

Genetic structure and differentiation of *Oryza sativa* L. in China revealed by microsatellites

Dongling Zhang · Hongliang Zhang · Meixing Wang · Junli Sun · Yongwen Qi · Fengmei Wang · Xinghua Wei · Longzhi Han · Xiangkun Wang · Zichao Li

Received: 12 November 2008 / Accepted: 13 July 2009 / Published online: 1 August 2009
© Springer-Verlag 2009

Abstract China is one of the largest centers of genetic diversity of *Oryza sativa* L. in the world. Using a genetically representative primary core collection of 3,024 rice landraces in China, we analyzed the genetic structure and intraspecific differentiation of *O. sativa*, and the directional evolution of SSR. The genetic structure was investigated by model-based structure analysis and construction of neighbor-joining phylogenetic tree. Comparison between genetic structure and predefined populations according to Ting's taxonomic system revealed a hierarchical genetic structure: two distinct subspecies, each with three ecotypes and different numbers of geo-ecogroups within each ecotype. Two subspecies evidently resulted from adaptation to different environments. The different cropping systems

imposed on the subspecies led to further differentiation, but the variation within each subspecies resulted from different causes. *Indica*, under tropical-like or lowland-like environments, exhibited clear differentiation among seasonal ecotypes, but not among soil-watery ecotypes; and *japonica* showed clear differences between soil water regime ecotypes, but not among seasonal ecotypes. Chinese cultivated rice took on evident directional evolution in microsatellite allele size at several aspects, such as subspecies and geographical populations. *Japonica* has smaller allele sizes than *indica*, and this may partly be the result of their different domestication times. Allele size was also negatively correlated with latitude and altitude, and this may be interpreted by different mutation rates, selection pressures, and population size effects under different environments and cropping systems.

Communicated by J. Yu.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1112-4) contains supplementary material, which is available to authorized users.

D. Zhang · H. Zhang (✉) · M. Wang · J. Sun · Y. Qi · F. Wang · X. Wang · Z. Li (✉)

Key Laboratory of Crop Genomics & Genetic Improvement of Ministry of Agriculture, and Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, People's Republic of China
e-mail: zhangl@cau.edu.cn

Z. Li
e-mail: lizichao@cau.edu.cn

X. Wei
China National Rice Research Institute, Hangzhou 310006, People's Republic of China

L. Han
Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, People's Republic of China

Introduction

China, with abundant rice genetic resources, is one of the centers of origin of *Oryza sativa* L. (Oka 1988), and one of centers of genetic diversity. China is exclusive in having developed large and nearly equal numbers of *indica* and *japonica* varieties. There were 56,020 accessions of cultivated rice in the ex situ germplasm collections of China until 1993 (ICGR CAAS 1996), 90.2% of which (50,526 accessions) were landraces, cultivated varieties (cultivars) grown and preserved by traditional farmers (Zeven 1998). Landrace varieties contain greater genetic diversity than elite cultivars (or commercial cultivars) and represent an intermediate stage in domestication between wild and elite cultivars (Londo et al. 2006). The abundant variation provides an important reservoir of genetic diversity and potential sources of beneficial alleles for rice breeding and improvement.

It is generally believed that the two primary subspecies of rice, *indica* and *japonica*, are the products of separate domestications from the ancestral species, *O. rufipogon* Griff. The evidence came from studies of biochemical traits (Second 1982) and hybrid sterility (Kato et al. 1928), subsequently supported by molecular analyses (Cheng et al. 2003), especially some new results on the evolutionary analysis of some domestication-related genes (Sweeney et al. 2007; Li et al. 2006; Konishi et al. 2006; see also review article by Kovach et al. 2007). However, studies using large numbers of cultivars indicated that some varieties do not belong to either of the two types, but they were generally related to differentiation of seasonal (Zhang et al. 2007a, b), soil-watery (Zhang et al. 2007a, b), or geographic ecotypes (Glaszmann 1987; Sano and Morishima 1992). Cai and Morishima (2000) even suggested that geographical differentiation might precede seasonal ecotype differentiation to produce groups like aman, aus, and boro. Using model-based structure analysis, Garris et al. (2005) divided *O. sativa* into five distinct groups, viz. *indica*, aus, aromatic, temperate *japonica*, and tropical *japonica*. This indicated that, in addition to the apparent differentiation between *indica* and *japonica*, other genetic structures or taxa, have evolved, but still have not been clearly identified and recognized. Therefore, it was of interest to (1) investigate the genetic structure of Chinese cultivated rice in relation to the process of intra-specific differentiation, and further, to understand the dynamics of intra-specific differentiation of cultivated rice, and (2) to gain information able to guide plant breeders in choosing parents for crossing between subspecies or ecotypes, and thus providing a more rational basis for expanding the gene pool and for identifying materials that harbor alleles of value for rice improvement.

The purpose of this study was (1) to investigate the population structure of 3,024 rice landrace accessions using SSRs, and (2) to examine the evolutionary relationships among intra-specific populations with emphasis on the intra- and inter-subspecific differentiation.

Materials and methods

Plant materials and predefined populations

The primary core collection of Chinese rice landraces, comprising 3,024 accessions, was used in this study. This collection was sampled from the basic collection of 50,526 accessions of rice landraces collected from all 32 provinces of China and represented 95% of the diversity in the basic collection at the morphological level (Li et al. 2003). In the database of Chinese Crop Germplasm Information System (http://icgr.caas.net.cn/cgris_english.html), all the

accessions of rice landraces were documented by Ting's five-level hierarchical taxonomy, i.e., two subspecies (*indica* or *hsien* versus *japonica* or *keng*), three photo-thermic ecotypes or seasonal ecotypes (early sown, medium sown, or late sown), two soil water regime ecotypes (hereafter called soil-watery ecotypes for simplicity) (lowland or upland), two endosperm types (waxy or non-waxy), and cultivars (Ting 1957). Studies of rice landraces from Yunnan and Guizhou have shown that the population structure of cultivated rice does not relate to endosperm types. According to the documented identities based on the first three levels of Ting's taxonomy, 3,024 accessions were grouped into 14 populations, which we called predefined populations. Table 1 shows the distribution of cultivars in predefined populations.

DNA extraction and SSR genotyping

Thirty-day seedlings of 10–20 individuals for each accession were collected and ground in liquid nitrogen. DNA was extracted from the ground tissues using the CTAB method. Thirty-six SSRs (three on each of the 12 chromosomes, Table S1) were randomly selected to analyze population structure and genetic diversity. SSR loci are particularly useful for the study of population structures and demographic histories of domesticated species in natural populations (Goldstein and Schlotterer 1999) because of their genomic abundance, conservation of distinctive flanking sequence across closely related species, apparent selective neutrality, and high variability, that facilitate the detection of greater diversity than by using equal numbers of RFLP, AFLP, or SNP loci (Akagi et al. 1997).

