Heredity (2008) 101, 527–535

www.nature.com/hdy

ORIGINAL ARTICLE

Genetic structure of *Oryza rufipogon* Griff. in China

MX Wang[1,5], HL Zhang[1,5], DL Zhang[1], YW Qi[1], ZL Fan[2], DY Li[3], DJ Pan[2], YS Cao[4], ZE Qiu[4], P Yu[1], QW Yang[4], XK Wang[1] and ZC Li[1]

[1]*Key Laboratory of Crop Genomics and Genetic Improvement of Ministry of Agriculture and Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing, China;* [2]*Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China;* [3]*Rice Research Institute, Guangxi Academy of Agricultural Science in Guangxi Province, Nanning, China and* [4]*Institute of Crop Sciences, CAAS, Beijing, China*

Oryza rufipogon Griff. (common wild rice; CWR) is the ancestor of Asian cultivated rice (*Oryza sativa* L.). Investigation of the genetic structure and diversity of CWR in China will provide information about the origin of cultivated rice and the grain quality and yield. In this study, we used 36 simple sequence repeat (SSR) markers to assay 889 accessions, which were highly representative of whole germplasm in China. The analysis revealed a hierarchical genetic structure within CWR. First, CWR has diverged into two ecotypic populations, a south subtropical population (SSP) and a middle subtropical population (MSP), probably owing to natural selection by the different climates. The distribution of specific alleles and haplotypes indicated that Chinese CWR had both *indica*-like and *japonica*-like variations; the SSP was an *indica*-like type, whereas the MSP was more *japonica*-like. The SSP and MSP further diverged into five (HN, GD-GX1, GX2, FJ and YN) and two (JX-HuN1 and HuN2) geographical populations, respectively. The genetic data suggest the isolation by distance, although water systems also appear to play an important role in the formation of homogenous populations, and occasionally landscape was also involved. The population GD-GX1, which grew widely in Guangdong and Guangxi provinces, was the largest geographical population in China. It had a high level of genetic diversity (GD) and the closest genetic relationship with other inferred populations. The population HN, with the smallest SSR molecular weights and the highest level of GD, may be the most ancestral population.
Heredity (2008) **101,** 527–535; doi:10.1038/hdy.2008.61; published online 1 October 2008

Keywords: China; common wild rice; SSR (simple sequence repeat); genetic structure

Introduction

Effective use of genetic resources is important for improving the quality and yield of rice, an important staple crop for the world. Common wild rice (CWR, *Oryza rufipogon* Griff.), as the ancestor of cultivated rice (Oka, 1974), contains a high level of genetic diversity (GD) and represents a rich genetic resource. However, a great part of the genetic information in it has not been uncovered. Only 60–70% of its genetic variation has been found in cultivated rice (Sun *et al*., 2001). Its abundant genetic variation in factors such as disease resistance, yield and quality (Yuan *et al*., 1989; Xiao *et al*., 1996) should be of major relevance to breeders.

Genetic structure is the basis of management, research and utilization of germplasm resources (Waples, 1995) and is critical to ecological conservation and studies of evolutionary relationships among populations (Frankham, 2003). China is one of the largest centers of GD for CWR in the world (Wang *et al*., 2004). The genetic structure and diversity of Chinese CWR is of interest to the scientists and has been investigated using different markers and populations (Gao *et al*., 2000a, b, 2002a, b; Zhou *et al*., 2003; Song *et al*., 2003a). All of these studies indicated that Chinese CWR is rich in GD, with obvious and complicated population subdivisions. Although these studies contributed much to our understanding of its GD and structure, the samples were inadequate to measure fine-scale GD and genetic structure over the entire population. CWR grows in 113 counties spread across eight provinces in China, but even the most representative of these studies sampled only 21 natural populations from 21 counties across provinces (Gao *et al*., 2000a). As they used different samples, these studies also reached different conclusions about the primary factors affecting the genetic structure of CWR in China (Gao *et al*., 2000b, 2002a; Zhou *et al*., 2003; Cai *et al*., 2004).

In China, CWR grows in eight provinces: Guangdong, Guangxi, Hainan, Jiangxi, Hunan, Fujian, Yunnan and Taiwan. Between 1988 and 1993, 5571 accessions in 113 counties across eight provinces were collected (Sheng and Huang, 1991; Huang and Sheng, 1996). Most of these are now planted in the germplasm gardens of wild rice in Nanning, Guangxi province and Guangzhou, Guangdong province. To examine the fine-scale GD and genetic structure of such a large CWR population, it is essential to obtain a representative sample. The concept of core collection (Frankel, 1984), which aims to settle the conflict between genetic representation and population scale, could be used to choose the smallest representative population with the largest amount of diversity. In our

Correspondence: Professor Z Li, College of Agronomy and Biotechnology, China Agricultural University, No. 2 Yuanmingyuan Xi Lu, Haidian District, Beijing 100193, China.
E-mail: lizichao@cau.edu.cn
[5]*These authors contributed equally to this work.*
Received 6 October 2007; revised 26 February 2008; accepted 13 May 2008; published online 1 October 2008

study, the research accessions obtained from the primary core collection represent 90.5% of the diversity of the entire Chinese CWR (Yu *et al*., 2003).

Simple sequence repeat (SSR) markers have proved to be effective and appropriate tools for studying GD (Davierwala *et al*., 2000) and detecting population structure (Garris *et al*., 2005). The objectives of our study were: (1) to identify the genetic structure of CWR in China (2) to find out how the genetic structure of CWR was formed and (3) to determine the genetic relationships among CWR populations.

Materials and methods

Plant material

According to Yu *et al*. (2003), the primary core collection of CWR in China was based on phenotypic data. It contained 920 accessions; of these, 860 were collected as a logarithmic proportion from the provinces where they grew and by the use of a clustering method based on a simple matching coefficient. The remaining 60 accessions were special phenotypes collected subjectively. The primary core collection represented 90.5% of the diversity of 5571 CWR. Among the 889 accessions used in this study (Supplementary Table S1), 864 were taken from the primary core collection at the gardens of wild rice in Nanning and Guangzhou. (The remaining 56 accessions out of the 920 that make up the core collection do not exist at these gardens.) These 864 accessions represented 98 counties across six provinces: Hainan (88), Guangdong (268), Guangxi (256), Fujian (59), Hunan (107) and Jiangxi 110 (86). A further 25 accessions were collected from Jinghong (2) and Yuanjiang (23) in Yunnan province, making the total of 889. The accessions could be grouped as prostrate (445), slant (255), semierect (126) and erect (63) according to the growth habits. All the accessions were perennial types. To investigate whether Chinese CWR diverged into *indica*-like and *japonica*-like types, we sampled 410 accessions of cultivated rice (*Oryza sativa* L.) (Supplementary Table S2), including 260 *Indica* accessions and 150 *Japonica* accessions from the same seven provinces where CWR was sampled.

