

# **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  $\leq$ 5 bp, and 5.6% were  $\geq$ 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, markerassisted 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; Mc-Couch 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 vet 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-

Genetics and Molecular Research 12 (4): 5226-5235 (2013)

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 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).

Genetics and Molecular Research 12 (4): 5226-5235 (2013)

## 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- $\mu$ L reaction volume containing 50-100 ng template DNA, 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M dNTP, 0.2  $\mu$ 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 Jiangshu 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  $F_2$  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 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).

Genetics and Molecular Research 12 (4): 5226-5235 (2013)

Y.X. Zeng et al.

Chr	0.1	Chro.z	0110.3	Children	Critto.b		CHIO.0	Child.7	Chiro.8	Crito.9	Child. 10	CAPO.11	Chiro.12
	0000	-	10	10-0-Date	10D	. 100	The suffrage DERIS	6-0-09	14-dh-2001	10-0-081	m-dh-2001	285-025111	3-1-0.30
216 -	DIGIA	112 - CO204	118-0-03034		100	201A				5		201-1-110	260130
- 60 F	D1013	163 - 10232 029 - 12233A	557 00018	575 DH02	527 - D	1602 6	29	53			50	40KELO2	12-0-20
1071 >	- D9010	722 0285	9007	1067		10	HE - CONT	KH3	1122-0-2000	118	97-0-2800	131	1961 01204
1005	D102	1190 - CEM		1105-0-04008	1099 - 0	AGEN						1790	22
1014 -	- D1028	1566 TT 1220A	15.08	1092 D404	Mill D	2504 15	51D664	114-0-014	16.8-00-2024	1925	1406-19804	200	100.000.0038
2122	01020	2110 0220	2016 - 00134	2022 DH05	1543	20	00-000	2214	210-106				
2227 1	5: D103	2572 0296A	2518-00008		2074 0	2605 24	Th D606	2015 CTHA	200-1006		2009-0-21000	2008 6 F GH	2541-01388
2206	DICIA	2772	3031 - 0304	2621 D405		2	ED COLLA			260 0101	Taxana Land		00-400
200 2	1 DIDIC	30.00	Anter Contes	3212 D102	2564 0	2605	17 - CEM	2008	2214 2000				
3265 /	1 D1040	2054	3056	3757 D108	2017 - 0	2667		3664 C.T.M.			2012-0-2 007		201-0-0120
3672	prict	3960	4157 0005		2009 - 10	20 m	ao beev	#2F1 E/28	36/62020	363-087			SUS
4174	11 01050	4440		4158 D400A	200 MN2 - < 0	2000 40 2006A 40	92-0690	ATK	#1232020		200-00-0100	4011 6110	
4347	E prosc	4020 - 01294	5811	4364		40	M 0015A	\$2><:File	201	and the second	474-2300	4540	100 0125
4872 8	2 Dress	5250 - 0229C	\$207-11-0906A	5001 D1000		40	03-1-0611	4712		458 090			5:3 0:36
6363	DICEA	5730	5745 03090	MAR - DITA	4600 - D	2610 60	21 00124	4005	6649-11-2012	408-000	400		5454
5472	E proso	00120 021100	6101-CO007A	6044 D112	4115 0	5515 55	0613	6413-0718	5940-1-2012	ALC: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	100-100		00 000
1822 2	2 D1060	6477 0211	6563	6239 - Di Di Di		79	40-0014		6×03				
6154 8	D'ICE	200 02110		0000	5606 - D	5612		F196	9529-2018	eeg	527-1-1492	761	60%D-ZB
6235	E DIGA	7997 - 021193	1008	6063 DH 130		64	64	验本经验	795			792	6920(2)
6640	DriceC	100	6452	100 0000	6200-0	810	10 - T - T - T - T - T - T - T - T - T -	A 10 10 10 10 10 10 10 10 10 10 10 10 10	100.000	64X 0913	6336	036	742905222
7809	D109	0150	8238 0000A	2909 D112E	6595 D	2514		ma	111-11-10-1	005-0014	5550-1107 K	100 - 1100	7989
7255	D1090	1619	8792 0909C			M	06-0617		8320	750-005	7944		BIGEDELINE
7856	DICOC	0091	S2000040	#755 DH13F DH95	1001	80	00	954-0-075	0602		100 100 1	Her Lines	
0073	D113	0081	6001 - T- 0016A	9345 - D114	7644	15 ACECC 2610	50D015A	9681 0730	300	1235 0815	1027-10-21072	998 - E # 02	9064 0428
610	D1110	9576 - 0214A	10000 - 00000	9394 - DI16A	NS +< 0	2154 86	04D610	1012	9411-11-2020		112-1-1-1-1	10431943	
6855	D110C	10010 × 0215A	10049 - 00154	9951 - D115 9992 - D115A	8242	917 51	24 0620	VIE 2	X681-2002	893	347-1371	10113	