The volume of the PCR reaction system was 15 μ l, containing 20 ng DNA template. The PCR procedure followed Panaud et al. (1996) with slight modifications:

Table 1 Distribution of predefined taxonomic identities for 3,024 accessions according to the first three levels of Ting's taxonomic system

	<i>Japonica</i>	<i>Indica</i>	Total
Early			
Lowland	214	327	541
Upland	93	70	163
Medium			
Lowland	344	408	752
Upland	168	91	259
Late			
Lowland	403	448	851
Upland	235	167	402
Unknown	15	41	56
Total	1,472	1,552	3,024

(1) initial denaturation at 94°C for 5 min; (2) 39 cycles with three steps—denaturing at 94°C for 1 min, annealing at 55–67°C (depending on the specific primer sequence) for 1 min, and extending at 72°C for 2 min; (3) a final extension at 72°C for 5 min after the above 39 cycles. The amplified products were denatured at 95°C for 5 min, cooled on ice, and subsequently run on 8% denatured polyacrylamide gel at 70 W. One of accessions with a particular allele was randomly selected as the check to identify other accessions with the same allele. All checks and a molecular weight marker (PUC19 DNA digested by MspI) were included in the rest runs. Finally, the molecular weights of all the alleles were estimated by a gel-run including all detected alleles and a standard molecular weight marker of a 10 bp DNA Ladder (Invitrogen, Carlsbad, California). Gels were visualized by a silver-staining method. In the case of non-amplification, PCRs were repeated to exclude failures. If both PCRs failed, we recorded a null allele. The non-simple genotype could be classified into two types here: one with two alleles and one with more than two alleles. The proportion of the former was $7.61 \pm 1.31\%$. The proportion of the latter was less than 1%. When more than two alleles occurred at a particular locus in one accession, they were amplified and run again. Two distinctly more stable and denser alleles in both experiments were selected for the structure analysis. If there were no distinctly more stable and denser alleles between two experiments, the genotype at that locus in the accession was recorded as missing.

Population structure and differentiation analysis

The model-based programs STRUCTURE (Pritchard et al. 2000; Falush et al. 2003) and INSTRUCT (Gao et al. 2007) were used to infer population structure using a burn-in length of 10,000, run length of 100,000. The former used a model allowing for admixture and correlated allele frequencies and the latter used a model allowing for population structure and population selfing rates. Ten independent simulations were run for each K (the number of clusters). The graphical display of the STRUCTURE results was generated using Distruct software (Rosenberg 2002; <http://www.cmb.usc.edu/noahr/distruct.html>). To determine the K value, we used both the LnP(D) value and Evanno's ΔK (Evanno et al. 2005). LnP(D) is the log likelihood of the observed genotype distribution in K clusters and can be output by STRUCTURE simulation (Pritchard et al. 2000). The inferred k is ideal when the highest LnP(D) occurs, but it is sometimes difficult to find the highest LnP(D) before a much larger K is examined. In such cases, a clear knee is recommended, but it is usually somewhat subjective. Evanno's ΔK took consideration of the variance of LnP(D) among repeated runs and usually can indicate the ideal K .

The suggested $\Delta k = M((L(k+1) - 2L(k) + L(k-1)) / S[L(k)])$, where $L(k)$ represents the k th LnP(D), M is the mean of 10 runs, and S their standard deviation. Principal coordinate analysis was made using software NTSYS 2.1 (Rohlf 1997) to investigate the tridimensional structure of landrace rice in China. Phylogenetic reconstruction was based on the neighbor-joining method using Nei's D_A , an efficient distance estimator in obtaining correct topologies (Takezaki and Nei 1996), and implemented in PowerMarker version 3.25 (Liu and Muse 2004; <http://www.powermarker.net>), because the neighbor-joining method is especially suited for datasets comprising lineages with largely varying rates of evolution (Saitou and Nei 1987). Using inferred populations, a two level AMOVA (analysis of molecular variance) was performed in Arlequin ver 3.11 (Excoffier et al. 2005) with 1,000 permutations and sum of squared size differences as molecular distance. To investigate the population differentiation, F_{ST} (Weir and Cockerham 1984) among populations was calculated and tested using FSTAT 2.9.3.2 (Goudet 2001).

Neutrality test

The Ewens–Watterson test of neutrality is able to detect deviations from a neutral equilibrium model. This is achieved by detecting whether there is either a deficit or an excess of expected homozygosity, estimated according to the observed allele frequencies ($Obs.F$) relative to that estimated according to null allele frequencies ($Nul.F$) by simulating random neutral samples (Ewens 1972; Watterson 1978). This test was performed using the program PopGene Version 1.31 (Yeh et al. 1999) and the algorithm given in Manly (1985). $L95$ and $U95$, which are, respectively, the lower and upper limits of the 95% confidence interval of $Nul.F$, were estimated through 1,000 simulations in this program. If $Obs.F$ is within this confidence interval, the locus is neutral; otherwise, it is not.

Summary statistics

PowerMarker was also used to calculate the average number of alleles (N_a), Nei's gene diversity index (H_{sk}) (Nei 1987), and polymorphism information content (PIC) (Anderson et al. 1993) values and stepwise mutation index. The average standardized molecular weight of the PCR products in each population was calculated as in Vigouroux et al. (2003). Allele richness per locus (R_s , an estimator independent of the sample size; Hurlbert 1971) was also estimated. R_s corresponds to the number of different alleles found when $2N$ gene copies are sampled at a particular locus. If $2N$ gene copies are totally examined at the locus, the expected number of different alleles in a sample size of $2n$ can be estimated with the formula:

$$R_s = \sum_{i=1} \left[1 - \frac{\binom{2N - N_i}{2n}}{\binom{2N}{2n}} \right],$$

where N_i represents the number of occurrences of the i th allele among the $2N$ sampled gene copies, and n is fixed as the number of individuals in the smallest group for the locus. The effective population sizes (N_e) for populations at different altitudes and latitudes were estimated using LDNe (Waples and Do 2008).

Statistical analysis

The statistical significance of the difference of the estimators measuring the genetic diversity (including N_a , R_s , H_{sk} , and PIC) and allele size was assessed using the t test and Wilcoxon paired test across loci. Using the inline applets of regression in Microsoft Excel, we also performed a regression analysis between genetic diversity (H_{sk} and N_a), effective population size (N_e), and latitude/altitude. Using TFPGA (Miller 1997), we fulfilled the Mantel test (Sokal 1979) to estimate the correlation between genetic differentiations and geographic distances (among administrative districts), and between genetic differentiations and differences of latitude/altitude.

Results

Inference of genetic structure of cultivated rice in China

Among the 36 SSR loci we used to infer the genetic structure of cultivated rice, 543 alleles were totally detected in 3,024 varieties. The number of alleles per locus varied widely among marker loci, ranging from 2 (RM60) to 31 (RM247) with an average of 15.1. When we ran the STRUCTURE simulation using all 3,024 accessions, the LnP(D) value increased with K from 1 to 15, but showed an evident knee at $k = 2$ (Fig. 1a). This implied that there must be two distinctly divergent populations. There was a sharp peak of Evanno's ΔK at $K = 2$, and a small peak at $k = 6$ (Fig. 1b–c). Due to the distinct divergence at $K = 2$ and the further increase of LnP(D), we performed further simulations using two inferred populations independently. Their ΔK indicated there were three subpopulations in each of the two inferred populations (Fig. 1e, g, and Fig. 2), corresponding to the six populations at $k = 6$ when simulation was run using all 3,024 accessions in one (Fig. S1). Simulations using INSTRUCT indicated that there was the lowest deviance information criterion at $K = 2$ and a secondary lower one at $k = 6$ (Fig. S2). The simulation using INSTRUCT and STRUCTURE gave consistent ancestry and structure patterns (Fig. S3). Thus, Chinese landrace

rice firstly differentiated into two distinct populations, which correspond to the subspecies: *indica* and *japonica* ($k = 2$, Fig. 2). The individual membership between the predefined populations and the inferred populations ($k = 3$ within each subspecies in Fig. 2 and Fig. S4, and K from 3 to 6 in Fig. S1) indicated that three inferred populations in *japonica* corresponded to Ting's soil-watery ecotypes, including lowland, upland, and medium (or intermediate) types; but those in *indica* corresponded to his seasonal ecotypes, including early, late, and medium (or intermediate) sown types (Fig. 2 and Fig. S1). Thus we defined the six inferred populations as six ecotypes: *japonica* lowland (*Jap.L*), *japonica* upland (*Jap.U*), *japonica* medium (*Jap.M*), *indica* early (*Ind.E*), *indica* late (*Ind.L*), and *indica* medium (*Ind.M*).

