

ORIGINAL ARTICLE

# Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm

C-H Wang<sup>1,3</sup>, X-M Zheng<sup>2,3,4</sup>, Q Xu<sup>1</sup>, X-P Yuan<sup>1</sup>, L Huang<sup>2</sup>, H-F Zhou<sup>2</sup>, X-H Wei<sup>1</sup> and S Ge<sup>2</sup>

Despite extensive studies on cultivated rice, the genetic structure and subdivision of this crop remain unclear at both global and local scales. Using 84 nuclear simple sequence repeat markers, we genotyped a panel of 153 global rice cultivars covering all previously recognized groups and 826 cultivars representing the diversity of Chinese rice germplasm. On the basis of model-based grouping, neighbour-joining tree and principal coordinate analysis, we confirmed the widely accepted five major groups of rice cultivars (*indica*, *aus*, *aromatic*, *temperate japonica* and *tropical japonica*), and demonstrated that *rayada* rice was unique in genealogy and should be treated as a new (the sixth) major group of rice germplasm. With reference to the global classification of rice cultivars, we identified three major groups (*indica*, *temperate japonica* and *tropical japonica*) in Chinese rice germplasm and showed that Chinese *temperate japonica* contained higher diversity than that of global samples, whereas Chinese *indica* and *tropical japonica* maintained slightly lower diversity than that present in the global samples. Particularly, we observed that all seasonal, drought-tolerant and endosperm types occurred within each of three major groups of Chinese cultivars, which does not support previous claims that seasonal differentiation exists in *Indica* and drought-tolerant differentiation is present in *Japonica*. It is most likely that differentiation of cultivar types arose multiple times stemming from artificial selection for adaptation to local environments.

*Heredity* (2014) **112**, 489–496; doi:10.1038/hdy.2013.130; published online 11 December 2013

**Keywords:** population structure; diversity; classification; *Oryza sativa*; rice germplasm

## INTRODUCTION

Asian cultivated rice (*Oryza sativa* L.) is one of the most important crops in the world. Rice not only serves as a primary food source for more than half of the world's population (Khush, 1997), but also provides an excellent model system for studying a wide range of biological questions (Shimamoto and Kyozuka, 2002; Wing *et al.*, 2005; Zhang *et al.*, 2008). As a result of both natural and artificial selection, rice has evolved into a crop containing high levels of morphological, physiological and genetic diversity, with >120 000 distinct rice varieties recognized worldwide (Oka, 1988; Khush, 1997; Sang and Ge, 2007; Vaughan *et al.*, 2008). An assessment and classification of this diversity is important for effective utilization of the germplasm and will provide a powerful evolutionary framework for many future biological studies using rice as a model.

To date, substantial investigations have been undertaken on the classification of rice cultivars (see reviews in Oka, 1988; Sang and Ge, 2007; Sweeney and McCouch, 2007). Although two major subspecies (*Indica* and *Japonica*) have been widely accepted without dispute, the genetic subdivision or substructure of the crop has been inconsistent (for example, Matsuo, 1952; Glaszmann, 1987; Ni *et al.*, 2002; Garris *et al.*, 2005; Ebana *et al.*, 2010). In a landmark study using 15 polymorphic enzyme loci on 1688 landraces, Glaszmann (1987)

identified six varietal groups (I to VI), with two of the largest groups (groups I and IV) corresponding to typical *Indica* and *Japonica* varieties (subspecies) (Khush, 1997). By screening a sample of 234 rice accessions using 169 nuclear simple sequence repeats (SSRs) and two chloroplast loci, Garris *et al.* (2005) detected five distinct groups and referred to them as *indica*, *aus*, *aromatic*, *temperate japonica* and *tropical japonica*, corresponding to Glaszmann's (1987) groups I, II, V and VI (*temperate japonica* and *tropical japonica*), respectively. This classification was subsequently supported by analyses of genome-wide single-nucleotide polymorphism (SNP) data (Caicedo *et al.*, 2007; Zhao *et al.*, 2010; Huang *et al.*, 2012). Nevertheless, a survey of whole-genome single-nucleotide polymorphisms in 140 diverse rice accessions from a world plus a Japanese core collection yielded slightly different result (Ebana *et al.*, 2010). This study identified seven rice groups, three of which were defined as *tropical japonica*, one as *temperate japonica*, two as *indica* and one as *aus* in terms of Garris *et al.*'s (2005) classification. Numerous studies on rice cultivars at a local scale have also been conducted and found different subdivisions of rice germplasm (de Oliveira Borba *et al.*, 2009; Zhang *et al.*, 2009, 2011b; Jin *et al.*, 2010; Li *et al.*, 2010). Thus, the genetic subdivision and classification of rice germplasm remain unclear, with three to seven major groups identified by different studies.

<sup>1</sup>State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, China and <sup>2</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing, China

<sup>3</sup>These authors contributed equally to this work.

<sup>4</sup>Current address: National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

Correspondence: Professor S Ge, State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing 100093, China.

E-mail: gesong@ibcas.ac.cn

or Professor X-H Wei, State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, China.

E-mail: hxzwei@gmail.com

Received 20 May 2013; revised 15 September 2013; accepted 14 October 2013; published online 11 December 2013

One problem arising from previous studies is that some groups of major varieties (for example, groups III and IV recognized by Glaszmann, 1987) were not included in all analyses of global rice germplasm (Garris *et al.*, 2005; Ebana *et al.*, 2010; Zhao *et al.*, 2010; Huang *et al.*, 2012). This may have led to unreliable estimate of the substructure and lineage affinities among rice cultivars. Another issue is that almost all previous studies on global rice germplasm did not include sufficient samples from China, which might have biased the estimates of both genetic diversity and population structure given that 71 970 rice accessions are recognized as present in China (Wei *et al.*, 2000). For example, of 1688 accessions surveyed by Glaszmann (1987), only 263 (15.6%) were Chinese and classified exclusively into groups I and VI. In the study by Garris *et al.* (2005), only 15 accessions from mainland China were included and these were grouped into just two groups (*indica* and *temperate japonica*). Recent studies have provided convincing evidence that more than two distinct groups were present in Chinese rice germplasm (Zhang *et al.*, 2009, 2011b; Huang *et al.*, 2010, 2012; Jin *et al.*, 2010), implying that greater sampling of Chinese rice cultivars is required for a better resolution of rice population genetics. Finally, as the largest rice gene pool, the Chinese rice germplasm has been a focus of investigations for a long time. However, different studies have obtained inconsistent opinions and thus yielded conflicting conclusions regarding cultivar grouping and subdivisions (for example, Ting, 1957; Cheng, 1985; Zhang *et al.*, 2009, 2011b; Huang *et al.*, 2010; Jin *et al.*, 2010). Hence, there is a need to clarify the classification of rice cultivars in China and to place this within a global classification of rice germplasm.