Genome DNA extraction and analysis with microsatellite markers

DNA from every accession was extracted from silica gel-dried leaf tissues using the CTAB (cationic detergent hexadecyltrimethylammonium bromide) method (Scott and Bendich, 1988). The 36 SSR loci (Supplementary Table S3) were randomly distributed over the 12 rice chromosomes. Amplification reactions were performed in a final volume of 15 ul comprising 0.9 unit Taq enzyme, 1.5 ul 10 × buffer, 22.2 mM $MgCl_2$, 46 ng SSR primers, 1.8 mM dNTP and 10 ng total DNA. The PCR amplification program was: (1) pre-denature for 5 min at 95 °C, (2) denature for 0.5 min at 95 °C, (3) annealing for 1 min at 55 °C (annealing temperature varied with actual primers), (4) extending for 1.5 min at 72 °C, (5) repeating cycles 2–4 30 times and (6) final extension for 10 min at 72 °C. The amplified products were denatured at 95 °C for 5 min, then cooled on ice, and subsequently run on 8% denatured polyacrylamide gel at 70 W. One check was randomly designated from accessions with certain alleles. For the same marker, all runs after the first run included not only the samples but also the checks and a standard molecular weight marker—PUC19 DNA digested by MspI. When all the samples were completely run, all the checks were run with another standard molecular weight marker—10 bp DNA ladder from Invitrogen (Catalog No. 10821-015), and the molecular weight for each allele was estimated. All the gels were stained using the silver method (Bassam *et al*., 1991). In the case of null alleles in these species, PCR amplifications were repeated to exclude failed PCR reactions. In some accessions, there were more than two alleles per locus. In these cases, we amplified and ran them again, and then selected the stable alleles—those that occurred in both replications. If there were more than two stable alleles, we randomly selected two alleles to form the genotype.

Statistical analyses

The structural analysis was carried out using STRUCTURE version 2 (Pritchard *et al*., 2000; Falush *et al*., 2003; http://pritch.bsd.uchicago.edu), which implements a clustering method for inferring population structure using genotype data. We ran the simulation 10 times independent for each k (the number of partitions) value in the range 1–15, using the method allowing for the admixture, correlated allele frequencies and no prior population information, with burn-in length 10 000 and run length 10 000. The graphical display of the STRUCTURE results was generated using Distruct software (Rosenberg, 2002; http://www.cmb.usc.edu/noahr/distruct.html). Evanno *et al*. (2005) reported that in most cases, the estimated 'log probability of data' did not provide a correct estimation of cluster number (k-value), and argued that an *ad hoc* statistic ΔK based on the rate of change in the log probability of data between successive K-values could accurately detect true K. The suggested $\Delta k = \mathrm{m}(|\mathrm{L(k+1)}-2\,\mathrm{L(k)}+\mathrm{L(k-1)}|)/\mathrm{s[L(k)]}$, where L($k$) represents the kth LnP(D), m is to the mean of 10 runs and s their standard deviation. We used the method of Evanno *et al*. (2005) to estimate the number of populations. Regression analyses between genetic distance and geographic distance of CWR in different latitude–longitude sections were made using the inline applets of regression in Microsoft Excel. To investigate whether water systems have an effect on genetic relations among CWR populations, accessions from 35 counties with more than five accessions were collected in Guangdong and Guangxi. The UPGMA tree of 35 counties was constructed using PowerMarker version 3.25 (Liu and Muse, 2004; http://www.powermarker.net) based on Nei's genetic distances (Nei *et al*., 1983). Allele number, observed heterozygosity, biased GD, genotype number, polymorphism information content index and Nei's genetic distance (Nei *et al*., 1983) were all calculated in PowerMarker version 3.25. Using program HP-rare 1.0 (Kalinowski, 2005), the allelic richness (an estimator independent of the sample size; Hurlbert, 1971) of inferred populations was investigated by rarefaction methods. To investigate the directional differentiation of allele size among the inferred populations, we calculated the average standardized allele size of each CWR population using 17 SSR loci with stepwise mutation (Supplementary Table S4). A stepwise mutation index

was calculated in PowerMarker version 3.25, and we treated those SSR loci with a stepwise mutation index higher than 0.9 as having stepwise mutation, according to Vigouroux *et al*. (2003). Their significances were assessed by *t*-test. Differences in allele frequency between *indica* and *japonica* were investigated using X^2-test and G^2-test. We used the terms, *indica*-specific alleles and *japonica*-specific alleles, to describe alleles with a significantly ($P<0.0001$) higher frequency in *indica* and those with a significantly ($P<0.0001$) higher frequency in *japonica*. Frequency distributions of specific alleles in inferred CWR populations, longitudinal regions (with intervals of 5°) and latitudinal regions (with intervals of 1°) were investigated using the *t*-test, and regression analyses were used to determine the frequency distribution along longitude and latitude. The index of genetic differentiation between *indica* and *japonica* for each SSR locus was calculated using the formula: $D=(GD_t-(GD_i+GD_j)/2)/GD_t$, where GD_t represents biased GD of total cultivitated rice, Gd_i and GD_j represent those of *indica* and *japonica*, respectively (Supplementary Table S5). Using four SSR loci with D higher than 0.3, the haplotype frequencies in different populations were calculated. Differences in the significance of haplotype frequencies between *indica* and *japonica* were examined using the X^2-test. We used the term indica-specific haplotypes to describe haplotyes with a significantly ($P<0.05$) higher frequency in *indica*, and *japonica*-specific haplotypes to describe those in reverse.