The principal coordinate analysis (Fig. S5) and neighbor-joining tree (Fig. S6) for 3,024 accessions both indicated that differentiation between two subspecies was the most apparent. Differentiation between soil-watery ecotypes within *japonica* was more apparent than within *indica*, and differentiation among seasonal ecotypes within *indica* was more apparent than within *japonica*. The more apparent differentiation among seasonal ecotypes within *indica* than within *japonica* was also confirmed by the differences in days to heading between three locations using 84 accessions (Table S2). This will be discussed later. More individuals of the predefined upland types in both predefined subspecies seemed to have admixed genomes; that is, among the predefined upland accessions, *japonica* contained a higher proportion of inferred *indica* genome (red circle at $K = 2$, Fig. 2).

The continuous increase of LnP(D) after $k = 3$ within the inferred *indica* and *japonica* populations implied there are subtle sub-structures within the six ecotypes (Fig. 1d, f). Therefore, we performed further independent simulations within each of the six ecotypes. These revealed different numbers of geo-ecogroups within each of the six ecotypes. Investigation of these subtle differentiation within ecotypes will be a good subject for future research.

Genetic diversities and differentiation of inferred populations and their geographical distributions

The genetic diversity in each inferred population and subpopulation was measured by four estimators, viz. allele number (N_a), allelic richness (R_s), gene diversity (H_{sk}), and PIC (Table 2). *Japonica* showed more variation in allelic richness ($z = 1.69$, $P < 0.05$), but lower heterogeneity as measured by H_{sk} ($z = 1.73$, $P < 0.05$) than *indica*. Within *japonica*, lowland rice showed significantly lower H_{sk} than upland rice ($t = 2.044$, $P < 0.05$); and within *indica*, early varieties showed significantly lower H_{sk} than late varieties ($t = 3.331$, $P < 0.01$). The other estimators did not show

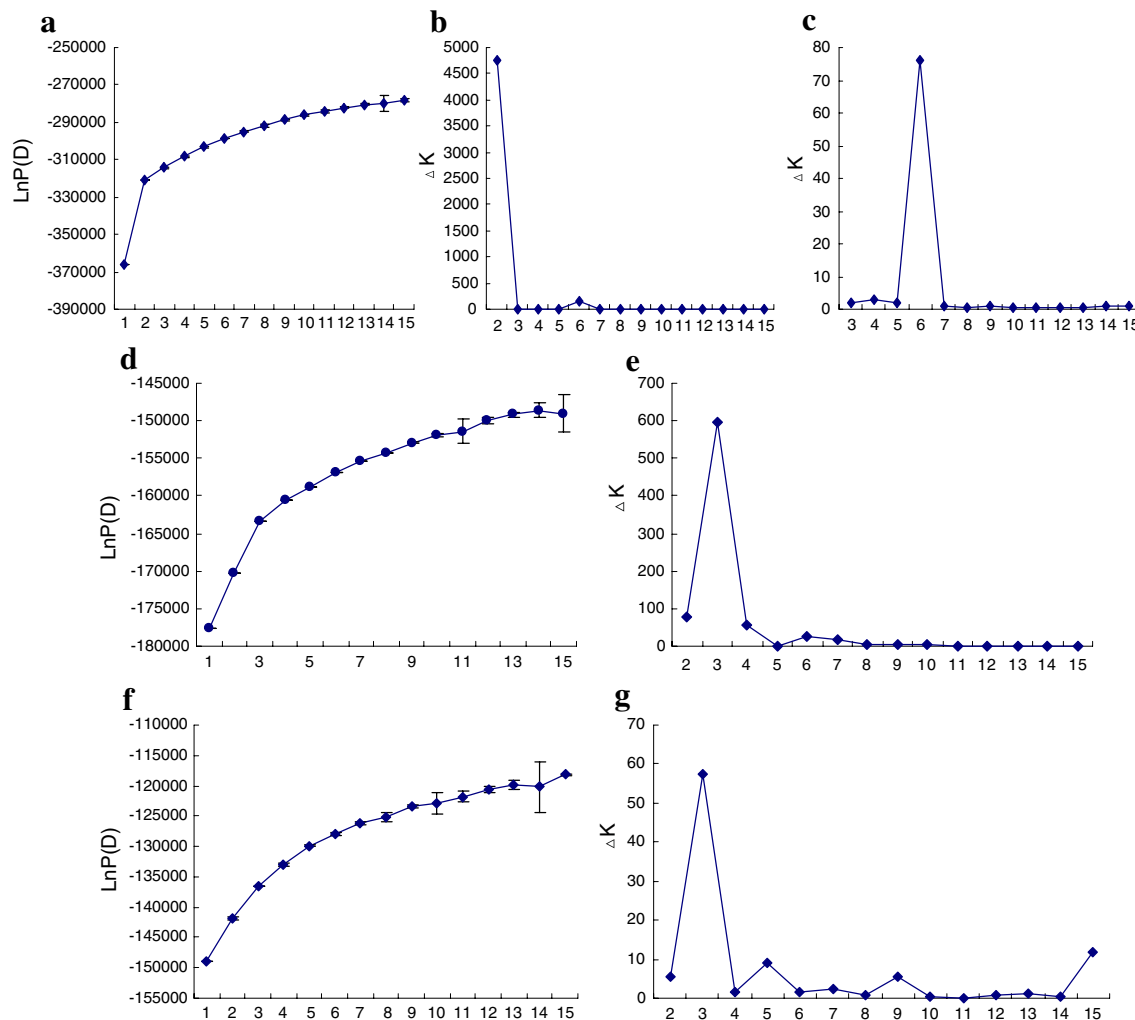


Fig. 1 The average LnP(D) and ΔK over 10 repeats of STRUCTURE simulations. **a–c** LnP(D) with $k = 1–15$, ΔK with $k = 2–15$, and ΔK with $k = 3–15$ for simulations using all 3,024 accessions; **d–e** LnP(D)

with $k = 1–15$ and ΔK with $k = 2–15$ for inferred *indica* population; **f–g** LnP(D) with $k = 1–15$ and ΔK with $k = 2–15$ for inferred *japonica* population

significant differences between ecotypes within each subspecies. Genetic diversity was negatively correlated with latitude (N_a , $r^2 = 0.5696$, $P < 10^{-5}$; H_{sk} , $r^2 = 0.5554$, $P < 10^{-5}$) and altitude (N_a , $r^2 = 0.7325$, $P < 10^{-5}$; H_{sk} , $r^2 = 0.5714$, $P < 0.001$) (Fig. 3). To eliminate the effect of subspecies (there were more *japonica* at high latitudes and altitudes than *indica*), we also examined these relationships within each subspecies and obtained similar results (Fig. S7).

AMOVA conducted using inferred subspecies as groups, and inferred ecotypes within subspecies as populations indicated that 7.58% of the variation in landrace rice in China could be attributed to differentiation among ecotypes (Table 3), and 36.65% was attributed to the differentiation between *indica* and *japonica*. F_{ST} among ecotypes showed that the differentiation among ecotypes was significant (Table 4). Both F_{ST} (Table 4) and principal coordinate

analysis (Fig S5) showed clearer differentiations among ecotypes within *japonica* than within *indica*. Mantel tests between genetic distances and geographical differences (measured by latitude ($r = 0.3783$, $P < 0.01$), altitude ($r = 0.7323$, $P < 0.001$), and geographical distance ($r = 0.4423$, $P < 0.01$)) exhibited significantly positive correlations (Fig. 4). Altitude had a large impact on genetic differentiation.

