Theor Appl Genet (2003) 108:131–140
DOI 10.1007/s00122-003-1400-3

ORIGINAL PAPER

**S. B. Yu · W. J. Xu · C. H. M. Vijayakumar · J. Ali · B. Y. Fu · J. L. Xu · Y. Z. Jiang · R. Marghirang · J. Domingo · C. Aquino · S. S. Virmani · Z. K. Li**

# Molecular diversity and multilocus organization of the parental lines used in the International Rice Molecular Breeding Program

Received: 21 April 2003 / Accepted: 27 June 2003 / Published online: 6 September 2003


**Abstract** One hundred and ninety three parental lines obtained from 26 countries for an international rice molecular breeding program were evaluated using 101 well-distributed simple sequence repeat (SSR) markers. An overall genetic diversity of 0.68 and an average of 6.3 alleles per locus were revealed, indicating a high level of genetic variation in these lines. Cluster analysis of the 193 accessions showed three major groups and nine subgroups. Group I corresponded to the classical indica subspecies, whereas groups II and III belong to the japonica subspecies. Indica and japonica differentiation accounted for only 6.5% of the total variation in the entire sample and 93.5% was due to within-subspecies diversity. Differentiation among eco-geographic regions accounted for 24% of the diversity within the subspecies. Larger amounts of the eco-geographical differentiation were resolved within japonica than within indica. The largest indica-japonica differentiation based on the single locus level was detected by markers on chromosomes 9 and 12, while the smallest differentiation was detected by markers on chromosomes 4 and 8. Furthermore, genetic differences at the single-locus and two-locus levels, as well as components due to allelic and gametic differentiation, were revealed between indica and japonica and among the main geographic regions. The multilocus analysis in genetic diversity showed a higher proportion of variation caused by predominant non-random associations of different loci within and among the classified subspecies and geographic subdivisions. The results suggest that selection for eco-geographical adaptation on multilocus associations was largely responsible for the maintenance of the extensive variation in the primary gene pool of rice.

**Keywords** Microsatellite variation · Genetic diversity · Linkage disequilibrium · Multilocus structure · Rice (*Oryza sativa* L.)

Communicated by Q. Zhang

S.B. Yu and W.J. Xu contributed equally to the work

S. B. Yu · W. J. Xu · C. H. M. Vijayakumar · J. Ali · B. Y. Fu · J. L. Xu · Y. Z. Jiang · R. Marghirang · J. Domingo · C. Aquino · S. S. Virmani · Z. K. Li (✉)
Plant Breeding, Genetics, and Biochemistry Division,
International Rice Research Institute,
DAPO Box 7777, Metro Manila, The Philippines
e-mail: z.li@cgiar.org
Tel.: +63-2-8450563
Fax: +63-2-8911292

C. H. M. Vijayakumar
The Directorate of Rice Research,
Hyderabad, India

B. Y. Fu · J. L. Xu · Y. Z. Jiang · Z. K. Li
Chinese Academy of Agricultural Sciences,
12 South Zhong-Guan-Cun St., Beijing, China

*Present address:*
S. B. Yu, National Key Laboratory of Crop Genetic Improvement,
Huazhong Agricultural University,
Wuhan, China

## Introduction

Cultivated rice (*Oryza sativa* L.) is one of the most important staple foods for more than half of the world's population. Rice is composed of two major subspecies, indica and japonica, and several ecotypes, which are adapted to diverse environmental conditions ranging from upland to lowland and from tropical to temperate zones. With such a wide range of geographic distribution, rice is well-known for its rich genetic resources maintained worldwide with more than 250,000 germplasm collections of more than 20 cultivated and wild species (Evernson et al. 1998). In the past decades, many efforts have been made to assess the genetic diversity within *O. sativa* at both the phenotypic and molecular levels (Second 1982; Glaszmann 1987; Zhang et al. 1992; Yang et al. 1994; Cho et al. 2000; Liu et al. 1996; Zhu et al. 1998; Li et. al. 2000; Ni et al. 2002).

Although this large germplasm collection has provided a rich source of genetic diversity for rice improvement in the past, the proportion of the germplasm collections used in breeding programs has been very limited (Vaughan and

Jackson 1995). The under-use of this germplasm has resulted from breeder's tendency to choose parental lines based on phenotype and insufficient knowledge about the genetic diversity in the germplasm collections. The limited use of germplasm in breeding has led to a significant reduction in genetic diversity in commercially grown rice cultivars worldwide and the increased vulnerability of the crop to pests and abiotic stresses.