Results

Genetic structure of CWR in China

The STRUCTURE simulation demonstrated that the LnP(D) value showed no clear peak in *K* between 1 and 15 (Supplementary Figure S1). Thus, it was difficult to determine the true *K* (number of populations). The magnitude change of LnP(D) relative to the standard deviation, called Δ*K* by Evanno *et al*. (2005), showed the highest peak at two; and there were two smaller peaks at seven and nine (Figure 1). We also checked the structure patterns in 10 repeats of each *k* from two to nine (Supplementary Figure S2). A clear and stable population structure could be found at two, three and seven, where a major structure pattern could be found in at least half of the simulations. Figure 2 shows the membership of accessions to the populations identified by STRUCTURE, the growth habits as classified by Pang *et al*. (1995) and Wang *et al*. (2004) and the geographic origins at $k=2$ and $k=7$. The membership indicated that the inferred genetic structure of Chinese CWR was accorded with the geographic origins (Figures 2b and d), but not with the growth habits (Figures 2a and c). CWR could be grouped into two model-based populations (MB-populations) and seven model-based subpopulations (MB-subpopulations) at $k=2$ and $k=7$, respectively. Two MB-populations at $k=2$ were isolated by the Nanling mountains, which separate the south subtropical climatic area from the middle subtropical area. One of these MB-populations grew in the south subtropical area (with the exception of the subpopulations from Hainan Island, which is tropical) and could be called a south subtropical population (SSP); the other grew in the middle subtropical area and thus could be called a middle subtropical population (MSP). The SSP included five MB-subpopulations and covered five provinces: Hainan, Guangdong, Guangxi, Fujian and Yunnan (Figure 3). Three of the five MB-subpopulations contained most of the accessions from Hainan, Fujian and Yunnan and were temporarily denoted as HN, FJ and YN, respectively. Most of the accessions from Guangdong and Guangxi were clustered into one MB-subpopulation, GD-GX1. The exception to this was the accessions from Fusui County in Guangxi, which formed another MB-subpopulation, GX2. MSP was divided into two MB-subpopulations and covered Hunan and Jiangxi provinces (Figure 3). One of these, from the Jiangxi and Chaling counties of Hunan, was called JX-HuN1, and the other, which included most of the accessions from the Jiangyong county of Hunan, was denoted as HuN2. Most of the MB-subpopulations were isolated by natural barriers or wide spaces (Figure 3). HN was isolated by the Qiongzhou channel; FJ was separated from GD-GX1 by the Wushan mountains and from JX-HuN1 and HuN2 by the Wuyi mountains. There were wide spaces from YN to GD-GX1 and GX2, and from JX-HuN1 to HuN2. Regression between genetic distance and geographical distance indicated that geographical distance was significantly correlated with genetic distance (Supplementary Figure S3).

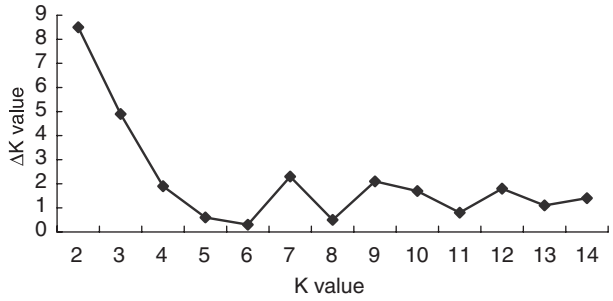

Figure 1 Magnitude of Δ*K* for each *K* value.

In addition to the apparent isolations among MB-subpopulations, we found seven populations distributed in five water systems: GD-GX1 and GX2 in the Zhujiang river system; HN in the river system of south Guangxi, Leizhou Byland and Hainan Island; YN in the river system of Yuanjiang-Lancangjiang (with two accessions in the Lancangjiang and Mekong river valley); FJ in the river system of coastal southeast China; JX-HuN1 in the Yangtze river system; and HuN2 between the Zhujiang river system and the Yangtze river system. GD-GX1 grew in a wide region within 20°N–25°N and 106°E–116°E, where the Zhujiang, a big river with several branches, runs (Figure 3 and Supplementary Figure S4). A phylogenetic tree was constructed based on different counties where more than five accessions were collected (Figure 4). The results indicated that populations from the same branch tended to be clustered together (Supplementary Figure S5). The differentiation between HN, YN, FJ, HuN2 and GD-GX1 could be explained by geographical isolation and water system, but there were exceptions. GX2, whose habitat was surrounded by that of GD-GX1 with no significant isolation clustered into a

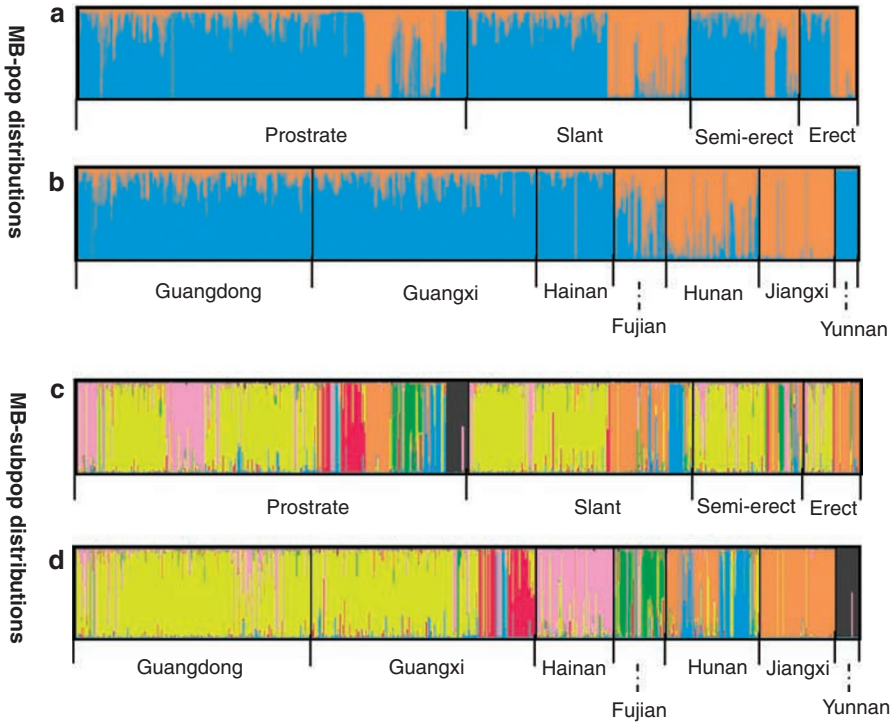

Figure 2 Membership of CWR individuals in the model-based populations when $k=2$ (**a** and **b**) and $k=7$ (**c** and **d**), and in predefined groups according to growth habits (**a** and **c**) and original provinces (**b** and **d**). Each individual was represented by a thin vertical line, the predefined groups were separated by black lines. At $k=2$ (**a** and **b**), the lake blue corresponding to south subtropical population and the orange corresponding to middle subtropical population. Blue—south subtropical population, orange—middle subtropical population. At $k=7$ (**c** and **d**), different colors represented the following model-based subpopulations, respectively: Pink—HN, yellow—GD-GX1, red—GX2, green—FJ, black—YN, blue—HuN2, Orange—JX-HuN1.