Neutrality test within different populations

Although SSR variation itself is mostly neutral, it has been reported that some SSR loci showed selective sweep reflecting selection or adaptation (Ellegren 2004). To detect whether there is selective sweep within each inferred population, we made neutrality tests using the program PopGene. Among the 16 loci that deviated from the null

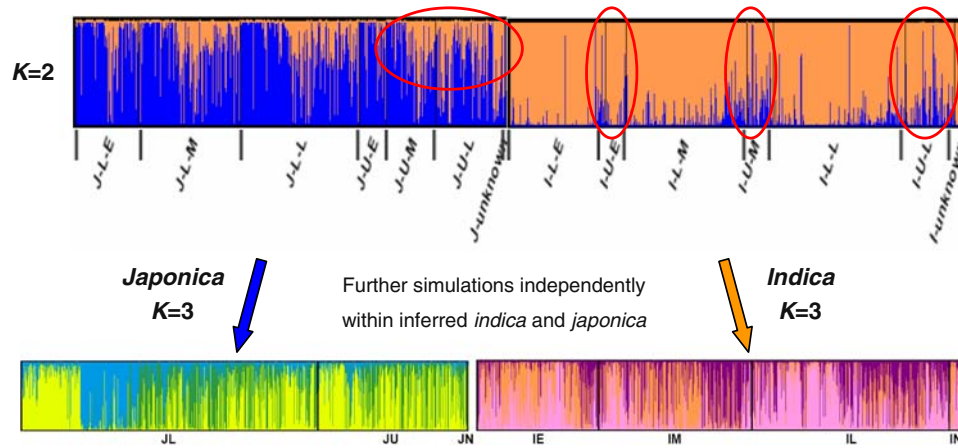


Fig. 2 Relationship between the predefined and inferred populations. The two inferred clusters ($k = 2$) resulted from simulation using all 3,024 accessions in one and correspond to *indica* and *japonica*, respectively. Then three clusters ($k = 3$) were inferred within inferred *indica* and *japonica* independently. Naming convention of the predefined populations when $k = 2$: the leftmost symbols, $I = indica$,

$J = japonica$; the symbols between two dashes, L lowland, U upland; the rightmost symbols: E early, M medium, L late. The labels when $k = 3$: JL *japonica* lowland, JU *japonica* upland, JN *japonica* unknown lowland or upland, IE early seasonal *indica*, IM medium seasonal *indica*, IL late seasonal *indica*, IN – *indica* unknowing seasonal type. Each vertical line represents one individual

Table 2 Genetic diversity of inferred ecotypes

Ecotypes	N	N_a	R_s	H_{sk}	PIC
<i>Jap.L</i>	477	10.4	8.4154	0.5517	0.5251
<i>Jap.U</i>	603	12.4	8.3457	0.6018	0.5758
<i>Jap.M</i>	362	11.6	8.3795	0.6356	0.6098
<i>Ind.E</i>	583	12	8.3802	0.6249	0.5941
<i>Ind.L</i>	583	12.6	8.3705	0.6808	0.6497
<i>Ind.M</i>	416	12	8.3446	0.6637	0.6358

N The number of accessions, N_a the number of alleles, R_s allele richness (Hurlbert 1971), H_{sk} Nei's unbiased gene diversity (1987), PIC polymorphic information coefficient

hypothesis of neutrality at $P = 0.05$, only four were shared by *indica* and *japonica* ecotypes; four were detected in one or two *indica* ecotype(s), and eight were detected in one or two *japonica* ecotype(s) (Table 5); more than half were detected in upland *japonica*, only three were detected in early *indica*.

Allele size in inferred populations

Directional evolution of microsatellites was reported in both animals (Rubinsztein et al. 1995) and plants (Vigouroux et al. 2003). We investigated the Stepwise Mutation Indices of the 36 loci (Table 5). Most of them showed evidence of stepwise mutation: 26 loci with $SMI > 0.80$ (the threshold used by Garris et al. 2005), 18 loci with $SMI > 0.95$, and 7 loci with $SMI < 0.7$. Among 13 loci common to the study of Garris et al. (2005) and the present work, only *rm219* and *rm25* showed dramatically different stepwise mutation properties (Table 5). The average standardized allele sizes in

japonica were significantly smaller ($t = 39.18$, $P \geq 0$) than those in *indica* except for the two intermediate ecotypes which were not significantly different from each other (Fig. 5).

Regression between allele size and altitude showed that allele size significantly decreased with increasing altitude (Fig. 6a). Similarly, allele size significantly decreased with increasing latitude, especially in the range 28° – $34^\circ N$ (Fig. 6b). To eliminate the effect of the distribution of subspecies, we studied the regression between allele size and altitude or latitude within each subspecies (Fig. S8). The correlations among latitudes were weaker than when using all accessions, and the correlation between allele size and altitude disappeared. In addition, the allele sizes in *japonica* were significantly smaller than those in *indica* at the same altitudes and latitudes (-0.34 vs. 0.46 , $t = 15.08$, $P < 10^{-30}$).

Discussion

Genetic structure and reclassification of *Oryza sativa* L. in China

It is well known that cultivated rice has differentiated into two subspecies, *indica* and *japonica* (Oka 1988; Sano and Morishima 1992). In addition to the evident differentiation of subspecies, it is commonly accepted that there are other possible intra-specific differentiations. For example, Ting's five-level taxonomic system (Ting 1957) described a hierarchical differentiation of cultivated rice in China. Although this system was commonly accepted in China

Fig. 3 Regressions between genetic and geographic estimators. **a** Number of alleles and latitude, **b** Nei's unbiased gene diversity and latitude, **c** number of alleles and altitude, **d** Nei's unbiased gene diversity and altitude

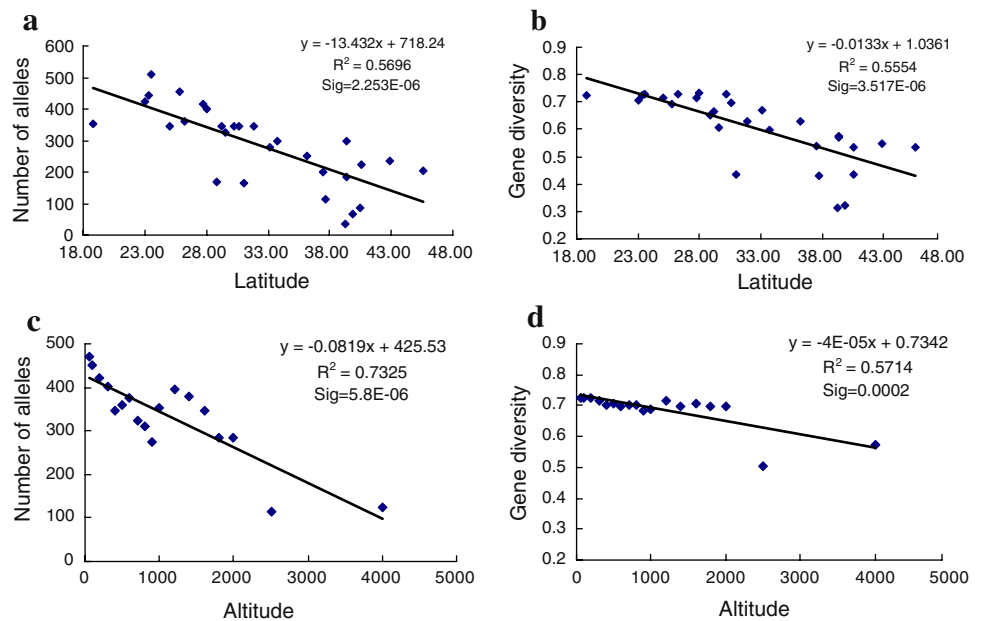


Table 3 Analysis of molecular variance (AMOVA) among inferred populations

Source of variation	df	Variance component	Percent of variation
Among groups ^a	1	606.6	36.65
Among populations ^b	4	125.5	7.58
Among individuals	3,018	675.2	40.80
Within individuals	3,024	247.6	14.96
Total	6,047	1,654.9	

