

## Genetic structure and eco-geographical differentiation of cultivated *Hsien* rice (*Oryza sativa* L. subsp. *indica*) in China revealed by microsatellites

ZHANG DongLing<sup>1</sup>, ZHANG HongLiang<sup>1\*</sup>, QI YongWen<sup>1</sup>, WANG MeiXing<sup>1</sup>, SUN JunLi<sup>1</sup>, DING Li<sup>2</sup> & LI ZiChao<sup>1\*</sup>

<sup>1</sup>Key Laboratory of Crop Heterosis and Utilization of Ministry of Education, and Beijing Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China;

<sup>2</sup>Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Received June 21, 2012; accepted July 2, 2012; published online October 29, 2012

*Indica* is not only an important rice subspecies widely planted in Asia and the rest of the world, but it is also the genetic background of the majority of hybrid varieties in China. Studies on genetic structure and genetic diversity in *indica* germplasm resources are important for the classification and utilization of cultivated rice in China. Using a genetically representative core collection comprising 1482 Chinese *indica* landraces, we analysed the genetic structure, geographic differentiation and diversity. Model-based structure analysis of varieties within three ecotypes revealed nine eco-geographical types partially accordant with certain ecological zones in China. Differentiation of eco-geographical types was attributed to local ecological adaption and physical isolation. These groups may be useful for developing heterotic groups of *indica*. To facilitate the identification of different ecotypes and eco-geographical types, we identified characteristic SSR alleles of each ecotype and eco-geographical type and a rapid index of discrimination based on characteristic alleles. The characteristic alleles and rapid discrimination index may guide development of heterotic groups, and selection of hybrid parents.

### population structure, *indica*, SSR

**Citation:** Zhang D L, Zhang H L, Qi Y W, et al. Genetic structure and eco-geographical differentiation of cultivated *Hsien* rice (*Oryza sativa* L. Subsp. *indica*) in China revealed by microsatellites. *Chin Sci Bull*, 2013, 58: 344–352, doi: 10.1007/s11434-012-5396-4

Rice is one of the most important crops in the world. China, with its abundant rice genetic resources, is one of the original centers of Asian cultivated rice (*Oryza sativa* L.), and one of the centers of genetic diversity [1] within the species. There were 56020 varieties of cultivated rice *in ex situ* germplasm collections in China until 1993 (ICGR CAAS 1996), and 67% of them were subspecies *indica*. *Indica* accounts for 70% of total Chinese rice area and 80% of the *indica* area is occupied by hybrid varieties. Some studies showed that *indica* has higher genetic diversity than *japonica* [2,3]. *Indica* is the genetic background of the majority of hybrid varieties in China and therefore plays an important

role in rice research and production. However, due to the large area of improved rice varieties derived largely from three genealogies, namely Aizizhan, Nantehao and Peta, the genetic base continues to be narrow.

Several classification systems of cultivated rice have been put forward. In these systems, it is generally believed that the two primary subspecies of cultivated rice, *indica* and *japonica*, were separately domesticated from the ancestral species *O. rufipogon* Griff. Genetic differentiation and classification within subspecies are controversial. Ting put forward a five-level hierarchical taxonomy, consisting of two subspecies, three seasonal or photothermic ecotypes, two soil-watery or soil water regime ecotypes, two endosperm types, and cultivars [4]. Cheng et al. [5] suggested a

\*Corresponding authors (email: lizichao@cau.edu.cn; zhangl@cau.edu.cn)

different five-level hierarchical taxonomy, viz., species, subspecies, ecological populations, ecotypes and cultivars. These taxonomies were primarily based on morphological traits, which were mostly quantitative and controlled by polygenes, and were easily influenced by the environment. Using SSR markers and the primary core collection of rice germplasm resources in China, Zhang et al. [6] divided *indica* into three ecotypes: *Indica* Early (*Ind.E*), *Indica* Late (*Ind.L*) and *Indica* Medium (*Ind.M*). Their results also indicated some subtle sub-structures within the three ecotypes.

Heterosis between *indica* and *japonica* is stronger than that between varieties within each subspecies. However, it is difficult to directly use the heterosis between the subspecies due to various problems, such as partial fertility of the  $F_1$  hybrids [7]. Heterosis within subspecies, especially within *indica*, has been widely used in breeding. In order to improve heterosis within subspecies, heterotic groups have been established to broaden the genetic base, and to offer predictability of superior hybrid combinations. Thus, most heterosis has been achieved from different ecological types within subspecies. Numerous studies on heterotic groups in maize have been reported [8,9], but there are few reports from rice [10,11]. Determination of genetic structures within subspecies is important for studying evolution and differentiation in subspecies, and can provide evidence for heterotic groups within rice.

Molecular markers have been widely applied to investigate the genetic structure, evolution and genetic diversity of rice germplasm resources [6,12–15]. It is feasible to study the genetic structure and ecotypic classification within *indica*. In the present study, we examined the genetic variation among 1482 *indica* varieties from the primary core collection of Chinese cultivated rice using 36 microsatellite markers located on 12 chromosomes. Our objectives were to (1) investigate the genetic structure and diversity within *indica* in China, and (2) examine the geographical differentiation and classification within the *indica* subspecies. The results will help management of germplasm collections in rice breeding programs and in heterotic group research.

## 1 Materials and methods

### 1.1 Plant materials

The 1482 varieties used in this study were *indica* landraces from the primary core collection of Chinese rice landraces. They were sampled from 50526 landrace accessions collected from all 32 provinces or cities of China and represents 95% of the total diversity [16].

### 1.2 DNA extraction and SSR analysis

Fresh leaves of 30-d seedlings were collected and ground in liquid nitrogen; DNA was extracted from the ground mate-

rial using the CTAB method [17]. Thirty-six SSRs (3 on each of the 12 chromosomes) were randomly selected to analyze the population structure and genetic diversity. The volume of the PCR system was 15  $\mu$ L, containing 50 ng DNA template. The PCR procedure was as follows: initial denaturing at 95°C for 5 min; 30 cycles with three steps: denaturing at 94°C for 1 min, annealing at 55–67°C (depending on the primer sequence) for 0.5 min, extension at 72°C for 1 min, and a final extension at 72°C for 10 min. The amplified products were denatured at 95°C for 5 min, cooled on ice, and subsequently run on 8% denatured polyacrylamide gels at 70 W. Gels were visualized by silver-staining [18]. The molecular weights of all allelic variants were estimated by a gel-run including all identified alleles and standard molecular weight markers, 10-bp DNA Ladder (Invitrogen, Carlsbad, California).

