

DNA methylation polymorphism and stability in Chinese *indica* hybrid rice

PENG Hai^{1,2†}, JIANG GuangHuai^{1†}, ZHANG Jing², ZHANG Weixiong^{2,3} & ZHAI WenXue^{1*}

¹Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China;

²Institute for Systems Biology and College of Life Science, Jiangnan University, Wuhan 430056, China;

³Department of Computer Science and Department of Genetics, Washington University in St. Louis, MO 63130, USA

Received August 2, 2013; accepted September 30, 2013

Conventional rice breeding has long focused on exploiting the DNA sequence diversity. However, epigenetic diversity, reflected particularly in DNA methylation, can also contribute to phenotypic variation and should not be overlooked in rice breeding. In this study, 20 parental lines of *indica* rice, which are widely used in hybrid rice breeding in China, were analyzed to investigate variations of DNA methylation and its inheritance. The results revealed a wide diversity in DNA methylation among these breeding lines. A positive correlation was seen between DNA methylation and genetic diversity. Furthermore, some of the methylated DNA was inherited in the subsequent generation, regardless of whether they were produced by selfing or hybrid-crossing. This study provides insight into the methylation patterns in rice, and suggests the importance of epigenetic diversity in rice breeding.

epigenetic diversity, genetic diversity, rice breeding line

Citation: Peng H, Jiang G H, Zhang J, et al. DNA methylation polymorphism and stability in Chinese *indica* hybrid rice. *Sci China Life Sci*, 2013, 56: 1097–1106, doi: 10.1007/s11427-013-4576-z

Rice (*Oryza sativa* L.) is the staple food crop for more than half of the world's population. Breeding of hybrid rice, especially the *indica* varieties, has been successful in China since the 1970s. Vigor of hybrid rice is defined as the improvement of a hybrid offspring relative to its inbred parents. As this depends on the genetic diversity between maternal and paternal lines, enormous effort has been devoted to elucidating genetic diversity among the parental lines used in breeding.

Presently, little is known about the epigenetic diversity, including variation of DNA methylation among parental breeding lines. Variation of DNA methylation can affect plant phenotypes [1] and impact some important agricultur-

al traits, such as resistance to rice bacterial blight (BB) [2,3], plant height [4] and yield [5]. Allelic variation in DNA methylation appears to be inheritable in subsequent generations [6–8], and inheritance of DNA methylation may be due to RNAi [9]. In *Arabidopsis*, large regions of contiguous methylation remain as stable DNA sequences across 30 generations [1]. In rice, the demethylation of the promoter region of the *Xa21G* gene activates transcription of *Xa21G* and subsequently confers resistance to rice BB. The methylation pattern and corresponding resistance to BB were stably inherited for at least nine generations [3]. This is sufficient for the commercial exploitation of hybrid rice because the average commercial life of a hybrid rice variety is about five years. Based on the stable inheritance of DNA methylation and its effects on phenotype, we have previously shown that DNA methylation is valuable for the de-

†Contributed equally to this work

*Corresponding author (email: wxzhai@genetics.ac.cn)

velopment of new varieties [10]. However, several questions remain to be answered regarding the role of DNA methylation in rice breeding. Is there a diversity of DNA methylation among parental lines used for rice breeding, and, if so, how widespread is it? Is the diversity of DNA methylation independent of DNA sequence diversity? Can DNA methylation variations be stably inherited across multiple generations?

In this study, we investigated the diversity of DNA methylation and patterns in 20 rice parental lines widely adopted in *indica* hybrid rice breeding programs in China. Our aim was to make the initial step in the establishment of a theoretical foundation for the application of DNA methylation alleles in hybrid rice breeding programs.

1 Materials and methods

1.1 Plant materials

Twenty *indica* rice parent lines (#1–#20; Table 1) were chosen for the study, including 11 restorer lines and nine maintainer lines. The restorer and CMS (cytoplasm male sterile) lines have been used as female and male parents in hybrid rice breeding programs, and varieties from these parental lines are widely cultivated in China (<http://www.ricedata.cn/variety/index.htm>). All 20 parental lines were planted in Hubei Province in the central region of China from April to September in 2009. Their selfing progenies were planted in Hainan Province in southern China from

November 2009 to April 2010. The progenies grown in Hainan were selfed, and used to investigate the inheritability of DNA methylation.

Jinke 1A ('#21') [11] is a male-sterile equivalent of #19 ('Jinke 1B'). These two lines possess the same nuclear genome but different cytoplasm. They were used to investigate the effects of cytoplasm on DNA methylation variation. #22 ('Jinke 1A'/'R8377') is the hybrid of #21 and #20, and was used to investigate the inheritance of methylation from the maternal and paternal lines. #23–#42 (Table 1) were used to investigate the segregation of DNA methylation patterns in the F₂ generation, and are the selfing progenies from an F₁ hybrid (#22).

1.2 Seedling culture and DNA isolation

Only healthy seeds with no visible defects or signs of disease were selected. Seeds were surface sterilized in 70% ethanol, rinsed twice in deionized water, and placed in a clean floating tray (DSC00667, HengSiTe Packaging Co., Ltd., Taizhou, China) filled with quartz for germination and growth. The seedlings were grown in a climax chamber (Fuma Test Equipment Co., Ltd., Shanghai, China) at 30°C/20°C (day/night), 80% relative humidity, and a 13/11 h light/dark photoperiod with an intensity of approximately 180 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Hoagland nutrient solution (electric conductivity 2.2–2.5 ms cm^{-1}) [12] was used to provide nutrition for plant growth. With the exception of the F₂ genera-