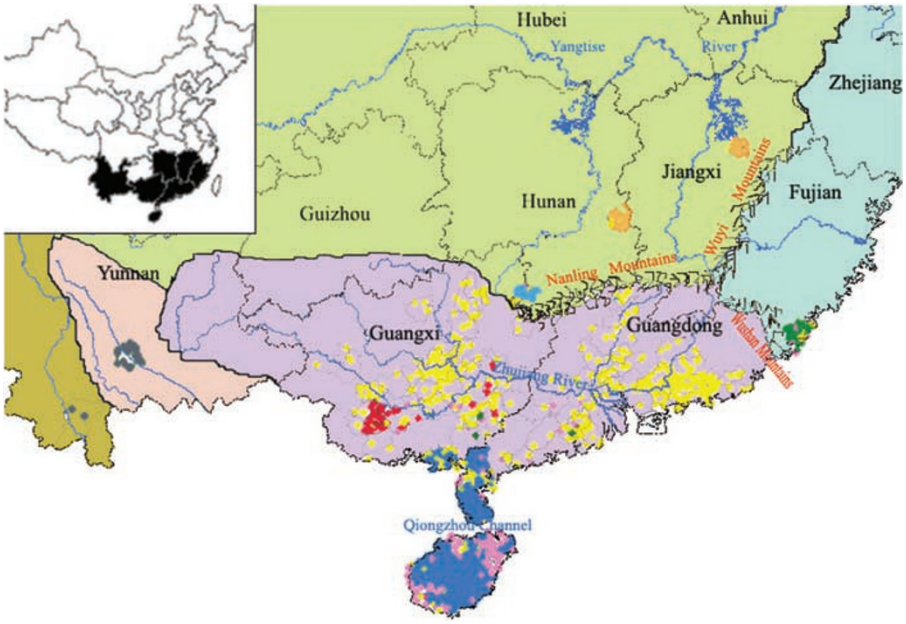

Figure 3 Geographical distributions of the inferred populations. The colored dots represented inferred subpopulations of CWR in China (same to the color in Figure 2). The colored background represented different water systems of seven provinces. Grass green—Yangtze river valley, plum—Zhujiang river valley, lake blue—Southeast water system of China, light pink—Yuanjiang and Honghe river valley, wheat—water system in south Guangxi, Leizhou Byland and Hainan, gold—Lancangjiang and Mekong river valley.

different population. We suspect that the factors besides geographical isolation and water system have contributed to its differentiation.

Genetic diversity and relationships of inferred populations

A total of 805 alleles were detected at 36 SSR loci in the 889 accessions. The number of alleles at each locus

ranged between 8 and 53, with an average of 22.36. Gene diversity ranged 0.2817–0.9648, with an average of 0.7946 (detailed diversity information for SSR loci is listed in Supplementary Table S3). Table 1 shows that SSP has a higher level of GD than MSP. Among MB-subpopulations, GD-GX1 contained the largest number of alleles and the most genotypes among the inferred populations, followed by HN. However, allele and genotype numbers in the other five MB-subpopulations were obviously lower than those in GD-GX1 and HN. Similarly, the majority (90.8%) of population-specific alleles were in HN and GD-GX1. Gene diversity and heterozygosity decreased from HN to JX-HuN1 with the increase in latitude, with the exception of YN, which had the lowest level of gene diversity and heterozygosity. Allelic richness followed a similar distribution.

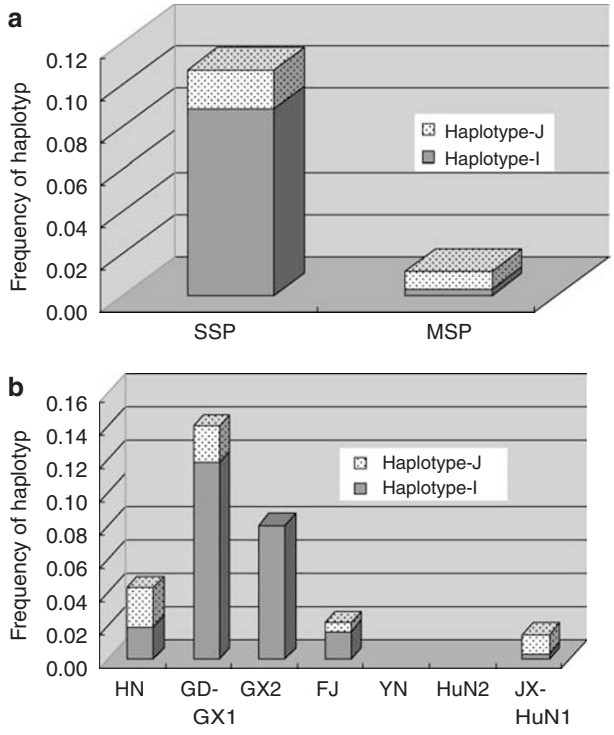

Figure 4 Distributions for *indica*-specific and *japonica*-specific haplotypes in two model-based populations (**a**) and seven model-based subpopulations (**b**). Haplotype-I—*indica* haplotype; Haplotype-J—*japonica* haplotype.

Pair-wise genetic distance (Nei *et al*, 1983) of seven MB-subpopulations indicated that GD-GX1showed smallest genetic difference with all of others and YN showed the largest genetic difference with all of the others (Table 2). Other research on the directional evolution of microsatellites (Rubinsztein *et al*., 1995; Vigouroux *et al*., 2003) has indicated that the size of SSR alleles in modern types is increased relative to ancestral types. We also investigated the distribution of allele size in the CWR populations for 17 stepwise SSR loci (Supplementary Table S4). The average standardized allele molecular weight of different MB-populations was estimated as follows: YN > > FJ > GD-GX1 > HuN2 > JX-HuN1 > > GX2 > HN. The average allele size of YN was significantly larger than that of the other inferred populations (Table 3).

Possibility of *indica*-like or *japonica*-like differentiation within *Oryza rufipogon* Griff.

