Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 19 (2009) 1241–1252
Progress in Natural Science
www.elsevier.com/locate/pnsc

Efficient *indica* and *japonica* rice identification based on the InDel molecular method: Its implication in rice breeding and evolutionary research

Bao-Rong Lu *, Xingxing Cai, Xin Jin

Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, Fudan University, Shanghai 200433, China

Received 29 December 2008; accepted 4 January 2009

Abstract

An efficient molecular method for the accurate and efficient identification of *indica* and *japonica* rice was created based on the polymorphisms of insertion/deletion (InDel) DNA fragments obtained from the basic local alignment search tool (BLAST) to the entire genomic sequences of *indica* (93-11) and *japonica* rice (Nipponbare). The 45 InDel loci were validated experimentally by the polymerase chain reaction (PCR) and polyacrylamide gel electrophoresis (PAGE) in 44 typical *indica* and *japonica* rice varieties, including 93-11 and Nipponbare. A neutrality test of the data matrix generated from electrophoretic banding patterns of various InDel loci indicated that 34 InDel loci were strongly associated with the differentiation of *indica* and *japonica* rice. More extensive analyses involving cultivated rice varieties from 11 Asian countries, and 12 wild *Oryza* species with various origins confirmed that *indica* and *japonica* characteristics could accurately be determined *via* calculating the average frequency of *indica*- or *japonica*-specific alleles on different InDel loci across the rice genome. This method was named as the "InDel molecular index" that combines molecular and statistical methods in determining the *indica* and *japonica* characteristics of rice varieties. Compared with the traditional methods based essentially on morphology, the InDel molecular index provides a very accurate, rapid, simple, and efficient method for identifying *indica* and *japonica* rice. In addition, the InDel index can be used to determine *indica* or *japonica* characteristics of wild *Oryza* species, which largely extends the utility of this method. The InDel molecular index provides a new tool for the effective selection of appropriate *indica* or *japonica* rice germplasm in rice breeding. It also offers a novel model for the study of the origin, evolution, and genetic differentiation of *indica* and *japonica* rice adapted to various environmental changes.

Keywords: Cultivated rice; Wild *Oryza*; InDel; Molecular marker; *indica–japonica* identification; Neutrality test

1. Introduction

Asian cultivated rice (*Oryza sativa* L.) was domesticated approximately 7000–8000 years before the present (BP) in the middle to lower parts of the Yangtze River region [1]. Since the onset of its domestication and followed by its long-term cultivation and dissemination across the globe, Asian cultivated rice has experienced significant genetic differentiation, adapting to different ecological conditions under both natural and human selection. Genetic differentiation generated abundant genetic diversity in rice, such as *indica* and *japonica* ecotypes [2–4], lowland and upland ecotypes [5,6], and non-glutinous and glutinous quality types [7,8], which have provided tremendous rice varieties for human consumption all over the world. Rice varieties also serve as important genetic resources for their further improvement [9]. The most significant type of genetic differentiation in rice is *indica–japonica* differentiation. As

* Corresponding author. Tel./fax: +86 21 65643668.
E-mail address: brlu@fudan.edu.cn (B.-R. Lu).

1002-0071/$ - see front matter © 2009 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2009.01.011

a consequence, Asian cultivated rice evolved into two major ecotypes (also referred to as subspecies) that have already developed considerable reproductive isolation [2,10].

Usually, the *indica* rice ecotype is mainly found in tropical and subtropical rice planting regions, with either low latitudes or altitudes, whereas the *japonica* rice ecotype is mostly cultivated in temperate regions with high latitudes. In some rice planting regions of low latitude, *japonica* rice can also be cultivated in mountainous areas with high elevations, such as the Yunnan and Guizhou Provinces of China, Laos, Myanmar, and Vietnam, as well as many other Southeast Asian countries. Adapted to various ecological conditions, *indica* and *japonica* rice has considerable differences in morphology (e.g., plant height and pubescence), agronomical characteristics (e.g., length/width ratio and persistence of grains), and physiological–biochemical features (e.g., winter hardiness and starch types in grains). Because of such significant differences in many aspects of *indica* and *japonica* rice, some rice scientists treated these two ecotypes as two subspecies, i.e., *Oryza sativa* subsp. *indica* Kato and *O. sativa* subsp. *japonica* Kato [2,10].

The significant genetic differentiation, for example, in agronomic characteristics and physiological–biochemical features makes *indica* and *japonica* extremely unique and valuable for genetic improvement and research on rice from different localities. Usually, *indica* and *japonica* rice varieties are distantly related in terms of their genetic relationships, therefore, the inter-subspecies hybridization between *indica* and *japonica* rice will result in remarkable genetic recombination and variation, which offers greater opportunities for rice breeders in selecting ideal variation types for rice improvement. Importantly, the hybridization between *indica* and *japonica* rice varieties will generate exceedingly strong hybrid vigor (heterosis). The *indica–japonica* hybrid rice may have great potential in rice production, given that the problems of between-subspecies sterility can be resolved [11,12]. In addition, the origins of *indica* and *japonica* rice and the mechanisms of *indica–japonica* differentiations are very attractive issues that have received international attention for research [2]. A great amount of scientific research data, although with considerable discrepancy, on the origin of Asian cultivated rice has been generated in the past decades. Nevertheless, questions such as the concerned ancestral species of Asian cultivated rice, the pathway and origins of *indica* and *japonica* rice, and their evolutionary processes are still under debate and are of great interest to be determined [13–17]. The exploration of the procedure and mechanism of genetic differentiation between *indica* and *japonica* rice provides important information for better understanding of adaptive evolution in plant species under the changing environments. In addition, it helps the rice breeders to move away from the dilemma that limits the utilization of super heterosis from inter-subspecies hybridization between *indica* and *japonica* with strong sterility in inter-subspecies hybrids, which will offer a novel strategic method for resolving practical problems in rice breeding.

The key prerequisite of utilizing rice germplasm, either for the study of *indica–japonica* genetic differentiation during the domestication process or for genetic improvement of rice varieties, is to accurately identify the *indica* and *japonica* rice varieties used. The traditional method for identifying *indica* and *japonica* rice varieties relies essentially on variation in morphological characters (e.g., plant height, plant type, status of pubescence of plants, and type of grains) in combination with some physiological and biochemical features (e.g., winter hardiness and phenol response in rice grains) [10,18]. Currently, the most common method for identification of *indica* and *japonica* rice varieties is "Chen's index" that examines six key characters of rice samples: (1) lemma hairiness, (2) phenol response in rice grains, (3) inter-node length of panicle axes, (4) color of grain husks, (5) hairiness of leaf-blades, and (6) length/width ratio of grains [19,20]. However, these morphological and physiological characters can greatly be influenced by the change of environmental conditions. As a consequence, the final results from the same rice samples obtained under different environmental conditions may vary considerably, which affects the accuracy and consistency of *indica–japonica* identification among regions. In addition, it is necessary to measure the concerned characters from mature rice plants when this morphological-based *indica* and *japonica* identification method is used. Therefore, it is time consuming to generate the results. As a general rule, the measured samples (seeds) are required to be cultivated in the field, and it takes a minimum period of 3–4 months (from sowing seeds to harvesting mature plants) for the examined samples to be ready for the measurement. In addition, the morphological-based measurement requires a minimum sample size with sufficient number of rice plants, which will use a large number of rice materials for identification, causing unnecessary squander. The above-mentioned disadvantages of the traditional Chen's index method have considerably limited its wide application for *indica* or *japonica* rice identification.