Table 1 Parental lines used in this study

ID	Name	Breeding history	Type
1	Fuhui 838	⁶⁰ Coγ radiated Minghui 63/glutinous rice 226	Restorer line
2	Yihui 3551	CDR22/81136/MingHui78	Restorer line
3	Duoxi No.1	Minghui 63/Tetep	Restorer line
4	R527	1318/88-R336	Restorer line
5	9311	⁶⁰ Coγ radiated Yangdao No.4/3021	Restorer line
6	Gui 99	Longye5-3//IR661/IR2061	Restorer line
7	Enhui 58	Minghui 63/Miyang46	Restorer line
8	CDR22	IR50/ Minghui 63	Restorer line
9	R128	XiangQing/261//XiangQing	Restorer line
10	Minghui 63	IR63/Gui630	Restorer line
11	II-32B	ZhenShan97/IR665	Maintainer line
12	Mianxiang 5B	XiangGai B/II-32B	Maintainer line
13	NongXiang 1B	Not known	Maintainer line
14	Xieqinzao	AiBai/ZhuJun//XieZhen No.1	Maintainer line
15	K17B	Zhong 83-49/BoHuiZhan	Maintainer line
16	Zhong 9B	YouIB/L301B	Maintainer line
17	D62B	Hongtu No.31/D297	Maintainer line
18	Jin 23B	Huangjin No.3/Feigai B//RuanMi M	Maintainer line
19	Jinke 1B	Gang 46B/Jin 23B	Maintainer line
20	R8377	R527/Mianfei 725	Restorer line
21	Jinke 1A	The male sterile line of Jinke 1B	Male sterile line
22		Jinke 1A/R8377	F ₁ generation
23–42		Jinke 1A/R8377	F ₂ generation

tion (#23–#42; Table 1), in which there was only one seed for each accession, the experimental design for accessions (#1–#22; Table 1) was a completely randomized design of a single factorial set of 22 accessions, with three replications for each treatment (10 seedlings per accession). At the three-leaf stage, the middle parts of the penultimate leaves from normally growing plants were collected for DNA extraction. Plant genomic DNA was isolated, using an extraction kit (DP305, Tiangen Biotech Co., Ltd., Beijing, China), following the manufacturer's instructions.

1.3 Methylation-sensitive amplified polymorphism (MSAP) and inter-simple sequence repeat (ISSR) assays

Genomic DNA methylation assay procedures, including isoschizomer digestion, adapter ligation, pre-amplification, and selective amplification, were described by Keyte et al. [13]. All primers and adapters used are listed in Table 2. Electrophoretic separation of amplification products was

performed as described by Keyte et al. [13]. The gel was stained with 0.2% AgNO₃, photographed using a digital camera (D90, Nikon, Japan), and scored visually.

Twenty-four ISSR primers (Table 3) were selected for the genetic assay of the 20 rice lines. The annealing temperature was optimized for each ISSR primer based on a temperature gradient experiment (Table 3). Other steps were performed as described by Ajibade et al. [14]. An aliquot (7 µL) of each amplified PCR product was separated using a 2% agarose gel, and electrophoresis was performed in a 0.5×TBE buffer at 5 V cm⁻¹ for 2.5 h. The gel was stained for 15 min in 0.5 µg mL⁻¹ solution of ethidium bromide, and photographed using an image system (Bio Boc-It 220, UVP, California, USA).

At least two replicate reactions using the same DNA starting material were performed for the entire MSAP and ISSR processes from digestion to gel scoring, for all accessions and all primer sets, in addition to the biological replicates mentioned above.

Table 2 Primers and adapters used for the DNA methylation assay

<i>EcoR</i> I		<i>Msp</i> I/ <i>Hpa</i> II
Adapter		
1	5'-GACTGCGTACC-3'	5'-GACGATGAGTCCTGAG-3'
2	3'-CTGACGCATGGTTAA-5'	3'-TGCTACTCAGGACTCAT-5'
PreAmp primer	5'-AGACTGCGTACCAATT+C-3'	5'-GACGATGAGTCCTGAGTAA+A-3'
Selective primer		
1	5'-AGACTGCGTACCAATTC+AAC-3'	5'-GACGATGAGTCCTGAGTAA+AGG-3'
2	5'-AGACTGCGTACCAATTC+ACA-3'	5'-GACGATGAGTCCTGAGTAA+ACA-3'
3	5'-AGACTGCGTACCAATTC+ACC-3'	5'-GACGATGAGTCCTGAGTAA+ATG-3'
4	5'-AGACTGCGTACCAATTC+AGT-3'	5'-GACGATGAGTCCTGAGTAA+AGA-3'
5	5'-AGACTGCGTACCAATTC+CAA-3'	5'-GACGATGAGTCCTGAGTAA+ACT-3'
6	5'-AGACTGCGTACCAATTC+CAC-3'	5'-GACGATGAGTCCTGAGTAA+ACC-3'
7		5'-GACGATGAGTCCTGAGTAA+AGC-3'
8		5'-GACGATGAGTCCTGAGTAA+ATA-3'
9		5'-GACGATGAGTCCTGAGTAA+AAT-3'
10		5'-GACGATGAGTCCTGAGTAA+AAA-3'

Table 3 Primers and their annealing temperatures used in the ISSR assay^{a)}

UBC No.	Sequence	Annealing temperature (°C)	UBC No.	Sequence	Annealing temperature (°C)
808	5'-AGAGAGAGAGAGAGAGC-3'	58.0	846	5'-GAGAGAGAGAGAGAGAC-3'	56.8
811	5'-GAGAGAGAGAGAGAGAC-3'	49.2	847	5'-CACACACACACACACARC-3'	56.8
812	5'-GAGAGAGAGAGAGAGAA-3'	47.4	855	5'-ACACACACACACACACYT-3'	56.8
817	5'-CACACACACACACACAA-3'	49.2	856	5'-ACACACACACACACACYA-3'	56.8
825	5'-ACACACACACACACACT-3'	51.5	857	5'-ACACACACACACACACYG-3'	47.4
826	5'-ACACACACACACACACC-3'	59.5	861	5'-ACCACCACCACCACCACC-3'	65.7
834	5'-AGAGAGAGAGAGAGAGYT-3'	54.1	868	5'-GAAGAAGAAGAAGAAGAA-3'	51.5
835	5'-AGAGAGAGAGAGAGAGYC-3'	49.2	880	5'-GGAGAGGAGAGGAGA-3'	51.5
836	5'-AGAGAGAGAGAGAGAGYA-3'	49.2	885	5'-BHBGAGAGAGAGAGAGA-3'	54.1
840	5'-GAGAGAGAGAGAGAGAYT-3'	51.5	888	5'-BDBCACACACACACACA-3'	59.5
841	5'-GAGAGAGAGAGAGAGAYC-3'	54.1	889	5'-DBDACACACACACACAC-3'	62.0
842	5'-GAGAGAGAGAGAGAGAYG-3'	54.1	891	5'-HVHTGTGTGTGTGTGTG-3'	56.8

a) Sequences were provided by University of British Columbia, Canada. Y, pyrimidine; R, purine; B, minus A; D, minus C; H, minus G; V, minus T.