To solve this problem, two key questions remain to be answered regarding how much of this genetic diversity, particularly for complex phenotypes, is "useful" to breeders, and whether there is a link between genetic variation at the phenotypic level and at the molecular level. To answer these questions and to broaden the genetic base of rice cultivars in the major rice growing-areas of Asia, we started an international rice molecular breeding program in 1998 and selected 193 germplasm accessions worldwide as the parental lines of this program. We report here molecular characterization of the genetic diversity of the 193 parental lines using a well-distributed set of SSR markers at both single-locus and two-locus levels. Specifically, we would like to determine the geographic distribution and multilocus organization of this diversity as the first step for better understanding and use of these materials in the molecular breeding program.

## Materials and methods

### Rice materials

Table 1 summarizes the 193 accessions of *O. sativa* used in this study, which included 83 landraces, 15 breeding lines and 95 improved varieties from 26 countries of eight major rice-growing regions of the world. These included 8 accessions from northern Asia (Japan and Korea), 46 from East Asia (China, including Taiwan), 63 from Southeast Asia (Thailand, Vietnam, Bangladesh, Myanmar, Indonesia, Malaysia and The Philippines), 56 from South Asia (India, Pakistan, Nepal and Sri Lanka), 6 from Middle Asia (Iran), 6 from the Americas (USA, Suriname and Peru), 4 from Africa, and 4 from Europe. For easy analysis, the samples from Southeast Asia were further divided into two zones designated SEA1 (Thailand, Vietnam, Bangladesh and Myanmar) and SEA2 (Indonesia, Malaysia and The Philippines). Most accessions were obtained from the Rice Germplasm Center at the International Rice Research Institute (IRRI) and the others from the collaborative institutes or partners of the International Rice Molecular Breeding Program directly.

### SSR assay

DNA was extracted from fresh seedling leaves of each accession following a modified CTAB method (Saghai-Maroof et al. 1984). One hundred-and-one SSR markers were selected to represent the entire rice genome at about a 10–20-cM interval based on the published framework map (Akagi et al. 1996; Temnykh et al. 2000). Of these, 24 were derived from GenBank sequences and 77 identified from genomic libraries. Information regarding chromosomal localizations and repeat types of the SSRs can be found in the Web database (www.gramene.org/microsat).

PCR amplification was carried out in 20-$\mu$l reaction mixtures, each containing 50 ng of template DNA, 0.4 $\mu$M of each primer, 200 $\mu$M of each dNTPs, 1 x reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% Gelatin) and 2 units of *Taq* DNA polymerase. An MJ Research (Model 96) or Perkin Elmer (PE 9600) thermo-cycler was used along with the following PCR profile: an initial denaturation step of 5 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 55°C, 2 min at 72°C, and then a final extension at 72°C for 5 min. The PCR products were resolved on a 4% sequencing gel followed by silver staining following the standard protocol (Temnykh et al. 2000).

Silver-stained gels were scanned to capture digital images of the gels after air drying. The allele size of the amplified band for each SSR locus was determined based on its migration relative to the 50-bp DNA ladder and four check varieties (Shennong 89366, Manawthukha, IR64 and Babaomi) that were loaded in each gel as references. The capital letters A–K were used as codes of different alleles (based on size) detected in each locus. For genetic similarity estimates and cluster analysis, the allelic data were converted to a binary format, with the presence of a specific allele (band) scored as unity and its absence as zero.

### Data analyses

Genetic variation at each SSR locus was expressed as the genic diversity index or genic diversity: $H=(1-\sum P_i^2)$, where $P_i$ is the frequency of the $i$th allele of the locus. For a subdivided sample, the total diversity of the sample can be partitioned into its components resulting from differentiation among populations and diversity within populations (Nei 1973). Such partitioning was carried out in three ways: based on the clustered subspecies (e.g., indica and japonica), geography, and cross-classified groups in which the total sample was first classified into two subspecies and then each subspecies was subdivided into geographic subpopulations. The statistical significance of the differences in diversity and the number of alleles between groups over loci were assessed using the Wilcoxon signed-rank test (StatSoft 1991).

Genetic similarities were estimated from the allele binary format dataset using the Dice method (Nei and Li 1979). The 193 accessions were clustered based on the matrix of genetic similarities using the unweighted pair group method with arithmetic averages (UPGMA). The cluster analysis and dendrogram construction were performed with NTSYS-PC (version 1.8). The

**Table 1** Geographic origins of the 193 *O. sativa* L. accessions sampled for SSR assay

| Geographic regions | Countries | Number of accessions | | |
|---|---|---|---|---|
| | | Indica | Japonica | Total |
| North Asia (NA) | Japan, Korea | 3 | 5 | 8 |
| East Asia (EA) | China | 31 | 15 | 46 |
| Southeast Asia (SEA1) | Thailand, Vietnam, Bangladesh, Myanmar | 29 | 6 | 35 |
| Southeast Asia (SEA2) | Indonesia, Malaysia, Philippines (IRRI) | 26 | 2 | 28 |
| South Asia (SA) | India, Pakistan, Nepal, Sri Lanka | 39 | 17 | 56 |
| Middle Asia (ME) | Iran | 2 | 4 | 6 |
| Africa | Egypt, Guinea, Guinea-Bissau, Madagascar | 1 | 3 | 4 |
| America | USA, Suriname, Peru | 1 | 5 | 6 |
| Europe | France, Hungary, Italy, Portugal | 0 | 4 | 4 |
| Total | 26 | 132 | 61 | 193 |

varietal groups determined from the cluster analysis were used as subpopulations for further investigation.