### 1.3 Statistical analyses

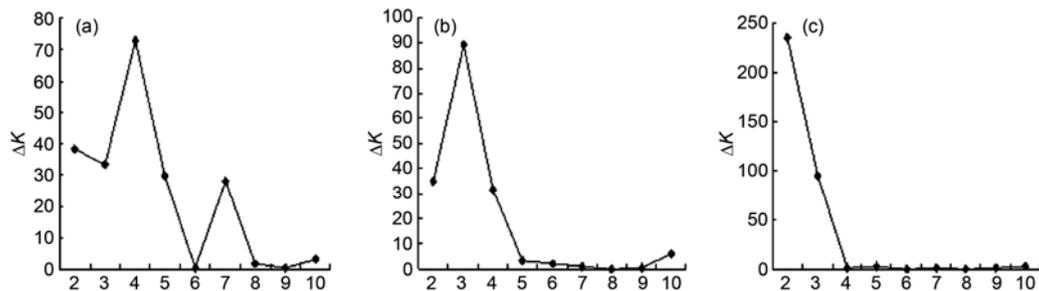
Nei's gene diversity index ( $H_e=1-\sum p_i^2$ ) [19] was calculated. The model-based program STRUCTURE was used to infer population structure [20,21]. Five independent simulations were run for each  $K$ . Phylogenetic reconstruction was based on the neighbor-joining method from the distance matrix using Nei's  $D_A$  [22], implemented in PowerMarker version 3.25 [23]. Average standardized molecular weights of the PCR products in each population were calculated as in Vigouroux et al. [24]. For examining population differentiation,  $G_{st}$  ( $G_{st}=D_{st}/H_T$ ) [25] among populations was calculated and tested. SPSS11.0 was used to carry out correlation analyses and tests of significance on differences.

## 2 Results

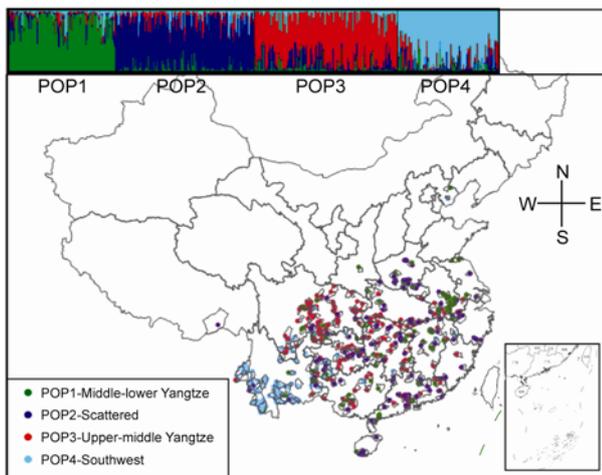
### 2.1 Population structure of *indica*

In our earlier study of population structure in rice landraces, the STRUCTURE simulation program divided the *indica* population into 3 ecotype groups, viz. early (*Ind.E*), late (*Ind.L*) and intermediate (*Ind.M*) ecotypes [6]. However, the gradually increasing  $\text{Ln}P(D)$  values after  $K=3$  indicated subtle sub-structures in the 3 ecotypes [6]. In the present study, we conducted further independent simulations within each of the ecotypes to examine the possibility of more subtle sub-structures. Evanno's  $\Delta K$  was used to determine the number of clusters [6,26]. Distinct peaks of the  $\Delta K$  value were observed at  $K=4$  in *Ind.E*, at  $K=2$  in *Ind.L* and at  $K=3$  in *Ind.M* (Figure 1). Comparison of the predefined and inferred populations revealed differentiation of different numbers of eco-geographical types within each *indica* ecotype (Figures 2–4).

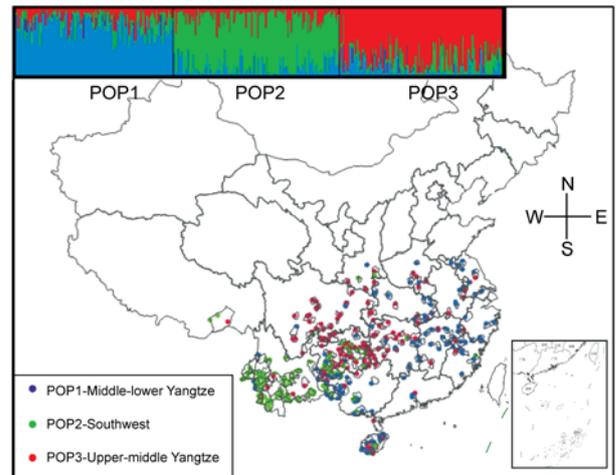
Among the four eco-geographical types of *Ind.E* (Figure 2), POP1 was mainly from the middle-lower Yangtze, POP3 from upper-middle Yangtze, POP4 from southwest China and POP2 was scattered. We named POP1 as the Middle-



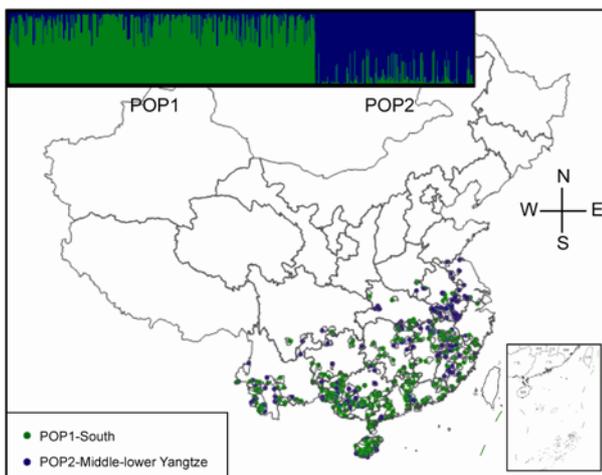
**Figure 1**  $\Delta K$  with  $K=2$  to 10 for three inferred *indica* ecotypes. (a) *Ind.E*; (b) *Ind.L*; (c) *Ind.M*.