Here, we used 84 SSR markers to genotype 153 rice accessions representing all major groups resolved in previous studies plus an additional 826 accessions representing Chinese rice cultivars. Our specific objectives were (1) to investigate the population genetic structure of global rice cultivars by sampling all known cultivar groups and sufficient Chinese accessions, so as to provide a better understanding of classification of the global rice germplasm; (2) to elucidate the population structure of Chinese rice germplasm and to re-assess its subdivision and classification with reference to the grouping of global rice cultivars; and (3) to evaluate the genetic diversity and differentiation of major groups of Chinese cultivars. In this way, we aimed to obtain an improved understanding of rice germplasm at both global and local scales, which will enhance its exploitation and utilization in rice genetics and breeding.

## MATERIALS AND METHODS

### Plant materials

A total of 826 accessions of Chinese rice cultivars were sampled, including 194 accessions from the rice core collection of the National Mid-term Genebank for Rice (Hangzhou, China; Supplementary Table S1). Of them, 480 were labelled as *Indica* and 346 as *Japonica*. The cultivars included 747

landraces and 79 modern varieties, and represent different seasonal (early, medium and late), drought-tolerant (lowland and upland) and endosperm (waxy and non-waxy) types (Table 1, Supplementary Table S1). In addition, we constructed a set of 153 'global' accessions that included 43 representative Chinese accessions (recognized in Zhang *et al.*, 2009; 2011a) and 110 accessions representing rice germplasm from countries other than China. The latter accessions consisted of the major rice groups identified by Glaszmann (1987) and Garris *et al.* (2005) and were obtained from the Germplasm Resource Center of the International Rice Research Institute (Supplementary Table S2). Of these, 31 accessions were selected based on Glaszmann's (1987) classification, including 7 from group II (*aus*), 8 from group III (*ashina*), 11 from group IV (*rayada*) and 5 from group V (*aromatic*) accessions. The remaining 79 accessions were chosen based on Garris *et al.*'s (2005) classification, including 25 *indica*, 14 *aus*, 10 *aromatic*, 16 *temperate japonica* and 14 *tropical japonica* accessions.

### DNA extraction and SSR genotyping

Total genomic DNA was extracted from tender plant shoots following a modified sodium dodecyl sulphate 'mini-extraction' protocol developed by Zheng *et al.* (1995). DNA extracts were quantified using a spectrophotometer, checked by 1.5% agarose gel electrophoresis and finally dissolved in  $0.1 \times$  Tris-EDTA buffer solution and stored at  $-20^\circ\text{C}$ . On the basis of available information on genome-wide SSR markers in rice (McCouch *et al.*, 2002), we initially screened a total of 125 SSR primer pairs and finally chose 84 primer pairs that were consistently amplified in our analysis. The SSR loci are randomly distributed throughout the 12 rice chromosomes and their detailed information is provided in Supplementary Table S3. Primers were synthesized by ABI (Applied Biosystems, Foster City, CA, USA), with the forward primers labelled with blue (FAM), yellow (NED), red (PET) and green (VIC) fluorophores, respectively.

DNA amplification was carried out using a 2720 thermal cycler (Applied Biosystems) in a  $10 \mu\text{l}$  reaction mixture. Each reaction contained  $10 \times$  buffer  $1.0 \mu\text{l}$ ,  $2 \text{ mmol l}^{-1}$  dNTPs  $1.0 \mu\text{l}$ ,  $25 \text{ mmol l}^{-1}$   $\text{MgCl}_2$   $1.0 \mu\text{l}$ ,  $0.6 \mu\text{l}$  each of the forward and reverse primer ( $10 \mu\text{mol l}^{-1}$ ),  $5 \text{ U } \mu\text{l}^{-1}$  Taq polymerase  $0.1 \mu\text{l}$  and  $20 \text{ ng}$  template DNA. The PCR cycle profile was:  $94^\circ\text{C}$ , 2 min; 35 cycles of  $94^\circ\text{C}$ , 45 s,  $55^\circ\text{C}$  ( $50^\circ\text{C}$  for RM6091,  $61^\circ\text{C}$  for RM161 and RM185,  $67^\circ\text{C}$  for RM119, RM132, RM169 and RM176, respectively), 45 s, and  $72^\circ\text{C}$ , 1 min; and  $72^\circ\text{C}$ , 8 min.

PCR products ( $2 \mu\text{l}$ ) were diluted in  $8 \mu\text{l}$  of ultra pure water, purified in the solution mixed with ethanol and sodium acetate for 30 min, and scoured in 100% alcohol for 15 min. The diluted DNA was dried and mixed with highly deionized-formamide (Applied Biosystems), and then submitted for fragment analysis by capillary electrophoresis using an Applied Biosystems 3130xl DNA analyser (Applied Biosystems). DNA fragment size analysis and allele calling were performed using GeneScan and GeneMapper software (Applied Biosystems), followed by the manual allele binning. The matrix standard kit was used to generate the 'multicomponent matrix' when analyzing the FAM-, NED-, PET-, VIC- and Liz-labelled DNA fragments with the ABI3130 series systems. The multicomponent matrix was used by Data Collection Software (Applied Biosystems) for the instrument to automatically analyse the five different coloured fluorescent dye-labelled samples in a single capillary. In addition, reference lanes were further verified by polyacrylamide gel electrophoresis. In the case of non-amplification, PCR was repeated to exclude technical failure and a null allele was recorded if both PCRs failed.