Genetic variation among and within subpopulations was analyzed by the software package POPGENE (version 1.31) (Yeh et al. 1997). The number of alleles, genic diversity, genetic distance among subpopulations, linkage disequilibrium (LD) and multilocus structure, were estimated by the software. LD between pairs of SSR alleles was estimated using the standardized disequilibrium coefficient: $D'=D/D_{max}$. $D=P_{ij}-p_ip_j$ is the linkage disequilibrium coefficient, where $p_i$ is the frequency of the $i^{th}$ allele of the first locus; $p_j$ is the frequency of the $j$th allele of the second locus; $P_{ij}$ is the frequency of the gametic type with the $i$th allele at the first locus and the $j$th allele at the second. D has values between −1 to 1, and $D_{max}$ is the highest possible value of the D, which equals the min $[p_i(1-p_j), (1-p_i)p_j]$. Chi-square tests were used to determine the significances of these estimates (Weir 1990). The multilocus structure of the sample was evaluated using Ohta's (1982) method in which population subdivision was determined by partitioning the total variance ($D_{IT}^2$) of pairwise two-locus data into two components: one due to allelic differentiation ($D_{ST}^2$) and another corresponding to LD within subpopulations ($D_{IS}^2$). Alternatively, the total variance ($D_{IT}^2$) can be viewed as composed of a portion accounted for by gametic differentiation at paired loci among subpopulations ($D'_{IS}^2$), and a portion representing LD in the whole population ($D'_{ST}^2$). Brown's analysis for multilocus associations among and within subpopulations was also performed with the POPGENE program. The multilocus association variances were partitioned into single-locus and two-locus components (Brown and Feldman 1981).

## Results

### SSR diversity in the sample

All 101 SSR loci were polymorphic and produced a total of 628 alleles among the 193 accessions assayed. The average number of alleles per locus was 6.25, ranging from 2 to 11. The average genic diversity over all SSR loci was 0.68, ranging from 0.13 (RM230 on chromosome 8) to 0.88 (RM21 on chromosome 11) (Table 2). The genic diversity at each SSR locus was significantly correlated with the number of alleles detected ($r$=0.624, $P$<0.001). The majority (73/101) of the SSRs are $(GA)_n$ di-nucleotide repeats. The number of repeats within the SSRs was not significantly correlated ($r$=0.195) with the genic diversity, but was significantly correlated ($r$=0.400, $P$<0.01) with the number of alleles at a locus.

### Classification of the sample

The 193 accessions were classified into nine clusters within three major groups by the UPGMA algorithm based on the genetic similarity matrix (Fig. 1). Group I contained 132 accessions from eight geographic regions, which could be further divided into three subgroups, GI-1, GI-2 and GI-3. Group II contained 50 entries that could be further clustered into three subgroups (GII-1, GII-2 and GII-3) plus one solitary genotype (Baobaomi from Yunnan, China). Group III had nine accessions, all originating in Northeast Asia (Japan, Korea and China) except one (Basmati 370) from India. Interestingly, a deepwater variety, Jalmagna from India, could not be classified into any group mentioned above and thus formed another single-solitary genotype.

Grouping of varieties revealed by this analysis generally agreed with the previous classification of varietal groups and/or subspecies based on the isozyme variation (Glaszmann 1987; Li and Rutger 2000). For example, GI-1 includes most modern indica cultivars (such as IR64, IR72, PSBR28 and Teqing), important parental lines of hybrid rice (such as Minghui63 and IR58025B), and several varieties derived from indica/japonica crosses (such as Milyang 23 and Zhong413). The entries in Group II apparently belonged to japonica and could be clearly grouped into different subclusters (e.g., GII-1, GII-2 and GII-3) representing several geographic regions (Iran, USA and Europe), except for two entries that appeared to have a different identity based on their phenotypes. Group III included japonica types from China and Northeast Asia. Thus, Group I corresponded to indica, and Groups II and III together comprised japonica for the subsequent analyses.

### Genetic diversity and differentiation among clustered groups

On average, the within-group diversity and among-group differentiation accounted for 68.2% and 21.8% of the total genetic diversity, respectively (Table 3). Both indica and japonica showed a high level of within-group diversity. The indica group had 6.1 alleles per locus and a diversity value of 0.641, and the japonica group had 5.6 alleles per locus and a diversity value of 0.661. The Wilcoxon test indicated that the overall diversity level did not differ significantly between the indica and japonica groups (Z=0.73, $P$=0.467).