Bands of interest on the gels were purified using gel purification kit (GMS20008.1, GenMed, Shanghai, China), and sequenced at Sangon Biotech Company. Sequences were searched against the NCBI database of the *O. sativa* (*japonica* cultivar-group) genome.

1.4 Statistical analysis

Gene diversity was calculated with Popgene version 1.32 [15], using the ‘diploid data’ function setting of the ‘Dominant’ submenu. The statistical package NTSYSpc version 2.10e (Applied Biostatistics, Setauket, NY, USA) was used for the Mantel test [16]. UPGMA trees were constructed and 1000 bootstrap resamplings were performed using the software package MEGA 5.10 [17].

2 Results

2.1 The MSAP band patterns

The restriction enzymes *Msp* I and *Hpa* II are isoschizomers that show different sensitivities to the methylation status of two adjacent cytosine residues. They both recognize the same tetranucleotide sequence, 5'-CCGG. *Hpa* II is inactive if one or both cytosines in 5'-CCGG are fully methylated (methylation of both strands), and the enzyme can cleave the hemimethylated version of this sequence (only one of the strands is methylated). In contrast, *Msp* I cleaves C^{5m}CCGG but not ^{5m}CCGG [18]. Thus, the presence or absence of fragments in gels loaded with the amplification products from *Eco*R I+*Msp* I and *Eco*R I+*Hpa* II digests can be used to classify the methylation status of a particular 5'-CCGG site into four types (Table 4). Most studies classi-

fy type-II and type-III as methylated and the other two types as unmethylated (Table 4). We adopted this classification in the current study.

However, the presence of a type-I or type-IV band does not completely rule out the possibility of methylation (Table 4). Therefore, the methylation levels reported in this report and many previous ones on MSAP might be underestimated.

2.2 Variations of DNA methylation across parental rice lines

Forty-eight pairs of selective primers were used to detect cytosine methylation at 5'-CCGG in the 20 parental lines (#1–#20; Table 1). In all, 1842 fragments were amplified. Figure 1A shows a representative gel profile of the amplification products from the 20 parental lines. On average, 275.25 fragments out of 1842 fragments were differentially amplified between the digestion of *Eco*R I+*Msp* I and the digestion of *Eco*R I+*Hpa* II, due to either full methylation (type-II) or hemimethylation (type-III). Therefore, 14.78% (275.25/1842) of the 5'-CCGG sites in the genomes of the 20 rice parental lines were methylated. This proportion is slightly smaller than that reported by Xiong et al. [19] (16.3%) and Ashikawa [20] (16.2%). However, these two previous studies only considered two rice accessions and any fragment absent from both accessions would not be counted. In this study, as long as one band was present in any one of the 20 parental lines, it would be counted as a type-I (unmethylated) site in other lines. Therefore, the number of type-I sites was greater than that in the two previous studies. The proportion of methylation sites varied among the 20 parental lines from 13.48% to 16.92%, with a CV (coefficient of variation) of 6.87% (Table 5).

Table 4 Methylation types of the 5'-CCGG site as determined from gel images

Type	<i>Msp</i> I ^{a)}	<i>Hpa</i> II ^{a,b)}	Possible methylated cytosine residue
I	–	–	The external cytosines of both strands
II	+	–	The internal cytosines of both strands
III	–	+	The external cytosine of only one of the two strands
IV	+	+	No cytosine or the internal cytosine of only one of the two strands

a) *Msp* I and *Hpa* II represent the amplification products from the *Eco*R I+*Msp* I digest and *Eco*R I+*Hpa* II digest, respectively. + and – represent the presence and absence of the fragment in the gel image. b) *Hpa* II cuts hemimethylated 5'-CCGG 50 times slower than unmethylated DNA [18].

Table 5 Basic statistical parameters of the MSAP band patterns of the 20 parental lines of hybrid rice

Type	Average	Maximum	Minimum	Coefficient of variation
I	571.20	653	539	6.00%
II	159.90	228	129	13.46%
III	115.35	148	87	16.91%
IV	995.55	1040	892	3.78%
Methylated loci ^{a)}	275.25	315	251	6.87%
Ratio of methylated loci(%) ^{b)}	14.78	16.92	13.48	6.87%

a) Methylated loci=type II+type III. b) Ratio of methylated loci (%)=100×methylated loci/1842.