**Table 1** Number (percentage) of accessions belonging to different cultivar types in three groups of Chinese germplasm used in this study

Group	Sample size <sup>a</sup>	Seasonal type			Drought-tolerant type		Endosperm type	
		Early	Medium	Late	Lowland	Upland	Waxy	Non-waxy
<i>indica</i>	495	43 (8.7%)	245 (49.5%)	207 (41.8%)	401 (81.0%)	94 (19.0%)	72 (14.5%)	423 (85.5%)
<i>temperate japonica</i>	201	23 (11.4%)	77 (38.3%)	101 (50.3%)	178 (88.6%)	23 (11.4%)	83 (41.3%)	118 (58.7%)
<i>tropical japonica</i>	82	3 (3.7%)	27 (32.9%)	52 (63.4%)	10 (12.2%)	72 (87.8%)	45 (54.9%)	37 (45.1%)
Total	778	69 (8.9%)	349 (44.8%)	360 (46.3%)	589 (75.7%)	189 (24.3%)	200 (25.7%)	578 (74.3%)

<sup>a</sup>Admixed accessions among three groups were excluded.

### Statistical analysis

The model-based program STRUCTURE (Pritchard *et al.*, 2000; Falush *et al.*, 2003) was used to calculate the genetic component for each accession using a burn-in length of 10 000, run length of 100 000, and a model allowing for admixture and correlated allele frequencies. To infer population structure, we used both the LnP(D) value and Evanno's  $\Delta K$  (Evanno *et al.*, 2005) based on five independent simulations. Genetic distance was calculated using the LogSharedAllele distance,  $D_{LS}$  (Chakraborty and Jin, 1993), and phylogenetic reconstruction was based on the neighbour-joining (NJ) method implemented in PowerMarker (Liu and Muse, 2005). We also used NTSYTS-pc version 2.10e (Rohlf, 1997) to perform the principal coordinate analysis (PCA), which summarized the major patterns of variation in multi-locus data set. To investigate the group relationship,  $F_{ST}$  value was calculated and tested using FSTAT (Goudet, 2001). PowerMarker (Liu and Muse, 2005) was used to calculate the number of alleles per locus ( $N_a$ ) and gene diversity per locus ( $H_e$ ) as well as polymorphism information content (PIC).

## RESULTS

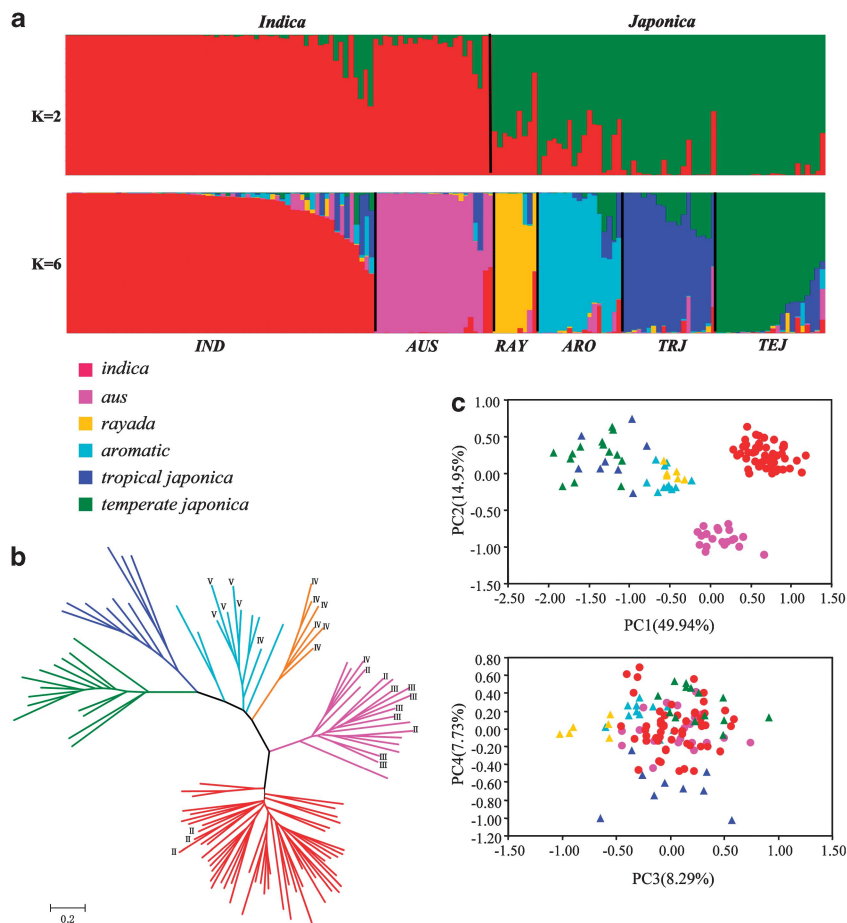
### Genetic structure and subdivision of global rice cultivars

We sampled 153 diverse global rice cultivars representing all major groups identified in previous studies (Glaszmann, 1987; Garris *et al.*, 2005; Zhang *et al.*, 2009, 2011a) to analyse the population genetic structure and the genetic affinities among major sub-populations of rice. We first used STRUCTURE, a model-based approach to calculate the genetic component of each accession and subdivide all cultivars into clusters. We found that the LnP(D) value increased with  $K$  from 1 to 10, but showed peaks of Evanno's  $\Delta K$  at  $k=2$  and  $k=6$ , with

very similar grouping results at  $K>6$  (Supplementary Figure S1). Thus, we chose  $K=2$  and  $K=6$  for the final analysis of the global accessions. At  $K=2$ , two major groups corresponding to two rice subspecies (*Indica* and *Japonica*) are apparent; whereas six well-clustered groups are evident at  $K=6$ , with two groups under *Indica* and four under *Japonica* (Figure 1a). It is noted that all five groups determined by Garris *et al.* (2005), that is, *indica*, *aus*, *aromatic*, *temperate japonica* and *tropical japonica*, were confirmed, corresponding to groups I, II, V and VI of Glaszmann (1987). As 39 (25%) accessions were admixed individuals (with the membership  $<80\%$  in any one group), we ran STRUCTURE again excluding these admixed accessions and the results remained unchanged (data not shown).