SSR markers on individual chromosomes showed varied levels of diversity and subspecific differentiation (Table 3). The highest diversity (0.77) was detected for the SSRs on chromosome 6 and the lowest (0.52) on chromosome 8. The indica group showed a consistently greater number of alleles per locus than the japonicas, whereas the differences in genic diversity between the two subspecies appeared to vary from one chromosome to another. Partitioning of the genetic diversity showed a strong differentiation among the clustered groups within the two subspecies. The differentiation ($G_{ST}$) was much greater (21.5%) among different clusters within the japonica group than within the indica group (12.1%). This among-cluster differentiation also varied considerably across different chromosomes. The overall subspecific differentiation also varied across SSRs on different chromosomes, with those on chromosomes 9 and 12 showing the greatest $G_{ST}$ values, and those on chromosomes 8 and 4 showing the smallest (see Table 5).

Rare alleles with frequencies less than 0.05 accounted for a significant proportion (23.7%) of the total alleles detected and apparently contributed to the observed subspecific and geographic differentiations in the sampled accessions. For instance, a total of 31 alleles at 27 loci

**Table 2** Chromosomal location, number of alleles ($N_A$), genetic diversity (H), and among varietal group differentiation ($G_{ST}$) at each of 101 SSR loci in 193 rice accessions