It would be interesting to know which CWR have diverged into *indica*-like and which into *japonica*-like

Table 2 Pair wise Nei *et al*.'s (1983) genetic distances among seven model-based CWR subpopulations

	HN	*GD-GX1*	*GX2*	*YN*	*FJ*	*HuN2*
GD-GX1	0.1203					
GX2	0.3753	0.2815				
YN	0.7929	0.7845	0.8440			
FJ	0.3333	0.2481	0.3855	0.8401		
HuN2	0.3943	0.3067	0.4565	0.8957	0.4174	
JX-HuN1	0.3025	0.1881	0.3744	0.8210	0.3286	0.3028

Table 3 *T*-test of differences in average standardized allelic size for pairs of model-based CWR subpopulations

	HN	*GD-GX1*	*GX2*	*YN*	*HuN2*	*FJ*
GD-GX1	−3.0362*					
GX2	−0.206	5.9968*				
YN	−16.9870*	−39.7240*	−20.4600*			
HuN2	−3.1170*	1.9447	−3.7450*	21.1780*		
FJ	−4.8602*	−1.2687	−5.6752*	15.9243*	2.4106	
JX-HuN1	−2.8604*	4.5777*	−3.6354*	37.3018*	−0.8898	3.7287*

Notes: The numbers in the table were the t-values between populations in the column and those in the row. So the positive value represented a higher allele size of the population in the column than that in the row, and in reverse for a negative value. *Indicate significant differences at $P = 0.05$.

Table 1 Summary statistics for model-based CWR populations and subpopulations

	SSP					*MSP*	
	HN	*GD-GX1*	*GX2*	*YN*	*FJ*	*HuN2*	*JX-HuN1*
Sample size	114	479	49	25	46	44	132
Allele no.	602	743	261	135	343	315	347
Genotype no.	1294	2649	360	131	446	431	604
Gene diversity	0.8348	0.7879	0.6416	0.2178	0.68	0.5991	0.6006
Heterozygosity	0.4357	0.4023	0.3645	0.1078	0.2935	0.2551	0.2153
Allelic richness	8.46	7.84	4.9	2.07	6.75	6.04	5.16
Specific allele no.	23	95	0	0	0	3	1

Notes: Specific alleles—alleles predominantly existed in one population; SSP—south-subtropical population; allelic richness—estimated with 18 loci as sample size of locus bootstrap; MSP—middle-subtropical population. Significance of differences for all values was listed in Supplementary Table S6.

Table 4 Average frequency of specific alleles and their *t*-test in model-based subpopulations, longitude regions and latitude regions

	Japonica-specific alleles	*Indica*-specific alleles	*T Stat*
Inferred populations			
HN	0.1139	0.1029	0.5842
GD-GX1	0.1222	0.1185	0.1543
GX2	0.1518	0.0964	1.7141
FJ	0.1203	0.1167	0.1266
YN	0.0464	0.1079	−1.4440
HuN2	0.1436	0.0751	2.0148*
JX-HuN1	0.1834	0.0674	3.5576*
Latitude regions			
18–19 °N	0.3949	0.1206	−1.3795
19–20 °N	0.4295	0.1082	0.0141
20–21 °N	0.4299	0.1286	−1.7545
21–22 °N	0.4629	0.1218	−0.5963
22–23 °N	0.4671	0.1320	−1.6399
23–24 °N	0.4980	0.1140	−0.1583
24–25 °N	0.5250	0.1181	0.2325
25–26 °N	0.5120	0.1339	−1.4773
26–27 °N	0.5184	0.1615	−2.6019*
28–29 °N	0.5347	0.1923	−3.6104*
Longitude regions			
102–103 °E	0.0483	0.0993	−1.2543
107–108 °E	0.1394	0.1068	1.4359
108–109 °E	0.1259	0.1049	1.1384
109–110 °E	0.1162	0.1204	0.1336
110–111 °E	0.1243	0.1011	1.4640
111–112 °E	0.1350	0.0895	1.4404
112–113 °E	0.1185	0.1114	0.6547
113–114 °E	0.1362	0.1038	1.4044
114–115 °E	0.1283	0.1133	1.1046
115–116 °E	0.1379	0.1201	1.1287
116–117 °E	0.1836	0.0716	3.3668*
117–118 °E	0.1214	0.1240	−0.0971

Notes: Japonica-specific alleles—alleles with significantly ($P<0.0001$) higher frequency in *japonica* than *indica; Indica*-specific alleles—alleles with significantly ($P<0.0001$) higher frequency in *indica* than *japonica*. *Indicates significant differences at $P=0.05$.

types, and how those two types are distributed in China. To answer this question, we detected 72 *japonica*-specific alleles and 61 *indica*-specific alleles from cultivated rice using the X^2-test and the G^2-test. Distribution of those two types of alleles within each MB-subpopulation indicated that there were more *japonica*-specific than *indica*-specific alleles in JX-HuN1 and HuN2, but no significant difference in other populations (Table 4). Regression of specific-allele frequency within geographic regions (longitude and latitude) indicated that the frequency of *japonica*–specific alleles significantly increased with increased latitude ($P=0.009$; Supplementary Figure S6); the frequency of *indica*-specific alleles, however, showed no significant correlation with latitude. The frequencies of both *japonica*-specific and *indica*-specific alleles showed no significant correlation with longitude. However, significant differences between the frequencies of those two types of specific alleles were found in the longitude region of 116–117°E and the latitude regions of 26–27°N and 28–29°N (Table 4), where the accessions were mostly came from JX-HuN1. In summary, *japonica* property increased from south to north. MB-subpopulations in SSP showed no significant divergence between *indica* and *japonica*, demonstrating both *indica* and *japonica* properties; MB-subpopulations in MSP were evidently *japonica*-like.

Table 5 Genetic distance (Nei *et al*., 1983) between inferred CWR populations and *indica*, *japonica*

	HN	*GD-GX1*	*GX2*	*FJ*	*YN*	*HuN2*	*JX-HuN1*
Japonica	0.3191	0.2486	0.4150	0.4220	0.5368	0.4061	0.3654
Indica	0.3199	0.2433	0.4106	0.4127	0.5124	0.4216	0.3973

We also investigated the distribution of haplotypes composed of four loci, rm296, rm71 rm267 and rm25, which showed a higher *indica-japonica* differentiation index (>0.3) than the other 32 (<0.2; Supplementary Table S5). There was a total of 1287 haplotypes in cultivated rice and CWR, of which 74 were in both cultivated rice and CWR, 47 were only in the cultivated rice and 1166 were only in CWR. The X^2-test of 74 haplotypes in cultivated rice revealed 38 *indica*-specific haplotypes (haplotype-I) and 11 *japonica*-specific haplotypes (haplotype-J). Both haplotype-I and haplotype-J were much more distributed in HN, GD-GX1, GX2 and FJ, which were MB-subpopulations in SSP, than in JX-HuN1, which was an MB-subpopulation in MSP. Neither type was found in YN and HuN2. The MB-subpopulations in SSP had a higher frequency of haplotype-I than haplotype-J; however, the reverse was the case for the MB-subpopulaton in MSP (Figure 4).