Our NJ tree and PCA further confirmed the STRUCTURE results. When admixed accessions were removed, six groups were clearly identified in NJ tree (Figure 1b). Note that all *rayada* rice accessions form a distinct cluster close to *aromatic* rice. In the PCA, the first two eigenvectors clearly separate the two major subspecies, *Indica* and *Japonica*. The former consists of two well-separated groups (*indica* and *aus*); whereas subdivision under *Japonica* is not apparent. In contrast, the third and fourth eigenvectors divide the *Japonica* into four groups, that is, *rayada*, *aromatic*, *temperate japonica* and *tropical japonica* (Figure 1c).

Taken together, all three approaches detected deep population structure of rice germplasm and demonstrated the presence of six distinct groups within Asian cultivated rice, that is, *indica* (I), *aus* (II),



**Figure 1** Population structure of 153 world-wide rice samples. (a) Model-based population assignment using STRUCTURE analysis. (b) NJ tree based on  $D_{LS}$  genetic distance. (c) Principal component analysis. Capital roman numbers represent the groups identified by Glaszmann (1987).

*rayada* (IV), *aromatic* (V), *temperate japonica* (VI) and *tropical japonica* (VI) (roman numerals in parentheses are the groups recognized by Glaszmann (1987); Figure 1c). Importantly, for the two groups (III and IV) of Glaszmann (1987) that were not sampled in previous studies, group III (*ashina*) seems to belong to the same group as *aus*; whereas group IV (*rayada*) represents a unique lineage (Figure 1; Supplementary Figure S1).

### Genetic structure and subdivision of Chinese rice cultivars

We first ran STRUCTURE on a panel of 826 Chinese rice cultivars to optimize the number of groups ( $K$ ). By comparing the  $\text{LnP}(D)$  and Evanno's  $\Delta K$  values by increasing  $K$  from 1 to 10, we found that  $\text{LnP}(D)$  values increased with  $K$ , with the highest log likelihood score at  $K=2$ , whereas a peak of Evanno's  $\Delta K$  occurred at  $K=5$  (Supplementary Figure S2). This indicated that the genetic structure of 826 Chinese accessions had two solutions at  $K=2$  and  $K=5$ , supporting the occurrence of two subspecies (*Indica* and *Japonica*) and five groups under two subspecies. This substructure was also supported by the NJ tree analysis (Supplementary Figure S3). It is noteworthy from the STRUCTURE analysis that a high proportion of admixed accessions was present in Chinese cultivars, particularly at  $K>3$ . Thus, we found 56 (6.7%) admixed individuals at  $K=2$ , but 447 (54.1%) admixed accessions at  $K=5$ , meaning that for these accessions,  $<80\%$  of their inferred ancestry arose from one of the five model-based populations. These results suggest that the majority of Chinese rice cultivars are derived from admixture among groups within either *Indica* or *Japonica* rather than between the two subspecies. After excluding all admixed accessions detected when  $K=5$ , we performed the STRUCTURE, NJ and PCA analyses on the remaining 379 accessions. As shown in Figure 2, two genetically distinct groups (*Indica* and *Japonica*) and five groups across the subspecies are again evident. However, the subdivisions within subspecies are less clear, particularly within *Indica*.

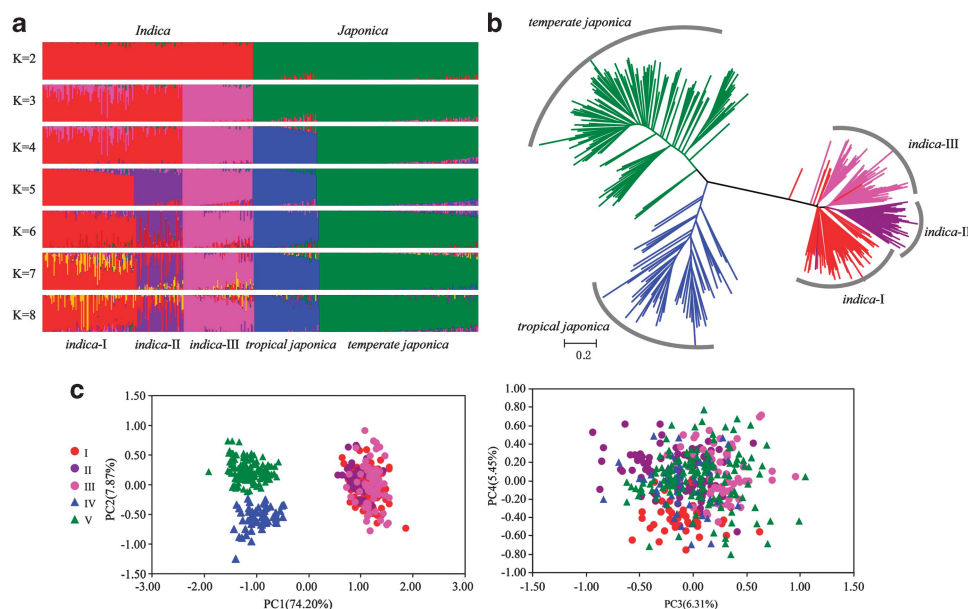
To determine the identity and genetic affinity of Chinese rice cultivars with reference to the global rice groups, we combined the 114 global and 379 Chinese cultivars with membership probabilities

over 80% (that is, accessions deemed not to be admixed) and constructed a NJ tree (Figure 3a). We found that three out of six groups identified for the global accessions (*indica*, *temperate japonica* and *tropical japonica*) are present in Chinese rice germplasm, whereas three minor groups identified in the global germplasm (*aus*, *aromatic* and *rayada*) are absent from Chinese germplasm (Figure 3a). When all 826 accessions (that is, admixed and non-admixed) were analysed, we found that 13 Chinese accessions clustered with either *aus* (8 accessions) or *aromatic* (5 accessions) (Supplementary Figure S4). However, these accessions were all admixed (Supplementary Table S1) and were probably introduced into China during the process of modern breeding practice. PCA gave similar results (Figure 3b), suggesting that Chinese rice germplasm consists mainly of three major groups as defined for the global cultivars, that is, *indica*, *temperate japonica* and *tropical japonica*.