| Locus | Chr. | $N_A$ | H | $G_{ST}$ | Locus | Ch. | $N_A$ | H | $G_{ST}$ |
|---|---|---|---|---|---|---|---|---|---|
| RM84 | *1* | 9 | 0.76 | 0.146 | RM30 | *6* | 5 | 0.69 | 0.381 |
| RM220 | *1* | 6 | 0.41 | 0.137 | RM51 | *7* | 6 | 0.77 | 0.306 |
| RM1 | *1* | 4 | 0.59 | 0.164 | RM11 | *7* | 6 | 0.80 | 0.351 |
| RM23 | *1* | 6 | 0.73 | 0.153 | OSR22 | *7* | 5 | 0.63 | 0.128 |
| RM259 | *1* | 6 | 0.60 | 0.268 | OSR4 | *7* | 8 | 0.55 | 0.073 |
| RM243 | *1* | 5 | 0.61 | 0.412 | RM10 | *7* | 4 | 0.58 | 0.260 |
| RM9 | *1* | 9 | 0.78 | 0.188 | RM234 | *7* | 7 | 0.83 | 0.225 |
| RM5 | *1* | 7 | 0.78 | 0.192 | RM18 | *7* | 6 | 0.77 | 0.145 |
| RM34 | *1* | 5 | 0.62 | 0.214 | OSR30 | *8* | 5 | 0.61 | 0.128 |
| RM237 | *1* | 5 | 0.65 | 0.076 | RM38 | *8* | 8 | 0.79 | 0.176 |
| RM200 | *1* | 5 | 0.74 | 0.241 | RM25 | *8* | 9 | 0.54 | 0.136 |
| OSR27 | *1* | 4 | 0.73 | 0.383 | RM42 | *8* | 4 | 0.36 | 0.175 |
| RM14 | *1* | 6 | 0.58 | 0.192 | RM256 | *8* | 3 | 0.27 | 0.034 |
| RM211 | *2* | 6 | 0.76 | 0.203 | RM80 | *8* | 8 | 0.85 | 0.08 |
| RM233 | *2* (*5*) | 3 | 0.52 | 0.516 | OSR7 | *8* | 4 | 0.58 | 0.17 |
| RM236 | *2* | 4 | 0.64 | 0.327 | RM230 | *8* | 3 | 0.13 | 0.03 |
| RM8 | *2* | 5 | 0.78 | 0.378 | RM219 | *9* | 9 | 0.83 | 0.261 |
| OSR8 | *2* | 5 | 0.72 | 0.18 | RM257 | *9* | 7 | 0.76 | 0.279 |
| OSR9A | *2* | 8 | 0.50 | 0.146 | RM242 | *9* | 8 | 0.85 | 0.281 |
| RM27 | *2* | 4 | 0.68 | 0.176 | RM201 | *9* | 4 | 0.70 | 0.543 |
| RM263 | *2* | 8 | 0.84 | 0.286 | OSR28 | *9* | 8 | 0.82 | 0.337 |
| RM221 | *2* | 4 | 0.66 | 0.458 | RM215 | *9* | 4 | 0.53 | 0.131 |
| RM6 | *2* | 6 | 0.77 | 0.375 | RM205 | *9* | 6 | 0.70 | 0.270 |
| RM208 | *2* | 6 | 0.75 | 0.250 | OSR12 | *9* | 7 | 0.66 | 0.380 |
| RM213 | *2* | 6 | 0.43 | 0.082 | RM222 | *10* | 5 | 0.72 | 0.186 |
| RM48 | *2* | 7 | 0.60 | 0.159 | RM244 | *10* | 4 | 0.48 | 0.101 |
| RM60 | *3* | 4 | 0.23 | 0.147 | RM216 | *10* | 6 | 0.66 | 0.109 |
| RM22 | *3* | 4 | 0.67 | 0.151 | RM239 | *10* | 3 | 0.54 | 0.246 |
| RM231 | *3* | 7 | 0.80 | 0.253 | OSR33 | *10* | 5 | 0.70 | 0.145 |
| RM7 | *3* | 8 | 0.77 | 0.304 | RM258 | *10* | 4 | 0.72 | 0.176 |
| RM218 | *3* | 11 | 0.87 | 0.152 | RM228 | *10* | 8 | 0.63 | 0.132 |
| RM232 | *3* | 9 | 0.76 | 0.083 | OSR1 | *11* | 4 | 0.41 | 0.294 |
| RM16 | *3* | 7 | 0.76 | 0.250 | RM167 | *11* | 7 | 0.72 | 0.238 |
| OSR31 | *3* | 4 | 0.55 | 0.282 | RM120 | *11* | 4 | 0.67 | 0.227 |
| RM55 | *3* | 5 | 0.69 | 0.234 | RM202 | *11* | 6 | 0.64 | 0.214 |
| RM227 | *3* | 6 | 0.66 | 0.161 | RM209 | *11* | 8 | 0.83 | 0.275 |
| RM261 | *4* | 7 | 0.71 | 0.254 | RM229 | *11* | 7 | 0.79 | 0.181 |
| RM252 | *4* | 7 | 0.83 | 0.183 | RM21 | *11* | 10 | 0.88 | 0.083 |
| RM241 | *4* | 6 | 0.73 | 0.203 | RM206 | *11* | 10 | 0.86 | 0.083 |
| RM255 | *4* | 5 | 0.42 | 0.273 | RM254 | *11* | 7 | 0.76 | 0.130 |
| OSR34 | *5* | 6 | 0.56 | 0.365 | OSR6 | *11* | 9 | 0.87 | 0.110 |
| RM122 | *5* | 6 | 0.73 | 0.158 | RM224 | *11* | 8 | 0.82 | 0.210 |
| RM13 | *5* | 6 | 0.74 | 0.168 | RM123 | *11* | 6 | 0.56 | 0.187 |
| RM163 | *5* | 9 | 0.84 | 0.170 | RM20 | *12* | 5 | 0.56 | 0.358 |
| RM164 | *5* | 8 | 0.84 | 0.176 | RM4 | *12* | 6 | 0.77 | 0.184 |
| OSR19 | *6* | 9 | 0.84 | 0.168 | RM19 | *12* | 7 | 0.79 | 0.252 |
| RM204 | *6* | 8 | 0.73 | 0.145 | RM247 | *12* | 7 | 0.55 | 0.257 |
| RM225 | *6* | 7 | 0.75 | 0.173 | OSR20 | *12* | 3 | 0.63 | 0.275 |
| RM253 | *6* | 7 | 0.78 | 0.291 | OSR32 | *12* | 7 | 0.83 | 0.158 |
| RM217 | *6* | 9 | 0.83 | 0.272 | RM17 | *12* | 7 | 0.79 | 0.390 |
| RM3 | *6* | 7 | 0.74 | 0.194 | | | | | |
| Average | | 6.2 | 0.68 | 0.218 | | | | | |

appeared to be subspecies-specific, excluding very rare alleles that occurred in less than six accessions in the entire sample (data not given). Of those, 29 alleles at 25 loci occurred only in the indica accessions, whereas two alleles at two loci (RM48 and RM20) were observed only in the japonica lines. Geographically, seven unique alleles at RM9, OSR9A, RM42, ORS7, RM222, RM202 and RM20 detected in accessions from EA, SEA1, SEA2 and SA were not found in the Chinese accessions which had 11 alleles at RM5, RM48, RM227, RM225, OSR7, RM219, RM205, OSR12, RM222, RM20 and OSR32 that were not detected in the samples from Southeast Asia and South Asia.