**Figure 2** Inferred eco-geographical types within *Ind.E* and their geographical distributions.



**Figure 4** Inferred eco-geographical types within *Ind.M* and their geographical distributions.



**Figure 3** Inferred eco-geographical types within *Ind.L* and their geographical distributions.

lower Yangtze type, POP2 as the Scattered type, POP3 as the Upper-middle Yangtze type and POP4 the Southwest type, respectively.

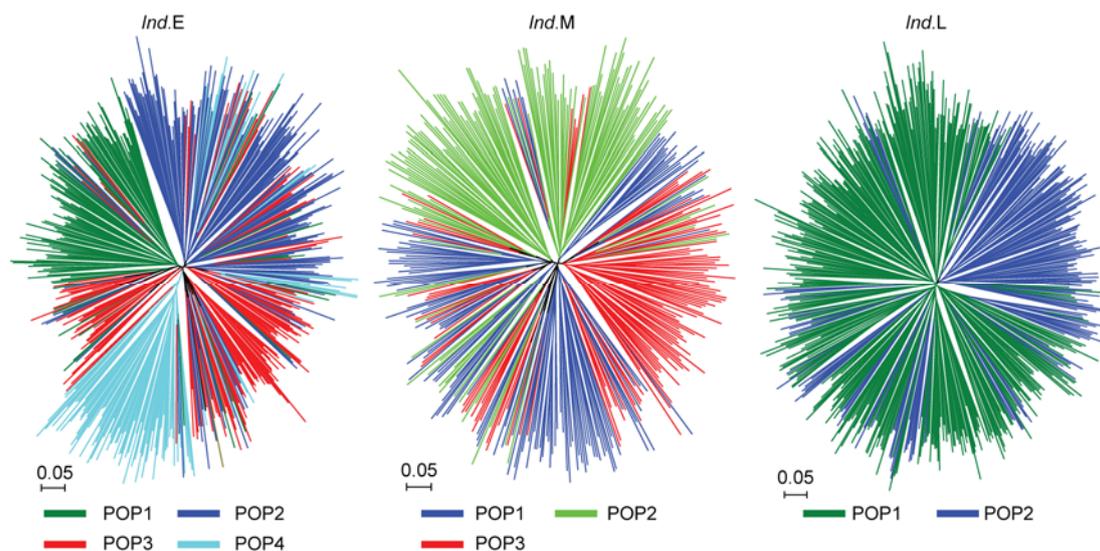
Among the two eco-geographical types in *Ind.L* (Figure 3), POP1 was from south China, and POP2 was from the middle-lower Yangtze. We named POP1 as South type, and POP2 as the middle-lower Yangtze type.

Among 3 eco-geographical types in *Ind.M* (Figure 4), POP1 was from the middle-lower Yangtze, POP2 was from southwest China, and POP3 was from the upper-middle Yangtze. We named POP1 as the Middle-lower Yangtze type, POP2 as the Southwest type and POP3 as the Upper-middle Yangtze type.

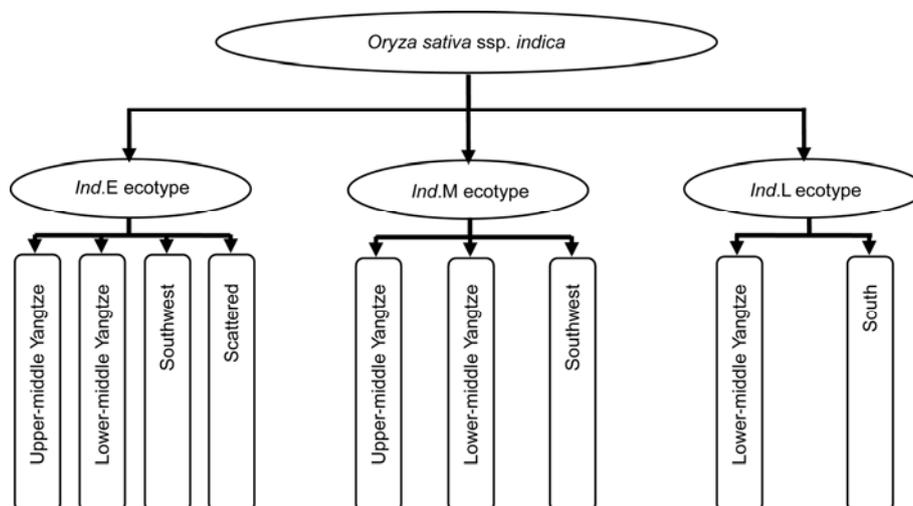
The distance-based neighbor-joining trees within the three ecotypes depicted different eco-geographical types that were almost in accord with those in the STRUCTURE simulation (Figure 5). Both methods indicated that there was not only clear differentiation of ecotypes within *indica* but also clear differentiation of eco-geographical types within ecotypes. We therefore suggested the taxonomy depicted in Figure 6; that is, *indica* firstly differentiated into three ecotypes and further differentiated into eco-geo-graphical types.

## 2.2 Genetic differentiation and diversity of different eco-geographical types

$F_{st}$  and Nei's genetic distances among the nine model-based populations showed that differentiation among eco-geo-graphical types varied significantly (Table 1). As a whole, *Ind.M* represented a larger genetic distance than *Ind.E* or *Ind.L*. Obviously, the genetic distance among eco-geo-graphical types within the same ecotype was smaller



**Figure 5** Unrooted neighbor-joining tree of three *indica* ecotypes based on Nei's  $D_A$  estimated by 36 nuclear SSRs. The colors correspond to model-based populations in STRUCTURE analysis.



**Figure 6** Proposed taxonomic system for *indica* rice in China.

than that among eco-geographical types of different ecotypes. Within the *Ind.E* ecotype, POP2 (the scattered type) had a closer genetic distance than the other three eco-geographical types. AMOVA among eco-geographical types within each ecotype indicated that variation attributed to differentiation among eco-geographical types in *Ind.E* was the highest (5.77%), and that in *Ind.L* was the lowest (1.87%). The average variation (3.81%) attributed to differentiation among eco-geographical types within ecotypes accounted for about half of that attributed to differentiation among ecotypes (7.58%) [6].