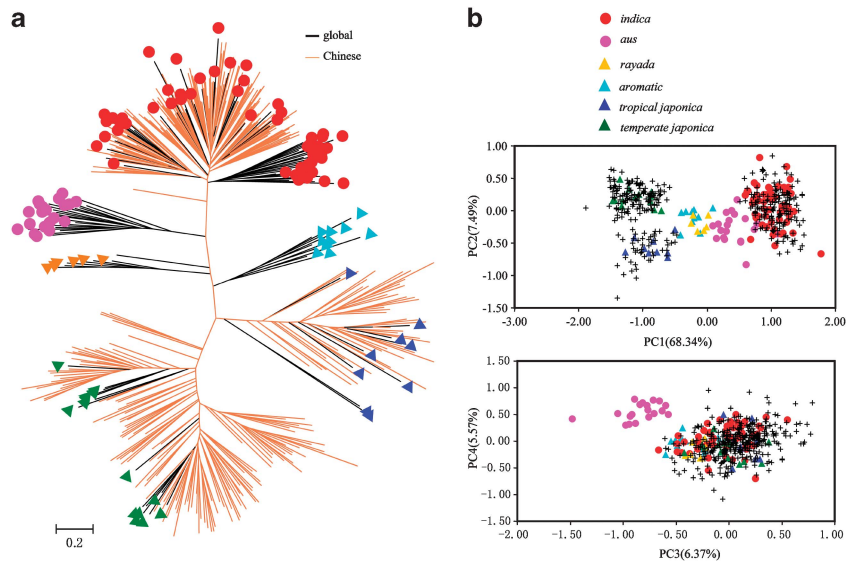
To better understand differentiation within each group in terms of seasonal, drought-tolerant and endosperm types, we categorized 778 Chinese cultivars according to their known records by excluding 48 admixed accessions among three groups (Table 1, Supplementary Table S1). It is evident that most Chinese cultivars are medium (44.9%) and late (46.3%) sown types, grow in lowland (75.7%) and are non-waxy (74.3%). However, the three major groups exhibit different characteristics for drought-tolerant type, with more lowland cultivars found in *indica* (81%) and *temperate japonica* (88.6%) while more upland cultivars are found in *tropical japonica* (87.8%). Importantly, the differentiation of cultivar types within subspecies claimed by previous studies (Ting, 1957; Zhang *et al.* 2009) is not evident. Under *Japonica*, all cultivar types are found in both *temperate japonica* and *tropical japonica* (Table 1). Similarly, all types are present in each of three *indica* groups, with no correlation found between cultivar types and genetic subdivision within subspecies (Supplementary Table S1).

### Level and organization of genetic diversity of rice cultivars

To evaluate the genetic diversity and differentiation of major groups of Chinese cultivars, we calculated several population genetic



**Figure 2** Population structure of 379 Chinese rice samples. (a) Model-based population assignment using STRUCTURE analysis. (b) NJ tree based on  $D_{L5}$  genetic distance. (c) Principal component analysis.



**Figure 3** Population structure of the combined samples of global and Chinese rice cultivars by excluding the admixed accessions. (a) NJ tree based on  $D_{LS}$  genetic distance. (b) Principal component analysis.

**Table 2** Genetic diversity of major groups of global and Chinese rice cultivars inferred by 84 SSR loci (s.d. are presented in parentheses)

Group/subspecies	N (%)			Na			He			PIC		
	Global	Chinese total	Chinese CC*	Global	Chinese total	Chinese CC	Global	Chinese total	Chinese CC	Global	Chinese	Chinese CC
<i>indica</i>	62 (40.5)	495 (63.7)	124 (67.4)	8.8 (4.9)	13.0 (7.5)	10.2 (5.8)	0.624	0.623	0.615	0.593	0.595	0.587
<i>aus</i>	24 (15.7)	—	—	6.2 (3.5)	—	—	0.604	—	—	0.568	—	—
<b><i>Indica</i></b>	<b>83</b>	<b>495</b>	<b>124</b>	<b>10.5 (6.0)</b>	<b>13.0 (7.5)</b>	<b>10.2 (5.8)</b>	<b>0.694</b>	<b>0.623</b>	<b>0.615</b>	<b>0.665</b>	<b>0.595</b>	<b>0.587</b>
<i>aromatic</i>	17 (11.1)	—	—	5.5 (2.7)	—	—	0.607	—	—	0.572	—	—
<i>rayada</i>	9 (5.9)	—	—	3.6 (1.9)	—	—	0.506	—	—	0.465	—	—
<i>temperate japonica</i>	22 (14.4)	201 (25.8)	43 (23.4)	5.7 (3.0)	11.7 (6.5)	7.9 (4.3)	0.573	0.608	0.609	0.542	0.585	0.583
<i>tropical japonica</i>	19 (12.4)	82 (10.5)	17 (9.2)	6.1 (2.7)	9.5 (5.3)	5.5 (2.7)	0.639	0.621	0.587	0.608	0.596	0.556
<b><i>Japonica</i></b>	<b>70</b>	<b>283</b>	<b>60</b>	<b>10.6 (5.7)</b>	<b>12.6 (7.0)</b>	<b>9.1 (5.0)</b>	<b>0.731</b>	<b>0.643</b>	<b>0.638</b>	<b>0.702</b>	<b>0.619</b>	<b>0.611</b>
Total	153	778	184	13.0 (7.4)	14.0 (8.0)	12.3 (7.0)	0.769	0.747	0.737	0.738	0.712	0.704

Abbreviations: CC, core collection; N, sample size; Na, the number of alleles of per locus; He, gene diversity per locus; PIC, polymorphism information content; SSR, simple sequence repeat. Two subspecies and their estimates of the parameters were bold faced.