**Fig. 1** Dendrogram of 193 rice accessions clustered based on the Dice genetic similarity matrix by UPGMA analysis, which showed three main groups and nine subclusters at the similarity coefficient of 0.35. The *horizontal axis* represents scales of the similarity coefficient. The names, origins, IRGC (International Rice Germplasm Center) accession numbers, and subspecific information (I = indica and J = japonica) of the 193 accessions were linked with the dendrogram, which is available as supporting information on the web site (http://link.springer.de)

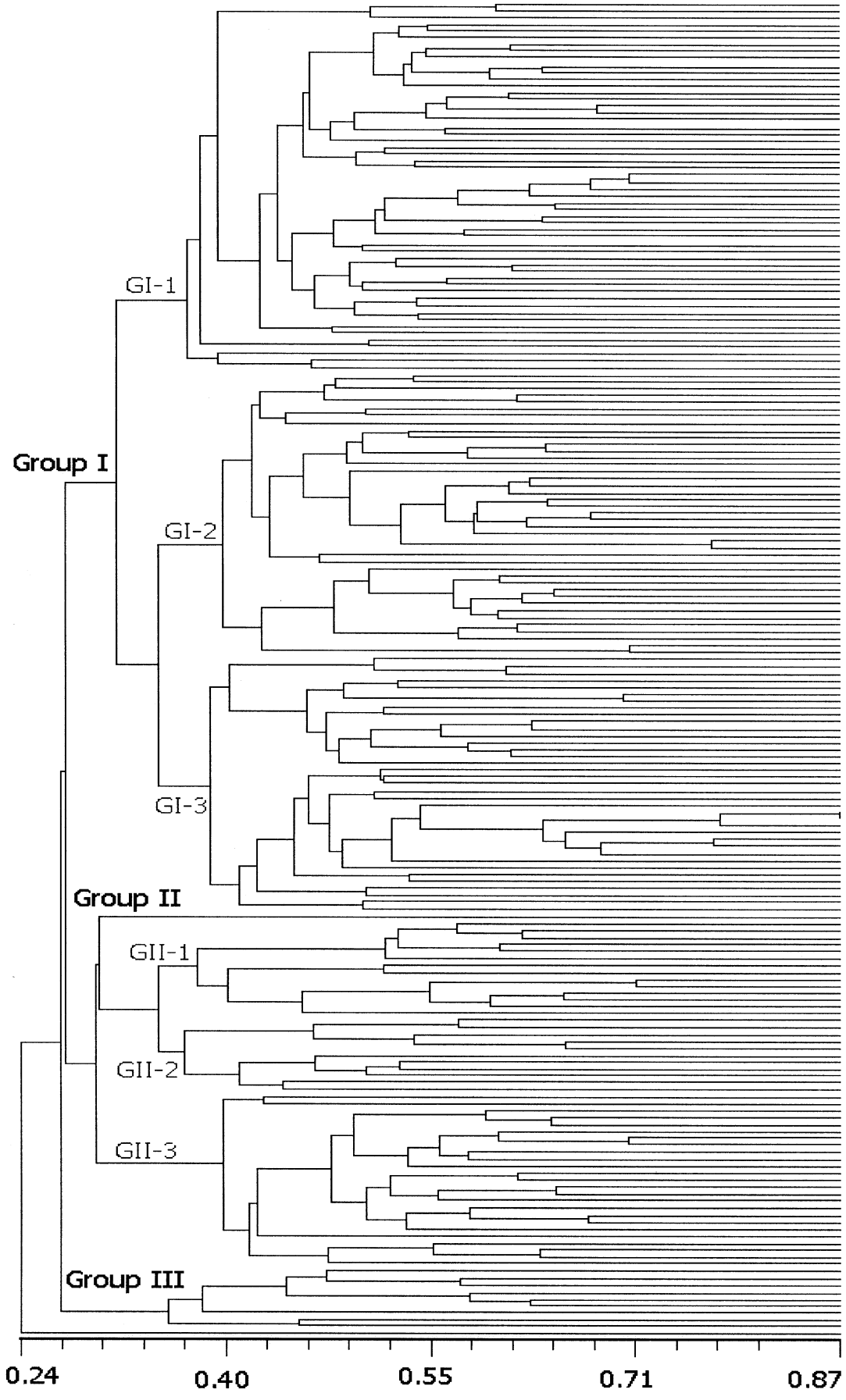
Group I
GI-1
GI-2
GI-3
Group II
GII-1
GII-2
GII-3
Group III
0.24
0.40
0.55
0.71
0.87

**Table 3** The number of alleles ($N_A$), genic diversity (H), and among-cluster genetic differentiation ($G_{ST}$) of SSR loci on 12 rice chromosomes within indica (I) and japonica (J), and the total sample (T)

