GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genome-wide polymorphisms between the parents of an elite hybrid rice and the development of a novel set of PCR-based InDel markers

K. Wang, J.Y. Zhuang, D.R. Huang, J.Z. Ying and Y.Y. Fan

Chinese National Center for Rice Improvement/State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, China

Corresponding author: Y.Y. Fan
E-mail: fanyeyangcnrri@163.com

Genet. Mol. Res. 14 (2): 3209-3222 (2015)
Received April 8, 2014
Accepted November 24, 2014
Published April 10, 2015
DOI http://dx.doi.org/10.4238/2015.April.10.33

ABSTRACT. Genome-wide re-sequencing of the Zhenshan 97 (ZS97) and Milyang 46 (MY46) parents of an elite three-line hybrid rice developed in China resulted in the generation of 9.91 G bases of data with an effective sequencing depth of 11.66x and 11.51x, respectively. Detection of genome-wide DNA polymorphisms, single nucleotide polymorphisms (SNPs), short insertions/deletions (InDels; 1-5 bp), and structural variations (SVs), which is an invaluable variation resource for genetic research and molecular marker-assisted breeding, was conducted by comparing whole-genome re-sequencing data. A total of 364,488 SNPs, 61,181 InDels and 6298 SVs were detected in ZS97 and 364,179 SNPs, 61,984 InDels and 6408 SVs were detected in MY46 compared to the 9311 reference sequence. Synteny analysis of the variation revealed a total of 77,013 identical and 181,737 different SNPs and 15,021 identical and 1205 different InDels between ZS97 and MY46, respectively. A total of

180 InDels 3-8 bp in length between ZS97 and MY46 were selected for experimental validation; 160 polymerase chain reaction products were efficiently separated on 6% non-denaturing polyacrylamide gels. Identification of genome-wide variation among the parents of the elite hybrid as well as the set of 160 polymerase chain reaction-based InDel markers will facilitate future genetic studies and the molecular breeding of hybrid rice.

Key words: Breeding; Hybrid rice; InDels; Polymorphisms; Single nucleotide polymorphisms

INTRODUCTION

A wide range of naturally occurring variation exists in rice cultivars. These variations of polymorphisms in DNA sequence are the basis of genetic diversity and can be exploited as genetic molecular markers for genetic research and molecular breeding. Numerous types of molecular markers have been developed, such as random amplified polymorphic DNA, amplified fragment length polymorphism, and simple sequence repeats (Jones et al., 2009). Advances in re-sequencing technology have enabled identification of genome-wide genetic variations through large-scale re-sequencing, including millions of single nucleotide polymorphisms (SNPs), the most frequent polymorphism in the genomes of most organisms, and insertions/deletions (InDels) (Shen et al., 2004). The discovery of SNPs in plant and animal gene pools through genome-scale re-sequencing has enabled the development of high-throughput SNP markers and SNP genotyping platforms. Unlike SNPs, which have been studied extensively, other forms of nature genetic variation, such as short InDels (1-5 bp), remain largely undetermined.

China's rice breeders began hybrid development in 1964 using a three-line hybrid-breeding system. By 1976, China started large-scale commercial production of the three-line hybrid rice. This technology has contributed to improving food security in China; the increased yield of hybrid rice has helped China feed an extra 60 million people every year (Li et al., 2009a). In the three-line hybrid rice-breeding system, the cytoplasmic male sterility line was crossed with the restorer line (R line) to produce F_1 hybrid rice, and with the maintainer line (B line) for self-reproduction. Shanyou 10, an elite hybrid that dominated hybrid rice production in South and Central China with 2.35 million hectares of coverage in 1990-2005, was bred from the Zhenshan 97A (ZS97A) and Milyang 46 (MY46) parents (Wan, 2010). These are the most frequently used parents in China, either directly as parents for other hybrids or as crossing donors for parental breeding. ZS97A, an elite cytoplasmic male sterility line, is the most frequently used female parent of widely planted hybrids in China. Zhenshan 97 (ZS97), the B line of ZS97A, is a typical early season *indica* variety from south China. MY46 is an elite R line derived from a cross involving three International Rice Research Institute varieties (IR8, IR24 and IR262) (Xu et al., 1989). These 2 varieties are representative parents from 2 heterotic pools (early season *indica* varieties in South China as female parents and low latitude *indica* varieties from International Rice Research Institute or from other Southeast Asian countries as

male parent) identified for the three-line hybrid rice in China. Next-generation sequencing technology allowed the discovery of a large number of SNPs and InDels by comparing high-quality re-sequenced whole genome of individuals. Some studies have investigated the genome-wide DNA polymorphisms between *indica* and *japonica* cultivars (Shen et al., 2004), among *japonica* (Yamamoto et al., 2010; Arai-Kichise et al., 2011) or among *indica* R lines (Li et al., 2012). Until recently, however, little information was available regarding genome-wide polymorphisms between elite hybrid parents, particularly the parents of large-scale planted elite hybrids in China. Further studies examining the genetic diversity of elite hybrid parents can improve our understanding of the 2 heterotic pools of hybrid rice breeding and promote the improvement of R lines and B lines for super-hybrid rice breeding.

Many genes/quantitative trait loci (QTL) have been cloned by QTL mapping using *indica*/*indica* dual-parent segregation population (Fan et al., 2006; Xue et al., 2008). *Indica*/*indica* crossing is a common pattern observed in molecular marker-assisted selection (MAS) breeding in *indica* variety breeding. Limited marker information among *indica*/*indica*, however, is a restricting factor in genes/QTL mapping and MAS breeding. To further investigate the utility of short InDels (3-5 bp) among elite three-line hybrid parents in QTL mapping or MAS breeding, we verified the availability of a set of short InDels as a new polymerase chain reaction (PCR)-based molecular marker resource.

The objective of the present study was to identify whole genome-wide DNA polymorphisms between the B line (ZS97) and R line (MY46) of an elite 3-line *indica* hybrid (Shanyou 10) in China as well as investigating the possible use of short InDels. The whole-genome re-sequencing analysis of ZS97 and MY46 was completed using the Solexa sequencing technology. The sequence-generated reads were mapped to the high-quality 9311 genomic sequence, and genome-wide variations were identified through comprehensive detection of SNPs, InDels, and structural variations (SVs) across the genome. The discovery of these genetic variations provides vital clues for unraveling the genetic basis underlying heterosis in hybrid rice and for improving hybrid rice. A new set of 160 PCR-based short InDel markers developed in this study is a new economically attractive DNA marker resource for genetic research and MAS breeding.