The fast development of molecular technology encourages the exploration of different types of molecular markers, such as random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), and microsatellite (also referred to as simple sequence repeats, SSR), for identifying *indica* and *japonica* rice varieties efficiently [3,4,21–23]. However, these molecular markers frequently give quite diverse results when used in identifying *indica* and *japonica* rice varieties, even though these molecular markers can sometimes be useful to cluster the two subspecies into two separate groups. This is because these molecular markers are not developed specifically based on differences between the two subspecies of rice. In addition, the *indica–japonica* differentiation in rice is a continuous process, from which rice varieties differentiate not only into the typical *indica* and *japonica*, but also into different types either close to *indica* or close to *japonica*, as well as into

intermediate types without considerable *indica–japonica* differentiation. It is extremely difficult to accurately identify these different types with various degrees of *indica–japonica* differentiation. For that reason, the development of a useful technique that can efficiently, quickly, and accurately identify *indica* and *japonica* rice varieties is urgently necessary.

The completion between DNA sequences of the total rice genomes and the rapid development of comparative genomics in rice provide tremendous biological information for evolutionary studies and genetic breeding of cultivated rice. Using the basic local alignment search tool (BLAST) to thoroughly compare the total genomic DNA sequences of *indica* variety 93-11 with those of *japonica* variety Nipponbare, Shen et al. [24] identified a great number of polymorphic DNA fragments and single nucleotide polymorphisms (SNP) that are different between the two varieties. The acquirement of these polymorphic DNA fragments or loci provided a novel thought and tool for studying *indica–japonica* differentiation of cultivated rice (even for closely related wild *Oryza* species) and also offered an important methodology for establishing efficient identification of *indica* and *japonica* rice varieties. Usually, differences in the insertion/deletion (InDel) DNA fragments between 93-11 and Nipponbare are about 25–75 base pairs (bp). So, it is very easy to detect the InDel fragments by a designed experiment using molecular methods. Hypothetically, it is highly feasible and valuable to develop such molecular markers for identification of *indica* and *japonica* rice varieties. Given the fact that the detected differences in DNA fragments (InDel) between 93-11 and Nipponbare would not necessarily represent the differences between the two subspecies, rather than differences between the two varieties or even between the two individuals used for the total genomic sequencing, it is necessary to validate the InDel loci for their effectiveness in *indica* and *japonica* identification. The validation needs to combine methods of experimental biology and population genetics, involving an experimental population comprising known *indica* and *japonica* rice varieties that are identified by the traditional Chen's index method. Only these InDel loci that truly reflect the genetic variation and differentiation of *indica* and *japonica* subspecies will be used for the identification. Furthermore, it is also important to develop a standard analytical protocol based on the InDel methodology for the accurate and effective identification of *indica* and *japonica* characteristics of rice. The objectives of this study were to (1) validate the InDel loci that should truly reflect *indica–japonica* differentiation in cultivated rice based on the known *indica* and *japonica* samples as an experimental population identified by Chen's index; (2) develop a standard experimental procedure and calculation method for efficient identification of *indica* or *japonica* characteristics of rice varieties; (3) select the validated InDel molecular markers to determine *indica* or *japonica* characteristics of rice varieties and wild *Oryza* species from a wide source and study *indica–japonica* genetic differentiation. This method provides a novel concept for studying the evolution of cultivated rice, particularly for its *indica–japonica* differentiation, and a new technical tool for rice genetic breeding.

2. Materials and methods

2.1. Plant materials

Four types of plant materials were included in this study: (1) the *indica* 93-11 and *japonica* Nipponbare that were used for the total genomic DNA sequencing (Table 1); (2) 21 typical *indica* and 21 typical *japonica* rice varieties from China that were identified by the traditional Chen's index (Table 1); (3) randomly selected 42 rice varieties from China and 11 countries in East, South, and Southeast Asia, for which the *indica* or *japonica* characteristics were unknown (Table 2); (4) 12 wild *Oryza* species (36 populations) containing different genomes (AA, BB, CC, EE, CCDD, and GG) (Table 3).

2.2. DNA extraction

DNA was extracted from seedlings of rice samples from a single individual, because our previous experiment indicated that there is no considerable variation among different individuals of the same varieties based on the insertion/deletion (InDel) molecular method. Seeds were germinated in an incubator at 37 °C and then transferred into a glasshouse at around 25 °C. About 0.5 g of fresh leaf samples were collected from seedlings at about the 3–4-leaf stage (sometimes, 0.2 g of dried leaf samples were also collected in the case of no fresh seedlings). DNA samples were extracted following the CTAB method [25], and the procedure of Song et al. [26].

2.3. PCR and electrophoretic analysis

From more than 50 InDel primer pairs obtained by Shen et al. [24] based on the BLAST between the total genomic DNA sequences of *indica* (93-11) and *japonica* (Nipponbare) rice varieties, 45 that consistently produced clear electrophoresis bands on polyacrylamide gel (PAGE) after amplification by the polymerase chain reaction (PCR) were selected for inclusion as candidate primers to identify *indica* and *japonica* rice varieties (Table 4). DNA samples from all the cultivated rice varieties and wild *Oryza* species (populations) were included for PCR analysis and electrophoresis using the 45 InDel primer pairs. The PCR reactions were performed in a PTC-200 thermocycler (Bio-Rad Laboratories, Inc.). A denaturation period of 4 min at 94 °C was followed by 36 cycles of 4 min at 94 °C, 30 s at 55 °C and 40 s at 72 °C, and then 10 min at 72 °C for final extension. Reactions were carried out in a 20-μl volume containing 2 mM buffer (including $MgCl_2$), 1.6 mM of dNTP, 10 mM of InDel primers (Shanghai Sangon Biological

Table 1
Origins of the typical *indica* and *japonica* rice varieties included for validating the efficiency of the insertion/deletion (InDel) molecular markers for *indica* and *japonica* rice identification.

Variety	Origin	Chen's ID[a]	InDel's ID[b]	F_i^c	Variety	Origin	Chen's ID[a]	InDel's ID[b]	F_i^c
93-11	China	*indica*	Typical *indica*	1.00	Nipponbare	Japan	*japonica*	Typical *japonica*	1.00
Yongnialushuibai	Hebei, China	*indica*	Typical *indica*	1.00	Fushun Xiaobairen	Hebei, China	*japonica*	Typical *japonica*	1.00
Gailiangwu	Zhejiang, China	*indica*	Typical *indica*	0.94	Weinan Hongmangdao	Hebei, China	*japonica*	Typical *japonica*	1.00
Sanbailituo	Zhejiang, China	*indica*	Typical *indica*	0.96	Zhululaozu	Hebei, China	*japonica*	Typical *japonica*	1.00
Bangqiunuo	Anhui, China	*indica*	Typical *indica*	0.91	Miziwan	Shanghai, China	*japonica*	Typical *japonica*	1.00
Qiuqianbai	Anhui, China	*indica*	Typical *indica*	1.00	Baomanuo	Shanghai, China	*japonica*	Typical *japonica*	1.00
Hubeizao	Jiangxi, China	*indica*	Typical *indica*	0.94	Jiangyinzao	Jiangsu, China	*japonica*	Typical *japonica*	0.96
Shangraozao	Jiangxi, China	*indica*	*indica*	0.85	Manyedao	Jiangsu, China	*japonica*	Typical *japonica*	0.91
Qingshuidao	Shandong, China	*indica*	Typical *indica*	0.93	Xueliqing	Zhejiang, China	*japonica*	Typical *japonica*	1.00
Chiminei	Guangdong, China	*indica*	Typical *indica*	1.00	Zihong	Zhejiang, China	*japonica*	Typical *japonica*	0.91
Lucaihao	Guangdong, China	*indica*	Typical *indica*	1.00	Tianzuozao	Anhui, China	*japonica*	Typical *japonica*	0.94
Guangluai-4	Zhejiang, China	*indica*	Typical *indica*	1.00	Jianyangzao	Jiangxi, China	*japonica*	Typical *japonica*	1.00
Guidongxian	Hunan, China	*indica*	Typical *indica*	0.94	Cuanzhanchi	Guangdong, China	*japonica*	Typical *japonica*	0.94
Xiaogu	Sichuan, China	*indica*	Typical *indica*	0.94	Gaoliangzao	Hunan, China	*japonica*	Typical *japonica*	1.00
Xiangu	Sichuan, China	*indica*	*indica*	0.79	Zhushituo	Sichuan, China	*japonica*	Typical *japonica*	1.00
Baikeshangu	Guizhou, China	*indica*	*indica*	0.82	Guiyangbang	Sichuan, China	*japonica*	Typical *japonica*	0.91
Yanzan	Guizhou, China	*indica*	Typical *indica*	0.97	Hongzangu	Guizhou, China	*japonica*	Typical *japonica*	0.94
Shiyangnuo	Guizhou, China	*indica*	*indica*	0.75	Majiugu	Shaanxi, China	*japonica*	Typical *japonica*	1.00
BG951	Sri Lanka	*indica*	Typical *indica*	0.94	Madaozi	Shaanxi, China	*japonica*	Typical *japonica*	0.94
IAC1300	Brazil	*indica*	Typical *indica*	0.91	Sichenggou Jiangmidao	Shaanxi, China	*japonica*	Typical *japonica*	0.97
IRAT10	Cote d'ivoire	*indica*	Typical *indica*	0.96	ITA408	Nigeria	*japonica*	Typical *japonica*	0.97
Tutounuodao	Jiangsu, China	*indica*	*indica*	0.82	ITA182	Nigeria	*japonica*	Typical *japonica*	0.97