Table 5 indicates that GD-GX1 is genetically closest to both *indica* and *japonica*. Comparison of population genetic difference between CWR inferred populations and cultivated rice from different provinces (Table 6) indicates that GD-GX1 has the closest genetic relationship with all rice cultivars. So if cultivated rice was partially domesticated in China, GD-GX1 might be the ancestry population.

Discussion

What contributed to the population subdivision within *Oryza rufipogon* Griff. in China?

Oryza rufipogon Griff. is the direct ancestor of cultivated rice (Oka, 1974). To understand the genetic differentiation, evolution, conservation and use of wild rice, its genetic structure was studied by many researchers, especially in China (Gao *et al*., 2000a, b, 2002a, b; Zhou *et al*., 2003; Song *et al*., 2003a). Despite this, the genetic structure of the whole germplasm resource in China has not been identified. Using the STRUCTURE program to infer population structure and Evanno's method (Evanno *et al*., 2005) to estimate the number of clusters, we found a two-level hierarchical structure composed of two populations—a SSP and a MSP—further divided into seven subpopulations. What contributed to these hierarchical divisions? In theory, mutation, isolation, selection and genetic drift can all influence genetic divergence (Thingsgaard, 2001; England *et al*., 2002; Primmer *et al*., 2006). For the CWR population, several possible factors have been put forward, including spatial or physical isolation and local adaptation (Gao *et al*., 2000b, 2002a; Zhou *et al*., 2003; Cai *et al*., 2004). However, we are far from reaching a consensus about which factors took effect and their relative significances. The hierarchical structure might indicate that there were hierarchical

Table 6 Population genetic distance (Nei *et al*., 1983) between inferred CWR populations and cultivated rice collected from different provinces in China

Samples of rice landraces	*Common wild rice*						
	HN	*GD-GX1*	*GX2*	*FJ*	*YN*	*HuN2*	*JX-HuN1*
Guangdong	0.1434	0.1033	0.2472	0.2361	0.4799	0.2722	0.2678
Guangxi	0.1114	0.0702	0.2039	0.1914	0.4323	0.2272	0.2156
Fujian	0.1065	0.0752	0.2196	0.2067	0.4993	0.2420	0.2306
Hainan	0.1135	0.0885	0.2167	0.1918	0.5611	0.2453	0.2251
Hunan	0.1283	0.0924	0.2210	0.2162	0.5109	0.2434	0.2191
Jiangxi	0.1257	0.0932	0.2314	0.2107	0.4836	0.2611	0.2553
Yunnan	0.1144	0.0855	0.2079	0.1976	0.4933	0.2282	0.2069

factors in the formation of the population subdivision. The highest ΔK and the most stable structure pattern at $k=2$ suggest that the differentiation that took place at the first level was the essential one within Chinese CWR. Most of the SSP grew in the south subtropical climate, whereas the MSP grew in the middle subtropical climate. This suggests that the first and most important factor was climate; that is, that natural adaptation caused the first level of differentiation within the Chinese CWR population. The role of climate is also supported by the *japonic*-like character of the MSP and the *indica*-like character of the SSP (see below). On this point, therefore, we differ from Cai *et al*. (2004), who argued that genetic differences among natural populations were the result of spatial isolation and not local adaptation. The most important factor at the second level of the hierarchy is undoubtedly isolation by space or physical barrier, which could explain all the divergences among each pair of seven subpopulations, with the exception of that between GX2 and GD-GX1. This observation is supported by previous studies, as summarized in the introduction section. In Guangdong and Guangxi, CWR grew widely and two populations in Jiangxi and Hunan, despite being located far away, formed the populations GD-GX1 and JX-HuN1, respectively. Thus, there must have been a way of achieving efficient gene flow. Given that CWR grows mainly beside or near rivers, seed and rhizome would have been easily dispersed by the water flow (Pang and Chen, 2002). The water system could compensate for the effect of spatial isolation by increasing the gene flow among individuals and populations living on the same water system. The relationship between population structure and water system has proved this. Furthermore, it has been observed that CWR in the same branches of the Zhujiang river tend to be clustered together. Although the bootstrap values are not very high, it is reasonable, given that they all belong to the same water system, to surmize that gene flow had occurred easily among the accessions in different branches. Underestimating the influence of the water system may be one of the causes of disagreement on the role of isolation in population subdivision, for example, the positive opinion of Gao *et al*. (2000b, 2002a) and the negative view taken by Zhou *et al*. (2003).

GX2 was an interesting population. It was not isolated from GD-GX1 by any evident natural barrier and both were within the same water system. Possible explanations include local adaptation and genetic drift. We prefer local adaptation as most of the accessions grew in the pond where there was no standing water after the rainy season (Zhengbin Chen, Rice 400 Research Institute, Guangdong Academy of Agricultural Sciences; private communication). The divergence of GX2 and GD-GX1 may have been a result of the particular landscape in Fusui county. The influence of local ecological adaptation on genetic structure has been demonstrated by others (Semon *et al*., 2005; Coulon *et al*., 2006).