The genetic diversity in each eco-geographical type was measured by four estimators, viz. allele number ( $N_a$ ), allelic richness ( $R_s$ ), genetic diversity ( $H_e$ ) and polymorphic information content ( $PIC$ ) (Table 2). Within the *Ind.E* ecotype, all estimators except  $N_a$  for the southwest eco-geographical

type were the highest; within the *Ind.L* ecotype, all estimators of the south eco-geographical type were the highest; and within the *Ind.M* ecotype, all estimators of the south eco-geographical type were the highest. We calculated the average standardized molecular weight ( $SMW$ ) in each eco-geographical type to investigate the directional evolution (Table 2). The *Ind.M* ecotype was significantly smaller in molecular weight than the others. Among all 9 eco-geographical types, the Upper-middle Yangtze and Southwest types in the *Ind.M* ecotype were significantly smaller in molecular weight than the others.

Regressions between genetic estimators (number of alleles, gene diversity and  $F_{st}$ ) and geographic estimators (latitude, difference in latitude and difference in geographic distance) showed that genetic diversity declined with increasing latitude, and genetic differentiation became increasingly

**Table 1** Nei's standard distance and Pairwise  $F_{st}$  among the model-based populations within each *indica* ecotype. Nei's standard distances are below the diagonal and pairwise  $F_{st}$  values are above the diagonal

Ecotype	Eco-geographical type	<i>Ind.E</i>				<i>Ind.M</i>			<i>Ind.L</i>	
		POP1	POP2	POP3	POP4	POP1	POP2	POP3	POP1	POP2
<i>Ind.E</i>	POP1		0.0693	0.0549	0.0784	0.1429	0.1395	0.1194	0.1409	0.0739
	POP2	0.0820		0.0336	0.0767	0.1077	0.1153	0.0960	0.0650	0.0704
	POP3	0.0865	0.0590		0.0797	0.1423	0.1335	0.0963	0.1209	0.1171
	POP4	0.1396	0.0718	0.1062		0.1262	0.0784	0.1250	0.1135	0.1149
<i>Ind.M</i>	POP1	0.1945	0.1598	0.1982	0.1906		0.0433	0.0455	0.0937	0.0980
	POP2	0.1983	0.1652	0.1895	0.1396	0.0838		0.0589	0.1068	0.1123
	POP3	0.1731	0.1370	0.1333	0.1797	0.0782	0.0906		0.1392	0.1367
<i>Ind.L</i>	POP1	0.1314	0.1058	0.1082	0.1136	0.1202	0.1488	0.1552		0.0388
	POP2	0.0976	0.1111	0.1504	0.1479	0.1483	0.1708	0.1960	0.0672	

**Table 2** Genetic diversity and molecular weight of inferred populations<sup>a)</sup>

Ecotype	Eco-geographical type	$N_a$	$R_s$	$H_e$	PIC	SMW
<i>Ind.E</i>	Lower-middle Yangtze	7.9	8.24	0.5733	0.5399	0.4074
	Scattered	9.8	8.18	0.6158	0.5851	-0.1524
	Upper-middle Yangtze	8.7	8.36	0.5520	0.5175	0.1478
	Southwest	8.8	8.36	0.6247	0.5925	-0.2013
	Whole ecotype	12	8.38	0.6249	0.5941	0.4306
<i>Ind.M</i>	Lower-middle Yangtze	9.7	8.18	0.6495	0.6175	-0.3144
	Southwest	10.3	8.28	0.6690	0.6409	-0.5513
	Upper-middle Yangtze	9.2	8.26	0.5978	0.5675	-0.6004
	Whole ecotype	12	8.34	0.6637	0.6358	-0.0645
<i>Ind.L</i>	South	11.8	8.4	0.6702	0.6377	0.2089
	Lower-middle Yangtze	10.1	8.21	0.6643	0.6326	0.4837
	Whole ecotype	12.6	8.37	0.6808	0.6497	0.6653

a)  $N_a$  = number of alleles;  $R_s$  = allele richness;  $H_e$  = Nei's unbiased gene diversity; PIC = polymorphic information coefficient; SMW = average standardized molecular weight.

distinct with increasing geographic distance (Figure 7). Localized eco-geographical types and distinct regressions between genetic estimators and geographic estimators implied that selection within the local ecological environment and spatial isolation were two primary factors driving genetic differentiation within *indica*.

### 2.3 Molecular characteristics and molecular discrimination between ecotypes and eco-geographical types

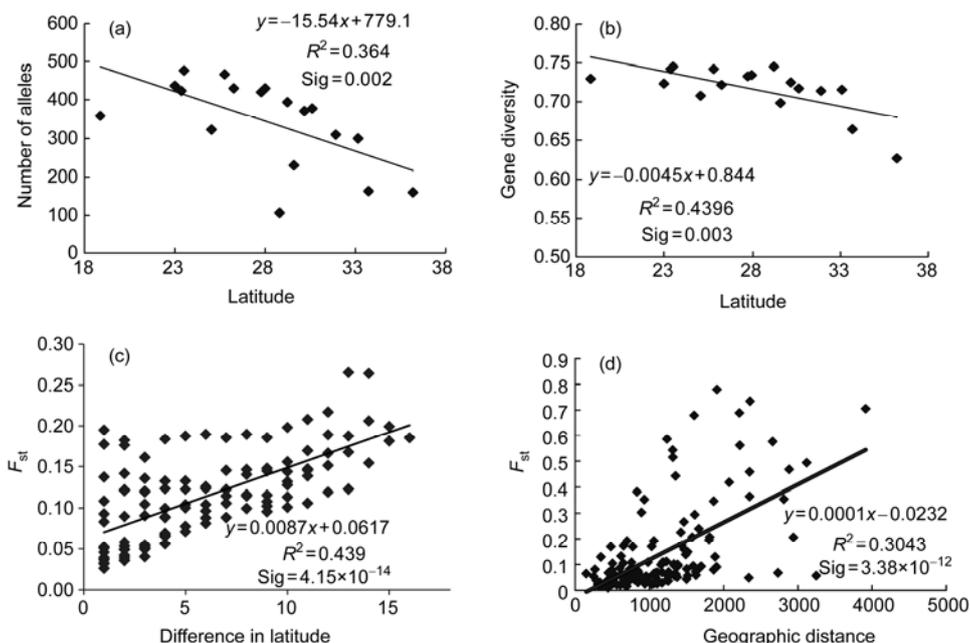
In order to identify the ecotype and eco-geographical type for individual accession among the 1482 varieties included in the current work, the molecular characteristics of the three ecotypes and nine eco-geographical types were assayed and their molecular discriminants were developed. 'Typical varieties' ( $Q \geq 0.95$ , i.e. the probability that the inferred ancestry derived from one of the model-based populations) were selected from each ecotype and eco-geographical type. Evidently, alleles were not evenly distributed among ecotypes and eco-geographical types. For comparison, we re-

garded the allele with a frequency higher than 50% in one ecotype or eco-geographical type as the characteristic allele for that ecotype or eco-geographical type (Figure 8).