MATERIAL AND METHODS

Sample preparation and sequencing

Genomic DNA was extracted from the leaf tissues of an individual plant using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The DNA from each line was then fragmented randomly. After electrophoresis, DNA fragments of the desired length were gel-purified. Adapter ligation and DNA cluster preparation were performed and subjected to Solexa sequencing (Illumina, Inc., San Diego, CA, USA).

Read mapping

The raw pair-end (PE) sequencing reads were aligned to the 9311 reference genome

sequence using SOAPaligner (Li et al., 2008) using the following criteria: if an original read could not be aligned to the reference sequence, the first base of the 5'-end and the last 2 bases of the 3'-end were deleted and then realigned to the reference. If alignment could still not be achieved, 2 more bases from the 3'-end were deleted. This procedure was repeated until the alignment was successful or the modified read was less than 27 bp. The average re-sequencing depth and coverage were calculated based on the alignment results.

Assembly of consensus sequences and detection of SNPs

Based on the alignment results and considering factors such as data characters, sequencing quality, and some other factors from experiments, a Bayesian model was applied to compute the likelihood of genotypes with the actual data. The genotype with the maximum likelihood was selected as the genotype of the re-sequencing individual at a specific base, and a quality value was designated accordingly to reveal the accuracy of the genotype. Polymorphic loci against the reference sequence were selected and filtered if specific criteria were met (e.g., the quality value >20 and at least 2 supporting reads) using SOAPsnp (Li et al., 2009b).

Detection of InDels and SVs

Mapped reads that met the PE requirements and contained gaps at only 1 end were used to detect short InDels (≤5 bp). The lengths of the detected InDels were within the range 1-5 bp. Gaps supported by ≥3 PE reads were retained using SOAPindel (Li et al., 2013). According to the principle of PE sequencing, 1 PE read should be aligned to the forward sequence and another should be aligned to the reverse in normal situations. The distance between the 2 aligned positions on the reference sequence should be in accordance with the size of the insert. If the directions or spans of the alignments of the 2 paired reads were different from what was expected, the region may contain SVs. Abnormal PE alignments observed in our analysis were further analyzed by clustering and compared with previously defined SVs. The SVs were detected using SOAPsv (http://soap.genomics.org.cn) with support from ≥3 abnormal PE reads.

Variations of SNPs and InDels between ZS97 and MY46

Variations of SNPs and InDels detected for the 2 parental lines were further compared to identify the identical and different SNP/InDel variation. Only variations with ≥3 effective sequence reads were mapped for 2 individuals and selected for further comparison.

Development of short InDel markers

InDels with insertion/deletion size ≥3 bp between ZS97 and MY46 were selected for further marker design. The Oligo 7.57 software (Molecular Biology Insights, Inc., Cascade, CO, USA) was used to design PCR primers, with a constraint of generating products of 70-180 bp. Genomic DNA was extracted from the leaves of ZS97 and MY46 (Zheng et al., 1995). PCR amplification was conducted as described (Chen et al., 1997). PCR products were separated by electrophoresis (6% non-denaturing polyacrylamide gel) and visualized by silver staining.

RESULTS

Genome sequencing

Whole-genome sequencing was conducted on the genomic DNA of ZS97 and MY46 using the Solexa sequencing technology. Two DNA libraries were constructed and 9.91-G bases were generated. The alignment of reads was used to build consensus genome sequences for each rice accession. Approximately 84.04% high-quality raw databases were aligned with the reference sequence of the 9311. The resulting consensus sequence of the 2 rice accessions covered 88.69 and 88.58% of the reference genome; on average, effective sequencing depths of 11.66x and 11.51x were achieved for ZS97 and MY46, respectively (Table 1).

Table 1. Summary of original re-sequencing data.

Sample	Insert size (bp)	Bases (G)	Mapped bases (G)	Depth (%)	Coverage (%)	Mismatch rate (%)
ZS97	477	4.98	4.19	11.66	88.69	0.77
MY46	472	4.93	4.14	11.51	88.58	0.66

Identification and distribution of variation across the rice genome

A total of 364,488 SNPs, 61,181 InDels, and 6298 SVs were detected between ZS97 and the 9311 reference sequence; 364,179 SNPs, 61,984 InDels, and 6408 SVs were detected between MY46 and the 9311 reference sequence (Table 2). The total polymorphisms detected in the 2 parental lines varied across different chromosomes. The largest number of polymorphisms (52,639 and 56,817) was observed on chromosomes 1 and 5 for ZS97 and MY46, respectively, while chromosomes 5 and 11 showed the lowest number of polymorphisms for ZS97 and MY46, respectively.

The genomic distribution of DNA polymorphisms (SNPs, InDels, and SVs) between the 2 accessions and the 9311 reference genome were examined by calculating the frequency of polymorphisms observed for each 500-kb interval along the chromosome. The average density of DNA polymorphisms detected per 500 kb across the genome in ZS97 and MY46 were similar to the 486.6 and 486.2 SNPs, 81.7 and 82.7 InDels, and 8.4 and 8.6 SVs, respectively (Table 2 and Figure 1). The frequency of SNPs within the genome with chromosomes 11 and 5 showed the highest (721.0) and the lowest densities (368.7) per 500-kb interval in ZS97, while chromosomes 11 and 5 showed the lowest (313.0) and the highest densities (765.0) per 500-kb interval in MY46; similar frequency results with SNPs were observed for InDels. The frequency of SVs within the genome with chromosomes 11 and 7 showed the highest (12.0) and the lowest densities (5.7) in ZS97, and MY46 showed chromosomes 5 and 2 with the highest (11.5) and the lowest densities (6.0) (Table 2).

Of the total short InDels (1-5 bp) detected between samples and the 9311 reference sequence, 30,411 insertions and 30,770 deletions were detected in ZS97 and 30,920 insertions and 31,064 deletions were detected in MY46. A skewed distribution was observed for InDel length, and most (72.5 and 72.1%) were mononucleotides, while 11.6 and 12.0% were 3-5-bp insertions or deletions in ZS97 and MY46, respectively (Table 3).

Table 2. Variations detected in each sample.