| Chr. | No. of loci | $N_A$ | | | H | | | $G_{ST}$ (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | I | J | T | I | J | T | I | J | T |
| 1 | 13 | 5.8 | 5.5 | 5.9 | 0.600 | 0.688 | 0.660 | 12.7 | 19.9 | 21.3 |
| 2 | 13 | 5.3 | 5.0 | 5.5 | 0.612 | 0.643 | 0.665 | 16.8 | 25.6 | 27.2 |
| 3 | 10 | 6.4 | 5.6 | 6.5 | 0.659 | 0.645 | 0.676 | 10.8 | 22.5 | 20.2 |
| 4 | 4 | 6.3 | 5.8 | 6.3 | 0.615 | 0.719 | 0.676 | 12.7 | 20.1 | 22.8 |
| 5 | 5 | 6.8 | 6.4 | 7.0 | 0.741 | 0.688 | 0.742 | 11.4 | 23.4 | 20.7 |
| 6 | 7 | 7.4 | 7.1 | 7.4 | 0.722 | 0.682 | 0.766 | 9.3 | 23.5 | 23.2 |
| 7 | 7 | 6.0 | 5.7 | 6.0 | 0.683 | 0.665 | 0.706 | 12.6 | 22.4 | 21.3 |
| 8 | 8 | 5.5 | 4.8 | 5.5 | 0.477 | 0.543 | 0.516 | 4.6 | 11.8 | 11.6 |
| 9 | 8 | 6.6 | 5.5 | 6.6 | 0.701 | 0.599 | 0.730 | 18.8 | 30.9 | 31.0 |
| 10 | 7 | 5.0 | 4.4 | 5.0 | 0.600 | 0.603 | 0.634 | 8.3 | 14.5 | 15.7 |
| 11 | 12 | 7.2 | 6.5 | 7.2 | 0.699 | 0.741 | 0.735 | 9.9 | 18.5 | 18.6 |
| 12 | 7 | 5.9 | 5.4 | 6.0 | 0.625 | 0.713 | 0.703 | 15.2 | 24.4 | 26.8 |
| Mean | | 6.1 | 5.6 | 6.2 | 0.641 | 0.661 | 0.682 | 12.1 | 21.5 | 21.8 |

**Table 4** Geographic patterns of genic diversity over 101 SSR loci of 193 rice accessions

| Region | Group | | | | | | | | Totals | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Indica[b] | | | | Japonica | | | | | | | |
| | $N_p$[a] | n | $N_A$ | H | $N_p$ | n | $N_A$ | H | $N_p$ | n | $N_A$ | H |
| NA | 75 | 3 | 1.96Â±0.69 | 0.38Â±0.24 | 91 | 5 | 2.38Â±0.76 | 0.47Â±0.19 | 97 | 8 | 3.07Â±1.03 | 0.54Â±0.18 |
| EA | 99 | 31 | 4.87Â±1.64 | 0.60Â±0.20 | 101 | 15 | 4.11Â±1.30 | 0.62Â±0.15 | 101 | 46 | 5.42Â±1.74 | 0.66Â±0.16 |
| SEA1 | 101 | 29 | 4.91Â±1.63 | 0.61Â±0.18 | 83 | 6 | 2.60Â±1.14 | 0.45Â±0.25 | 101 | 35 | 5.29Â±1.75 | 0.63Â±0.17 |
| SEA2 | 101 | 26 | 4.49Â±1.40 | 0.59Â±0.17 | 66 | 2 | 1.68Â±0.47 | 0.34Â±0.23 | 101 | 28 | 4.78Â±1.53 | 0.61Â±0.16 |
| SA | 101 | 39 | 5.30Â±1.65 | 0.62Â±0.17 | 101 | 16 | 4.21Â±1.39 | 0.62Â±0.16 | 101 | 55 | 5.73Â±1.70 | 0.67Â±0.15 |
| ME | 83 | 4 | 2.29Â±0.82 | 0.45Â±0.23 | 75 | 4 | 2.02Â±0.76 | 0.36Â±0.24 | 98 | 6 | 3.00Â±0.94 | 0.55Â±0.17 |
| Africa | | | | | 73 | 3 | 1.92Â±0.69 | 0.37Â±0.24 | 87 | 4 | 2.33Â±0.83 | 0.45Â±0.22 |
| America | | | | | 92 | 5 | 2.53Â±0.81 | 0.48Â±0.20 | 95 | 6 | 2.83Â±0.84 | 0.52Â±0.18 |
| Europe | | | | | 70 | 4 | 1.91Â±0.75 | 0.33Â±0.24 | 70 | 4 | 1.91Â±0.75 | 0.33Â±0.24 |
| Total | 101 | 132 | 6.14Â±1.79 | 0.64Â±0.17 | 101 | 60 | 5.60Â±1.83 | 0.66Â±0.14 | 101 | 192[c] | 6.26Â±1.81 | 0.68Â±0.14 |

[a] $N_p$ = number of polymorphic loci among the accessions; *n* = sample size for each subgroup; $N_A$ = observed number of alleles per locus; H = Nei's (1973) gene diversity across loci
[b] Two accessions from Africa and America were included in the ME
[c] One unique variety (Jalmagna) was not included

## Genic diversity and its geographic distribution patterns

Table 4 shows the results of the genic diversity partitioned within and among different geographic regions. Based on the number of alleles per locus and the estimated genic diversity from the SSRs, East Asia (EA), Southeast Asia (SEA) and South Asia (SA) appeared to be highly diverse for the indica group, and EA and SA were highly diverse for the japonica group.