[a] Classification of *indica* or *japonica* as identified by Chen's index.
[b] Classification of *indica* or *japonica* as indicated by the InDel molecular index.
[c] Frequency of the *indica*-specific alleles.

Engineering Technology & Services Co., Ltd.), 20 ng of genomic DNA, and 0.6 U of *Taq* polymerase (TaKaRa Inc.). PCR products were resolved on a 4% denaturing polyacrylamide gel. After electrophoresis, bands were revealed using the silver-staining procedure described by Song et al. [26]. Electrophoretic banding patterns were recorded by photographs taken by a Sony digital camera. The null alleles were confirmed after several repetitions with different amplification conditions to ensure that no reaction failure existed.

2.4. Genotype score and allele frequency calculation

The InDel molecular markers were co-dominant; therefore, the banding patterns were scored as homozygous *indica* genotype (II), homozygous *japonica* genotype (JJ), or heterozygous *indica–japonica* genotype (IJ). The electrophoretic banding patterns of the sequenced *indica* rice variety (93-11) and *japonica* rice variety (Nipponbare) were used as standard references for determining the *indica* or *japonica* genotype, respectively, at a particular InDel locus.

Table 2
Origins of the randomly selected rice varieties worldwide for *indica* and *japonica* identifications using the insertion/deletion (InDel) molecular markers.

Variety	Origin	InDel's ID[a]	F_i[b]	Variety	Origin	InDel's ID[a]	F_i[b]
Yinqiunuo	Anhui, China	Typical *japonica*	0.09	Rodjolele	Indonesia	Close to *japonica*	0.26
Baimi Guanzhan	Jiangxi, China	*japonica*	0.21	Gundil	Indonesia	Typical *indica*	0.97
Nengshuibai	Jiangxi, China	Intermediate type	0.46	Chiyoda Wase	Japan	Close to *japonica*	0.29
Rinong-59	Shangdong, China	Typical *japonica*	0.03	Akihikari	Japan	Typical *japonica*	0.03
Zhangunuo	Hunan, China	Close to *indica*	0.74	Jangyeongdo	S Korea	Close to *japonica*	0.35
Guidongxian	Hunan, China	Typical *indica*	0.90	Bo Ri Byeo	S Korea	Typical *japonica*	0.04
Zagu	Sichuan, China	*indica*	0.76	Deng Kay	Lao PDR	*japonica*	0.24
Guiyangbang	Sichuan, China	Typical *japonica*	0.09	Leuang Kham	Lao PDR	*indica*	0.85
Duanmang Huangkenuo	Guizhou, China	*japonica*	0.12	Mahsuri	Malaysia	Typical *indica*	0.91
04Fs448	Shaangxi, China	Close to *japonica*	0.26	Padi Bubulangking	Malaysia	Close to *japonica*	0.31
Laomangdao	Shaangxi, China	*japonica*	0.24	Oolongman	Myanmar	Close to *japonica*	0.35
Muojiangdao	Yunnan, China	Typical *indica*	0.97	Yorkkama Kyaway Kyay	Myanmar	*japonica*	0.21
Taibai-8	Yunnan, China	*japonica*	0.15	Balasang	Philippines	Typical *indica*	0.91
Bailengshui	Yunnan, China	Close to *japonica*	0.29	Karabkab	Philippines	*japonica*	0.24
Heilengshui	Yunnan, China	Close to *japonica*	0.38	Puang Salawm	Thailand	Typical *indica*	0.94
Zhezhen-1	Zhejiang, China	Typical *indica*	0.97	TD57	Thailand	*japonica*	0.19
Zhenongda-104	Zhejiang, China	Typical *japonica*	0.06	Jhona20	India	Close to *indica*	0.68
58816	Bangladesh	*indica*	0.87	Bara Daddy	India	Typical *indica*	0.94
25901	Bangladesh	*indica*	0.76	Dhanya	India	Intermediate type	0.53
64893	Bhutan	*japonica*	0.22	Amteboro	India	Typical *indica*	0.96
72527	Bhutan	Intermediate type	0.51	Dahrdun Basmati	India	*indica*	0.88

[a] Classification of *indica* or *japonica* as indicated by the InDel molecular index.
[b] Frequency of the *indica* alleles.

Table 3
Wild rice species (populations) with different origins included for *indica* and *japonica* identifications using the insertion/deletion (InDel) molecular markers.

Species	IRGB No.[a]	Origin	F_i[b]	Species	IRGB No.[a]	Origin	F_i[b]
Oryza rufipogon	101974	India	0.50	*O. meridionalis*	82042	Australia	0.54
	106081	India	0.50		105303	Australia	0.65
	104643	Thailand	0.50	*O. glumaepatula*	100968	Brazil	0.50
	105568	Philippines	0.53		105668	Brazil	0.49
	105400	China	0.46		105561	Colombia	0.44
	105887	Bangladesh	0.51	*O. longistaminata*	103890	Senegal	0.52
	105698	Nepal	0.47		103905	Tanzania	0.48
	106145	Lao PDR	0.47		105076	Tanzania	0.47
	106502	Papua New Guinea	0.50		104977	Kenya	0.48
	106264	Papua New Guinea	0.56	*O. glaberrima*	102236	Liberia	0.52
O. nivara	104689	India	0.54	*O. barthii*	104132	Cameroon	0.53
	106348	Myanmar	0.51	*O. officinalis*	Yun-3[c]	Yunnan, China	0.56
	105386	Thailand	0.55		Yun-4[c]	Yunnan, China	0.57
	105433	Sri Lanka	0.52	*O. australiensis*	101467	Australia	0.43
	103821	China	0.53	*O. alta*	100161	Brazil	0.55
	105885	Bangladesh	0.46	*O. punctata*	104154	Cameroon	0.56
	105704	Nepal	0.45	*O. granulata*	Yun-1[c]	Yunnan, China	0.49
	105706	Nepal	0.60		Yun-2[c]	Yunnan, China	0.53

[a] Accession No. of the International Rice Gene Bank of the International Rice Research Institute (IRRI).
[b] Frequency of the *indica* alleles.
[c] Accession No. of Fudan University.