Chromosome	SNPs				InDels				SVs			
	ZS97		MY46		ZS97		MY46		ZS97		MY46	
	Count	No./500 kb	Count	No./500 kb	Count	No./500 kb	Count	No./500 kb	Count	No./500 kb	Count	No./500 kb
Chr1	43,538	460.4	42,801	452.6	8195	86.7	7915	83.7	906	9.6	916	9.7
Chr2	31,464	412.9	29,801	391.0	5709	74.9	5017	65.8	500	6.6	460	6.0
Chr3	40,430	482.6	38,526	459.9	7665	91.5	7732	92.3	730	8.7	708	8.5
Chr4	32,184	463.5	25,979	374.1	5111	73.6	4065	58.5	585	8.4	487	7.0
Chr5	23,034	368.7	47,797	765.0	3876	62.0	8301	132.9	404	6.5	719	11.5
Chr6	25,005	379.9	20,544	312.1	4355	66.2	3603	54.7	462	7.0	423	6.4
Chr7	25,024	447.5	29,992	536.4	3696	66.1	4870	87.1	320	5.7	418	7.5
Chr8	29,303	482.0	42,263	695.2	4825	79.4	6522	107.3	614	10.1	709	11.7
Chr9	28,413	653.0	19,717	453.1	4291	98.6	3324	76.4	326	7.5	327	7.5
Chr10	24,964	562.2	23,188	522.2	3907	88.0	3890	87.6	420	9.5	442	10.0
Chr11	33,219	721.0	14,418	313.0	5182	112.5	2151	46.7	551	12.0	297	6.4
Chr12	27,910	605.4	29,153	632.4	4369	94.8	4594	99.7	480	10.4	502	10.9
Total	364,488	486.6	364,179	486.2	61,181	81.7	61,984	82.7	6298	8.4	6408	8.6

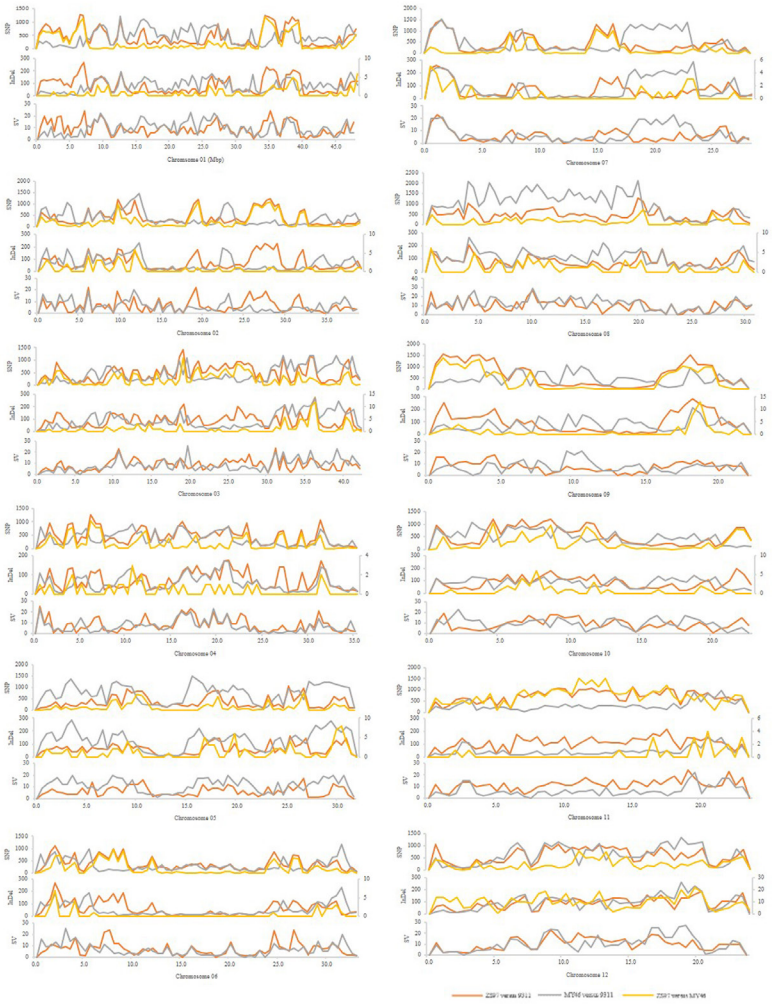

Figure 1. Distribution of DNA polymorphisms identified between 9311, ZS97 and MY46 along each chromosome. The x-axis represents the physical distance along each chromosome, split into 500-kb intervals. The first y-axis indicates the number of three kinds of DNA polymorphisms and the second y-axis indicates the number of InDel polymorphisms between ZS97 and MY46.

Table 3. Number of short InDels identified in the two re-sequencing samples.

InDel size (bp)	ZS97 *vs* 9311	MY46 *vs* 9311	ZS97 *vs* MY46
1	44,371	44,689	612
2	9714	9867	413
3	3675	3822	117
4	2388	2539	39
5	1033	1067	11
6	0	0	12
7	0	0	0
8	0	0	1
Total	61,181	61,984	1205

Genetic variation between ZS97 and MY46

DNA variations between ZS97 and MY46 may reflect the basis of genetic divergence of B lines and R lines of three-line hybrids. Synteny analysis of the variations revealed a total of 77,013 identical SNPs and 181,737 different SNPs as well as 15,021 identical InDels and 1205 different InDels between ZS97 and MY46 (Table 4 and **Table S1**). The frequency of SNPs between ZS97 and MY46 varied from 283.1 SNPs/Mb on chromosome 8 to 984.7 SNPs/Mb on chromosome 11, with an average of 485.2 SNPs/Mb; the frequency of InDels between ZS97 and MY46 varied from 1.4 InDels/Mb on chromosome 4 to 5.9 InDels/Mb on chromosome 12, with an average of 3.2 InDels/Mb (Table 4).

Table 4. Variations detected between the two re-sequencing samples.

Chromosome	SNPs			InDels		
	No. identical	No. different	No. different/Mb	No. identical	No. different	No. different/Mb
Chr1	9231	22,978	486.0	1921	157	3.3
Chr2	4281	18,282	479.8	1110	92	2.4
Chr3	9464	20,638	492.7	2001	180	4.3
Chr4	8203	14,283	411.4	1609	48	1.4
Chr5	5013	8890	284.6	1046	153	4.9
Chr6	3330	13,080	397.4	926	72	2.2
Chr7	6091	10,261	367.0	1069	56	2.0
Chr8	9863	8605	283.1	1682	120	3.9
Chr9	3785	18,452	848.1	723	89	4.1
Chr10	6436	11,949	538.1	1107	65	2.9
Chr11	3342	22,682	984.7	690	38	1.6
Chr12	7974	11,637	504.9	1137	135	5.9
Total	77,013	181,737	485.2	15,021	1205	3.2

Of the total 1205 different InDels detected between ZS97 and MY46, length varied in the range of 1-8 bp, indicating bias towards mononucleotide InDels, most (50.8%) of which were mononucleotides, 34.3% were of 2 bp variation, and 14.9% were of 3-8 bp variation (Table 3). InDels ≥3 bp, which can be observed easily on polyacrylamide gel electrophoresis, are potential variations that can be used for PCR-based marker design.