parameters for both global and Chinese samples. We analysed 778 non-admixed cultivars (Table 2) and found that, for the global cultivars, genetic diversity varied slightly among six major groups, with the lowest found for *rayada* ( $He = 0.506$ ;  $PIC = 0.465$ ) and the highest for *tropical japonica* ( $He = 0.639$ ;  $PIC = 0.608$ ). It is noted that two minor groups (*aromatic* and *rayada*) maintained comparable diversity to that of other groups despite smaller sample sizes (17 and 9, respectively). These estimates of genetic diversity for global accessions are slightly higher than that of Garris *et al.* (2005) except for *rayada* that was not evaluated previously. For Chinese cultivars, we obtained almost the same level of diversity to global samples for *indica*; whereas the Chinese *temperate japonica* possessed higher diversity than that of global samples, contrasting to *tropical japonica* in which slightly higher diversity was found for the global cultivars (Table 2). It is evident that at the subspecies level, *Japonica* shows higher diversity than *Indica* for both global ( $He = 0.731$  vs 0.694;  $PIC = 0.702$  vs 0.665) and Chinese ( $He = 0.643$  vs 0.623;  $PIC = 0.619$  vs 0.595) samples.

Table 3 shows the genetic differentiation among the recognized groups for both global and Chinese cultivars. Pairwise  $F_{ST}$  values

**Table 3** Pairwise  $F_{ST}$  values among six groups for the global (above the diagonal) and Chinese (below the diagonal) samples estimated based on 84 SSR loci

Group	<i>indica</i>	<i>aus</i>	<i>aromatic</i>	<i>rayada</i>	<i>temperate japonica</i>	<i>tropical japonica</i>
<i>indica</i>						
<i>aus</i>	—	0.191***	0.201***	0.234***	0.298***	0.256***
<i>aromatic</i>	—	—	0.189***	0.244***	0.291***	0.242***
<i>rayada</i>	—	—	—	0.182***	0.180***	0.133***
<i>temperate japonica</i>	0.305***	—	—	—	0.217***	0.217***
<i>tropical japonica</i>	0.290***	—	—	—	0.106***	0.129***

Abbreviation: SSR, simple sequence repeat. \*\*\* $P < 0.001$

showed highly significant differentiation among the six global groups, ranging from 0.129 (*temperate japonica* vs *tropical japonica*) to 0.298 (*indica* vs *temperate japonica*). For the Chinese cultivars,

differentiation between three groups varied from 0.106 (*temperate japonica* vs *tropical japonica*) to 0.305 (*indica* vs *temperate japonica*), paralleling the results for the comparisons of global cultivars. Note that significantly larger genetic differentiation was recorded between the two subspecies (*Indica* vs *Japonica*) at both global ( $F_{ST}=0.167$ ) and China ( $F_{ST}=0.285$ ) scales, consistent with the genetic structure and genealogical analyses. We also calculated the  $F_{ST}$  value between three subgroups within Chinese *Indica* and found relatively low differentiation between them ( $F_{ST}=0.076-0.132$ ), consistent with the finding that subdivision was less apparent within Chinese *Indica*.

We further estimated the genetic diversity of the 184 Chinese cultivars that were identified to belong to the core collection (Zhang *et al.* 2011a). We found that the number of alleles in the core collection is lower than that in total samples, particularly for *tropical japonica* (9.5 vs 5.5). Nevertheless, over 98% of the diversity in the total samples is retained in the core collection, as measured by heterozygosity and PIC values at either species or group levels (Table 2). This result indicates that the Chinese core collection maintained abundant diversity and efficiently represented most of the genetic diversity present in Chinese rice germplasm.

## DISCUSSION

### Six distinct and major groups within *O. sativa*

The classification of rice cultivars can be dated back to the beginning of the twentieth century when Kato *et al.* (1928) indicated the existence of two main varietal types, denoted as *Indica* and *Japonica*, corresponding to the *Hsien* and *Keng* rice in China (Ting, 1957; Second, 1982; Oka, 1988). On the basis of gross morphology, Matsuo (1952) recognized a third type of cultivar that was later referred to as *Javanica* by Morinaga and Kuriyama (1958) according to its geographical distribution. Although Oka (1958) considered *Japonica* and *Javanica* as tropical and temperate components of a single *Japonica* type, the *Indica*, *Japonica* and *Javanica* terms have been used by many rice workers to refer to three morphological types (Chang, 1976; Glaszmann, 1987; Khush, 1997). In recent decades, the classification of five major groups of rice cultivars identified by Garris *et al.* (2005) have been widely accepted, which provided a clearer description of the population structure and organization of rice germplasm. However, different subdivisions of rice cultivars have also been proposed (see reviews in Oka, 1988 and Khush, 1997) and implicated in recent molecular studies (for example, Ni *et al.*, 2002; Zhang *et al.*, 2009, 2011b; Ebana *et al.*, 2010; Jin *et al.*, 2010). Importantly, the *ashina* and *rayada* rice (groups III and IV of Glaszmann, 1987) have not been included in most previous studies such as Garris *et al.* (2005) and Huang *et al.* (2012), and few accessions from China were sampled in the analyses.

In this study, we investigated a panel of global rice cultivars that consisted of all previously recognized groups and included sufficient samples from China using multiple approaches including model-based grouping, NJ tree and PCA. We confirmed the five distinct groups recognized by Garris *et al.* (2005), that is, *indica*, *aus*, *aromatic*, *temperate japonica* and *tropical japonica*, corresponding to Glaszmann's (1987) groups I, II, V and VI (*temperate japonica* and *tropical japonica*), respectively. We also found that Glaszmann's (1987) group III (*ashina*) was mixed with group II (*aus*) and not a distinct group. Importantly, we demonstrated that Glaszmann's (1987) group IV (*rayada*) was unique in genealogy and maintained a high level of genetic diversity although it grows only in India and Bangladesh (Glaszmann, 1987). The *rayada* cultivars comprise a very particular rice sown in November–December and harvested as much as 12 months later. This group is able to tolerate cold and flooding, is

photoperiod sensitive and lacks secondary dormancy (Glaszmann, 1987; Vaughan *et al.*, 2008). Having these unusual characteristics and being genetically distinct, *rayada* rice should be considered as a new (the sixth) major group of rice germplasm. This rice group is not only an important genetic resource for rice breeding, but also provides materials for investigating rice domestication and adaptation in the complex environments of deepwater areas.