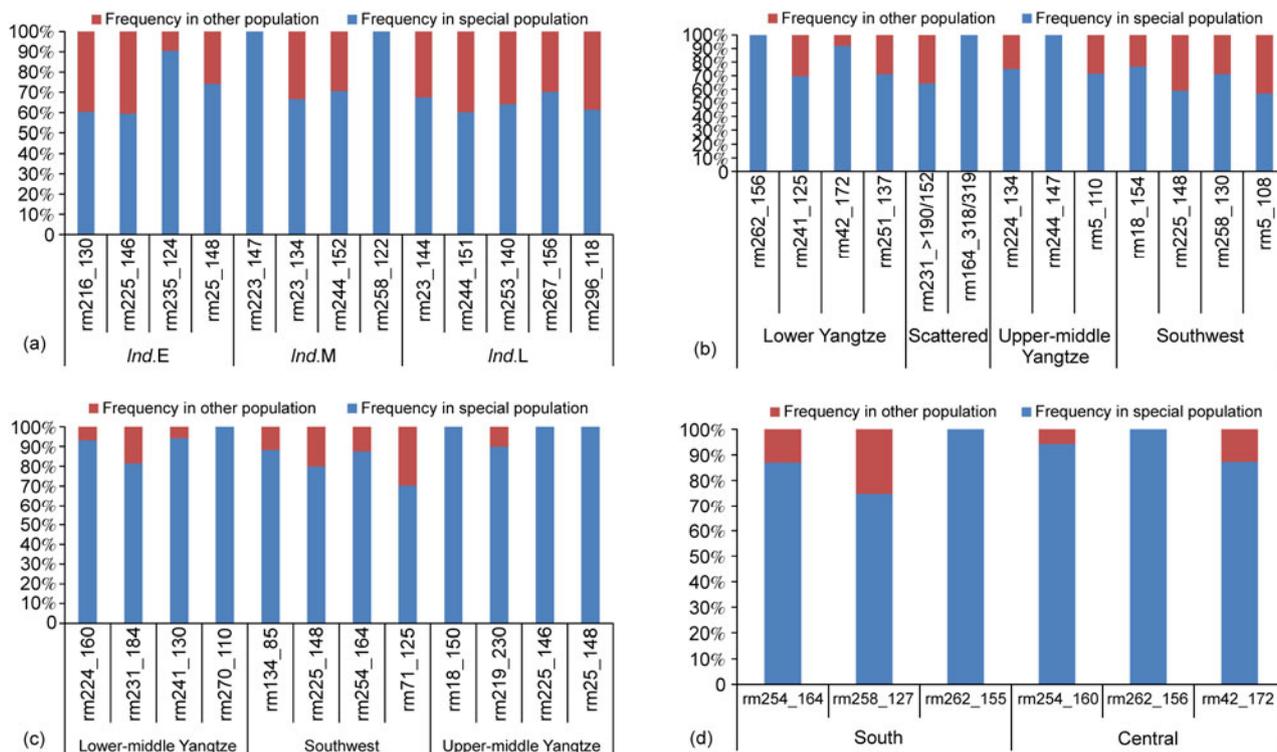
We established a molecular discriminant based on characteristic alleles to rapidly identify varieties belonging to each ecotype and eco-geographical type. The characteristic alleles of the three ecotypes in *indica* are listed in Figure 8a. The variety identity (*VI*) belonging to each ecotype of *indica* was determined by the discriminant index (*DI*). According to the macro function in Microsoft Excel, the relationship between *VI* and *DI* could be expressed by the following discriminant function:

$$VI(A, B, C) = \text{INDEX}(\{"A", "B", "C"\}, 0, \text{MATCH}(\text{MAX}(a, b, c), \{a, b, c\}, 0)),$$

where the format of the function INDEX() is INDEX (array, row\_num, [column\_num]), which returns the element of the given position in the given array. A, B and C represent *Ind.E*, *Ind.M* and *Ind.L*, respectively, and *a*, *b* and *c* are the discriminant indices of the three ecotypes, respectively. The



**Figure 7** Regressions between genetic estimators (number of alleles, gene diversity and  $F_{st}$ ) and geographic estimators (latitude, differences in latitude and differences in geographic distance).



**Figure 8** Characteristic alleles of 3 ecotypes and nine eco-geographical types. (a) The three primary ecotypes; (b) 4 eco-geographical types within *Ind.E*; (c) three eco-geographical types within *Ind.M*; (d) two eco-geographical types within *Ind.L*.

format of the function MATCH() is MATCH(lookup\_value, lookup\_array, [match\_type]), which returns the position of the element in the lookup\_array which matches exactly with the lookup\_value (when the match\_type is 0). Function MAX(array) returns the maximum value in the given array.

That is, if  $a > b$  and  $c$ , the variety belongs to the ecotype *Ind.E*; if  $b > a$  and  $c$ , the variety belongs to the ecotype *Ind.M*; if  $c > a$  and  $b$ , the variety belongs to the ecotype *Ind.L*. The discriminant indexes ( $a$ ,  $b$  and  $c$ ) are calculated as follows:

$$a = (0.6 \times rm216\_130 + 0.59 \times rm225\_146 + 0.91 \times rm235\_124$$

$$+0.74 \times \text{rm25\_148})/2.84,$$

$$b=(1 \times \text{rm223\_147}+0.67 \times \text{rm23\_134}+0.51 \times \text{rm244\_152}+1 \times \text{rm258\_122})/3.18,$$

$$c=(0.6 \times \text{rm244\_151}+0.64 \times \text{rm253\_140}+0.7 \times \text{rm267\_156}+0.61 \times \text{rm296\_118}+0.67 \times \text{rm23\_144})/3.22.$$

Taking the formula of discriminant index  $a$  as an example, where  $\text{rm216\_130}$ ,  $\text{rm225\_146}$ ,  $\text{rm235\_124}$  and  $\text{rm25\_148}$  are characteristic alleles of SSR markers,  $\text{rm216}$ ,  $\text{rm225}$ ,  $\text{rm235}$  and  $\text{rm25}$  are SSR markers, and the number following the underline is the molecular weight of the characteristic allele. If the molecular weight was equal to that of the characteristic allele, the characteristic allele in the formula was given the value 1; otherwise it was given the value 0. The number before the multiplication sign is the ratio of the frequency of the corresponding characteristic allele in a specific ecotype to that in the whole population. For example, 0.6 is the ratio of the frequency of  $\text{rm216\_130}$  in *Ind.E* and its frequency in the whole population. The number following the division sign was the sum of the frequency ratios of all characteristic alleles. For example, 2.84 was the sum of 0.6, 0.59, 0.91 and 0.74. Discriminant indices  $b$  and  $c$  are estimated similarly as  $a$ .