Potential of short InDels as PCR-based markers

Whole-genome sequencing technology allowed the detection of InDels as well as of SNPs. To validate the availability of InDels identified between the two 3-line restorer and maintainer accessions as novel PCR-based DNA markers, we selected short InDels ≥3 bp and converted them to PCR-based markers. A total of 180 InDels (3-8 bp), which were distributed across the 12 chromosomes varying from 30 InDels on chromosome 3 to 6 InDels on chromosomes 7 and 11, were selected between the ZS97 and MY46 genomes. Based on this selection, we designed primer pairs to amplify 70-176 bp surrounding the InDels. Following PCR analysis, all 180 primer pairs showed reliable amplification using genomic DNA of ZS97 and MY46 as the DNA template; 160 of the InDels (88.9%) identified were polymorphic between ZS97 and MY46 as determined using electrophoresis (6% non-denaturing polyacrylamide gel) and 20 were non-significant polymorphisms [Table 5, **Table S2 and Figure S1 (A-L)**].

Table 5. Sequence, product size (bp), InDel size (bp), Tm (°C) and other positional information of 160 novel markers.

Marker name	Chr	Position	InDel	Size	Tm (°C)	Forward primer (5'-3')	Reverse primer (5'-3')
ZM01-6166	1	6,166,310	4	113	55	GATCAATGTTAAAGCCGTCA	CTACACCTATGTTCCGTTC
ZM01-25592	1	25,592,392	3	132	55	CCACCATCGTACTTACTTCC	AGTCTTATATACCCCGTATCTG
ZM01-25795	1	25,795,785	3	128	55	TGCATTCTTACAACACCCAA	AAAAGATGAGATTGTCGCTA
ZM01-29697	1	29,697,596	4	107	55	TAAAAGATTGAGACGGGCATCACA	AGTCAAACCCCGTGAAATACACC
ZM01-31987	1	31,987,065	4	137	55	AAACAAACATAAGTACTGGC	TTTCCGTCATTGATAGCTTC
ZM01-33311	1	33,311,272	3	141	55	ATCTTGATACTTGCATGGGTT	GGCTGAATAGTCATACCATCC
ZM01-33599	1	33,599,594	3	127	50	ATTTAATAGTCCGGAGCATC	GCGTAATGAACACTATCAGC
ZM01-33635	1	33,635,552	4	129	55	ATCTTAAGTTTGTTACAGC	TCTTGCAGAGCAGGTACA
ZM01-33763	1	33,763,269	3	105	55	TATAAAACGTCGGGAACGGTCA	GTAGAGTTGTCTGCATCCAT
ZM01-34040	1	34,040,444	3	115	55	TACTAGAGTAGGCTATTCCA	TGTGAGCCCCTTATCAGA
ZM01-34496	1	34,496,507	3	96	55	CAAATGGAGCAAACATAACAGC	GCCAATTAGTTGCGTCGAAG
ZM01-34659	1	34,659,898	3	133	55	TAGTGTAAATCGAACTAGCACA	TTGTAACACCTATCTTTCGGA
ZM01-34926	1	34,926,445	6	122	55	CATCCGATATGTAGCTCT	TCTCTCGGATCATCGATT
ZM01-34939	1	34,939,626	3	144	55	GGATCTAAACTCGGCCTA	CCTAGCAGTACTTTCAGTC
ZM01-35109	1	35,109,705	3	140	55	ACAGTAAAGCTAACGTGGAA	TCCCACATGATTACCGTG
ZM01-37479	1	37,479,620	3	123	55	TTTTATTGAGCAAATGGATAGGAA	TCATACTTGGCAGTCTCC
ZM01-38175	1	38,175,533	3	106	55	CCGAGATGCTATAATCTAACTGG	CTTTATGCTTGGGATAGGTT
ZM01-45748	1	45,748,039	4	131	55	CCACACCACCCCATTTCGTTC	CAGCGTCCCTAGTCCGTCC
ZM01-47056	1	47,056,664	3	135	55	TTTTTGGGTGGTGCTAAACT	AACACGATGAGTTTTAACATT
ZM01-47134	1	47,134,640	3	123	50	CTCGTTAATTTATCTTGTCCA	AGTATGAGACATATCAGTCCA
ZM02-1019	2	1,019,438	3	129	55	CTTTATGAATTATATGAGAGCGAT	ACAGAATGCTAATTTGATTCGG
ZM02-1395	2	1,395,978	4	156	55	ATGAACAAACCTTGGCAAT	ATTGAAACCGCATAAGCA
ZM02-2943	2	2,943,514	3	123	50	CATGTCTAAATGTGTTGCCACT	AATATTATAGACCACTCACCCCTT
ZM02-3635	2	3,635,575	4	144	55	CTACCGAAGTACTAGCAT	CAATAAATCAATCTTGTCGGAA
ZM02-5606	2	5,606,693	4	169	55	GTTGTAAGGTCAAACTTCCGTA	TGAAACTTGTCGATGCACCA
ZM02-8282	2	8,282,532	3	154	55	GAGAATAACTAAAATGCCCTT	TACGTACTCACTCCGTCA
ZM02-9368	2	9,368,347	3	141	55	ATTTTCTCGGCAAATTCGT	TTAACCAGAAGCAGTAGCCA
ZM02-10987	2	10,987,902	6	176	55	CAAATCATCTGCAATTCTG	TGGGTGTAATTATATTAACCT
ZM02-11418	2	11,418,481	3	148	55	ATTAATATACTGCAAAGCTTGG	TTTGGTTTATCGTTAGCCCTG
ZM02-11702	2	11,702,737	4	172	55	CGTATGTAATTTAATATGTGCT	TTTTCAGTCTGTTTGCACT
ZM02-12054	2	12,054,902	6	104	55	TTTTATTCCGTCCAAGCTG	AAAATGCTCATACATGTGTC
ZM02-22054	2	22,054,114	3	140	55	ATATGCTGAAACATACTCCC	TGTCTATTCTCTCCGTCCA
ZM02-26794	2	26,794,752	4	172	55	AAAATATGTGCAAATGCC	TTTTCTCCAATTAATGTGGG
ZM03-4264	3	4,264,953	3	118	55	CACGGTTTACTGCTACCA	AAAATTACATTGGCCGTT
ZM03-4282	3	4,282,337	4	120	55	TATTCTGAACAGCCGCAAC	TTTTCTTTGCCAAGCTGACA
ZM03-7255	3	7,255,690	3	125	55	CATGGTCATATGCTGCTAC	TAATGTCTCAAGGCGTTT
ZM03-7384	3	7,384,543	3	137	50	ACTAGAGAAGAAAACGAT	CAATTTTAATTACCCCTG
ZM03-9504	3	9,504,322	5	141	55	TGTTAAAAGTCAACGGTGTC	TAAACTGACCTCTTTTTAATGC
ZM03-9566	3	9,566,204	3	138	55	CACATTTGAGATCGGATTGC	GAAGATAAAAGAGCCCTTCA
ZM03-9841	3	9,841,318	5	70	55	CCGGGAAATGAAGGTAACCAT	CTGTCCCGTATGTCCGTGTT
ZM03-10702	3	10,702,753	3	135	55	GCATGGGATTCATCAACA	TTACCCACTGCAATACACA
ZM03-10792	3	10,792,912	3	118	55	GTATGATTGGAGTACTTCCTG	ACATGACAATGAGATATTCGG
ZM03-12340	3	12,340,392	4	138	55	ATATTTACTAGTGGGTACCAA	CTAGTTGCCTTTACGTGA