Based on the reference, the genotype of an examined cultivar or wild *Oryza* sample was determined at a given locus (either as II, JJ, or IJ). If the banding pattern (position of bands) of the examined rice sample was identical to that of 93-11, it was determined as homozygous *indica* genotype (II) at the given InDel locus (Fig. 1, lanes I and 1–11). If the banding pattern of the examined sample was identical to that of Nipponbare, it was determined as homozygous *japonica* genotype (JJ) at the given InDel locus (Fig. 1, lanes J, 12, and 14–17). If the examined sample showed both bands that were identical to 93-11 and Nipponbare at a particular InDel locus, this locus was determined as heterozygous *indica–japonica* genotype (IJ) (Fig. 1, lane 13). The *indica* or *japonica* characteristics of a rice sample were determined based on the average allelic frequency (F) calculated from genotype data (II, JJ, and IJ) of the total number (N) of examined InDel loci. The calculation for the *indica* or *japonica* allelic frequency (F_i or F_j) of a particular rice

Table 4
Forty-five insertion/deletion (InDel) molecular markers (based on Shen et al. [24]) used for studying *indica* and *japonica* differentiations and identifying *indica* and *japonica* characteristics in rice varieties.

Locus	DNA sequence of primer (5′–3′)	DNA sequence of primer (5′–3′)	Difference in amplified DNA fragments between 93-11 and Nipponbare (bp)
R1M7	ATTCCTGGTTCTACATTACTTA	CGCCTCACTAGAATATCGGA	37
R1M20	TTGGAACAGGGAAGAAGC	AGGACATAGTTGTAATGGGTAG	42
R1M30	AAGGGGCCCTAATTTATCTAG	TGTTTACTTTGTTCTTGGACTG	49
R1M37	ATAGTTCGCCATCGTCAT	ACACGCCATAGCAAGGAA	53
R1M47	AATAGAATTACTGATGAAACCTTA	GCCCGTTACCGCTTATGT	51
R2M10	CCCAGTCTGCTGCCATCT	GAATGTATTTCAGTTCCAGTAAG	48
R2M24	GGGCAACAACGGCTCTG	AGGGAATAAGGCGATACGG	31
R2M26	GCAGCAAAGTGCGGAGTA	CAGGTGAATTGCCAATTT	38
R2M37	ACTGTTACCCAAACGCTA	ACGTGCACCTACTACAGAAA	65
R2M50	CCTGAAGGAAATGATAGCAATAG	GTTTTGTATGCTCTTCACTTGTC	42
R3M10	CCGAGTACCATTGCTTTC	CTGCCATAGTTACTGCTCTGTT	37
R3M23	TGCTTACAAGGGTCCAAT	GGAGGTGCCTACCAAGAG	36
R3M30	AGGCTAAGTGAAGAAATAATAAG	CTCCGTATTCATTACTGGTTG	24
R3M37	GCATTGAATTGTACTCTTATTATAT	ACGAATCAAAAGGAGACTAAAAT	56
R3M53	ACACTGGCTACGGCAAAG	TTTGTTCGGGAATAATGATGC	35
R4M13	TACACGGTAGACATCCAACA	ATGATTTAACCGTAGATTGG	32
R4M17	AGTGCTCGGTTTTGTTTTC	GTCAGATATAATTGATGGATGTA	51
R4M30	GCTTCTCCTGGTTGTATGC	AAAATAGGGAGGCAGATAGAC	40
R4M43	CTTGAACCTGAGTGAGTGG	CGATGAAAATGATGTCTA	34
R4M50	TTTTGTGAAACTTGACCCTC	GCGTCCATGTCTTTATTGTG	33
R5M13	GAGAAAGAGTGGAAGGAG	AGTATCGTCAGGAGGGTC	32
R5M30	CTCAATTTCACCCATCCC	CGCTCCGTCTCCAACCTC	46
R6M14	AAATGTCCATGTGTTTGCTTC	CATGTGTGGAATGTGGTTG	34
R6M30	CACAAGCCGTAGCAGAGC	TCACGAAAAAGACCCCAAG	34
R6M44	TTAGGAATAAAGGCTGGATA	TTACCGTTAATAGGTGGAA	34
R7M7	ACCTTCCCTCCCCTTTTGAT	AACTTGGTCTTCCTGTTTTATTG	67
R7M20	GTTTTGTGCATTCCTTTAC	TTTATGACATTTTGACCG	66
R7M37	CAGCCCTAAATCTAAATACCC	ACGTTGAGACAGGCGAGC	36
R8M23	CCTATTCACTCTACCGACAT	GTTTAGTTCCCATTGCTTT	36
R8M33	CGAAAGAGGAGAGGGGTAGT	CGAAAACGAGAAACAAATA	38
R8M46	CAGCAGAGTCCAGAGAAGAT	GCATAAGATGGCGAGTGA	30
R9M10	CTTTGGATTCAGGGGGA	AACTTGAAACGGAGGCAG	43
R9M20	ACTGCTTTGATGGCTTGTG	CTCCCCAAACTGAATCC	40
R9M30	CTCACCTACCTAAAACCCAAC	CCACCCAAATCTGATACTG	32
R9M42	CTATAAGACCAAAACGAAAACT	GAAAACCATTGTGTCACTGTA	48
R10M10	GAATACAACCCCCTAAAAAC	ATGGACCGTTGAGGAGAC	38
R10M17	TGAACAATAAACCACAGAAGCA	CCCTTTATTCCCTCCTTTG	31
R10M30	CCCTAAAAATAGAGCAACCT	ACCCATAATACTACCAATCAAC	19
R10M40	GTCCCTAGGCCATCTCTTG	GCGAATAGGGGTGGACAG	33
R11M23	AAGGTTGACAAGGACAGAAG	TCGCAGGAATGGATAAAA	42
R11M40	AAGAAAAATATCTATTGAGGAGTG	GGAGGACCATAAATGACGG	41
R12M10	ATCATTTCAGCCTGTGCC	AGCTTAATAGGGGGGACG	47
R12M27	ATTTCATTGCCATCAGTT	GTAATCTTCTATCCGTTCA	33
R12M33	TTGATGATAGTATTTGCTGATG	AGATAGTGTCGGCGGTGG	42
R12M43	CCGCCGAGAAGAAACAAAG	CCCAAGAACAGGATTACA	30

The 34 underlined InDel primer pairs in bold letters indicate the effective loci for *indica* and *japonica* identification.

sample at all examined InDel loci followed the formula below:

Frequency of *indica* alleles (F_i)

$$F_i = \frac{2\sum_1^N X_{ii} + \sum_1^N X_{ij}}{2N} \quad (1)$$

Frequency of *japonica* alleles (F_j)

$$F_j = \frac{2\sum_1^N X_{jj} + \sum_1^N X_{ij}}{2N} \quad (2)$$

where X_{ii} indicates the homozygous *indica* genotype (II) at a given InDel locus of a particular rice sample scored based on the electrophoresis banding pattern; X_{jj} indicates the homozygous *japonica* genotype (JJ) at a given InDel locus of a particular rice sample scored based on the electrophoresis banding pattern; X_{ij} indicates the heterozygous *indica–japonica* genotype (IJ) at a given InDel locus of a particular rice sample scored based on the electrophoresis banding pattern; N indicates the total number of InDel loci examined.