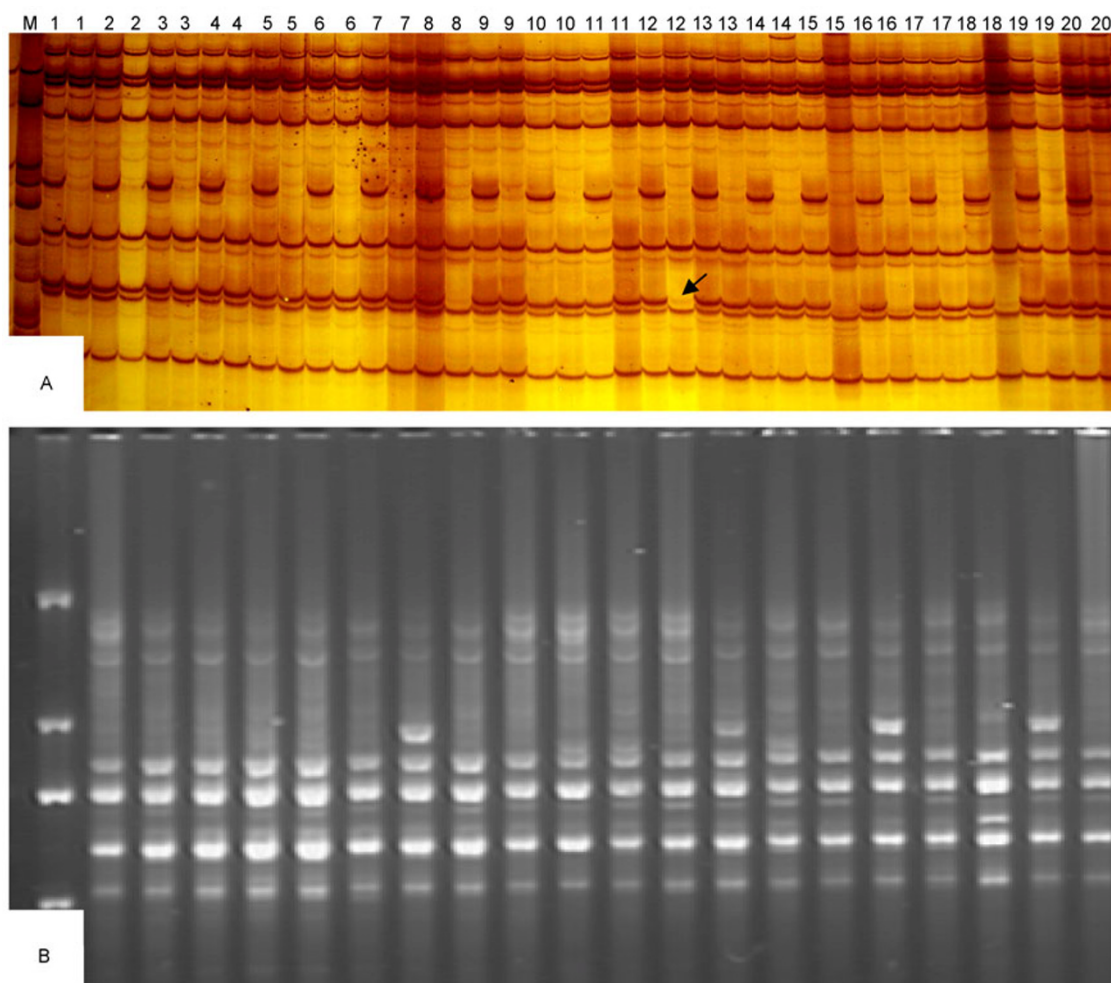


Figure 1 Examples of methylation (A) and ISSR genetic (B) profiles from 20 parental lines of hybrid rice. A, Methylation profile. M, marker. The selective primer combinations are 5'-AGACTGCGTACCAATTC+ACC-3' and 5'-GACGATGAGTCCTGAGTAA+ACT-3' (Table 2). Pairs of lanes with the same number above the figure refer to a single Chinese parental line of hybrid rice, as shown in Table 1. The former and latter lanes for each parental line represent the amplification products from the *EcoR* I+*Msp* I digests and the *EcoR* I+*Hpa* II digests, respectively. An arrow represents a DNA methylation site that is polymorphic among the 20 accessions. B, ISSR profile. The primer used was 825 (Table 3). The first lane contains the marker DL2000. The following 20 lanes represent the rice lines #1–#20 shown in Table 1.

2.3 Relationship between genetic and epigenetic diversities

Of the 1842 sites investigated, 821 (44.57%) were methylated in at least one of the 20 parental lines, among which 430 (52.37%) were polymorphic. The other 1412 sites were unmethylated or monomorphic in their DNA methylation status. Of those, 240 (17.00%) were genetically polymorphic, and could be used to investigate genetic diversity among the 20 rice accessions. The gene diversity [21], as measured by the average degree of gene differentiation (heterozygosity) for all analyzed loci among the population, was similar for sites with DNA methylation polymorphism and those with DNA sequence polymorphism (0.31 vs. 0.33; Table 6). In contrast, the proportion of DNA methylation polymorphic sites was clearly higher than that of DNA sequence polymorphic sites (52.37% vs. 17.00%), which indicated the relative plasticity of DNA methylation modifi-

cation in contrast to the more rigid DNA sequence. The methylation diversity (43%) in watermelon (*Citrullus lanatus*) heirlooms has also been reported to be higher than genetic diversity (19.8%) [22].

On a phylogenetic tree constructed from the 430 sites with DNA methylation polymorphism (Figure 2A), the 22 rice accessions formed two groups (I and II) with bootstrap values of 93% and 91%, respectively. Group I contains eight maintainer lines (#11–#15, #17–#19; Table 1), a male sterile (#21) equivalent of #19 and an F_1 line #22. Group II contains 11 restorer lines (#1–#10 and #20; Table 1) and a maintainer line #16. 'Jinke 1A' (#21) and 'Jinke 1B' (#19), which share the same nuclear genome but have different cytoplasms, could be clustered together with a relatively low bootstrap value of 65%. Differential DNA methylation patterns have been previously observed between reciprocal rice hybrids [23] that have different cytoplasms but the

same nuclear genome. Our results and those from previous studies indicated that the cytoplasm might, to some extent, influence DNA methylation of the rice nuclear genome. Epigenetic variation might account for the phenotypic variations, such as plant height and time to flowering, widely observed between a male sterile line and its maintainer line. The ‘Jinke 1A’/‘R8377’ (#22) hybrid has a close epigenetic relationship to its maternal parent, ‘Jinke 1A’ (Group I), but is remote from its paternal parent, ‘R8377’ (Group II, Figure 2A).