Continued on next page

Table 5. Continued.

Marker name	Chr	Position	InDel	Size	Tm (°C)	Forward primer (5'-3')	Reverse primer (5'-3')
ZM03-14936	3	14,936,465	3	139	55	AGATTAAACTCAACTTCGTGT	GGCCCATTACATTAGTCCA
ZM03-17320	3	17,320,113	4	91	55	GTATCTTGAACATAGACCCAT	AGTAGTTGTTGATAATATATTGCC
ZM03-27028	3	27,028,004	3	104	55	ATATAGGTGATTTGTGTATGTGA	CTGGCTTTCTAATTACATCC
ZM03-32256	3	32,256,144	8	100	55	AGACACGGACCTATGCTG	ATCCAATCGTTCATTTAGCAGG
ZM03-32568	3	32,568,894	3	105	55	ATCCAAAGGCTAGTATTCAGGT	GGAAAACCAATTGCCAAGTGT
ZM03-33520	3	33,520,353	3	140	55	CTAATCCCTACCATCCGAAC	TCCTGCAGATTCACGGAC
ZM03-33760	3	33,760,330	4	150	55	ATGGATTGGCATAAGTTT	TTCGTTGTACTATGATGGAT
ZM03-33886	3	33,886,043	3	119	55	CGGTAAAACGCCACTGAT	TTCTGAAGCATGGACCTCT
ZM03-33991	3	33,991,818	3	150	55	TGCTTGATCACGATTGGAGA	ATCCCAAGTTTGTGTTGTG
ZM03-34397	3	34,397,580	3	100	55	GTCAACAAGAGCATAAACGCTA	TTATGTTGCTCATTGATCGGT
ZM03-34783	3	34,783,271	3	143	55	TTTATATTAAGCATTGCCCAC	TAACTTCTTTGACCACCG
ZM03-35597	3	35,597,296	6	113	55	GGTGATTGATTCTTTAATTTGGG	CATCATAATACGCAAGCCTT
ZM03-38622	3	38,622,310	4	150	55	GCCTCTTGCTCTATTTATAGTGA	CCCCATACCTCAAGAAATGTGAC
ZM03-38953	3	38,953,914	4	91	55	AATAAAGACGTTGGCACCT	TCGCTGATCACTCCATGCT
ZM03-39077	3	39,077,173	4	101	55	TCTGTATTTTGGACCGCATC	CCTTGTTTACATCAGCGATT
ZM03-39332	3	39,332,326	4	97	55	TGCCGTTTAATAGGTAGCAC	AAACATGATCCACACAGACCACG
ZM03-39688	3	39,688,207	4	120	55	CTAGAAATAAGCAGCAAACCA	GGCCATATTCTGCATCGAAA
ZM03-39993	3	39,993,656	3	147	55	ATCGCATCCATTAAAATATACCTT	CTAACTGACACTAGGCGTGG
ZM03-40842	3	40,842,904	4	145	55	CTCTTGCATTTGGTAGGGTT	TTGTCCTTTCGTCACGGGTT
ZM04-3974	4	3,974,926	6	108	55	GTTCACCCTAATGTCCTC	TTTTCTGCAAGTTTCATGT
ZM04-5158	4	5,158,576	4	136	55	CCGGGTATTAAATGATAATAGC	CCTGGAAGTTTTGACTCG
ZM04-9340	4	9,340,960	3	128	50	ATTTACTATAAAGAATGCTCCC	AACATTTATGGTTTACTCCCC
ZM04-10111	4	10,111,231	3	128	55	TTCCCCGTCATTTACACA	TTAGTAATGGCAATCTAGGAA
ZM04-14055	4	14,055,277	3	128	55	TCAATATTCTAGTGGTGCTCT	CATTTATCCTCGGTTGCAT
ZM04-26146	4	26,146,421	3	93	55	TATTGGGCACTATAGCTTC	ATAAGCCAAAATAACTCACA
ZM04-30866	4	30,866,602	3	142	55	AACATCGATTAACTGAGCTCC	TCTTGGGCATCTGCGTCA
ZM05-1522	5	1,522,808	3	95	55	CAAGGGATCTAAATAATGCAG	TGTCACAGATTATTTGCTCC
ZM05-2926	5	2,926,961	3	128	55	CGGAGGTTTTCATTCCAAT	CCTATTCGGGATACATAAACTG
ZM05-3006	5	3,006,943	3	110	55	GAACTAACTTCATCAACACT	TCTTTCTGTTCATAAATTTGC
ZM05-3244	5	3,244,940	3	141	55	CTCTCCTTTACCTCTCGCATC	GAGAGCGAAACCATCACCC
ZM05-8478	5	8,478,173	4	131	55	GGAGCTTATCATAAGGCAAT	AAGTAGAGCAAAATTTAGGAA
ZM05-15915	5	15,915,956	3	149	55	TGCGTTAGGACTCTCACT	CCTTTTCCGATAATGGAG
ZM05-16346	5	16,346,648	3	94	55	GAAGTTTCCGTAAACGAT	CCCACACAAGAAGCATCC
ZM05-17183	5	17,183,861	4	82	55	GCCAGACCTCCAGTCCAG	GCACCAACGACAGGTTTACTCAC
ZM05-17524	5	17,524,431	3	75	55	GATCATTCACATTTCGTCCCA	TCCGATTTGTTAATGTAGACTCA
ZM05-17650	5	17,650,174	3	94	55	TAGCCCGTATCAAATCCGTA	AGTTATAACGGTACTAAGCTTG
ZM05-21785	5	21,785,819	3	127	55	ATGAATACCGATAAATTCCACT	TATATGTTTTCAACCCGCAGA
ZM05-22482	5	22,482,725	4	87	55	TGGACTTAATGACACATCCCTT	GCTGACGTCCTGCTCCCG
ZM05-23154	5	23,154,510	3	133	55	CCCTAGTTCTAATTAGTTGCCAA	CCTCCAACGTTTGCCTTT
ZM05-25028	5	25,028,363	3	102	55	TGAAGAGGTCCCAATATACTCC	AGGTAAGGGTAAAGAACCAT
ZM05-25743	5	25,743,359	3	131	55	TATGAATACCTGGTGTCAAAC	TTTCTTGAGGTAACCCAT
ZM05-27090	5	27,090,463	5	127	55	TCGTCATTTCCTTATCTGTG	TTATAGGTTACCATGTACAGC