9100	D111A	12500	10755 00110C		min - D	2010 00	44 0671		Xx80	40% CP18		1939 EFCS	808 - 218
9420	DING		19082	10040 - LL - 04150 10050 0150	8090 0	90 ACH 50	EI DEUSA	19539 - CT23A	10000-0000		008 2 500	10001-000	600- 0.00
2743	D1110	11158	11202 00126	10943 D115A	\$901 D	25.12 900	EX	910		104013	1002-01021		0.00
90807	D112		19958	1.204 Carl	M28-0	2234		000-11-004	7.68	N N 0520	100		1.04
90376	D1123	12502 -0217	10110 0013A	1.791 Dit 19		107	C230 - 19	5985-F778	198-308				1175 D-230
11136	D112C	12560 - 0213A	00100	12200 CO19A	1000		97 0024		1080-1107	1000	1000-1000		
11255	DIIIA	12767	12458-0224	12745-0121	16422 0	3624	10.0005	1940-0721	1007		107		000-002
11622	01130	13036	1000 W144	13856D(22		117	AGNI IN	1007-0-078	0.01 203	and There			
11879	01130 01130			MTI DOM	11600 - D	100	05-0400		1055-300	0.224	102-1021		-m Tone
12053	D114		14210 00145	13/45 0429	19671 - 0	125	AND - DOMA	uses cos	1000 - 1000	17354	1.20-1-1903		Grie
1111	D1148	11000	14739 - ODIS	14135 0124	titeld - D	1220 128	03 D637	10703 6730			100-100	THEAT OF BEICH	1000 07.74
12209	D115	14762	A100	1620-0405		512	56-0625	1431 C/21	9431	1007-010	1000-1001	-686 E4 C8	WXKDCIR
12631	D1158	15/87	15454 05450			07	20 - 002	1000	818-1-22	10674	1000-1000	454-0-6100	2017年187月
14201	01150	11717	10171 - 0014	15410 CH25A	12653 0	924 10	44 - 2000	16425			1903-0-1400		VOX D/28
14334	D1156	19390	outer being			141	61 - D632A	100-074	1007-00-203	0.4		12/38	Kal20128
14730	DIIBA	35430 0221A	100006	16145 DE27	12547 - 0	2825A 540	DO DETA		4212-1126	10646-0425	1100	1954CET102	102-0120
10257	01160	12208	17167 0847A	16564 0428	13745	25.21		175.H C/H	1074	1400 0101	0823-0900	12981 67100	NOR
1429	10117 D117A		1827 - 03670	17079-0129				NON CTRA				100124-0010	16564 D1242
10100	01178	17940 012200	10087 0016A		14213-0	2627		10681 - 1752	128	1404-0000	3014-0-2100	1084-0118	803-0-00
963732	0118	17990	100102 - CO1000	17667-D130	161/2-0	ASTR			1713		14/14	1004	TAN- DOM
10918	DIMA	1001	10075 0019	18201 0431	14995 D	9625 963	01 DE32	Sec. Los	10.05 - 10.00	10002 11 1000		196296103	17928 01245
19054	DINC	19070 - 05738	15416 0020	18042 DIGVA	1000	967	22 0633	2006		1003 0102	900	20141-0-01140	8.95 - 9.26
19405	D113	19214 0224 19482 - 0224A	19654 00326A	19140 DCD	15625	920A 171	ACC 00	2000 010	0.54		74644	2550	10.14 DO.16
10780	DIM	19500	20175 00209	19191 - D433A 19542 - D434	10000	112	-0034	2002 >	0.0	x0072 0189	1000 1000	2997	
20372	DINC	19030	20680	19848 DIDIA	Have 1 to	176	04 0435	2020	19673042	NRT 0014		240-000	000
20447	DIDBA	29/39	20052 0029A	20020-0435	H535 D	9631 180	H1-DE36		2008		8539		190401389
21156	01293	20049	21124 00210	20483 D430	17021 - 0	2635A	10.00	2246-1-0-00	20023	17902	1200-1202	Dist.	關本設置
21643	DIGIA	21430 - 0225A	21552 - 0023	20542 CA36A 20542 CA36A	17312	35.23 589	0030	22-00-07465	2002-104	17421-0106		2593 21108	1.8- 0.00
22118	DIDZA	21805 - 212200	21717 00200	21384 DE32A	12015	100	48-7 - DE38A	22601-0740			555-1-1 mil	22906	2014 - 01.04
22554	0028	31932 - 02230	22110 00200	2003-000	17054 0	75.25		2997	2100 100	NER DDF	1980	204	20er 0128
23852	01020	22NDF CO22A	225 18 C00200	22290D+40	18214 0	200 100	0 > < 000	22444 6747	2785-1-208	NGR	100-100	2005-01108	2321 井 22週
20574	0102	22962 02230	2211.01	20957 - 01454	18732			289-010	2005			2010 11100	2181-1028
25133	PERA	20064 0228	23549 03245	22801 D1408					7007	104 1018	1002	max- erts	2/2 0/28
24571	0038	20057 78005	24006	2002 - DHUA	19208	2000 2000 2000	01-0641	2001	2009	1942-0040	1008	22-412	경찰:바람했
21012	DIGA	24240 - 02200	24405	2540	HICKS - D	9539 114	200	2001-0-0-0	203-110				2061-01302
21835	Drus	20716 - 2220A	25039	2545 0445	18551-1-0	AUCH		25430	301	10514-0041		26215	28/20/20
2010	DIDISA DIDISA	1024802290	26457 00250	24450 C146A	28177-0	210	04	200-052		X97 0643	2010-120-1	2400-0105	2022-0124