The genetic variation of CWR populations could be summarized as follows: the smallest allelic size and highest GD are seen in HN, indicating that it might be the most ancestral population in China. GD-GX1, the largest population, had the most alleles, high GD and the closest genetic relationship with all other populations. The other populations, with the exception of YN, had lower genetic diversities and fewer alleles, and their allelic sizes were similar to that of GD-GX1. These results could be interpreted in two ways. In the first interpretation, the genetic variation represents an evolutionary relationship among seven CWR populations. Thus, CWR in China might have originated in Hainan and been dispersed to Guangdong and Guangxi and hence to other populations (with the exception of YN). The CWR from Yunnan, with an extremely distant genetic relationship and no common haplotype with other wild-rice populations in China, might be a south Asian population, as proposed by Sun *et al*. (2002). The HN population grew mainly in Hainan. As an island, Hainan has not been considered a likely candidate for the role of original center. However, it was joined with the mainland one million years ago; *Oryza rufipogon* Griff. appeared about seven million years ago (Second, 1985b). Thus, genes could have been exchanged between populations in Hainan and those in mainland of China one million years ago. If CWR had been introduced from south Asia, its first station would have been Hainan. This possibility needs further study with more CWR, including CWR from outside China. In the second interpretation, changes in the distribution of the CWR population were related to changes in the environment and other factors, such as human population. According to both the fossil rice phytoliths (Zhao and Piperno, 2000; Lu *et al*., 2002) and ancient Chinese documents (Huang *et al*., 1998), CWR may have been distributed further north than it is today. Cooling during the Younger Dryas epoch ($\sim$13 000–10 000 BP) and the pressure of population expansion during past 2000 years are other possible reasons for the loss of diversity in CWR at high latitude.

Although both interpretations could explain the decay in GD from the south to the north of China, only the first could explain the decrease in allele size from south to north, so we would argue that the first interpretation is the correct one.

Indica-japonica divergence of *Oryza rufipogon* Griff. in China

The question of whether differentiation between *indica* and *japonica* occurred in Chinese CWR has been widely discussed by researchers. Second (1982, 1985a) reported that it was *japonica*-like. Sun *et al*. (2002) compared germplasm collected from south Asia, southeast Asia and China and concluded that China had both *indica*-like and *japonica*-like accessions. Cai *et al*. (1996) also found that some *indica*-like accessions existed in southern China, and that those in the north were *japonica*-like. In our results, both *japonica*- or *indica*-like alleles and haplotypes indicated that there were both *japonica*-like and *indica*-like variants in Chinese CWR populations. These variants were more *japonica*-like in the MSPs and more *indica*-like in the SSPs. It was known that more *japonica* than *indica* rice was planted from north to south in China, and that gene flow may occur, especially from cultivated rice to CWR when the cultivated rice was planted near the CWR population (Messeguer *et al*., 2001; Song *et al*., 2003b). Was the *indica-japonica* divergence within CWR a result of gene flow from cultivated rice to CWR? No, in our opinion. First, it is well known that CWR in China grew in a region lower than 600 m, and the wild rice of JX-HuN1 and HuN2 in a region lower than 250 m. No *japonica* was planted near populations of wild rice in Jiangxi and Hunan. Thus, the possibility that *japonica* rice transferred genes to CWR is very slim. In addition, most of the research reported that the pollen of cultivated rice could be transferred only 30–40 m (Messeguer *et al*., 2001; Song *et al*., 2003b), and most of the individuals were sampled from a point at least 50 m away from the population edge. So, we deduced that *indica*-like and *japonica*-like divergence was caused mainly by natural adaptation to different climatic regions rather than the influence of gene flow from cultivated rice. There are two major hypotheses about the domestication of the two subspecies of Asian cultivated rice, *indica* and *japonica*. Recent research favors the multiple origin model (Second, 1982; Londo *et al*., 2006) over the single origin model (Ting, 1957; Oka, 1974; Oka and Morishima, 1982). The multiple origin model suggests that *japonica* was domesticated in China and *indica* in south Asia. If so, it is difficult to understand some of our results. We found that both *indica* and *japonica,* especially *indica,* have the least genetic distance with GD-GX1; all *indica* and *japonica* haplotypes could be found and the frequency of *indica* haplotypes was higher than that of *japonica* haplotypes in GD-GX1. Further research is needed to consider the possibility that Chinese *indica* was domesticated independently in China. If that proves to be the case then GD-GX1, with the closest genetic relationship with cultivated rice, might be the closest to its ancestral population, and *japonica* may have been domesticated, in turn, from the domesticated *indica* or from wild rice at high latitude.

Acknowledgements

This work was supported by the National Basic Research Program of China ('973' Program, 2004CB117201).

References

Bassam BJ, Caetano-Anolles G, Gresshoff PM (1991). Fast and sensitive silver staining of DNA in polyacrylamine gels. *Anal Biochem* **196**: 80–83.

Cai HW, Wang XK, Morishima H (2004). Comparison of population genetic structures of common wild rice (*Oryza rufipogon* Griff.), as 490 revealed by analyses of quantitative traits, allozymes and RFLPs. *Heredity* **92**: 409–417.

Cai HW, Wang XK, Pang HH (1996). Isozyme studies on the Hsien-Keng (indica-japonica) differentiation of the common wild rice (Oryza rufipogon Griff. In China. In: Wang XK, Sun CQ (eds). *Origin and Differentiation of Chinese Cultivated Rice*. China Agricultural University Press: Beijing, pp 147–152.

Coulon A, Guillot G, Cosson JF, Angibault JM, Aulagnier S, Cargnelutti B *et al.* (2006). Genetic structure is influenced bylandscape features: empirical evidence from a roe deer population. *Mol Ecol* **15**: 1669–1679.

Davierwala AP, Chowdari KV, Kumar S, Reddy AP, Ranjekar PK, Gupta VS (2000). Use of three different marker systems to estimate genetic diversity of Indian elite rice varieties. *Genetica* **108**: 269–284.

England PR, Annette VU, Robert JW, David JA (2002). Microsatellite diversity and genetic structure of fragmented populations of the rare, fire-dependent shrub *Grevillea macleayana*. *Mol Ecol* **11**: 967–977.

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.

Frankel OH (1984). Genetic perspectives of germplasm conservation. In: Arber W, Llinmeasee K, Peacock WJ *et al*. (eds). *Genetic Manipulation: Impact on Man and Society*. Cambridge University Press: London, pp 161–170.

Frankham R (2003). Genetics and conservation biology in *Comptes rendus*. *Biologies* **326**: 22–29.

Gao LZ, Chen W, Jiang WZ, Ge S, Hong DY, Wang XK (2000a). Genetic erosion in the Northern marginal population of the common wild rice *Oryza rufipogon* Griff. and its conservation, revealed by the change of population genetic structure. *Hereditas* **133**: 47–53.

Gao LZ, Ge S, Hong DY, Lin R, Tao G, Xu Z (2002a). Allozyme variation and conservation genetics of common wild rice (Oryza rufipogon Griff.) in Yunnan, China. *Euphytica* **124**: 273–281.

Gao LZ, Hong DY, Ge S (2000b). Allozyme variation and population genetic structure of common wild rice *Oryza rufipogon* Griff. in China. *Theor Appl Genet* **101**: 494–502.