Continued on next page

Table 5. Continued.

Marker name	Chr	Position	InDel	Size	Tm (°C)	Forward primer (5'-3')	Reverse primer (5'-3')
ZM05-28440	5	28,440,781	3	120	55	AACTTTTGCAATAAAACATGACCC	TCTCGTCTTCGCGTGCTT
ZM05-28686	5	28,686,164	3	149	55	AGGAAATAACTAAAATTCGGA	TATATCCAAGCGAACTCCA
ZM05-28868	5	28,868,796	3	145	55	GCTCCAGACTCTTATCCA	CCAGTTGCGATATGGTTC
ZM05-29129	5	29,129,349	3	142	55	TATTATTTTGGAGTAGGCTCA	CCAGAATACTAGATATAACGCAT
ZM05-29275	5	29,275,136	5	150	55	AAATTTGGTTGTTAACACCT	ACTTATCTAACTAATAAGGCAT
ZM05-30095	5	30,095,224	3	104	55	TCAGGGACGTCTATACAGCA	CGCCTAATATACATTGCGTT
ZM05-30357	5	30,357,609	3	141	55	ATGTTTTACGCAAATATGGAC	GCAGCACATTAGTTCACC
ZM06-432	6	432,379	3	130	55	GTGATACAACATACCCCACT	CATTAGGCTAGTTTTGCTCA
ZM06-504	6	504,204	3	149	55	GAATGATACTTAGATGCTCA	TTACCTTTCCAAACCGATG
ZM06-783	6	783,406	3	150	55	TTCTTTCGTTCTGTTCTTGTC	TATCAGAAAAAGCACCGACAG
ZM06-1045	6	1,045,092	3	117	55	TGTACAGATATCTAGCAGT	CAACCTGCATATATTTGACA
ZM06-1750	6	1,750,770	4	135	55	GTTTAACTCATATGGACATGC	AATAAAAGCTAAAATGACCGA
ZM06-5000	6	5,000,764	3	96	55	CCATCCTATAGCCAAGCTC	ATATGACAGGAAATCGGCAA
ZM06-5515	6	5,515,110	3	109	55	AAGAATTCATAGCATGCAA	TCAAATTTATCCTTGAACCCA
ZM06-28630	6	28,630,843	3	139	55	CTTCTCCCAATCAATTGCAAA	TCAGCTACTTTAGAAAACACCA
ZM06-28869	6	28,869,677	3	139	55	TTTATAGGTTTCAGTGCCAT	AAATAAATACTACCAGGTCCA
ZM06-30218	6	30,218,566	6	133	55	CAGATAATTTTCCCGCAAA	TCCTAATTTAAAAGGTTGCAT
ZM06-30963	6	30,963,562	3	103	50	TCAATAGAACTTCGCATCGTC	CACAGCCCAATCTATATTTGT
ZM07-3639	7	363,970	4	138	55	CAATCGTATTTCTGCCCGTTG	TATAACAGTCAAGTTGGGCAT
ZM07-7975	7	7,975,469	5	91	55	GCGTTTATAGGAAGTTTAATGGAC	AATCTAGATGAAGTTCACCGT
ZM07-8178	7	8,178,566	3	98	55	AGTCCCATATTAAATAACCAG	GAATTATAAAAGCTGCCAAGT
ZM07-8451	7	8,451,549	4	146	55	CATAATCTAGACAAGGGGTT	TGAAATTAACATCACACCGACA
ZM07-22761	7	22,761,930	3	136	55	CAAGAGATCCAGCAAATCCG	AATAATTGGGGTTTGAATCTGACA
ZM07-22876	7	22,876,136	3	146	55	TATGATAGCCTGATTTGGAAC	TAACCCTGTAATCTGGCAA
ZM08-4246	8	4,246,371	3	91	55	TGTCTATAAATACCATGCGAGA	TCCTATTAATTGTTGCCAACC
ZM08-4843	8	4,843,691	3	131	55	CCTGCTTATTCATCTACAATTCCA	TGCTCTACCTTCTTCCGTCA
ZM08-6821	8	6,821,648	6	148	55	CACATGTCACTCATATGAAA	TGTCAGCCACATATAGAAA
ZM08-8654	8	8,654,159	3	86	55	CTTCTAGCGAATATTTAGACC	TTATAGCCTTTGACACGAG
ZM08-9537	8	9,537,442	3	99	55	CAATGAGATGAAAGCACCA	TGAACTCTGAAAATTGCAGT
ZM08-10668	8	10,668,087	5	134	55	CATGCTCCCCAATGTTCA	TTTGTCAATTCAGCTTGTCT
ZM08-11287	8	11,287,354	5	92	55	TCAAGATCTGATGTCCCA	TATAAAAGAACCTGAATCGTC
ZM08-11640	8	11,640,102	3	133	55	TGTGACTAATGAATAACCACT	CAATTTGTAGGTGGAATCCAA
ZM08-12105	8	12,105,218	6	107	55	ATGAAACAAGATATGGCGTGA	CATGCGAATTAAGGGGCTAT
ZM08-13161	8	13,161,662	3	95	55	ACTTCTTTTAAGCGATGTGG	TTTTTCGGACGGATGGAGTA
ZM08-13872	8	13,872,516	3	70	55	TCCACATTTTGTCTCAAGTGC	CTAGCAAATACTAGAATCAGCA
ZM08-15115	8	15,115,235	3	133	55	CCACTTGCATATTTACCCT	ATTTTCCCATACTTTTGACAT
ZM08-19485	8	19,485,711	3	98	55	TAATTAATGGGCTACTCG	ACGTGAATTTTATGCATT
ZM08-24816	8	24,816,483	4	114	55	ATAAGATAGAGGCTGCTAGGAA	AAAAGAAATGTACAGATACCCC
ZM09-16974	9	16,974,717	3	129	55	TTTCTTCATACAGGTTCTGC	GTAGGCCATAATTAAATCAGC
ZM09-17484	9	17,484,031	3	138	55	TGGTCATATTTATATAGGGTT	GAAGATGCTTAACAGTCA
ZM09-18092	9	18,092,369	3	93	55	TGTTAGTTCTGATTGGCATT	AAAGATTTATTTGTGGACGA
ZM09-18126	9	18,126,200	5	147	55	GAAAACCCCTAAAATTGAGT	CATATTGTCTGCAACTGTG