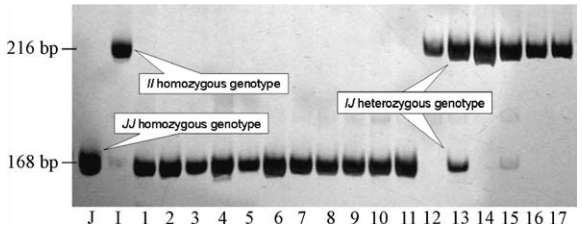

Fig. 1. PCR amplification and electrophoresis of 19 rice varieties using the insertion/deletion (InDel) primer pair R1M37, showing the homozygous *indica* genotype (II), homozygous *japonica* genotype (JJ), and heterozygous *indica–japonica* genotype (IJ). Lanes J, I: Nipponbare and 93-11; lanes 1–17: Jiangyinzhao, Fushun Xiaobairen, Zihong, Zhushituo, Sichengou Jiangmidao, Tianzuozao, Yingqiunuo, Manyedao, Weinan Hongmangdao, Madaozi, Hongzangu, Zahu, Panqiunuo (heterozygous genotype), Xiangu, Baikeshangu (heterozygous genotype), Huidongxian, and Hubeizao.

2.5. Determining InDel loci used for accurate indica and japonica rice identification

The particular molecular markers developed based on polymorphisms of InDel loci clearly reflected differences between the total genomic DNA sequences of *indica* rice 93-11 and *japonica* rice Nipponbare. However, these may not necessarily represent the differences between *indica* and *japonica* subspecies at the genomic level. Therefore, experimental validation for the effectiveness of the 45 InDel markers in identifying *indica* and *japonica* rice was performed, involving 22 typical *indica* rice varieties and 22 typical *japonica* rice varieties as an experimental population. Following the neutrality theory of Manly [27], if the InDel fragments (or their tightly linked functional genes) were under considerable selection during the evolutionary process of *indica–japonica* rice differentiation, the distribution of these fragments would show significant deviation from the neutrality (under selection) in the experimental population. The Ewens–Watterson Neutrality Test model [28,29] included in the analytical software for population genetics (POPGENE ver. 13.2 [30]) was applied to examine whether the InDel loci were under significant selection during *indica–japonica* differentiation. The above analysis involved the data matrix of genotypes obtained from PCR and electrophoresis of all the InDel primer pairs (loci) in the experimental population with 44 typical *indica* and *japonica* rice samples. Only those InDel loci that showed significant selection in the *indica–japonica* rice population were maintained as the key InDel markers for identification of all other rice varieties. The experimental validation of InDel markers for determination of *indica–japonica* differentiation was confirmed by principal component analysis (PCA) using the software MINITAB ver. 14.13 [31]. Only the InDel loci with high values contributing to the *indica–japonica* differentiation were maintained for the identification of *indica* and *japonica* rice varieties.

2.6. Standards for indica and japonica rice determination

The development of the InDel molecular index for identifying *indica* and *japonica* rice varieties was essentially based on the average allelic frequency (F) of multiple InDel loci across the total rice genome. The unique feature of this method for determining *indica* or *japonica* characteristics and studying the genetic relationships of *indica* and *japonica* rice varieties was to have a comprehensive judgment from multiple InDel loci. Therefore, this InDel index method could calculate the average frequency of *indica*-specific alleles (F_i) or *japonica*-specific alleles (F_j) at multiple loci across the whole genomes, which provides a guarantee for the accurate identification of *indica* and *japonica* rice. Different types of rice with various degrees of *indica* and *japonica* differentiations in the examined samples could accurately be determined following the established standard as indicated in Table 5. The standard was created based on the frequency of *indica*- or *japonica*-specific alleles at different InDel loci of the examined rice samples.

3. Results

3.1. PCR products from InDel primer pairs and their electrophoretic patterns

PCR amplification of DNA samples extracted from the sequenced *indica* variety (93-11) and *japonica* variety (Nipponbare) was performed using the selected 45 InDel primer pairs. The amplified DNA products were subjected to electrophoresis that visualized distinct and polymorphic (usually dimorphic) banding patterns from the two rice varieties (Fig. 1). This confirmed the existence of the detected DNA fragments and their variation at the 45 InDel loci obtained from BLAST across the total genomic sequences between 93-11 and Nipponbare (Fig. 1). Similarly, distinct electrophoretic patterns were obtained from PCR amplified DNA samples of the 42 typical *indica* and *japonica* rice varieties that were identified based on the traditional Chen's index. The InDel primer pair R5M20 failed to amplify PCR products in some rice varieties, and some other InDel primer pairs were not able to produce polymorphic electrophoretic bands from DNA samples of rice varieties after PCR amplification. The same results of PCR amplification and electrophoresis were obtained from 42 other randomly selected rice varieties from different countries or regions using the 45 InDel primer pairs. This also confirmed the existence of the detected DNA fragments and their variation at the 45 InDel loci in nearly all the rice varieties included in this study. When the 45 InDel primer pairs were used to amplify DNA samples of wild *Oryza* species with different genomes, similar results were obtained from PCR amplification and electrophoresis, although a greater variation in banding patterns was observed in the wild species. The 45 InDel primer pairs could amplify DNA samples of all the AA-genome *Oryza* species, but could not produce perfect results from DNA

Table 5
Classification standard between typical *indica* and typical *japonica* rice based on the frequency of *indica*-specific alleles (F_i) or *japonica*-specific alleles (F_j) calculated from InDel amplified products.

Allele frequency		Type of rice identified by the InDel index
indica-specific (F_i)	*japonica*-specific (F_j)	
>0.90	<0.10	Typical *indica*
0.75–0.89	0.11–0.25	*indica*
0.61–0.74	0.26–0.39	Close to *indica*
0.40–0.60	0.40–0.60	Intermediate type
0.26–0.39	0.61–0.74	Close to *japonica*
0.11–0.25	0.75–0.89	*japonica*
<0.10	>0.90	Typical *japonica*

samples of *Oryza* species with other genomes (e.g., BB, CC, EE, CCDD, and GG). For example, those InDel primer pairs, R9M42, R1M47, R5M13, R4M50, and R8M46 did not amplify any PCR products. In addition, DNA samples from wild *Oryza* species (populations) produced a much higher level of polymorphic bands (Fig. 2) than those from cultivated rice when the same InDel primer pairs were used for amplification. These results indicated that InDel primer pairs have significant roles in identifying *indica* or *japonica* characteristics of cultivated and wild rice species, as well as in studying *indica–japonica* genetic differentiation.

3.2. InDel loci associated with indica and japonica differentiation

Applying the Ewens–Watterson neutrality testing model, we analyzed the data matrix generated from PCR analysis and electrophoresis of the experimental population, consisting of 22 typical *indica* rice varieties (including 93-11) and 22 typical *japonica* rice varieties (including Nipponbare), using the 45 InDel primer pairs. Results from the analysis clearly indicated 34 InDel primer pairs that demonstrated significant selection under *indica–japonica* rice differentiation by showing evident deviation from neutrality in distribution. There were 11 InDel loci that did not have any association with the *indica–japonica* differentiation in rice by showing neutral selection and random distribution in the experimental population (Appendix A). Results from PCA analysis of the 45 InDel loci using the same data set, but transformed into a binary data matrix (0, 1), strongly supported the results from the neutrality test. In the PCA analysis, around 34 InDel primer pairs showed significantly high contribution to the separation of the 44 typical *indica* and *japonica* rice varieties (Table 4). The physical location of the 34 InDel loci was identified as distributed on 12 rice chromosomes (Fig. 3). The selected 34 InDel primer pairs were retained for further analysis of *indica* and *japonica* characteristics of DNA samples from cultivated rice varieties and wild rice species (populations) with a much wider distribution.