26778	Drissi Drize	20290 - 0230	20010 - 172764	24921 D145	20058 - 0	2541 225	11 DOM	2644-0-020	2408-11-200	2003-0040	1011	2002-0112	2000
27728	DICER	ACC10	242104	25304 D148	10011- 10	1000			30.022005				
27818	D127	20010 002308	20042 - 00227A			200	07-0645	2005-0-0-54	28581	2142	250-101		202 01Ni
20754	01275	20/07	27296 - 00270	20220 0448	21514	240 254	0646	- nd	200 - 200	1000	200 200		M2 - 0.30
25246	DIDBA	1000	27840	10712 Dial	10010	-	000 1000 0007*	27905	340			100	2007-0-248
30200	DI29A	117147	00000 00000	and an and a second	00111-0	240	004	2862 - 0787		2043 0540	2003-0980	20107	
X 590	01298	279.70 - 02 339	26459 0025	27277-0150	225.62 0	2545 244	08 - 064IA	10011-0-572	200-100	2003	2540-1961	2028 - 1 6103	A43
38242	DIJDA	26075	28804	27362 D451	22900 0	1545 240	(10 D0 A)	2965-0286				NON-CONF	2023年202
31892	0131	25505		20194- D152A		253	01 0050	1000 - 1000	27967	2003-0140	17-30-11-1482	X04	255 600
22114 21	D131A	20054 0234	and a second	arta futa	22430 0	2647 247	ANNO - NE	2008-0-0791	2010-11-2182	2010-0-040	2602-0-2602	现计计算机	<b>新新学校</b>
32800	D122	2007	29567	10125 Crist	22621 0	2543 256	08 0651						
31700	D1228	2000	20425-0031	20077 DISA	1000	202	90-0652						
30805 // 34254	1 D1334	30219 0235	20075 - 0025A	29/3 M	244730	240							
34736	D1330	XV20	31302 0030	NEWS DOM	24001-0	268	0000						
30055	DID4A	31217 - 022300	1054 003M	MITHE DATE		512	0004						
30505	01348	34345 - 0236		XETS-TO-DISCA	25411 0	210	- 1000						
X001 /	DISSA	32100	32360 0333	3/175-0454	295.62 0	#102 EP9	0050						
3634	DUN	32913	32059	2/2560200	100	260	HH D657						
37113	DIDEA	AND COLLA	11441	32228 D1080	20545-0	2553 100	000						
37854	DOD?	33334	23749	10132									
38381	00078	33557	34178 003340	20041	27053	98544							
3812	DISIBA	34000	34417 - 0035	20250-0-0-02		267	34 0660						
238	01383	34000	54616 - C 0036A 25015 - C 00356	30739-0-0900	29628-0	30	DE CENTA						
200	0129	10044	35430-10000C	NUTE Date	2010	202	-001						
40133	- D1358	35313 - 22102	26757 03365			300	05 - DEKIA						
40736	0143	A6420 BCCK	36073 03395	34681	20043-0	2007 114	0000						
41529	D1488	35855	3003	26121-0496	29641	2000 215 2000A 215	45						
41701	D141	NVSF JUST	74680	2600 DH67		-	00 - 0064						
42463	- D1410	X100 X 300A	27211-U-00376	264G-U-D908	20007 4,j D	ana							
42901	DHZA												
41334	DH20												
41818	- D543												
41508	05430												
44800	D943C												

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 <i>indica-japonica</i> 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.

Genetics and Molecular Research 12 (4): 5226-5235 (2013)

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  $\leq$ 5 bp in length, and 5.6% of the markers were  $\geq$ 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 *indicajaponica* 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 <u>Tables S1</u> to <u>S12</u>. 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).

Genetics and Molecular Research 12 (4): 5226-5235 (2013)



Y.X. Zeng et al.



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 markers. Therefore, we could not increase the polymorphic marker rate by developing markers with longer InDel sequences.

Genetics and Molecular Research 12 (4): 5226-5235 (2013)

**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

Genetics and Molecular Research 12 (4): 5226-5235 (2013)

Y.X.	Zeng	et	al.	
------	------	----	-----	--

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<sup>SG</sup>*, 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.

Genetics and Molecular Research 12 (4): 5226-5235 (2013)

- Shen YJ, Jiang H, Jin JP, Zhang ZB, et al. (2004). Development of genome-wide DNA polymorphism database for mapbased 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 *qSTV11<sup>TQ</sup>*, 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-thermosensitive 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.