To maintain heterozygosity in hybrids, breeders try to avoid gene exchange between the restorer and maintainer lines. This helps to keep them genetically independent from each other. The distinction of the two lines on the epigenetic phylogenetic tree (Figure 2A) suggested a possible correlation of epigenetic polymorphism with genetic polymorphism. To confirm this hypothesis, we constructed and examined a dendrogram with 240 genetically polymorphic

sites corresponding to unmethylated or monomorphic DNA methylation sites. Similar to the DNA methylation dendrogram, most of the maintainer and restorer lines were clustered into two different groups, with bootstrap values of 93% (Figure 2B). A Mantel test revealed a positive correlation between the epigenetic (DNA methylation) and the genetic distances of the 22 rice accessions ($r=0.62$, $P=0.003$).

Because the presence of a type-I or type-IV band cannot rule out the possibility of variation in methylation, it was necessary to determine the strength of correlation between the variation in methylation and genetic polymorphism. We used 24 independent ISSR markers to detect genetic polymorphism. We obtained 181 amplified loci (e.g., Figure 1B), of which 23 (12.71%) were polymorphic. A Mantel test revealed a positive correlation between DNA methylation and ISSR genetic distance ($r=0.55$, $P=0.04$). The Mantel coefficient between the genetic distances based on ISSR

Table 6 Comparisons between the diversity of genetic and epigenetic (DNA methylation) sites

Type	Total No.	Polymorphic site		Gene diversity ^{a)}
		No.	Ratio	
DNA methylation site	821	430	52.37%	0.31
Genetic site	1412	240	17.00%	0.33

a) Gene diversity was calculated according to Nei [21], using only the polymorphic sites.

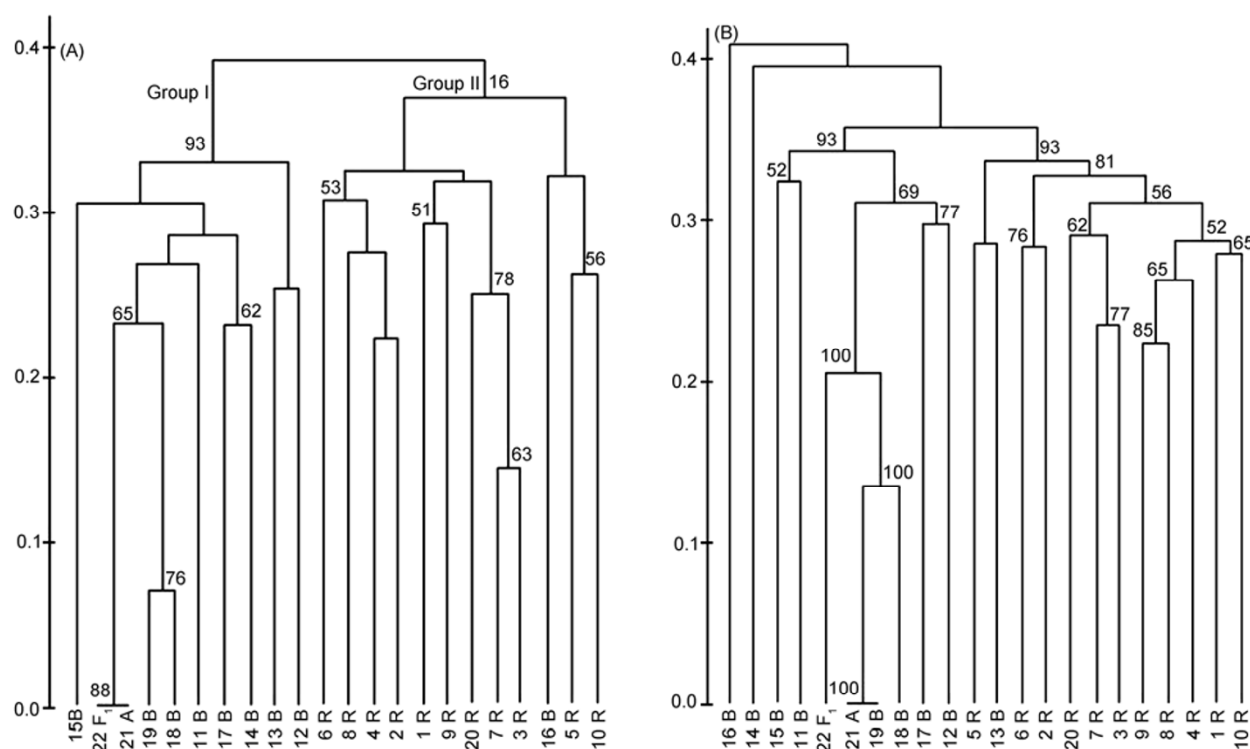


Figure 2 UPGMA dendrograms of Chinese parental lines of hybrid rice using DNA methylation (A) and classical genetic (B) polymorphic sites. (A), Numbers are the cultivar codes shown in Table 1. R, restorer line; M, maintainer line; A, male sterile line; F₁, F₁ generation. (B), Branching robustness is expressed as percentage reliability after 1000 bootstrap resamplings. Only bootstrap values >50 are shown.

markers and the unmethylated MASP markers reached 0.89 ($P=0.03$), suggesting that the two techniques were approximately equivalent for the assessment of genetic diversity.

2.4 Inheritance of DNA methylation alleles

To determine the inheritability of methylation across generations, we scrutinized variations in the polymorphic methylation sites between the 20 parental lines and their selfed progenies. One hundred and forty-eight randomly selected sites were investigated. An average of 31.00 (20.95%) sites exhibited changes in their DNA methylation status between parents and offspring, with a CV of 30.19% (Table 7). It is important to note that parental lines and their offspring were grown at different locations and different seasons (see Material and methods). This might contribute to the unstable transmission of DNA methylation status from parent to offspring. Caution should be taken when interpreting the breeding value of the remaining 79.05% 'stable' sites. A tracking study is required to determine the stability of those sites over several selfing generations or under different environments.

