mutations in the rice genome, provide suitable resources for designing primers with higher density than SSR markers. The higher frequencies of InDel and SNP markers in the rice genome make them preferable for fine mapping rice genes (Pan et al., 2008; Wu et al., 2011; Xu et al., 2011; Kwon et al., 2012; Zeng et al., 2013). It is worth noting that special equipment or hardware is indispensable for SNP genotyping, which has restricted the application of SNP markers. Although PCR-based SNP markers do not require special infrastructure, the development of PCR-based SNP markers is nonetheless difficult, and therefore more rare. InDel markers represent a balance between SSR and SNP markers. They provide a higher density than traditional SSR markers and utilize the same experimental procedure as used for SSR markers. Genome-wide InDel polymorphisms were detected soon after the release of the rice genome sequences (Shen et al., 2004; Arai-Kichise et al., 2011). However, the development of InDel markers with good coverage of the rice genome is time-consuming because not all InDel polymorphisms can be transferred to InDel markers; for example, sequences with high (or low) GC content are not suitable for primer design. Other restricting factors for InDel marker development include: secondary structure formed by primers, melting temperature of the forward and the reverse primers differ greatly, and too many SNPs occur at either side of the InDel candidate sequences. In this study, we copied more than 3500 InDel sequences into the PRIMER PREMIER software program and tried to transfer them to primers, but only 1047 markers were ultimately developed.

The 1047 InDel primer pairs were designed based on the insertion-deletion length polymorphisms between 93-11 and Nipponbare, aiming for utilization between *indica* and *japonica* rice. By analyzing these markers using two *indica-japonica* combinations, it was found that the polymorphic rate of the markers varied with different crosses, from 45.5% between Lemont and Yangdao4 to 59.5% between Taichung65 and IR8. We did not test the polymorphic rate of these markers in *indica-indica* or *japonica-japonica* combinations. Therefore, the potential of these markers within such combinations should be studied in the future.

We aimed to develop InDel markers with an average interval of ~500 kb between adjacent markers along the chromosomes so that the developed markers would have good

coverage of the rice genome. Of the 1047 InDel markers, 73 (7.0%) of them contained SSRs. Because some SSRs were linked with the InDel sequences, the SSRs were included when designing InDel primers. The second reason is that because there were no suitable InDel sequences for developing primers at specific positions of the rice genome, the SSRs were chosen. Finally, we were able to develop a set of InDel markers with an average interval of 392 kb between adjacent markers along the rice genome. This density is suitable for primary mapping of rice genes. At specific positions of the rice chromosomes, the marker density can be augmented according to the requirements of different studies using the marker development procedure reported in this study.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31101004, #31071080), the National Technical Key Project of Transgenic New Variety Breeding (Grant #2011ZX08001-002), a grant from the Zhejiang Province for Public Welfare (#2011C22002), the Basic Research Fund for Agricultural Academy/Institute (#2012ZL094), the “863” Program (#2012AA101201), and a special fund for technical innovation team in Zhejiang Province (#2010R50024).

[Supplementary material](#)

REFERENCES

- Akagi H, Yokozeki Y, Inagaki A and Fujimura T (1996). Microsatellite DNA markers for rice chromosomes. *Theor. Appl. Genet.* 93: 1071-1077.
- Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, et al. (2011). Discovery of genome-wide DNA polymorphisms in a landrace cultivar of Japonica rice by whole-genome sequencing. *Plant Cell Physiol.* 52: 274-282.
- Bassam BJ, Caetano-Anolles G and Gresshoff PM (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem.* 196: 80-83.
- Chen H, He H, Zou Y, Chen W, et al. (2011). Development and application of a set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 123: 869-879.
- Huang X, Feng Q, Qian Q, Zhao Q, et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res.* 19: 1068-1076.
- International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. *Nature* 436: 793-800.
- Kwon T, Lee JH, Park SK, Hwang UH, et al. (2012). Fine mapping and identification of candidate rice genes associated with *qSTV11^{SG}*, a major QTL for rice stripe disease resistance. *Theor. Appl. Genet.* 125: 1033-1046.
- Lander ES, Green P, Abrahamson J, Barlow A, et al. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181.
- Liu B, Wang Y, Zhai W, Deng J, et al. (2013). Development of InDel markers for *Brassica rapa* based on whole-genome re-sequencing. *Theor. Appl. Genet.* 126: 231-239.
- McCouch SR, Teytelman L, Xu Y, Lobos KB, et al. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res.* 9: 199-207.
- Nasu S, Suzuki J, Ohta R, Hasegawa K, et al. (2002). Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Res.* 9: 163-171.
- Pan CH, Li AH, Dai ZY, Zhang HX, et al. (2008). InDel and SNP markers and their applications in map-based cloning of rice genes. *Rice Sci.* 15: 251-258.
- Panaud O, Chen X and McCouch SR (1996). Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* 252: 597-607.
- Ren ZH, Gao JP, Li LG, Cai XL, et al. (2005). A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nat. Genet.* 37: 1141-1146.

- Shen YJ, Jiang H, Jin JP, Zhang ZB, et al. (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol.* 135: 1198-1205.
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, et al. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11: 1441-1452.
- Wu X, Zuo S, Chen Z, Zhang Y, et al. (2011). Fine mapping of *qSTV11¹⁰*, a major gene conferring resistance to rice stripe disease. *Theor. Appl. Genet.* 122: 915-923.
- Xu J, Wang B, Wu Y, Du P, et al. (2011). Fine mapping and candidate gene analysis of *ptgms2-1*, the photoperiod-thermo-sensitive genic male sterile gene in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 122: 365-372.
- Ye CR, Argayoso MA, Redoña ED, Sierra SN, et al. (2012). Mapping QTL for heat tolerance at flowering stage in rice using SNP markers. *Plant Breed.* 131: 33-41.
- Zeng YX, Ma LY, Ji ZJ, Wen ZH, et al. (2013). Fine mapping and candidate gene analysis of *LM3*, a novel lesion mimic gene in rice. *Biologia* 68: 82-90.