Gao LZ, Schaal BA, Zhang CH, Jia JZ, Dong YS (2002b). Assessment of population genetic structure in common wild rice *Oryza rufipogon* Griff. using microsatellite and allozyme markers. *Theor Appl Genet* **106**: 173–180.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**: 1631–1638.

Huang QG, Sheng JS (1996). *Catalog of Rice Germplasm Resources in China (Oryza rufipogon Griff.)*. Agriculture Publishing Company of China Press: Beijing.

Huang H, Xiaolan L, Siming W, Guoqin H (1998). Process and driving forces of adverse separation of the distributing area of wild rice and rice (*O. sativa*. L) during the past 2000 years I.

Relationship between the distributing area of wild rice and rice (*O. sativa*. L) and population distribution in ancient and modern times. *Acta Ecologica sinica* **18**: 119–126.
Hurlbert SH (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology* **52**: 577–586.
Kalinowski ST (2005). HP-RARE1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.
Liu K, Muse S (2004). PowerMarker: new genetic data analysis software, version 2.7 (http://www.powermarker.net).
Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *P Natl Acad Sci USA* **103**: 9578–9583.
Lu H, Liu Z, Wu N, Berne S, Saito Y, Liu B *et al.* (2002). Rice domestication and climatic change: phytolith evidence from East China. *Boreas* **31**: 378–385.
Messeguer J, Fogher C, Guiderdoni E, Marfà V, Català M, Baldi G *et al.* (2001). Field assessments of gene flow from transgenic to cultivated rice (*Oryza sativa* L.) using a herbicides resistance gene as tracer marker. *Theor Appl Genet* **103**: 1151–1159.
Nei M, Tajima F, Tateno Y (1983). Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* **19**: 153–170.
Oka HI (1974). Experimental studies on the origin of cultivated rice. *Genetics* **78**: 475–486.
Oka HI, Morishima H (1982). Phylogenetic differentiation of cultivated rice. XXIII. Potentiality of wild progenitor to evolve the *indica* and *japonica* types of rice cultivars. *Euphytica* **31**: 41–50.
Pang HH, Chen CB (2002). *Oryza rufipogon Griff. Germplasm Resource in China*. Science and Technology Publishing Company of Guangxi Province Press: Nanning.
Pang HH, Cai HW, Wang XK (1995). Morphological classification of common wild rice (*Oryza rufipogon* Griff.) in China. *China Acta Agron Sinica* **21**: 17–24.
Primmer CR, Veselov A, Zubchenko AJ, Poututkin A, Bakhmet I, Koskinen MT (2006). Isolation by distance within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River in northwest Russia. *Mol Ecol* **15**: 653–666.
Pritchard JK, Matthew S, Peter D (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
Rosenberg NA (2002). Destruct: a program for the graphical display of structure results (http://www.cmb.usc.edu/_noahr/distruct.html).
Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH *et al.* (1995). Microsatellites are generally longer in humans compared to their homologs in nonhuman-primates—evidence for directional evolution at the microsatellite loci. *Nat Genet* **10**: 337–343.
Scott OR, Bendich AJ (1988). Extraction of DNA from plant tissues. *Plant Mol Bio Ma* **A6**: 1–10.
Second G (1982). Origin of the genetic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Jpn J Gene* **57**: 25–57.
Second G (1985a). Evolutionary relationship in the *sativa* group of *Oryza* based on isozyme data. *Genet Sel Evol* **17**: 89–114.
Second G (1985b). Geographic origins, genetic diversity and the molecular clock hypothesis in the *Oryzeae*. *Nato As1 Series* **G5**: 41–56.
Sheng JS, Huang QG (1991). *Catalog of Rice Germplasm Resources in China*. Agriculture Publishing Company of China Press: Beijing.
Song ZP, Xu X, Wang B, Chen JK, Lu BR (2003a). Genetic diversity in the northernmost *Oryza rufipogon* Griff. populations estimated by SSR markers. *Theor Appl Genet* **107**: 1492–1499.
Song ZP, Lu BR, Zhu YQ, Chen JK (2003b). Gene flow from cultivated rice to the wild species *Oryza rufrpogon* under experimental field conditions. *New Phytologist* **157**: 657–665.
Sun CQ, Wang XK, Li ZC, Yoshimura A, Iwata N (2001). Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. *Theor Appl Genet* **102**: 157–162.
Sun CQ, Wang XK, Yoshimura A, Doi K (2002). Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). *Theor Appl Genet* **104**: 1335–1345.
Thingsgaard K (2001). Population structure and genetic diversity of the amphiatlantic haploid peatmoss *Sphagnum affine* (Sphagnopsida). *Heredity* **87**: 485–496.
Ting Y (1957). Origin and evolvement of Chinese cultivated rice in China. *China Acta Agro Sinica* **8**: 243–260.
Vigouroux Y, Matsuoka Y, Doebley J (2003). Directional evolution for microsatellite size in maize. *Mol Biol Evol* **20**: 1480–1483.
Wang XK, Sun CQ, Cai HW, Huang YH (2004). A study on classification and genetic diversity of common wild rice in different Asian countries. In: Yang QW, Cheng DZ (eds). *Studies and Applications of Wild Rice in China*. China Meteorological Press: Beijing, pp 7–117.
Waples RS (1995). Evolutionary significant units and the conservation of biological diversity under the endangered species. *Acta Am Fish Soc Symp* **17**: 8–27.
Xiao JH, Grandillo S, Ahn SN, McCouch SR, Tanksley SD (1996). Genes from wild rice improve yield. *Nature* **384**: 223–224.
Yu P, Li ZC, Zhang HL, Cao YS, Li DY, Wang XK (2003). Sampling strategy of primary core collection of common wild rice (*Oryza rufipogon* Griff.) in China. *J China Agr U* **8**: 37–41.
Yuan LP, Virmani SS, Mao CX (1989). Hybrid rice: achievements and further outlook. In: Box PO (ed). *Progress in Irrigated Rice Research*. International Rice Research Institute: Manila, The Philippines, pp 219–223.
Zhao Z, Piperno DR (2000). Late pleistocene/Holocene environments in the middle Yangtze River Valley, China and rice (*Oryza sativa* L.) domestication: the phytolith evidence. *Geoarchaeology* **15**: 203–222.
Zhou HF, Xie ZW, Ge S (2003). Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon* Griff.) in China. *Theor Appl Genet* **107**: 332–339.

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