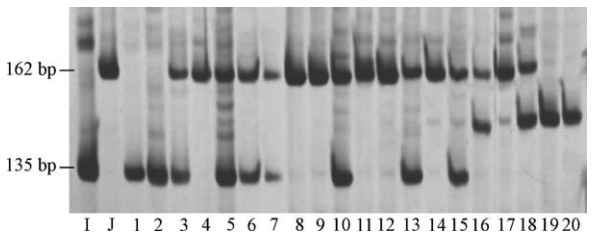

Fig. 2. PCR amplification and electrophoresis of *indica* 93-11 and *japonica* Nipponbare (lanes I, J), and 20 wild *Oryza* samples using the insertion/deletion (InDel) primer pair R10M30. Lanes 1–10: *Oryza rufipogon*; lanes 11–17: *O. nivara*; lanes 18–20: *O. longistaminata*. More heterozygous genotypes and specific alleles different from 93-11 or Nipponbare were observed in wild *Oryza* samples.

3.3. Determination of cultivated and wild rice for indica and japonica characteristics based on InDel molecular markers

Applying the 34 selected InDel primer pairs that showed a close association with *indica–japonica* rice differentiation, we calculated the *indica* or *japonica* allelic frequency (F_i and F_j) of 22 typical *indica* (including 93-11) and 22 typical *japonica* rice varieties (including Nipponbare), respectively, using the formulae established ((1) and (2)). When compared the results generated from the two methods of the InDel molecular index and the traditional Chen's index with the same set of rice samples, we found that the two independent methods produced nearly identical outcomes (Table 1). Based on the genotype-scoring results (II, JJ, and IJ) of the 34 InDel loci from other cultivated rice and wild *Oryza* species (populations) with a wider distribution, we calculated the average frequencies of the *indica*-specific (F_i) and *japonica*-specific (F_j) alleles of these rice samples using our established formulae ((1) and (2)). Results from the calculation indicated that the cultivated rice samples from South Asia (including India and Bangladesh) showed the characteristics of typical *indica* to intermediate types, whereas most of the cultivated rice samples from East Asia (including Japan and Korea) were *japonica* types. Interestingly, cultivated rice samples from Southeast Asian countries showed considerable variation in terms of their *indica–japonica* characteristics, ranging from typical *indica* to typical *japonica* varieties, although more rice types such as close to *indica* or close to *japonica* types were identified (Table 2). Results further confirmed that wild *Oryza* species (populations) with different genomes, regardless of their origins, were all intermediate types, lacking evident *indica–japonica* differentiation. The frequency (F_i and F_j) of *indica*- or *japonica*-specific alleles varied between 0.45 and 0.65 (Table 3).

4. Discussion

Comparative studies applying BLAST to the total genomic sequences of a typical *indica* rice variety (93-11) and *japonica* rice variety (Nipponbare) revealed substantial differences in DNA fragments, as either insertion or deletion (InDel) at specific loci, between the two rice subspecies. Experimental analysis based on PCR of DNA samples extracted from *indica* 93-11 and *japonica* Nipponbare, as

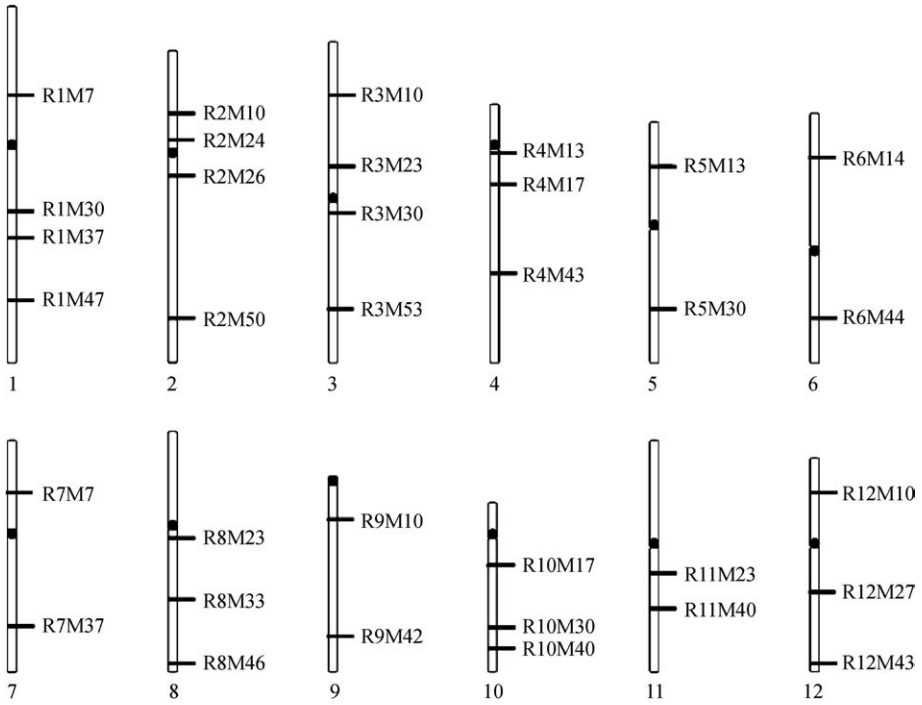

Fig. 3. The physical location of the 34 insertion/deletion (InDel) loci that showed a strong association with *indica–japonica* differentiation across the 12 rice chromosomes (black dots indicate the position of the centromere).

well as electrophoresis of the PCR products confirmed the differences in DNA sequences between 93-11 and Nipponbare at the 45 loci, obtained from BLAST. Applying the methods of experimental biology and population genetics, we performed PCR application of rice DNA samples using the primer pairs designed based on the differences in DNA fragments at the 45 InDel loci and conducted electrophoresis of the amplified PCR products. As a consequence, polymorphic (mostly dimorphic) banding patterns were generated from electrophoresis of DNA samples of 86 cultivated rice varieties and 11 wild *Oryza* species (populations) with different origins. This indicated that the 45 InDel primer pairs truly reflected differences at the 45 specific loci across the whole genomes of *indica* and *japonica* rice. Therefore, theoretically speaking, the InDel molecular markers can be used to determine the *indica* and *japonica* characteristics and the *indica–japonica* differentiation in rice varieties. Nevertheless, the DNA fragmental differences detected from BLAST analysis across the total genomic DNA of *indica* 93-11 and *japonica* Nipponbare may only represent the differences between the two varieties (93-11 and Nipponbare), rather than differences between *indica* and *japonica* subspecies. Therefore, the InDel molecular markers need a thorough validation for their efficiency in *indica* and *japonica* identification by including a large sample size of typical *indica* and *japonica* varieties as an experimental population for analysis. The key point of validation is to determine whether the specific InDel loci are closely associated with the genetic differentiation of *indica* and *japonica* rice during domestication and evolutionary processes. Only those InDel loci that are strongly associated with the *indica–japonica* differentiation can be applied for the effective determination of *indica* and *japonica* characteristics in rice varieties.