The identities of 431 varieties were discriminated using these formulas. The agreements between formula-based identity by discriminant index and model-based identity by STRUCTURE were higher than 82% (Table 3).

The characteristic alleles and discriminant indices of different eco-geographical types were developed in a way similar to three *indica* ecotypes (Figure 8(b)–(d), and Table 4). The agreements of 8 eco-geographical types, except for the scattered type, between formula-based identity by discriminant index and model-based identity by STRUCTURE were higher than 70% (Table 4).

### 3 Discussion

#### 3.1 Population structure and geographical differentiation within *indica*

It is generally accepted that there are two subspecies in Asian cultivated rice (*Oryza sativa* L.), viz. *indica* and *japonica*. This was confirmed by evidence from biochemical data, hybrid sterility and molecular analyses [6,27,28]. However, population structure and classification within subspecies have been debated over a long period. As described in the introduction, Ting put forward a five-level hierarchical taxonomy [4]. Qian et al [29] insisted that Chinese *indica* should be divided into early rice, late rice of central China and late rice of south China. Cheng et al [5] suggested that Chinese *indica* should be divided into early-middle rice, late rice and winter rice according to the season of cultivation, cultivation habit and ecological distribution. Our current results revealed a hierarchical genetic structure and differentiation within *indica*, which was different from both Ting and Cheng et al.

Subspecies *indica* obviously can be divided into different seasonal ecotypes, especially different photosensitivity groups.

**Table 3** Agreement between formula-based identity by discriminant index and model-based identity by STRUCTURE of three ecotypes

	<i>Ind.E</i> <sup>a)</sup>	<i>Ind.M</i> <sup>a)</sup>	<i>Ind.L</i> <sup>a)</sup>	Total
<i>Ind.E</i> <sup>b)</sup>	131	13	3	147
<i>Ind.M</i> <sup>b)</sup>	3	82	1	86
<i>Ind.L</i> <sup>b)</sup>	25	4	169	198
Total	159	99	173	
Agreement (%)	82.39	82.83	97.69	

a) by discriminant index; b) by STRUCTURE.

**Table 4** Discriminant indices and agreement between formula-based identity by discriminant index and model-based identity by STRUCTURE for nine eco-geographical types

Ecotype	Ecotypic discriminant <sup>a)</sup>	Eco-geographical type	Discriminant index	Agreement (%)
<i>Ind. E</i>	VI(D, E, F, G)	D: Lower-middle Yangtze	$d=(1.0 \times \text{rm262\_156}+0.7 \times \text{rm241\_125}+0.92 \times \text{rm42\_172}+0.71 \times \text{rm251\_137})/3.33$	100.00
		E: Scatted	$e=(0.64 \times \text{rm231\_}>190/152+1.0 \times \text{rm164\_318}/319)/1.64$	58.33
		F: Upper-middle Yangtze	$f=(0.75 \times \text{rm224\_134}+1.0 \times \text{rm244\_147}+0.71 \times \text{rm5\_110})/2.46$	100.00
		G: Southwest	$g=(0.77 \times \text{rm18\_154}+0.59 \times \text{rm225\_148}+0.71 \times \text{rm258\_130}+0.57 \times \text{rm5\_108})/2.64$	100.00
<i>Ind. M</i>	VI(H, I, J)	H: Lower-middle Yangtze	$h=(0.93 \times \text{rm224\_160}+0.8 \times \text{rm231\_184}+0.94 \times \text{rm241\_130}+1.0 \times \text{rm270\_110})/3.67$	72.41
		I: Southwest	$i=(0.88 \times \text{rm134\_85}+0.8 \times \text{rm225\_148}+0.7 \times \text{rm254\_164}+1.0 \times \text{rm71\_125})/3.38$	100.00
		J: Upper-middle Yangtze	$j=(1.0 \times \text{rm18\_150}+0.9 \times \text{rm219\_230}+1.0 \times \text{rm225\_146}+1.0 \times \text{rm25\_148})/3.9$	92.86
<i>Ind. L</i>	VI(K, L)	K: South	$k=(0.87 \times \text{rm254\_164}+0.75 \times \text{rm258\_127}+1.0 \times \text{rm262\_155})/2.62$	87.88
		L: Lower-middle Yangtze	$l=(0.94 \times \text{rm254\_160}+1.0 \times \text{rm262\_156}+0.87 \times \text{rm42\_172})/2.81$	89.47

a)  $VI(D,E,F,G)=INDEX(\{“D”,“E”,“F”,“G”\},0,MATCH(MAX(d,e,f,g),\{d,e,f,g\},0))$ ;  $VI(H,I,J)=INDEX(\{“H”,“I”,“J”\},0,MATCH(MAX(h,i,j),\{h,i,j\},0))$ ;  $VI(K,L)=INDEX(\{“K”,“L”\},0,MATCH(MAX(k,l),\{k,l\},0))$ .

However, there was a medium ecotype, which differed from the early and late ecotypes in genetic diversity, molecular weight, and allelic frequency, and had a smaller genetic distance relative to *japonica* [6]. This medium ecotype is mainly distributed in southwest China and the Yangtze valley. Similar variation also occurs in subspecies *japonica*, which is mainly planted in the north of this region, whereas *indica* is mainly planted in the south. Thus some *indica* and *japonica* varieties can grow normally, although not optimally, in intermediate environments. Some researchers have attributed the low differentiation of varieties in this region to genetic introgression between *indica* and *japonica* [30]. However, it is impossible for genetic introgression to have generated the different ecotypic structures in the two subspecies.