^a Groups were defined by two inferred subspecies

^b Populations were defined by inferred ecotypes

Table 4 F_{ST} among six ecotypes

Ecotypes	Jap.M	Jap.U	Jap.L	Ind.E	Ind.M	Ind.L
Jap.M		0.0710	0.1038	0.1855	0.1866	0.1693
Jap.U	*		0.0668	0.2463	0.1596	0.2286
Jap.L	*	*		0.2837	0.1918	0.2669
Ind.E	*	*	*		0.0776	0.0484
Ind.M	*	*	*	*		0.0838
Ind.L	*	*	*	*	*	

Pairwise F_{ST} estimates appear above the diagonal and the significant level of pairwise F_{ST} appear below the diagonal

* Significant at $P < 0.05$

and the differentiation between *indica* and *japonica* was proved at many levels, no molecular evidence has been found to support such intra-subspecific differentiation. Using varieties from parts of China and SSR markers, we earlier reported intra-subspecific differentiation patterns that differed from that of Ting (Zhang et al. 2007a, b).

Our current results confirmed the primary differentiation between *indica* and *japonica* within *O. sativa*. Structure simulations using STRUCTURE and INSTRUCT showed distinct structures at $K = 2$, at which the two subspecies were differentiated. The differentiation between the subspecies accounted for 36.6% of the genetic diversity in cultivated rice in China. Within the *indica* and *japonica* populations, however, the substructures were different from those indicated by Ting's (1957) and Cheng–Wang's (Cheng et al. 1984b) taxonomic systems, where seasonal ecotypes were firstly classified under both subspecies. In the present study, *japonica* appeared more distinct in terms of differentiation among soil-watery ecotypes (upland versus lowland), whereas *indica* was more clearly subdivided by seasonal ecotypes (early, medium versus late season). These different differentiations within the two subspecies may be attributed to their different growth environments and the corresponding cropping systems. It is well known that rice cultivars in temperate countries or regions like Japan, Korea, and northern China (roughly higher than 32°N) are exclusively *japonica*, whereas rice cultivars in tropical and subtropical regions are predominantly *indica*. In regions south of Yangtze River in China, both *indica* and *japonica* are grown, but *indica* is planted in the lowlands or valleys and *japonica* is grown on the hills (Xu et al. 1974; Khush 1997). For example, in Yunnan (21°8'32" N–29°15'8" N in southwest of China), *indica* varieties are mainly grown below 1,400 m, and *japonica* varieties mainly above 1,800 m (up to 2,200 m), with both types being grown between the two levels (Xu et al. 1974). This distribution led scientists to believe that *indica* is synonymous to tropical rice and *japonica* represents temperate rice.

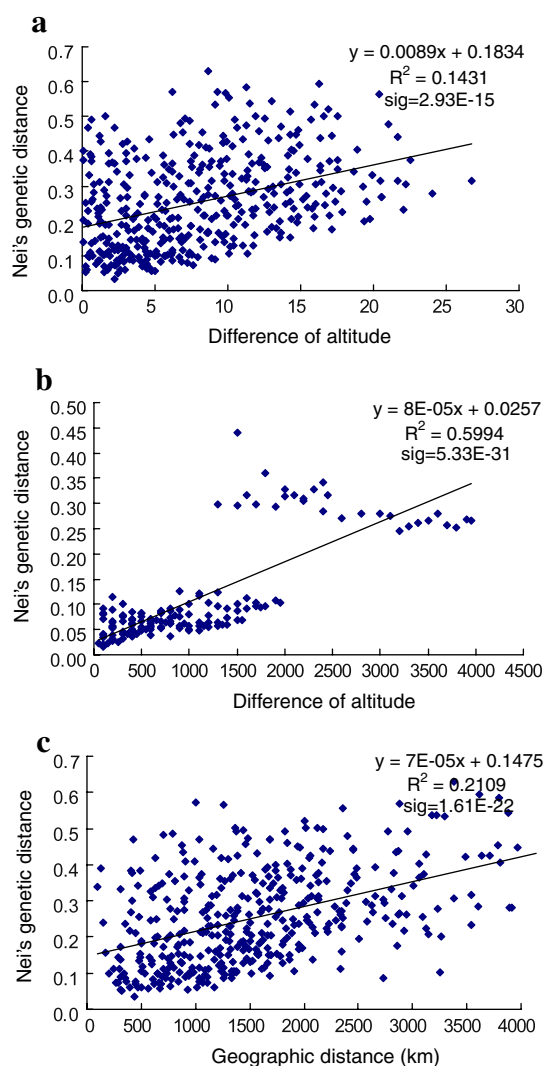


Fig. 4 Relationship between matrix of genetic distance and matrix of geographic estimators. **a** Nei's genetic distance and difference in latitude, **b** Nei's genetic distance and difference in altitude, **c** Nei's genetic distance and geographic distance

At low altitude or low latitude where *indica* prevails, water resources are generally ample, thus it is not necessary to develop water-economical varieties, i.e., upland rice. Consequently, there is no apparent soil-watery differentiation within *indica*. In fact, there are quite fewer upland varieties in *indica* than in *japonica* under Ting's taxonomic system. However, the bountiful heat and light resources available in many parts of south China provide conditions supporting the development of cropping systems with complicated seasonal patterns, such as double or triple cropping on an annual basis. Specifically adapted seasonal ecotypes are required to meet the needs of these intensive cropping systems. Such seasonal ecotypes derive their adaptation from different photothermic responses. The behavior of the various ecotypes are determined by three characteristics, viz the basic vegetative phase (BVP),

photoperiod sensitivity (PS), and temperature sensitivity (TS) (Cheng et al. 1984a). Studies on BVP and PS were reviewed in detail by Vergara and Chang (1985). TS has been supported by only a few reports (AGRTPR 1978; Cheng et al. 1984a; Nakagawa et al. 2005). In general, shorter day-lengths and higher temperature can shorten days to heading. To estimate the photoperiod sensitivity and temperature sensitivity of the six inferred subgroups, we investigated the days to heading of 84 accessions sown at different locations (Sanya, Hangzhou, and Beijing; information on sowing date, day length, and temperature is shown in Table S2, and Fig. S9). The results showed that the days to heading clearly varied between the inferred subgroups in *indica*, but not in *japonica* (Table S2). The early varieties were strongly temperature-sensitive and weakly photoperiod-sensitive. From Sanya to Hangzhou, the days to heading were not distinctly changed due to the counterbalance of longer day-lengths and higher temperatures (Table S2). From Hangzhou to Beijing, the days to heading were distinctly prolonged due to longer day-lengths and lower temperature. The late varieties were strongly sensitive to both temperature and photoperiod, but the temperature sensitivity takes effect only when the day-length is short (especially shorter than 12 h). Thus the days to heading are distinctly prolonged from Sanya to Hangzhou, as day-lengths get longer and temperatures become lower. But difference of days to heading is less between Hangzhou and Beijing than between Sanya and Hangzhou, due to the less difference in day-lengths between Hangzhou and Beijing than between Sanya and Hangzhou, and the weaker temperature sensitivities of late varieties under long day-lengths (longer than 13 h, Fig. S9). From Sanya to Beijing, the obviously longer day-lengths (4 h) caused distinctly prolonged days to heading. Similar to the sensitivities to both day-length and temperature for the middle varieties, their differences of days to heading among three places lied between those for early varieties and late varieties. As noted by Vergara and Chang (1985), the days to heading for rice are also affected by other factors (such as seedling vigor), so days to heading within inferred subgroups show large variances. In summary, *indica* shows more differentiation among seasonal ecotypes than among soil-watery ecotypes.