Continued on next page

Table 5. Continued.

Marker name	Chr	Position	InDel	Size	Tm (°C)	Forward primer (5'-3')	Reverse primer (5'-3')
ZM09-19238	9	19,238,963	3	119	55	GTGGCTACTTTTGTACACT	TCAAAAGCTGAACATTCCCT
ZM10-3611	10	3,611,607	3	108	55	CCTGAGATTTATAACATAGCCTT	GCTCCATACAGATCCCTT
ZM10-5325	10	5,325,532	4	144	55	AATTAATCACAGCCCGAGCAG	ATGAGTCAATGAAAATGCCTCT
ZM10-6389	10	6,389,180	3	130	55	GCCACAGTTAGACATATAGCC	TCGGGTTGCTCAATTTAACTCT
ZM10-6454	10	6,454,501	6	78	55	AGAGACCTAATAGTGTTGTGC	AGATATGTTTCTATTACAAGCCAT
ZM10-7448	10	7,448,659	3	134	55	TTTAGTTCTTGCGAAGTCC	GATAAGTTTCGTTGCCAAA
ZM10-7540	10	7,540,949	3	147	55	CCTTGTTTTCCTCGCGTA	CCTTAGCACATCATTAATTGAG
ZM10-7946	10	7,946,095	3	119	55	GTGATTGCCAATAATTAGACA	TTATTGATACTGTATGCACCA
ZM10-8064	10	8,064,852	4	105	55	CCTCGATTCATATTATCCTACTCC	TGTGTGTCTGTATGTCCGTC
ZM10-9944	10	9,944,177	3	138	55	GTCATTTCCCATTGGACT	CAAACCAGTAAGATTGAACA
ZM10-15528	10	15,528,270	3	95	55	CTGTGATGGCTTAGTTCT	GTTAAACAACTGTAGTGCAT
ZM10-19192	10	19,192,458	3	114	55	TTCTTCAGAGATGGGATT	ACTAACCGTAGAAATAGTCA
ZM11-1639	11	1,639,430	3	100	55	TGTTCTTAAGCAGTTACCAGC	CCATGTTACGCAGTATTCCC
ZM11-8415	11	8,415,061	3	149	55	TGTTGAAGTCCAGCTTAAGTG	TCTTGTCAAGCTGTTGTGA
ZM11-9545	11	9,545,870	3	121	55	ATTTTAAATATAAGACGTAACCT	TATTATTACCTCCACGCTT
ZM11-21467	11	21,467,722	3	95	55	TGACCCAAAACATTTAGCTG	ATATTATCGTAGCGTTAGCAC
ZM11-22862	11	22,862,451	6	126	55	TATTTTCAGAAAGCAAGCC	TTCTTTTGATAGAGAATCGC
ZM12-7977	12	7,977,815	4	103	55	CATGGATAATCGCCAAGACA	TATCTCAGTTGGTTTAGTGCAT
ZM12-8009	12	8,009,626	6	96	55	GTCAATTCAATGGCACCAGA	GAAATTTGGCCCAACACC
ZM12-8333	12	8,333,466	3	138	55	CTATACATTTCCTATGGCTT	CCCTTTGAAACATTGGAT
ZM12-12525	12	12,525,459	3	132	55	AATCAAGATTGGCAAAGAGCGAAA	GTTTCTACCATTGGAACAATCCCC
ZM12-15590	12	15,590,317	4	126	55	CTCGTCTTAGTAATTTGGCTT	TTTCAGCTCAGGCATAGAAGT
ZM12-16215	12	16,215,296	3	108	55	CCCTTATAACCGCCAAGA	GCTGGATCAACTAAATGAACTGA
ZM12-18223	12	18,223,407	3	107	55	TGTTTAGTCACGCATACACA	TATATGGGAGGAAAACATCTCT
ZM12-18741	12	18,741,745	3	88	55	TACTAATATACTAGCAACCCT	TTACCTGACATTATGATCGAG
ZM12-19010	12	19,010,585	3	108	55	TGCTAGCTGCTGATAGAAGT	GTGTAGCTTCAGCAATTCAGA
ZM12-19305	12	19,305,743	3	79	55	CTAATATATTCTGGCAGTCCT	TTGCAGAGAAAACAAGTTCC
ZM12-19384	12	19,384,452	5	122	55	TTCAGCTTGCTTCGATTGCTC	TTACCCGCTTTTAATGCCTT
ZM12-19410	12	19,410,822	4	144	55	TGATAATCACACGCTGTT	TGCCTATCTTGAAACCCA
ZM12-19577	12	19,577,585	3	111	55	AAGTAGAACAAATGGCAC	GCCTGAAACTCTATCTTGTTG
ZM12-22166	12	22,166,689	3	123	55	ACTAGATCCATACCCAACCTGC	TCTCCCTCAACCTGTACCCCT
ZM12-22296	12	22,296,479	4	145	55	AAGAAACTGTGTAATTGAGCC	GCCCAACTATGATGAACT
ZM12-22862	12	22,862,143	4	145	55	GTTGATCCGAGTTAAAATACCAC	GCTAGACTGCTCCTAATTCAGA

Chr = chromosome; Position = the physical map position according to the reference 9311 genome; InDel = InDel polymorphism size (bp) between ZS97 and MY46; Size = PCR product size (bp) for ZS97 or MY46 based on the re-sequencing data; Tm = optimized annealing temperature (°C).