The F_1 hybrid combination of 'Jinke 1A'/'R8377' and its

parental lines were used to analyze the pattern of DNA methylation inheritance in the crossing process. Figure 2A shows that the maternal parent 'Jinke 1A' has a closer relationship with the F_1 hybrid than that with the paternal parent 'R8377', as revealed in the clustering tree of DNA methylation alleles.

Sixteen methylated sites that differed between 'Jinke 1A' and 'R8377' were randomly selected to investigate the segregation patterns of F_1 methylation alleles in the F_2 population (Table 8; Figure S1 in Supporting Information). The methylation status of some alleles in F_1 hybrid was faithfully inherited to the F_2 population. For example, the unmethylated status at site 1 of the F_1 hybrid was replicated in all 20 F_2 individuals. The segregation ratio of four sites (#5, #6, #8 and #9; Table 8) fitted well with the 3:1 ratio expected for single-locus segregation. Nevertheless, the inheritance of some DNA methylation alleles in F_1 hybrids, such as #1, #2, #3, and #4, failed to obey Mendelian rules, indicating a complex inheritance that requires further study. Ten of the 16 sites were successfully sequenced. Four sites were located in protein coding regions, and the other six were in noncoding regions (Table 8).

Table 7 Inheritance of methylation alleles during the selfing process of the 20 hybrid rice parental lines

Type	Average	Maximum	Minimum	Coefficient of variation (%)
Changed sites	31.00	43	20	23.68
Unchanged sites	117.00	128	39	6.27
Ratio of changed sites (%) ^{a)}	20.95	29.05	13.51	30.19

a) Ratio of changed sites (%) = $100 \times \text{changed sites} / (\text{changed sites} + \text{unchanged sites})$.

Table 8 Segregation of methylation alleles in the F_2 generation

Site	Jinke 1A ^{a)}	R8377 ^{a)}	F_1 ^{a)}	F_2			Chromosome	Position	Best match
				Methylated	Unmethylated	Uncertain			
1	+	–	–	0	20	0	11	1100054–1100154	Sodium/hydrogen exchanger family protein
2	+	–	–	3	17	0	11	2827811–2827907	Non-coding region
3	+	–	+	11	9	0	2	30643422–30643519	Mitochondrial import inner membrane translocase
4	+	–	+	16	2	2			
5	+	–	–	5	15	0	7	29725356–29725450	Non-coding region
6	+	–	+	5	15	0	3	19218810–19218912	Non-coding region
7	+	–	–	8	12	0			
8	+	–	+	12	4	4			
9	–	+	–	15	5	0	10	12352012–12352111	Hypothetical protein
10	–	+	+	16	4	0	12	9201421–9201506	Non-coding region
11	–	+	–	8	11	1	2	793701–793789	Non-coding region
12	–	+	–	8	11	1			
13	–	+	–	8	11	1	7	1646702–1646800	Similar to resistance protein candidate
14	–	+	–	10	10	0	1	29779556–29779634	Non-coding region
15	–	+	–	8	11	1			
16	–	+	–	8	11	1			

a) +, methylated; –, unmethylated.

3 Discussion

3.1 MSAP method for DNA methylation detection

Although the MSAP method is inexpensive for genome-wide investigation of DNA methylation status, it has a low resolution, a high false positive rate, and is restricted to methylation at CCGG sites. Recently developed methods, such as bisulfate-based sequencing and MeDIP-seq, have a higher resolution of 1 [24,25] to 35 [26] base pairs. Furthermore, they can be used to profile all forms of single methylation polymorphisms (SMPs). This broad coverage is important because CG-, CHG-, and CHH-SMPs accounted for 23%, 13% and 64% of all SMPs, respectively, in a natural *Arabidopsis* population [27].

It is important to be cautious when interpreting ‘un-methylated’ sites in F_1 hybrids and F_2 segregation populations detected by MSAP (Table 8; Figure S1 in Supporting Information). On the gel image derived from the MSAP, two bands from two digestions (*EcoR* I+*Msp* I and *EcoR* I+*Hpa* II) are present for the unmethylated allele, while one of them is absent for the methylated allele. The unmethylated allele seems to be dominant and masks the methylated allele on the gel image. Therefore, the unmethylated status, judged from the MSAP gel images, can be derived from homozygous or heterozygous unmethylated alleles. This may partially explain the observed methylation loss in F_1 , and the violation of Mendelian rules in the F_2 population (Table 8; Figure S1 in Supporting Information).

3.2 DNA methylation and breeding

Variation is the raw material for breeding selection, and the inheritance of variation ensures that selection occurs. For epigenetic variation that occurs as a result of DNA methylation, the issues of most concern for breeders are whether there is a diversity of DNA methylation among breeding lines and whether this can be inherited in multiple subsequent generations.

The rates of SMP formation are typically orders of magnitude greater than that of spontaneous DNA sequence mutations. For example, 30000 DMPs (DNA methylation polymorphisms) have been reported, but only 30 SNPs [1]. This suggests a more prevalent epigenetic variation in the natural population. A recent study revealed that epigenetic variations are widespread within natural *A. thaliana* lines [27]. DNA methylation polymorphisms have also been observed between wild and cultivated rice [25], and between the *indica* and *japonica* subspecies of rice [23].

In this study, we found that the level of natural variation in DNA methylation in a set of *indica* rice parental lines was greater than that of the variation in DNA sequence. This is consistent with previously reported results [1,22]. The high levels of epigenetic polymorphism in existing rice parental lines may provide new resources for breeding selection, and may help compensate for the gradual decrease

of genetic diversity observed in rice population since the 1970s [28].

An important issue to consider when exploiting methylation diversity for breeding is the correlation between DNA methylation and DNA sequence mutations. If this correlation is too strong, for example, if a methylation variation frequently (or always) occurs when a mutation in the nucleotide sequence occurs, then the methylation would simply accompany the genetic polymorphism, along with any changes in frequency associated with selection pressures on the DNA sequence. This would suggest that specifically selecting for DNA methylation during breeding might not contribute to additional phenotypic diversity, and therefore, might not be beneficial in breeding programs.