### Three major groups present in Chinese rice germplasm

As one of the centres of rice domestication and the largest country of rice production (Normile, 1997; Liu *et al.*, 2007), China maintains a huge amount of rice germplasm, with 53 263 rice landraces and 5633 modern varieties in the Chinese National Crop Genebank (Han and Cao, 2005). Although great efforts have been made to study Chinese rice germplasm, varied numbers of subgroups under two subspecies have been reported (Ting, 1957; Cheng, 1985; Zhang *et al.*, 2009, 2011b; Huang *et al.*, 2010; Jin *et al.*, 2010; Li *et al.*, 2010). For example, Jin *et al.* (2010) used 100 genome-wide SSRs to investigate the population genetic structure of 416 Chinese rice accessions and grouped them into seven groups, with six groups belonging to *Indica* and only one to *Japonica*. By re-sequencing 517 rice landraces, however, Huang *et al.* (2010) found that both subspecies contained three subgroups. This discord might arise from the employment of different genetic markers and analytical methods as well as different samples used in the studies. More importantly, all previous studies on Chinese rice germplasm focused exclusively on Chinese accessions. Without inclusion of collections from areas outside China, these studies were unable to make clear the number of cultivar groups and their genetic affinities in terms of the widely recognized systems proposed by Glaszmann (1987) and Garris *et al.* (2005).

Here we demonstrated that out of six groups identified for the global rice germplasm, only three major groups (*indica*, *temperate japonica* and *tropical japonica*) were found in China, with *indica* being the largest group. Although 13 cultivars that were sampled mainly from western China clustered with *aus* and *aromatic* on the genealogical tree, model-based STRUCTURE analyses indicated that they were all admixed accessions among groups and of limited significance. It is worthwhile mentioning that Chinese cultivars exhibited higher diversity than that of global samples for *temperate japonica* but slightly lower diversity than global samples for *tropical japonica*, probably reflecting the higher proportion of *temperate japonica* in Chinese rice germplasm (63.6%).

### Diversity and differentiation of Chinese rice cultivars

Chinese farmers have traditionally recognized two major groups of rice cultivars, *Hsien* and *Keng*, based on grain types and sensitivity to temperature and photoperiod (Ting, 1957; Cheng, 1985). These two groups actually correspond to the two subspecies *Indica* and *Japonica* (Glaszmann, 1987) and have been widely accepted in China (Ting, 1957; Cheng, 1985; Zhang *et al.*, 2009). According to six diagnostic characters, Cheng (1985) further classified Chinese rice into *Hsein*, *Hsein-cline*, *Keng-cline* and *Keng*, whereas Ting (1957) developed a five-level taxonomic system for Chinese germplasm, in which rice cultivars were divided into *Hsien* and *Keng* subspecies, two seasonal types (early/medium and late sown), two drought-tolerant (or watery regime) types (lowland and upland) and two endosperm types (waxy and non-waxy) and other cultivars. In addition to the *Hsien* and *Keng* classification, the other four types of divisions were mainly based on ecological or habitat features and were not supported by subsequent investigations, particularly molecular approaches (Zhang *et al.*, 2009, 2011a). By screening 3024 Chinese landraces, for example, Zhang

*et al.* (2009) found clear differentiation within two subspecies, that is, the subdivision among seasonal types within *Indica* and differentiation between drought-tolerant types within *Japonica*. They attributed such intra-subspecies differentiation to their different growth environments and cropping systems.

In this study, we found that all seasonal, drought-tolerant and endosperm types of cultivars occurred within each of three major groups. Within either *temperate japonica* or *tropical japonica*, almost all cultivar types are present (Supplementary Table S4). Furthermore, within *indica*, the three subgroups (*indica*-I, *indica* II and *indica*-III) are not correlated with differentiation of cultivar types (Supplementary Table S4). These observations do not support the findings of Zhang *et al.* (2009) that claimed seasonal differentiation in *Indica* and drought-tolerant differentiation in *Japonica*. Therefore, it is most likely that differentiation of cultivar types arose multiple times to adapt to local environments under artificial selection.

Interestingly, we observed that the number of alleles per locus ( $N_a$ ) was significantly lower in Chinese core collection than in the total Chinese germplasm at the group levels although the core collection retained very high percentage of total genetic diversity (over 98%). This finding suggested that numerous rare alleles might not be captured in the core collection. In recent decades, many elite genes or alleles have been discovered from landraces and utilized in rice breeding, including the semi-dwarf genes (Monna *et al.*, 2002; Sasaki *et al.*, 2002), resistant/tolerant genes to diseases and pests, and genes for improving grain yield and quality (Fukuoka *et al.*, 2009; Ishimaru *et al.*, 2013). It is likely, therefore, that three minor groups of rice germplasm (*aus*, *aromatic* and *rayada*) that have a limited presence in Chinese germplasm may be important genetic resources for future rice genetics and breeding in China given their unique genetic and phenotypic characteristics (Glazmann, 1987; Khush, 1997).

## DATA ARCHIVING

Data were deposited in the Dryad repository: doi:10.5061/dryad.m8jj7. dryad.55425.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank Mr Daolong Ming (China National Rice Research Institute, China) for assistance in growing samples in greenhouse and to Richard Abbott for reading and improving the manuscript. We also thank the International Rice Research Institute (Los Banos, Philippines) for providing seed samples. This work was supported by the National Natural Science Foundation of China (91231201 and 31070195) and the Basic Research Budget of China National Rice Research Institute (2009RG001-3).