DISCUSSION

Genetic diversity within parental lines of hybrid rice is the foundation of heterosis utilization in hybrid rice breeding. In the present study, genome analysis based on re-sequencing of 2 parental lines, ZS97 and MY46, of an elite 3-line *indica* hybrid was used for comprehensive identification of SNP, InDel, and SV variations. This information provides valuable clues for understanding heterosis and improving hybrid rice breeding.

Compared to the 9311 reference sequence, a total of 364,488 SNPs, 61,181 InDels, and 6298 SVs were detected in ZS97; 364,179 SNPs, 61,984 InDels, and 6408 SVs were detected in MY46. Additionally, 77,013 identical SNPs and 181,737 different SNPs as well as 15,021 identical InDels and 1205 different InDels were observed between ZS97 and MY46. The total number of different SNPs and InDels estimated between ZS97 and MY46 was much lower compared to that reported for more diverse landrace populations, as well as between *indica* and *japonica* cultivars (Shen et al., 2004; Huang et al., 2010; Subbaiyan et al., 2012). This may be because of the absence of SNPs/InDels from 1 sample, resulting in lower re-sequencing depth and coverage, as well as the strict filter requirement (at least 3 supporting reads for every variation in both 2 samples) in comparative analysis and the closed relationship of the 2 samples. However, the total genetic variation detected between ZS97 (B line) and MY46 (R line) was higher compared to the R lines of cytoplasmic male sterility and among *japonica* (Yamamoto et al., 2010; Arai-Kichise et al., 2011; Li et al., 2012), which is consistent with the expected genetic divergence between the B and R lines of three-line hybrid and was higher compared to R lines. These variations improve the understanding of the complicated genetic basis of the 2 heterotic pools of the three-line hybrid rice in China.

Simple sequence repeat and SNP systems are the 2 most widely used markers in genetic research and MAS breeding of rice. Because of the limited number and non-functional variation of simple sequence repeat and strict platform requirement for SNP detection, InDels have become a valuable resource for genetic research (Shen et al., 2004; Subbaiyan et al., 2012). InDels with moderate size differences are preferable for their ease of genotyping and have been used widely in genes/QTL mapping (Zhang et al., 2012) and rice MAS breeding (Hayashi et al., 2006). There are 2 subspecies of cultivated rice, *indica* and *japonica*, and most early development of InDel markers was based on sequence comparison between the 2 subspecies with distant genetic divergence (Shen et al., 2004; Wang et al., 2005); most of these InDel markers cannot reflect the genetic divergence within subspecies. Recent advance in next-generation of re-sequencing technology have made it possible to identify InDel markers within the same subspecies and even accessions with close genetic relationships. Both increased and saturated availability of DNA markers between *indica* cultivars are vitally important for efficient 3-line *indica* hybrid breeding and genetic research based on genetic populations derived from *indica*/*indica* crosses. In our study, of 1205 different short InDels between ZS97 and MY46, 180 were 3-8 bp in length, which was easily detected by polyacrylamide gel electrophoresis, and may be the preferred method for developing PCR-based markers. We converted these InDels into PCR-based markers; 160 of the tested InDels produced amplified fragments and showed polymorphisms between ZS97 and MY46, which can be used as a new source of markers for genetic studies based on *indica*/*indica*-derived populations and molecular improvement of three-line hybrid rice breeding.

ACKNOWLEDGMENTS

Research supported by the China National High-Tech Research and Development Program (#2012AA101102), the Chinese High-Yielding Transgenic Program (#2013ZX08001-004), and the Research on Key Teamwork Construction in Technology Innovations of Zhejiang Province, China (#2010R50024).

Supplementary material

REFERENCES

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.

Chen X, Temnykh S, Xu Y, Cho YG, et al. (1997). Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor*. *Appl*. *Genet*. 95: 553-567.

Fan C, Xing Y, Mao H, Lu T, et al. (2006). *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor*. *Appl*. *Genet*. 112: 1164-1171.

Hayashi K, Yoshida H and Ashikawa I (2006). Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. *Theor. Appl. Genet*. 113: 251-260.

Huang X, Wei X, Sang T, Zhao Q, et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42: 961-967.

Jones N, Ougham H, Thomas H and Pasakinskiene I (2009). Markers and mapping revisited: finding your gene. *New Phytol*. 183: 935-966.

Li J, Xin Y and Yuan L (2009a). Hybrid rice technology development: Ensuring China's food security. International Food Policy Research Institute (IFPRI), Washington, DC.

Li R, Li Y, Kristiansen K and Wang J (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics* 24: 713-714.

Li R, Li Y, Fang X, Yang H, et al. (2009b). SNP detection for massively parallel whole-genome resequencing. *Genome Res*. 19: 1124-1132.

Li S, Wang S, Deng Q, Zheng A, et al. (2012). Identification of genome-wide variations among three elite restorer lines for hybrid-rice. *PloS One* 7: e30952.

Li S, Li R, Li H, Lu J, et al. (2013). SOAPindel: Efficient identification of indels from short paired reads. *Genome Res*. 23: 195-200.

Shen Y, Jiang H, Jin J, Zhang Z, et al. (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol.* 135: 1198-1205.

Subbaiyan GK, Waters DL, Katiyar SK, Sadananda AR, et al. (2012). Genome-wide DNA polymorphisms in elite *indica* rice inbreds discovered by whole-genome sequencing. *Plant Biotechnol. J*. 10: 623-634.

Wan JM (2010). Chinese rice genetics and breeding and breed pedigree: 1986-2005. China Agricultural Press, Beijing (in Chinese).

Wang X, Zhao X, Zhu J and Wu W (2005). Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). *DNA Res*. 12: 417-427.

Xu XZ, Chen KR, Tong HJ and Cai QH (1989). The development and utilization of restorer line Milyang 46. *Hybrid Rice* 4: 22-24 (in Chinese).

Xue W, Xing Y, Weng X, Zhao Y, et al. (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet*. 40: 761-767.

Yamamoto T, Nagasaki H, Yonemaru J, Ebana K, et al. (2010). Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics* 11: 267.

Zhang Z, Wang K, Guo L, Zhu YJ, et al. (2012). Pleiotropism of the photoperiod-insensitive allele of *Hd1* on heading date, plant height and yield traits in rice. *PloS One* 7: e52538.

Zheng KL, Huang N, Bennett J and Khush GS (1995). PCR-based marker-assisted selection in rice breeding. IRRI Discussion Paper Series No. 12. International Rice Research Institute, Los Baños.