Correlations or similar phylogenetic trees between epigenetic and genetic diversities have been observed in watermelon heirlooms [22] and rice [29]. In this study, the correlation between genetic variation and epigenetic (CG DNA methylation) variation was estimated to be 62%. This suggests that some methylation sites do not overlap with sites that carry nucleotide polymorphisms. The selection of these sites might contribute to improvement of rice varieties. It has been reported that maize hybrid vigor, for the agronomic traits of numbers of kernels per row and rows per ear, exhibited correlations with DNA methylation, but not with genetic polymorphism [30]. Furthermore, narrow genetic diversity at the DNA sequence level is reported in watermelon despite broad phenotypic diversity in growth habits, fruit traits, and resistance to biotic and abiotic stresses [31]. This apparent contradiction can be explained by considering DNA methylation as a contributor to phenotypic diversity [22]. Our work, and that of the two previous studies, has shown that there is at least partial independence of DNA methylation from that of genetic polymorphism. This suggests that DNA methylation can contribute to phenotypic diversity and therefore, to breeding.

Given that variation in DNA methylation can be influenced by the environment [32], a concern in rice breeding is whether DNA methylation can be transmitted reliably from one generation to the next. There is a particular concern in the Northern region of China over the inheritance of DNA methylation alleles from parents to their selfing progeny because inbred *japonica* rice varieties are popular in this region. Over one million (<10 %) CG SMPs and 2485 CG-DMRs have been identified in *Arabidopsis* plants propagated by single-seed selfing descent for 30 generations [1,33]. Most (~80%) methylation alleles in this study were estimated to be stable and could be used to generate inbred rice. When the individual breeding histories of watermelon varieties were analyzed, a majority (~90%) methylation alleles were found to be stably inherited [22]. A larger genomic regions of contiguous methylation have been reported to be as stable as DNA sequence mutations in a breeding history study of 30 generations of *A. thaliana* [1]. A further concern for inbred rice breeders is the segregation patterns

of heterozygous methylation alleles. In heterozygous methylation alleles of tomato (*Solanum lycopersicum*) [34] and *Arabidopsis* [35], both consistency and violation of Mendelian rules have been observed; this was also the case in this study (Table 8; Figure S1 in Supporting Information). Therefore, it is not expected that the frequency of heterozygous methylation alleles in the segregation generation will be the same as that of the genetic alleles.

Rice breeding in Southern China mainly makes use of hybrid vigor. In hybrids, interactions between allelic variants are thought to be involved in vigor [36–38]. The recessive allele of *Brassica* *S*-locus protein 11 genes (*SP11*) is methylated *in trans* by RdDM, which is guided by siRNAs produced from the *Smi* gene on the dominant *SP11*-allele, resulting in the expression repression of recessive allele [39]. These alteration modes of DNA methylation and expression have also been documented in a genome-wide study of *Arabidopsis* parental lines and their hybrids [35]. These examples suggest that DNA methylation might be involved in hybrid vigor through the regulation of gene expression or phenotype. We found that more methylation sites in F_1 hybrid rice had a status similar to that in the maternal parent, rather than that in the paternal parent. This suggests that more attention should be paid that in the maternal parent when exploiting epigenetic allelic differences in hybrid rice breeding. A caveat of the current study is that only one hybrid combination was investigated. Other inheritance patterns of DNA methylation alleles might also exist, as in the case of interstrain hybrids of sorghum (*Sorghum bicolor* L.) [40]. Furthermore, there may also be a difference in methylation among maize F_1 individuals [41]. Further studies are required to investigate the uniformity of DNA methylation in rice F_1 populations as genetic instability in the F_1 population would inhibit the commercial use of hybrid rice varieties.

In summary, our study has revealed polymorphic patterns in DNA methylation among breeding lines of Chinese *indica* rice. It showed that these polymorphisms could be inherited in the next generation, in progeny generated by selfing or through hybrid crosses. These findings provide a foundation for the use of DNA methylation alleles in rice breeding programs. As this area is still a virgin territory for rice breeders, it constitutes exciting but challenging possibilities for plant breeders [10].

This work was supported by the National Natural Science Foundation of China (31071379), the Post-Doctoral Foundation of China (20090450616), and a grant from “Yellow Crane” Special Talent Program of Wuhan.

- 1 Becker C, Hagmann J, Muller J, et al. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, 2011, 480: 245–249
- 2 Sha A H, Lin X H, Huang J B, et al. Analysis of DNA methylation related to rice adult plant resistance to bacterial blight based on methylation-sensitive AFLP (MSAP) analysis. *Mol Genet Genomics*,