In the current research, many varieties predefined as upland *indica* were identified as *japonica* through simulation of STRUCTURE; hence, more blue lines (red-circled) in the predefined populations I-U-E (*indica*, upland, early), I-U-M (*indica*, upland, medium), and I-U-L (*indica*, upland, late) than other predefined *indica* populations ($K = 2$, Fig. 2). This accords with the facts that varieties found in upland areas in Thailand, Burma, India, and other tropical Asian countries, usually show characteristics of *japonica*. Glaszmann (1987) found that most upland

Table 5 Neutrality test and the stepwise Mutation Index of 36 SSR loci

Locus	Neutrality test							SMI ^b	SMI ^c
	<i>Jap.M</i>	<i>Jap.U</i>	<i>Jap.L</i>	<i>Ind.E</i>	<i>Ind.M</i>	<i>Ind.L</i>	Gao ^a		
rm134								0.8175	
rm164							√	0.5928	N
rm18								1.0000	
rm190								1.0000	Y
rm206	*	*	*	*	*	*	√	0.6141	N
rm211							√(J)	0.7025	
rm216								0.9036	
rm219								0.6090	Y
rm223			*				√	0.9979	
rm224					*	*	√	0.7106	
rm225							√	0.9524	Y
rm23								0.9997	
rm231	*	*					√	0.6392	N
rm235					*	*	√	0.9162	
rm241	*	*	*		*	*	√	0.6334	
rm242	*					*	√	0.9706	
rm244							√	0.8080	N
rm247	*	*	*	*	*	*	√(I)	1.0000	Y
rm249			*				√	0.6954	
rm25							√	0.9993	N
rm251		*	*					0.8839	Y
rm253							√	1.0000	
rm254								0.9765	
rm255							√	0.9945	
rm258		*	*				√(J)	0.6913	
rm262								0.8089	
rm267		*						1.0000	
rm270				*				1.0000	Y
rm296								1.0000	Y
rm335		*						0.9957	
rm42						*		0.7503	
rm5								0.9881	Y
rm60								1.0000	
rm71		*						0.8459	
rm81a								0.9817	
rm82								0.8619	
Locus number	5	9	6	3	5	7	17(3)	36	13

* Indicates loci selective at $P = 0.05$ in neutrality test

^a Gao—the marked loci were also used by Gao and Innan (2008), and (I) or (J) indicate non-neutrality in *indica* or *japonica*

^b SMI—Stepwise mutation index estimated by us

^c SMI—Stepwise mutation index, Garris et al. (2005), N—SMI lower than 0.8, Y—SMI equal to or higher than 0.8

cultivars belonged to an enzymic group corresponding to *japonica*. Furthermore, a survey of native or traditional varieties of *O. sativa* in Africa by Kochko (1987) using isozymes (37 presumed loci) again indicated that varieties grown in upland field conditions tended to have genotypes

like *japonica*. These reports demonstrate that the *japonica* type is more often associated with upland conditions, and implies that the origin of the *japonica* type might be related to adaptive selection under upland or non-standing water conditions, where limited resources of water, heat, and

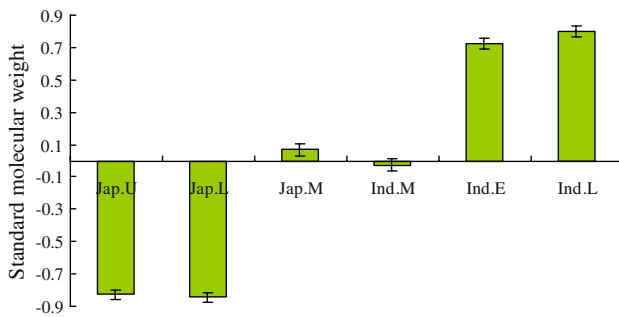


Fig. 5 Average standardized molecular weights of microsatellites and standard errors for the six inferred ecotypes estimated using 26 SSR markers with a stepwise mutation index higher than 0.80

light required the development of varieties with drought tolerance. Other varieties planted in standing water evolved into lowland varieties. At the same time, insufficient light and heat restricted the development of diverse seasonal ecotypes within *japonica*. Nevertheless, because of the wide distribution from low to high latitudes and thus distinct differences in day length, *japonica* also showed weak differentiations for seasonal adaptation, but they are nearly accordant with geographical distribution. For example, *japonica* in south China mainly consists of late varieties, but northern varieties are predominantly early or medium types. In addition, the clearer differentiations among *japonica* ecotypes than among *indica* ecotypes, as revealed by F_{ST} , principal coordinate analysis and selective SSR number, also indicate that *japonica* underwent more selection than *indica* during the domestication process. In the present study (Fig. 2), the majority of predefined upland *japonica* was inferred as upland type, whereas about half of the predefined lowland *japonica* was inferred as upland. This might reflect the different types of selection that were imposed on them. The upland type (in fields without standing water) is under directional selection in soil water regime, thus accumulating specific genotypes adapted to upland conditions. The lowland type (in fields with standing water) is not under directional selection, and thus includes genotypes that are both the same or different from upland genotypes. Thus varieties with lowland

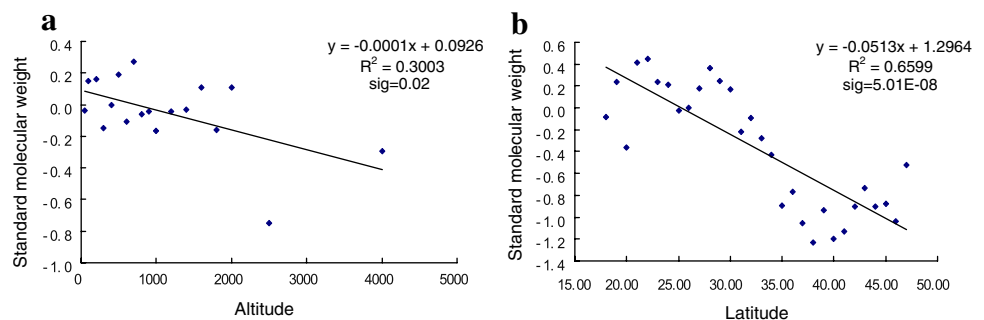
genotypes are seldom collected in upland fields, but varieties with upland genotypes do occur in lowland fields. In summary, *japonica* appeared more differentiations among soil-watery ecotypes than among seasonal ecotypes.

Our results therefore indicate a hierarchical differentiation and taxonomic system for cultivated rice (at least in China). The intra-specific differentiation of cultivated rice is as follows: *indica* and *japonica* were domesticated independently in tropical lowland areas and temperate upland areas, respectively. Complicated cropping systems and ample heat and light resources in tropical-like or lowland-like areas permitted the development of distinct seasonal ecotypes in *indica*, but in contrast, simple cropping systems and the restricted soil water regimes in temperate-like and/or upland-like areas drove the development of distinct soil-watery ecotypes with less clear seasonal ecotypes in *japonica*. This was consistent with other results using Chinese rice landraces from a part of China (Zhang et al. 2007a, b). Thus the hierarchical taxonomic system for cultivated rice should be subspecies (*indica* versus *japonica*), followed by ecotypes (distinct seasonal ecotypes for *indica* and distinct soil-watery ecotypes for *japonica*), and finally eco-geographical types. Further studies on the ecotypes and eco-geographical types may provide information on the utilization of intra-subspecific heterosis.