In this study, we applied the neutrality test – one of the powerful population genetic analytical tools and PCA to validate the 45 InDel primer pairs including 22 typical *indica* and 22 typical *japonica* rice varieties that have confidently been determined using the traditional Chen's index as an experimental population. Among the 45 InDel loci, 34 showed a strong association with the genetic differentiation between *indica* and *japonica* rice. Experimental data further indicated that the 34 InDel specific primer pairs can accurately determine *indica* and *japonica* characteristics of rice varieties collected from a wide range of origins. Comparing the results with the Chen's index method [19,20] used to identify *indica* and *japonica* rice varieties based essentially on morphological characters, the InDel molecular index established in this study can provide even more accurate results for the identification of *indica* and *japonica* rice varieties. In addition, the InDel molecular index only requires a minimum amount of rice DNA samples (<100 mg) for use in identifying *indica* and *japonica* rice samples and does not rigidly require DNA samples from a specific organ or tissues of rice plants. Therefore, the InDel molecular index has an excellent feature of fast (usually obtains results within a week), simple (any laboratory with PCR facilities and electrophoresis can do the identification), and highly effective in identifying *indica* and *japonica* rice samples. Interestingly, the InDel molecular index can circumvent the procedure of seed germination and seedling culture. If it is combined with the protocol of

direct DNA extraction from rice seeds [32], the method of the InDel molecular index can largely speed up the identification process (only two days) for *indica* and *japonica* rice samples. More importantly, because the InDel specific loci are distributed on different rice chromosomes, the InDel molecular index can determine *indica* and *japonica* characteristics of rice samples synthetically on the basis of multiple loci across the entire rice genome. This provided a unique method to determine *indica–japonica* characteristics of a wide range of rice varieties, such as those that differentiate into typical *indica* or *japonica*, and those that have less *indica–japonica* differentiation, and even for those that do not have clear *indica–japonica* differentiation (Tables 2 and 3). The InDel molecular index not only provides a highly accurate and effective method for *indica–japonica* rice identification, but more importantly, it also can be used to analyze samples of all the wild *Oryza* species and wild relatives in the tribe Oryzeae, significantly extending the range of research materials and research scopes. Thus, the method of the InDel molecular index has its ideal implications in identifying *indica* and *japonica* rice varieties and analyzing cultivated and wild rice species in terms of their *indica* and *japonica* differentiations.

Rice breeding is a unique procedure for rice germplasm innovation through genetic manipulation. Therefore, one of the most important tasks in rice breeding is to select appropriate rice germplasm in breeding programs. One of the strategies in rice breeding is to utilize the abundant genetic recombination resulting from hybridization between *indica* and *japonica* subspecies. As a prerequisite, selection of appropriate *indica* and *japonica* germplasm as parents included in hybridization becomes very critical. In traditional rice breeding programs, the determination of *indica* or *japonica* rice germplasm as parents for making a hybrid combination is essentially based on its agronomic characteristics and pedigree information, which creates a lot of uncertainty. The application of the InDel molecular index in combination with agronomic characteristics will increase the accuracy and efficacy for identifying *indica* and *japonica* rice varieties, which essentially resolves the problems of correct selection for *indica* and *japonica* rice germplasm used in rice breeding [33]. In addition, the breeding of "super rice" by utilizing the super heterosis (hybrid vigor) from *indica–japonica* hybridization is another important rice breeding strategy [11,12]. However, it is very difficult to utilize the super heterosis between *indica–japonica* hybrids because of their high degree of sterility [34] that immediately reduces the productivity of *indica–japonica* hybrid rice and become the major constraint for utilizing super heterosis when developing *indica–japonica* hybrid rice. The application of the InDel molecular index method may not only facilitate the accurate identification of *indica* and *japonica* germplasms with various degrees of differentiation, but also assist in studying mechanisms of genetic differentiation in *indica* and *japonica* rice varieties. The InDel molecular index offers a novel thought and useful tool for selecting most suitable *indica* and *japonica* parents for hybrid combinations by exploring the intimate relationships between *indica–japonica* differentiation and fertility in different *indica–japonica* hybrid combinations. This procedure can be illustrated by selecting the most "fit" *indica* and *japonica* parents based on the frequency of their *indica*-specific (F_i) or *japonica*-specific (F_j) alleles of the InDel loci, which may improve the situation of compatibility between the parents, as well as fertility in the sub-specific hybrids. This novel method may contribute to the successful breeding of "super rice" by optimizing the super hybrid vigor, and at the same time, improving fertility of the hybrids between *indica–japonica* subspecies.

This study applied the newly developed method of the InDel molecular index to determine the *indica* or *japonica* characteristics of a set of samples including typical *indica* and typical *japonica* rice varieties that have been confirmed by the traditional Chen's index, as well as the randomly selected rice varieties from a wide range of origins and different wild *Oryza* species (populations). Results from this study showed that most of the cultivated rice samples had considerable *indica* and *japonica* differentiation, but the wild *Oryza* species (populations) did not have such differentiation, i.e., the wild rice species showed intermediate characteristics in terms of *indica* or *japonica* variation. This is particularly true for wild rice species containing the other genomes rather than the AA-genome. It is true that there exists no *indica–japonica* differentiation in wild *Oryza* species or populations. This is a very interesting question that deserves a follow-up study. It would be very important to analyze more wild *Oryza* species, particularly the ancestral species of cultivated rice – common wild rice (*Oryza rufipogon*) and its close wild relatives with a wide range of distribution, to examine their *indica* or *japonica* characteristics for understanding their genetic differentiation. Results from such analyses applying the InDel molecular index, from a completely new angle, explain the single origin (monophyletic) or multiple origins (polyphyletic) of Asian cultivated rice, which is still an unresolved scientific question receiving long-term debate [2,15–17]. Following our classification standard (Table 5), data from this study further showed that most of the rice varieties from South Asia represented the typical *indica*, *indica*, close to *indica*, or intermediate types, whereas most of the rice varieties from Japan, Korea, and northern China were identified as typical *japonica* or *japonica*. Rice varieties from Southeast Asia and southern China represented a wide range of variation in terms of *indica–japonica* differentiation. This finding indicated that the InDel molecular index is a very valuable tool that can be used not only to determine the *indica* or *japonica* characteristics of cultivated rice and wild *Oryza* species, but also to investigate *indica–japonica* differentiation in rice varieties and wild *Oryza* species/populations. This tool will be of great significance to interpret the origin and domestication of cultivated rice from its wild ancestor and to investigate how *indica–japonica* differentiation occurred during the domestication process by adapting to human and natural selection under different geographical and environmental conditions.

Acknowledgements

This work was supported by the National Basic Research Program of China (Grant No. 2006CB100205), the National Natural Science Foundation of China (Grant No. 2007ZD001), and the Shanghai Commission of Science and Technology (Grant No. 08JC1401700).

Appendix A

The Ewens–Watterson Neutrality Test [29,30] of the 45 InDel primer pairs under selection during *indica–japonica* differentiation of cultivated rice.