- 2005, 273: 484–490
- 3 Akimoto K, Katakami H, Kim H J, et al. Epigenetic inheritance in rice plants. *Ann Bot (Lond)*, 2007, 100: 205–217
- 4 Fieldes M A. Heritable effects of 5-azacytidine treatments on the growth and development of flax (*Linum usitatissimum*) genotrophs and genotypes. *Genome*, 1994, 37: 1–11
- 5 Qi X, Li Z H, Jiang L L, et al. Grain-yield heterosis in *Zea mays* L. shows positive correlation with parental difference in CHG methylation. *Crop Sci*, 2010, 50: 2338–2346
- 6 Habu Y, Kakutani T, Paszkowski J. Epigenetic developmental mechanisms in plants: Molecules and targets of plant epigenetic regulation. *Curr Opin Genet Dev*, 2001, 11: 215–220
- 7 Kakutani T. Epi-alleles in plants: Inheritance of epigenetic information over generations. *Plant Cell Physiol*, 2002, 43: 1106–1111
- 8 Guo Z, Zeng L, Li M, et al. Inheritance of DNA methylation in DH and its backcrossed lines of *Brassica napus*. *Afr J Biotechnol*, 2011, 10: 7736–7745
- 9 Teixeira F K, Heredia F, Sarazin A, et al. A role for RNAi in the selective correction of DNA methylation defects. *Science*, 2009, 323: 1600–1604
- 10 Peng H, Zhang J. Plant genomic DNA methylation in response to stresses: Potential applications and challenges in plant breeding. *Prog Nat Sci*, 2009, 14: 265–271
- 11 Xie Q, Peng H. Breeding and utilization of good quality *Indica* CMS line Jinke 1A in rice (in Chinese). *Hybrid Rice*, 2007, 22: 11–13
- 12 Hoagland D R, Arnon D I. The water-culture method for growing plants without soil. Berkeley: College of Agriculture, University of California, 1950
- 13 Keyte A L, Percifield R, Liu B, et al. Intraspecific DNA methylation polymorphism in cotton (*Gossypium hirsutum* L.). *J Hered*, 2006, 97: 444–450
- 14 Ajibade S R, Weeden N F, Chite S M. Inter simple sequence repeat analysis of genetic relationships in the genus *Vigna*. *Euphytica*, 2000, 111: 47–55
- 15 Yeh F, Yang R, Boyle T, et al. POPGEN Ver. 1.32. The user-friendly software for population genetic analysis. Molecular Biology and Bio-technology Center, University of Alberta, Alberta, Canada, 1997
- 16 Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*, 1967, 27: 209–220
- 17 Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, 28: 2731–2739
- 18 McClelland M, Nelson M, Raschke E. Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res*, 1994, 22: 3640–3659
- 19 Xiong L Z, Xu C G, Saghai Maroof M A, et al. Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. *Mol Gen Genet*, 1999, 261: 439–446
- 20 Ashikawa I. Surveying CpG methylation at 5'-CCGG in the genomes of rice cultivars. *Plant Mol Biol*, 2001, 45: 31–39
- 21 Nei M. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA*, 1973, 70: 3321–3323
- 22 Nimmakayala P, Vajja G, Gist R A, et al. Effect of DNA methylation on molecular diversity of watermelon heirlooms and stability of methylation specific polymorphisms across the genealogies. *Euphytica*, 2011, 177: 79–89
- 23 He G, Zhu X, Elling A A, et al. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell*, 2010, 22: 17–33
- 24 Cokus S J, Feng S, Zhang X, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature*, 2008, 452: 215–219
- 25 Li X, Zhu J, Hu F, et al. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics*, 2012, 13: 300

- 26 Zhang X Y, Yazaki J, Sundaresan A, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, 2006, 126: 1189–1201
- 27 Schmitz R J, Schultz M D, Urich M A, et al. Patterns of population epigenomic diversity. *Nature*, 2013, 495: 193–198
- 28 Liu C G, Zhang G Q. SSR analysis of genetic diversity and the temporal trends of major commercial inbred *indica* rice cultivars in south China in 1949–2005 (in Chinese). *Acta Agron Sin*, 2010, 36: 1843–1852
- 29 Takata M, Kishima Y, Sano Y. DNA methylation polymorphisms in rice and wild rice strains: Detection of epigenetic markers. *Breeding Sci*, 2005, 55: 57–63
- 30 Yang W, Yu X, Liu B. Parental epigenetic difference in DNA methylation-level may play contrasting roles for different agronomic traits related to yield heterosis in maize. *Afr J Biotechnol*, 2011, 10: 9253–9263
- 31 Levi A, Thomas C E, Newman M, et al. ISSR and AFLP markers differ among American watermelon cultivars with limited genetic diversity. *J Am Soc Hortic Sci*, 2004, 129: 553–558
- 32 Peng H, Xi T, Zhang J, et al. Stability of stress-Induced DNA methylation in plant (in Chinese). *Acta Agron Sin*, 2011, 44: 2431–2438
- 33 Schmitz R J, Schultz M D, Lewsey M G, et al. Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, 2011, 334: 369–373
- 34 Messeguer R, Ganai M W, Steffens J C, et al. Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. *Plant Mol Biol*, 1991, 16: 753–770
- 35 Greaves I K, Groszmann M, Ying H, et al. Trans chromosomal methylation in *Arabidopsis* hybrids. *Proc Natl Acad Sci USA*, 2012, 109: 3570–3575
- 36 Birchler J A, Auger D L, Riddle N C. In search of the molecular basis of heterosis. *Plant Cell*, 2003, 15: 2236–2239
- 37 Swanson-Wagner R A, Jia Y, DeCook R, et al. All possible modes of gene action are observed in a global comparison of gene expression in a maize F₁ hybrid and its inbred parents. *Proc Natl Acad Sci USA*, 2006, 103: 6805–6810
- 38 Springer N M, Stupar R M. Allelic variation and heterosis in maize: how do two halves make more than a whole? *Genome Res*, 2007, 17: 264–275
- 39 Tarutani Y, Shiba H, Iwano M, et al. Trans-acting small RNA determines dominance relationships in *Brassica* self-incompatibility. *Nature*, 2010, 466: 983–986
- 40 Zhang M S, Yan H Y, Zhao N, et al. Endosperm-specific hypomethylation, and meiotic inheritance and variation of DNA methylation level and pattern in sorghum (*Sorghum bicolor* L.) inter-strain hybrids. *Theor Appl Genet*, 2007, 115: 195–207
- 41 Zhao X, Chai Y, Liu B. Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids. *Plant Sci*, 2007, 172: 930–938

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Supporting Information

Figure S1 The MSAP gel images of the parents, F₁ hybrid and F₂ population. Lane 1/lane 2, lane 3/lane 4, lane 5/lane 6 and lanes 7–46 are from the maternal line Jinke1A, the paternal line R8377, their F₁ hybrid and the 20 F₂ progenies, respectively. Every two lanes represent a sample. The odd and even number are the products digested by *EcoR* I+*Msp* I and *EcoR* I+*Hpa* II, respectively. The numbers on the left of the figure represent the site codes listed in Table 8.

The supporting information is available online at life.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.