Directional evolution of microsatellites in cultivated rice in China

The phenomenon that new microsatellite mutations tend to increase molecular size has been documented in plants, such as chickpea (*Cicer arietinum* L.) by Udupa and Baum (2001), and maize by Vigouroux et al. (2003), and in humans (Weber and Wang 1993; Primmer et al. 1996; Cooper et al. 1999). In a separate study (Wang et al. 2008), the genetic structure of common wild rice (*O. rufipogon* Giff.) was analyzed using the same 36 loci as in the present study and 19 loci were proved to be stepwise (with a threshold of SMI = 0.80). Among those 19 loci, only Rm231 showed a clear difference in SMI (0.8704 for wild rice, 0.6392 for cultivated rice). Cultivated rice (*O. sativa* L.)

Fig. 6 Regression between microsatellite allele size estimated using 26 SSR markers with SMI >0.80 and altitude (a) or latitude (b)



appeared to have significantly higher molecular weights than its ancestor (the average standardized molecular weights were 0.29 and -0.99 , respectively; $t = 39.38$, $P \geq 10^{-10}$). In the current study, cultivated rice had an evident further directional evolution in allele size in both subspecies and geography. The molecular weight of *japonica* was significantly smaller than *indica*, and also increased with decreased latitude and altitude. However, according to the hypothesis of directional evolution of microsatellite size, it is unclear whether directional evolution at SSR loci corresponds to the directional selection history of those populations, or, whether there are other interpretations for such results.

Do the populations with smaller allele size at SSR loci originate earlier than those with larger ones? Positive answers were given by others (Udupa and Baum 2001; Vigouroux et al. 2003; Weber and Wang 1993; Primmer et al. 1996; Cooper et al. 1999). However, for populations within cultivated rice, such as subspecies and other populations, the reasons might be different. Common wild rice was perennial with both sexual and asexual reproductive strategies; the asexual role was dominant under natural conditions (Oka 1992; Xie et al. 2001). Thus wild rice had fewer generations to evolve than its cultivated counterpart over the same time periods. If microsatellite mutations showed an almost 2:1 ratio of gain in size over losses (Banchs et al. 1994; Weber and Wang 1993), SSR alleles should be smaller in the later originating populations of cultivated rice than in those originating earlier, assuming they originated from the same or similar populations of common wild rice and that domestication events occurred in the time longer than several thousands years. If so, *indica* (with larger alleles) should have originated earlier than *japonica* in China, as was proposed by Ting (1957); and *indica* should be more similar to wild rice in many aspects such as morphology, genetic material, and growth habits. It is well known that the distribution of the two subspecies is related to altitude and latitude and allele size is correlated with altitude and latitude (Fig. 6, Fig. S8). Hence we need to determine how much of the difference in allele size between the two subspecies is related to their different occurring times of domestication and how much it relates to changes after domestication. The most influential environmental effects probably relate to geography (altitude and latitude), under which adaptation to the specific natural conditions occurred. Comparison of the different allele sizes between the two subspecies with those at the same altitudes and latitudes indicated that the environmental difference accounted for about 30% of the difference in allele size between the two subspecies: $\{[0.55 - (-0.61)] - [0.46 - (-0.34)]\} / [0.55 - (-0.61)]$. This was also indicated by the regression of allele size and altitude or latitude using all 3,024 collectively, or using

each subspecies separately (Fig. 6, Fig. S7). More evidence, especially from the wild rice, is needed to prove whether the other 70% of the difference in allele size between the two subspecies derives from their different times of domestication.

Temperature is the most obvious environmental factor that changed over different latitudes and altitudes. The general macroevolutionary patterns show that most higher taxonomic groups originate in the tropics where temperatures are high (Jablonski 1993), and that speciation rates decrease with decreasing temperatures from the equator to the poles (Flessa and Jablonski 1996; Stehli et al. 1969). Higher temperature is expected to result in higher mutation rate (Bleiweiss 1998; Gillooly 2005). We observed a negative correlation between allele number and altitude/latitude (Fig. 3a, c and Fig. S7). It is possible that populations at lower altitude/latitude subjected to higher temperature exhibit higher mutation rates, and consequently have larger SSR allele sizes. A negative correlation between allele size and altitude was also found in maize (Vigouroux et al. 2003).

In addition, we speculate that selection was a contributing factor to the directional evolution of some SSR loci. Under more rigid conditions at higher altitude/latitude, newly occurring variations are seldom inherited to the next generation due to their rare frequency, and thus increase in the gain in allele size is slower than at lower altitudes/latitudes. In our current study, the number of alleles and Nei's diversity index at low altitudes/latitudes were higher than those at high altitudes/latitudes. Selection or founder effects are two main ones among the possible reasons for these phenomena. At high latitudes, the fewer variations (Fig. S7) and smaller effective population size (Fig. S10) might imply that the decrease in genetic diversity with latitude mainly results from the founder effect occurring during dispersal of cultivated rice from low latitude to high latitudes (The dispersal has been argued by researchers, see review by Khush 1997). However, at higher altitudes, the fewer variations but larger effective population sizes (Fig. S10) imply that SSR variations suffer from more intense selection under more rigid conditions, such as lower temperature, less heat, and lower water availability. Given that most SSR variations are neutral, the argument for selectivity of SSR may make some researchers uncomfortable, but it has been reported that some SSR loci could reflect selection or adaptation effects (Ellegren 2004). Tam et al. (2005) reported that 18.75% of SSR loci are non-neutral in a tomato collection whereas all SSR loci were neutral in a pepper collection. In the current work, the proportion of non-neutral SSR loci was 44% across all ecotypes, and ranged from 8 to 25% within ecotypes. Seventeen of the SSR markers were also used by Gao and Innan (2008). Among 10 non-neutral loci, 2 (rm247 and rm258) were

selective in both studies, but 7 additional loci were detected only by us, and 1 by Gao and Innan. The different neutralities for the same loci detected in different studies could result from different selective events occurring in different populations. Of course, it may be too early to conclude a “selective sweep” for a specific locus if we simply detect a deviation from neutrality.

Acknowledgments We thank Professor Robert A McIntosh, University of Sydney, for suggested revisions to the manuscript. This research was supported by the National Basic Research Program of China (“973” Program, 2004CB117201), Program for Changjiang Scholars and Innovative Research Team in University, and Program of Introducing Talents of Discipline to Universities (111-2-03).