InDel locus	*n*	*k*	Obs *F*	Min *F*	Max *F*	Mean	SE	L95	U95
R1M7	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R1M20	88	2	0.60	0.50	0.97	0.81	0.02	0.50	0.97
R1M30	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R1M37	86	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R1M47	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R2M10	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R2M24	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R2M26	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R2M37	88	2	0.73	0.50	0.97	0.80	0.02	0.50	0.97
R2M50	88	2	**0.50**	0.50	0.97	0.79	0.02	0.50	0.97
R3M10	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R3M23	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R3M30	88	2	**0.50**	0.50	0.97	0.79	0.02	0.50	0.97
R3M37	82	2	0.56	0.50	0.97	0.80	0.02	0.50	0.97
R3M53	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R4M13	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R4M17	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R4M30	84	2	0.67	0.50	0.97	0.81	0.02	0.50	0.97
R4M43	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R4M50	82	2	**0.50**	0.50	0.97	0.79	0.02	0.50	0.97
R5M13	88	2	**0.50**	0.50	0.97	0.79	0.02	0.50	0.97
R5M30	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R6M14	88	2	**0.50**	0.50	0.97	0.79	0.02	0.50	0.97
R6M30	88	2	0.63	0.50	0.97	0.80	0.02	0.50	0.97
R6M44	88	2	**0.50**	0.50	0.97	0.79	0.02	0.50	0.97
R7M7	86	2	**0.51**	0.50	0.97	0.80	0.02	0.50	0.97
R7M20	88	2	0.60	0.50	0.97	0.80	0.02	0.50	0.97
R7M37	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R8M23	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R8M33	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R8M46	86	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R9M10	88	2	**0.53**	0.50	0.97	0.79	0.02	0.50	0.97
R9M20	88	2	0.52	0.50	0.97	0.80	0.02	0.50	0.97
R9M30	88	2	0.60	0.50	0.97	0.80	0.02	0.50	0.97
R9M42	88	2	**0.50**	0.50	0.97	0.81	0.02	0.50	0.97
R10M10	88	2	0.53	0.50	0.97	0.79	0.02	0.50	0.97
R10M17	88	2	**0.50**	0.50	0.97	0.81	0.02	0.50	0.97
R10M30	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R10M40	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R11M23	88	2	**0.50**	0.50	0.97	0.81	0.02	0.50	0.97
R11M40	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R12M10	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97
R12M27	88	2	**0.50**	0.50	0.97	0.79	0.02	0.50	0.97

(*continued on next page*)

Appendix A (*continued*)

InDel locus	*n*	*k*	Obs *F*	Min *F*	Max *F*	Mean	SE	L95	U95
R12M33	88	2	0.55	0.50	0.97	0.81	0.02	0.50	0.97
R12M43	88	2	**0.50**	0.50	0.97	0.80	0.02	0.50	0.97

The 34 InDel loci with underlines and in bold letters showed significant departure from neutrality. *n*, sample size; *k*, number of alleles per locus; Obs *F*, distribution of observed frequencies; Min *F*, the minimum values of frequencies; Max *F*, the maximum values of frequencies; Mean, average values; SE, standard error; L95, the values of the lower-limit at the significance level; U95, the values of the upper-limit at the significance level.

References

[1] Zong Y, Chen Z, Innes JB, et al. Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. Nature 2007;449:459–63.

[2] Vaughan DA, Lu BR, Tomooka N. The evolving story of rice evolution. Plant Sci 2008;174:394–408.

[3] Lu BR, Zheng KL, Qian HR, et al. Genetic differentiation of wild relatives of rice as referred by the RFLP analysis. Theor Appl Genet 2002;106:101–6.

[4] Liu KD, Zhang QF, Zhang DP, et al. Genetic variation and *indica–japonica* differentiation in Yunnan indigenous rice. Acta Bot Sin 1995;37:719–24, [in Chinese].

[5] Wang YP, Wei XH, Hua L, et al. Genetic diversity in upland rice germplasm from different geographical regions. Acta Agron Sin 2007;33:2034–40, [in Chinese].

[6] Gupta PC, O'Toole JC. Upland rice: a global perspective. Los Baños, The Philippines: The International Rice Research Institute; 1986.

[7] Olsen KM, Caicedo AL, Polato N, et al. Selection under domestication: evidence for a sweep in the rice waxy genomic region. Genetics 2006;173:975–83.

[8] Zeng Y, Zhang H, Li Z, et al. Evaluation of genetic diversity of rice landraces (*Oryza sativa* L.) in Yunnan, China. Breed Sci 2007;57:91–9.

[9] Bellon MR, Brar DS, Lu BR, et al. Rice genetic resources. In: Dwoling NG, Greenfield SM, Fischer KS, editors. Sustainability of rice in the global food system. Manila: Pacific Basin Study Center and IRRI; 1998. p. 251–83.

[10] Wang XK, Li RH. Determination and classification of subspecies of Asian rice and their inter-subspecies hybrids. Chin Sci Bull 1997;42:2596–603, [in Chinese].

[11] Khush G. Green revolution: the way forward. Nat Rev 2001;2:815–22.

[12] Peng SB, Laza RC, Visperas RM, et al. Rice: progress in breaking the yield ceiling. In: Proceedings of the fourth International Crop Science Congress, September 26–October 1, 2004, Brisbane, Australia, Published on CD.

[13] Lin Z, Griffith ME, Li X, et al. Origin of seed shattering in rice (*Oryza sativa* L.). Planta 2007;226:11–20.

[14] Sweeney MT, Thomson MJ, Pfeil BE, et al. Caught red-handed: Rc encodes a basic helix–loop–helix protein conditioning red pericarp in rice. Plant Cell 2006;18:283–94.

[15] Londo JP, Chiang YC, Hung KH, et al. Phylogeography of Asian wild rice reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc Natl Acad Sci USA 2006;103:9578–83.

[16] Sweeney MT, Thompson MJ, Cho YG, et al. Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genet 2007;3:1418–24.

[17] Garris AJ, Tai TH, Coburn J, et al. Genetic structure and diversity in *Oryza sativa* L. Genetics 2005;169:1631–8.

[18] Oka HI, Chang WT. Rice varieties intermediate between wild and cultivated forms and the origin of the Japonica type. Bot Bull Acad Sin 1962;3:109–31.

[19] Cheng KS, Wang XK, Lu YX, et al. Studies on the indigenous rices in Yunnan and their utilization. II. A revised classification of Asian cultivated rice. Acta Agron Sin 1984;10:271–9, [in Chinese].

[20] Cheng KS. Identification in *indica* and *japonica* subspecies of Asian cultivated rice. Kunming: Yunnan Science and Technology Press; 1993, [in Chinese].

[21] Long WH, Xu MH. RAPD-based genetic difference between *indica* rice and *japonica* rice. J Yunnan Agric Univ 2002;17:245–7, [in Chinese].

[22] Zhu MY, Wang YY, Zhu YY, et al. Estimating genetic diversity of rice landraces from Yunnan by SSR assay and its implication for conservation. Acta Bot Sin 2004;46:1458–67.

[23] Wang SW, Liu X, Wang Y, et al. Genome polymorphisms between *indica* and *japonica* revealed by RFLP. Sci Agric Sin 2006;39:1038–43, [in Chinese].

[24] Shen YJ, Jiang H, Jin JP, et al. Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. Plant Physiol 2004;135:1198–205.

[25] Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 1980;8:4321–5.

[26] Song ZP, Xu X, Wang B, et al. Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. Theor Appl Genet 2003;107:1492–9.

[27] Manly BFJ. The statistics of natural selection. London: Chapman and Hall; 1985.

[28] Ewens WJ. The sampling theory of selectively neutral alleles. Theor Popul Biol 1972;3:87–112.

[29] Watterson GA. The sampling theory of selectively neutral alleles. Adv Appl Probab 1974;6:463–88.

[30] Yeh FC, Yang RC, Boyle T. POPGENE, version 1.32. Software Microsoft Window-Based Freeware for Population Genetic Analysis; 1997.

[31] MINITAB. MINITAB statistical software release 14 for windows. MINITAB Inc.; 1994.

[32] Kang HW, Cho YG, Yoon UH. A rapid DNA extraction method for RFLP and PCR analysis from a single dry seed. Plant Mol Biol Rep 1998;16:1–9.

[33] Yang ZY, Gao Y. Progress in research of utilizing heterosis between *indica* and *japonica* rice subspecies. Acta Agron Sin 1996;22:422–9, [in Chinese].

[34] Zhang GQ, Yang JC. Genetic studies of the hybrid sterility in cultivated rice (*Oryza sativa*) III. Allele differentiation of F1 pollen sterility in different types of varieties. J Genet Genomics 1993;20:541–51, [in Chinese].