Within each *indica* ecotype, STRUCTURE simulation indicated the presence of more subtle sub-structures. *Indica* ecotypes did not differentiate into distinct soil-watery ecotypes like *japonica* as Ting proposed [6], but showed different degrees of geographical differentiation. The distribution of eco-geographical types partly accorded with the cropping regionalization pattern of rice in China [30]. Southwest types in *Ind.E* and *Ind.M* were mainly distributed in ecological sub-zone I<sub>2</sub>, with single season cropping in river valleys and basins of South Yunnan, where the cumulative temperature ( $\geq 10^{\circ}\text{C}$ ) is 5800–7000 $^{\circ}\text{C}$  and total sunlight in the cropping season is 1000–1300 h; Middle-lower Yangtze types in *Ind.E* and *Ind.L* were mainly distributed in ecological sub-zone II<sub>1</sub>, double or single cropping region of the middle-lower Yangtze River, where the cumulative temperature ( $\geq 10^{\circ}\text{C}$ ) is 4500–5500 $^{\circ}\text{C}$  and total sunlight is 1300–1500 h; the Middle-lower Yangtze type in *Ind.M*, however, was distributed in ecological sub-zones II<sub>1</sub> and II<sub>3</sub>. The latter is a double cropping system in the hills and plains south of the Yangtze River, where the cumulative temperature ( $\geq 10^{\circ}\text{C}$ ) is 5300–6500 $^{\circ}\text{C}$  and the total sunlight length is 1200–1400 h; The Upper-middle Yangtze type in *Ind.M* mainly occurred in ecological sub-zones II<sub>2</sub> and III<sub>2</sub>, single season, but also twice sown on ridges and valleys of the Dian-Chuan plateau and in Chuan-Shan basin. Regionalization of rice cropping areas in China is according to local ecological environments, but including aspects of daylength, sunlight hours, temperature and rainfall [31]. It seems that local ecological conditions and the cropping system play important roles in the formation of eco-geographical types within each *indica* ecotype. In addition, the positive correlation between region and genetic differentiation implied that spatial isolation was also an important factor. In summary, selection in local ecological environments and spatial isolation were the primary factors causing genetic differentiation within *indica* rice.

### 3.2 Potential utilization of ecotypes and eco-geographical types within *indica* in heterosis

Over the last 20 years, it has been increasingly difficult to

develop hybrids with strong heterosis due to difficulties of finding parents with optimal genetic differences. To facilitate the screening of potential parents with optimal genetic differences, the concept of heterotic groups was put forward in maize [32,33]. In rice, Chen et al. [9] investigated the patterns of heterosis between early/middle/late ecotypes of *indica* and early/late *japonica*. Sun et al. [34] identified valuable heterotic groups by large-scale cross analysis. Liu et al. [35] showed that stronger heterosis could be obtained in crosses between ecotypes from different geographic origins. However, neither genetic differences nor geographic differences *per se* were correlated with heterotic magnitude. Xiao et al. [36] indicated that genetic distance based on RAPD and SSR markers could predict heterosis of hybrids within subspecies, but not hybrids between subspecies. Smith and Smith reported a weak correlation between genetic distance and heterosis in maize [37]. The hierarchical classification put forward in the present study may provide a scientific framework beneficial to the prediction and utilization of heterosis in rice.

We not only give a hierarchical classification that can reveal differentiation in both DNA and ecological adaptation, but also provide a discriminant method for identifying different ecotypes and eco-geographical types using SSR markers. According to Messmer et al. it is difficult to classify varieties as types being potentially valuable in hybrid development by geographic origin or endosperm type alone, because many varieties were mixed in origin, or were introgressed between each other [38]. In addition, pedigree relationships revealed by DNA markers could only predict the heterosis of hybrids developed by parents with certain levels of differentiation [36–40]. The given hierarchical classification reveals the differentiation at different levels (ecotypes and eco-geographical types) within *indica*, and may provide a framework to test heterosis at different levels. The most important is that, rather than just giving a framework of classification, we provide a method to identify these types using simple DNA markers. Thus, by using molecular data to search for heterotic groups on smaller numbers of potential heterotic patterns and to evaluate them intensively, breeders could have a more economic and solid approach to making decisions at the beginning of hybrid breeding programs.

The medium ecotype could provide ideal varieties as parents for utilization of heterosis between *indica* and *japonica* and for increasing heterosis within *japonica*. The yield superiority of *japonica* hybrid rice in China is only about 10%, significantly lower than that of *indica* hybrids [41]. In order to utilize restorer genes and elite genes in *indica* (e.g. eurytopic genes, resistance genes), the breeder usually needs to identify intermediate varieties as a bridge for crossing between *indica* and *japonica*. The most important criterion for a bridge variety is a moderate genetic background. If the background of *indica* is low, the purpose of widening genetic differences and increasing heterosis cannot be achieved; if the background is high, it is difficult

to develop varieties with adaptation to the ecological environments of *japonica*. In our study, there is a medium ecotype (*Ind.M*), with obvious differences from early and late ecotypes of *indica* in genetic diversity, molecular weight, and allele frequency, and they had a smaller genetic distance from *japonica* [6]. Moreover, this medium ecotype is distributed from the southwest to the middle Yangtze valley, where both *indica* and *japonica* are planted. Evidently, the medium ecotype is intermediate in both genetic composition and adaptability to the ecological environment. They may be preferred candidates for utilization of heterosis between the two subspecies. To facilitate identification of medium ecotypes, we also identified characteristic SSR alleles and a rapid discrimination method based on the characteristic alleles.

We thank Professor Robert A McIntosh, University of Sydney, for suggested revisions to the manuscript. This work was supported by the National Basic Research Program of China (2010CB125904, 2004CB117201), and the National Natural Science Foundation of China (30871506, 31171613).