References

- AGRTPR—a study group on the response to temperature and photoperiod of rice. (1978) Studies on the photoperiod and temperature ecology of rice cultivars in China. Science Press, Beijing, China, 154 pp (in Chinese)
- Akagi H, Yokozeki Y, Nagaki A, Fujimura T (1997) Highly polymorphic microsatellites of rice consist of AT repeats and a classification of closely related cultivars with these microsatellite loci. *Theor Appl Genet* 94:61–67
- Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36:181–186
- Banchs L, Bosch A, Guimera J, Lázaro C, Puig A, Estivill X (1994) New alleles at microsatellite loci in *CEPH* families mainly arise from somatic mutations in the lymphoblastoid cell lines. *Hum Mutat* 3:365–372
- Bleiweiss R (1998) Slow rate of molecular evolution in high-elevation humming birds. *Proc Natl Acad Sci USA* 95:612–616
- Cai HW, Morishima H (2000) Diversity of rice varieties and cropping system in Bangladesh deepwater areas. *Jpn Agric Res Q* 34:221–226
- Cheng KS, Lu YX, Luo J, Huang NW, Lin GR, Wang XK (1984a) Studies on indigenous rice in Yunnan and their utilization—the photo-thermo-response patterns of rices in Yunnan province and their relation to the classification of maturity groups. *Acta Agron Sin* 10(3):163–171
- Cheng KS, Zhou JW, Lu YX, Huang NW, Wang XK (1984b) Studies on the indigenous rices in Yunnan and their utilization II. Revised classification of Asian cultivated rice. *Acta Agron Sin* 10:271–280
- Cheng CY, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E (2003) Polyphyletic origin of cultivated rice: based on the interspersed pattern of SINES. *Mol Biol Evol* 20:67–75
- Cooper G, Burroughs NJ, Rand DA, Rubinsztein DC, Amos W (1999) Markov chain Monte Carlo analysis of human Y-chromosome microsatellites provides evidence of biased mutation. *Proc Natl Acad Sci USA* 96:11916–11921
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat Genet* 5:435–445
- 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
- Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3:87–112
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
- 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
- Flessa KW, Jablonski D (1996) The geography of evolutionary turnover: a global analysis of extant bivalves. In: Jablonski D, Erwin DH, Lipps JH (eds) *Evolutionary palaeobiology*. University of Chicago Press, Chicago, pp 376–397
- Gao LZ, Innan H (2008) Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. *indica* and ssp. *japonica*, demonstrated by multilocus microsatellites. *Genetics* 179:965–976
- Gao H, Williamson S, Bustamante CD (2007) An MCMC approach for joint inference of population structure and inbreeding rates from multi-locus genotype data. *Genetics* 176:1635–1651
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169:1631–1638
- Gillooly JF, Allen AP, West GB, Brown JH (2005) The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc Natl Acad Sci USA* 102:140–145
- Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. *Theor Appl Genet* 74:21–30
- Goldstein DB, Schlotterer C (1999) *Microsatellites: evolution and applications*. Oxford University Press, Oxford
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www.unil.ch/izea/software/fstat.html>
- Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577–586
- ICGRCAAS (Institute of Crop Germplasm Resources of China Academy of Agricultural Science) (1996) Catalogue of rice germplasm resources in China (1988–1993). China Agricultural Press, Beijing
- Jablonski D (1993) The tropics as a source of evolutionary novelty through geological time. *Nature* 364:142–144
- Kato S, Kosaka H, Hara S (1928) On the affinity of rice varieties as shown by the fertility of rice plants. *Bull Sci Fac Agric Kyushu Univ* 3:132–147
- Khush G (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* 35:25–34
- Kochko AD (1987) Isozyme variability of traditional rice in Africa. *Theor Appl Genet* 73:675–682
- Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. *Science* 312:1382–1396
- Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. *Trends Genet* 23:578–587
- Li ZC, Zhang HL, Cao YS, Qiu ZE, Wei XH, Tang SX, Yu P, Wang XK (2003) Studies on the sampling strategy for the primary core collection of Chinese indigenous rice. *Acta Agron Sin* 29:20–24
- Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. *Science* 311:1936–1939
- 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* L. *Proc Natl Acad Sci USA* 103:9578–9583
- Manly BFJ (1985) *The statistics of natural selection*. Chapman and Hall, London
- Miller MP (1997) Tools for population genetic analyses (TFPGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data

- Nakagawa H, Yamagishi J, Miyamoto N, Motoyama M, Yano M, Nemoto K (2005) Flowering response of rice to photoperiod and temperature: a QTL analysis using a phenological model. *Theor Appl Genet* 110:778–786
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Oka HI (1988) *Origin of cultivated rice*. Japan Science Society Press, Tokyo
- Oka HI (1992) Ecology of wild rice planted in Taiwan III. Differences in regenerating strategies among genetic stocks. *Bot Bull Acad Sin* 33:133–140
- Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet* 252:597–607
- Primmer CR, Ellegren H, Saino N, Moller AP (1996) Directional evolution in germline microsatellite mutations. *Nat Genet* 13:391–393
- 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, New York
- Rosenberg NA (2002) Distruct: 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, Margolis RL, Ross CA, Ferguson-Smith MA (1995) Microsatellite evolution—evidence for directionality and variation in rate between species. *Nat Genet* 10:337–343
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sano R, Morishima H (1992) *Indica-japonica* differentiation of rice cultivars viewed from the variation in key characters and isozymes with special reference to land races from the Himalayan hilly areas. *Theor Appl Genet* 84:266–274
- Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Jpn J Genet* 57:25–57
- Sokal RR (1979) Testing statistical significance of geographic variation patterns. *Syst Zool* 28:227–232
- Stehli FG, Douglas DG, Newell ND (1969) Generation and maintenance of gradients of taxonomic diversity. *Science* 164:947–949
- Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD et al (2007) Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet* 3:1418–1424
- Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399
- Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien M (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theor Appl Genet* 110:819–831
- Ting Y (1957) The origin and evolution of cultivated rice in China. *Acta Agron Sin* 8:243–260
- Udupa SM, Baum M (2001) High mutation rate and mutational bias at (TAA)_n microsatellite loci in chickpea (*Cicer arietinum* L.). *Mol Gen Genet* 265:1097–1103
- Vergara BS, Chang TT (1985) *The flowering response of the rice plant to photoperiod*, 4th edn. IRRI, Los Banos, Philippines, p 61
- Vigouroux Y, Matsuoka Y, Doebley J (2003) Directional evolution for microsatellite size in maize. *Mol Biol Evol* 20:1480–1483
- Wang MX, Zhang HL, Zhang DL, Qi YW, Fan ZL, Li DY, Pan DJ, Cao YS, Qiu ZE, Yu P, Yang QW, Wang XK, Li ZC (2008) Genetic structure of *Oryza rufipogon* Griff. in China. *Heredity* (<http://www.nature.com/hdy/journal/vaop/ncurrent/full/hdy200861a.html>)
- Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Mol Ecol Res* 8:753–756
- Watterson GA (1978) The homozygosity test of neutrality. *Genetics* 88:405–417
- Weber J, Wang C (1993) Mutation of short tandem repeats. *Hum Mol Genet* 8:1123–1128
- Weir BS, Cockerham CC (1984) Estimation F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Xie ZW, Lu YQ, Ge S, Hong DY, Li FZ (2001) Clonality in wild rice (*Oryza rufipogon*, Poaceae) and its implications for conservation management. *Am J Bot* 88:1058–1064
- Xu XH, Wang GC, Zheng XB, Wang HX (1974) A report on the vertical distribution of the rice varieties in Simao, Yunnan. *Acta Bot Sin* 16:208–222
- Yeh FC, Yang RC, Boyle T (1999) PopGene version 1.31, Microsoft Window-based freeware for population genetic analysis. (<http://www.ualberta.ca/~fyeh/>)
- Zeven AC (1998) Landraces: a review of definitions and classifications. *Euphytica* 104:127–139
- Zhang DL, Zhang HL, Wei XH, Qi YW, Wang MX, Sun JL, Ding L, Tang SX, Qiu ZE, Cao YS, Wang XK, Li ZC (2007a) Genetic structure and diversity of *Oryza sativa* L. in Guizhou, China. *Chin Sci Bull* 52:343–351
- Zhang HL, Sun JL, Wang MX, Liao DQ, Zeng YW, Shen SQ, Yu P, Wang XK, Li ZC (2007b) Genetic structure and phylogeography of rice landraces in Yunnan, China revealed by SSR. *Genome* 50:72–83