- Ebana K, Yonemaru J, Fukuoka S, Iwata H, Kanamori H, Namiki N *et al.* (2010). Genetic structure revealed by a whole-genome single-nucleotide polymorphism survey of diverse accessions of cultivated Asian rice (*Oryza sativa* L.). *Breeding Sci* **60**: 390–397.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Falush D, Stephens M, Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- Fukuoka S, Saka N, Koga H, Ono K, Shimizu T, Ebana K *et al.* (2009). Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* **325**: 998–1001.
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**: 1631–1638.
- Glazmann JC (1987). Isozymes and classification of Asian rice varieties. *Theor Appl Genet* **74**: 21–30.
- Goudet J (2001). *FSTAT, Version 2.9.3, A Program to Estimate and Test Gene Diversities and Fixation Indices*. Lausanne University: Lausanne, Switzerland.
- Han LZ, Cao GL (2005). Status of collection, conservation and propagation of rice germplasm in China. *J Plant Genet Res* **6**: 359–364.
- Huang XH, Wei XH, Sang T, Zhao QA, Feng Q, Zhao Y *et al.* (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* **42**: 961–976.
- Huang XH, Zhao Y, Wei XH, Li CY, Wang A, Zhao Q *et al.* (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet* **44**: 32–53.
- Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H *et al.* (2013). Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat Genet* **45**: 707–711.
- Jin L, Lu Y, Xiao P, Sun M, Corke H, Bao JS (2010). Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor Appl Genet* **121**: 475–487.
- Kato S, Kosaka H, Hara S (1928). On the affinity of rice varieties as shown by fertility of hybrid plants. *Agric Kyushu Univ* **3**: 132–147.
- Khush GS (1997). Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* **35**: 25–34.
- Li X, Yan W, Agrama H, Hu B, Jia L, Jia M *et al.* (2010). Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. *Genetica* **138**: 1221–1230.
- Liu KJ, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**: 2128–2129.
- Liu L, Lee GA, Jiang LP, Zhang JZ (2007). The earliest rice domestication in China. *Antiquity* **81**: 313.
- Matsuo T (1952). Genecological studies on cultivated rice. *Bull Nat Inst Agr Sci Series D3*: 1–111.
- McCouch SR, Temnykh L, Xu YB, Lobos KB, Clare K *et al.* (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res* **9**: 199–207.
- Monna L, Kitazawa N, Yoshino R, Suzuki J, Masuda H, Maehara Y *et al.* (2002). Positional cloning of rice semidwarfing gene, *sd-1*: rice 'Green revolution gene' encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res* **9**: 11–17.
- Morinaga T, Kuriyama H (1958). Intermediate type of rice in the subcontinent of India and Java. *Jpn J Breed* **7**: 253–259.
- Ni J, Colowit PM, Mackill DJ (2002). Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop Sci* **42**: 601–607.
- Normile D (1997). Yangtze seen as earliest rice site. *Science* **275**: 309.
- Oka HI (1958). Intervarietal variation and classification of cultivated rice. *Indian J Genet Pl Br* **18**: 79–89.
- Oka HI (1988). *Origin of Cultivated Rice*. Japan Scientific Societies Press, Tokyo, Japan.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Rohlf F (1997). *NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, Version 2.00*. Exeter Software: Setauket, NY, USA.
- Sang T, Ge S (2007). Genetics and phylogenetics of rice domestication. *Curr Opin Genet Dev* **17**: 533–538.
- Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D *et al.* (2002). Green revolution: a mutant gibberellin-synthesis gene in rice—new insight into the rice variant that helped to avert famine over thirty years ago. *Nature* **416**: 701–702.
- Second G (1982). Origin of the genetic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Jpn J Genet* **57**: 25–57.
- Shimamoto K, Kyozuka J (2002). Rice as a model for comparative genomics of plants. *Annu Rev Plant Biol* **53**: 399–419.
- Sweeney M, McCouch S (2007). The complex history of the domestication of rice. *Ann Bot-London* **100**: 951–957.
- Ting Y (1957). The origin and evolution of cultivated rice in China. *Acta Agron Sin* **8**: 243–260.
- Vaughan DA, Lu BR, Tomooka N (2008). The evolving story of rice evolution. *Plant Sci* **174**: 394–408.
- Wei XH, Tang SX, Yu HY, Jiang YZ, Qiu ZG (2000). Studies on methods of developing a core collection for China traditional japonica rice. *Chinese J Rice Sci* **14**: 237–240.

Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fiedel-Alon A, York TL *et al.* (2007). Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet* **3**: 1745–1756.

Chakraborty R, Jin L (1993). A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. *DNA Fingerprinting: State of the Science*. Birkhauser Verlag, Basel, Switzerland. pp 153–175.

Chang TT (1976). The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica* **25**: 425–441.

Cheng KS (1985). A statistical evaluation of the classification of rice cultivars into *hsien* and *keng* subspecies. *Rice Genet Newsl* **2**: 46–48.

de Oliveira Borba TC, Brondani RP, Rangel PH, Brondani C (2009). Microsatellite marker-mediated analysis of the EMBRAPA Rice Core Collection genetic diversity. *Genetica* **137**: 293–304.

- Wing RA, Ammiraju JSS, Luo MZX, Kim H, Yu YS, Kudrna D *et al.* (2005). The Oryza Map Alignment Project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol Biol* **59**: 53–62.
- Zhang DL, Zhang HL, Wang MX, Sun JL, Qi YW, Wang F *et al.* (2009). Genetic structure and differentiation of *Oryza sativa* L. in China revealed by microsatellites. *Theor Appl Genet* **119**: 1105–1117.
- Zhang HL, Zhang DL, Wang MX, Sun JL, Qi YW, Li JJ *et al.* (2011a). A core collection and mini core collection of *Oryza sativa* L. in China. *Theor Appl Genet* **122**: 49–61.
- Zhang P, Li J, Li X, Liu X, Zhao X, Lu Y (2011b). Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS One* **6**: e27565.
- Zhang QF, Li JY, Xue YB, Han B, Deng XW (2008). Rice 2020: a call for an international coordinated effort in rice functional genomics. *Mol Plant* **1**: 715–719.
- Zhao KY, Wright M, Kimball J, Eizenga G, McClung A, Kovach M *et al.* (2010). Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS One* **5**: e10780.
- Zheng KL, Subudhi PK, Domingo J, Magpantay G, Huang N (1995). Rapid DNA isolation for marker assisted selection in rice breeding. *Rice Genet News* **12**: 255–258.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)