- 1 Oka H I. Origin of Cultivated Rice. Tokyo: Japan Science Society Press, 1988
- 2 Sun X L, Cai H W, Wang X K. Diversity and nonrandom association of rice isozyme genes. *Acta Gent Sin*, 1996, 23: 276–285
- 3 Wei X H, Tang S X, Jiang Y Z, et al. Genetic diversity of allozymes associated with morphological traits in Chinese improved rice varieties. *Chin J of Rice Sci*, 2003, 17: 123–128
- 4 Ting Y. The origin and evolution of cultivated rice in China. *Acta Agron Sin*, 1957, 8: 243–260
- 5 Cheng K S, Zhou J W, Lu Y X, et al. Study on the indigenous rices in Yunnan and their utilization II. A revised classification of Asian cultivated rice. *Acta Agron Sin*, 1984, 10: 271–280
- 6 Zhang D L, Zhang H L, Wang M X, et al. Genetic structure and differentiation of landrace rice (*Oryza sativa* L.) in China revealed by microsatellites. *Theor Appl Genet*, 2009, 119: 1105–1117
- 7 Zou J S, Lv C G. Practice and thinking on rice breeding for high yield. *Acta Agron Sin*, 2005, 31: 254–258
- 8 Stuber C W, Lincoln S E, Wolff D W, et al. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics*, 1992, 132: 823–839
- 9 Melchinger A E, Lee M, Lamkey K R, et al. Genetic diversity for restriction length polymorphisms and heterosis for two diallele sets of maize inbreds. *Theor Appl Genet*, 1990, 80: 488–496
- 10 Chen L Y, Dai K G, Li G T, et al. A comparative study of different categories of F<sub>1</sub> hybrids between *indica* and *japonica*. *Hybrid Rice*, 1992, 4: 35–38
- 11 Sun C Q, Jiang T B, Chen L, et al. Studies on the relationship between heterosis and genetic differentiation in hybrid rice (*Oryza sativa* L.). *Acta Agronom Sin*, 2000, 26: 641–649
- 12 Garris A J, Tai T H, Coburn J, et al. Genetic structure and diversity in *Oryza sativa* L. *Genetics*, 2005, 169: 1631–1638
- 13 Wang M X, Zhang H L, Zhang D L, et al. Genetic structure of *Oryza rufipogon* Griff. in China. *Heredity*, 2008, 101: 527–535
- 14 Wang M X, Zhang H L, Zhang D L, et al. Geographical genetic diversity and divergence of common wild rice (*O. rufipogon* Griff.) in China. *Chin Sci Bull*, 2008, 53: 3559–3566
- 15 Zhang D L, Zhang H L, Wei X H, et al. Genetic structure and diversity of *Oryza sativa* L. in Guizhou, China. *Chin Sci Bull*, 2007, 52: 343–351
- 16 Li Z C, Zhang H L, Cao Y S, et al. Studies on the sampling strategy for primary core collection of Chinese indigenous rice. *Acta Agronom Sin*, 2003, 29: 20–24
- 17 Saghai-Marouf M A, Soliman K M, Jorgensen R A, et al. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci USA*, 1984, 81: 8014–8018
- 18 Panaud O, Chen X L, McCouch S R. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet*, 1996, 252: 597–607
- 19 Nei M. Analysis of diversity in subdivided populations. *Proc Natl Acad Sci USA*, 1973, 70: 3321–3323
- 20 Pritchard J K, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*, 2000, 155: 945–959
- 21 Falush D, Stephens M, Pritchard J K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 2003, 164: 1567–1587
- 22 Nei M, Tajima F, Tatenos Y. Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol*, 1983, 19: 153–170
- 23 Liu K, Muse S. PowerMarker: New Genetic Data Analysis Software, Version 2.7 (<http://www.powermarker.net>). 2004
- 24 Vigouroux Y, Matsuoka Y, Doebley J. Directional evolutionary for microsatellite size in maize. *Mol Biol Evol*, 2003, 20: 1480–1483
- 25 Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 1978, 89: 583–590
- 26 Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software Structure: A simulation study. *Mol Ecol*, 2005, 14: 2611–2620
- 27 Chang T T. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica*, 1976, 25: 425–441
- 28 Morishima H, Sano Y, Oka H I. Evolutionary studies in cultivated rice and its wild relatives. *Oxf Surv Evol Biol*, 1992, 8: 135–184
- 29 Qian Y W, Liu J Z, He K M. Analysis of genetic relationships of rice by esterase isozymes. *Guangdong Agri Sci*, 1983, 6: 12–14
- 30 Mei F Q, Wu X Z, Yao C X, et al. Rice cropping regionalization in China. *Chin J Rice Sci*, 1988, 2: 97–110
- 31 Zhao K Y, Wright M, Kimball J, et al. Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS One*, 2010, 5: 1–11
- 32 Stuber C W, Lincoln S E, Wolff D W, et al. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics*, 1992, 132: 823–839
- 33 Melchinger A E, Lee M, Lamkey K R, et al. Genetic diversity for restriction length polymorphisms and heterosis for two diallele sets of maize inbreds. *Theor Appl Genet*, 1990, 80: 488–496
- 34 Sun C Q, Chen L, Li Z C, et al. Studies on heterotic ecotypes for two line hybrid rice. *Hybrid rice*, 1999, 14: 34–38
- 35 Liu W, Shi Y L, Ma H W, et al. Study on heterotic ecotypes of *japonica* rice based on the heterosis. *Acta Bot Boreal Occident Sin*, 2005, 25: 64–69
- 36 Xiao J, Li J, Yuan L, et al. Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. *Theor Appl Genet*, 1996, 92: 637–643
- 37 Smith O S, Smith J S C. Measurement of genetic diversity among maize inbreds; a comparison of isozymic, RFLP, pedigree, and heterosis data. *Maydica*, 1992, 37: 53–60
- 38 Messmer M M, Melchinger A E, Boppenmaier J, et al. Relationships among early European maize (*Zea mays* L.) inbreds: I. Genetic diversity among flint and dent lines revealed by RFLPs. *Crop Sci*, 1992, 32: 1301–1309
- 39 Melchinger A E, Lee M, Lamkey K R, et al. Genetic diversity for restriction fragment length polymorphisms: Relation to genetic effects in maize inbreds. *Crop Sci*, 1990, 30: 1033–1040
- 40 Smith O S, Smith J S C, Bowen S L, et al. Similarities among a group of elite maize inbreds as measured by pedigree, F<sub>1</sub> heterosis, and RFLPs. *Theor Appl Genet*, 1990, 80: 833–840
- 41 Deng H F, He Q, Shu F, et al. Status and technical strategy on development of *japonica* hybrid rice in China. *Hybrid Rice*, 2006, 21: 1–6