The japonica accessions showed a much-more pronounced geographic differentiation, although the mean genic diversity within japonica was the same as that within indica. The relative contribution of among-region diversity to the total diversity within the subspecies was 24.2%, ranging from 11.5% on chromosome 7 to 34.8% on chromosome 11 (Table 5). The average subspecific differentiation across the chromosomes was only 6.5%. Among-region variation within indica accounted for 15.3% of the total variation, and that within japonica accounted for 30.1% of the total variation. The japonica group showed a greater geographic diversity than the indica group (Table 5).

## LD between SSR loci within subspecies

Table 6 shows the results of LD analyses. The total number of significant LDs ($P<0.001$) detected in the 193 lines was 3,407 for indica and 2,344 for japonica, respectively, which was 59-times and 39-times higher than that expected by chance. More significant LDs were detected in the indica than in the japonica group. LDs also differed among geographic regions. Binomial tests showed that the significant LDs detected between alleles at different loci were much more frequent than expected by chance, indicating prevalent non-random associations between alleles at different loci of the genome in the sample. A careful check of the LDs significant at 0.001 indicated that the average number of LDs between loci on the same chromosome was only 26 for the indica group and 14 for the japonica group, suggesting that linkage was not the main causes of the observed LDs. Interestingly, strong associations between many alleles specific to subspecies were frequently detected, such as associations of allele B at RM219 with allele B at RM202, allele F at RM231 with allele F at RM224, and allele E at RM7 with allele B at RM219.

**Table 5** Patterns of genic diversity of SSR markers on different chromosomes for the geographically subdivided rice accessions

| Chr. | $H_T$[a] | $G_{ST}$ (%) | Indica | | | Japonica | | |
|---|---|---|---|---|---|---|---|---|
| | | | $H_S$ | $H_C$ | $G_{CS}$ (%) | $H_S$ | $H_C$ | $G_{CS}$ (%) |
| 1 | 0.696 | 6.3 | 0.599 | 0.513 | 14.0 | 0.688 | 0.472 | 31.3 |
| 2 | 0.674 | 8.1 | 0.622 | 0.518 | 16.9 | 0.619 | 0.443 | 28.2 |
| 3 | 0.677 | 4.6 | 0.657 | 0.556 | 15.4 | 0.641 | 0.462 | 29.0 |
| 4 | 0.688 | 4.0 | 0.621 | 0.520 | 15.8 | 0.691 | 0.476 | 31.5 |
| 5 | 0.712 | 5.2 | 0.711 | 0.609 | 14.3 | 0.653 | 0.474 | 28.5 |
| 6 | 0.748 | 8.9 | 0.734 | 0.620 | 15.6 | 0.656 | 0.479 | 26.1 |
| 7 | 0.704 | 4.7 | 0.696 | 0.614 | 11.5 | 0.652 | 0.440 | 32.6 |
| 8 | 0.524 | 4.3 | 0.472 | 0.383 | 15.6 | 0.517 | 0.367 | 27.3 |
| 9 | 0.694 | 9.8 | 0.700 | 0.602 | 14.2 | 0.574 | 0.428 | 24.2 |
| 10 | 0.622 | 7.0 | 0.584 | 0.511 | 13.2 | 0.569 | 0.372 | 34.4 |
| 11 | 0.749 | 5.1 | 0.685 | 0.564 | 17.4 | 0.731 | 0.477 | 34.8 |
| 12 | 0.733 | 9.0 | 0.643 | 0.544 | 14.6 | 0.683 | 0.488 | 28.9 |
| Overall | 0.686 | 6.5 | 0.641 | 0.542 | 15.3 | 0.642 | 0.449 | 30.1 |

[a] $H_T$, $H_S$ and $H_C$ represent the mean genic diversity of the total, within-group, and within-region samples; $G_{ST}$ and $G_{CS}$ represent the relative contribution of between-subspecies diversity to the total variation and that of among-region diversity to total within-subspecies variation, respectively

**Table 6** Number of significant ($P$<0.001) linkage disequilibrium (LD) of alleles between different SSR loci for subdivided populations, and for the entire samples

| Region | Indica | | | Japonica | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|
| | N[a] | n | n/N (%) | N | n | n/N (%) | N | n | n/N (%) |
| NA | 9,779 | 0 | 0.0 | 17,960 | 0 | 0.0 | 29,048 | 0 | 0.0 |
| EA | 44,601 | 1,273 | 2.9 | 45,017 | 598 | 1.3 | 59,485 | 1,515 | 2.5 |
| SEA1 | 46,662 | 1,508 | 3.2 | 16,985 | 0 | 0.0 | 54,374 | 1,650 | 3.0 |
| SEA2 | 41,247 | 1,049 | 2.5 | 6,095 | 0 | 0.0 | 44,841 | 1,384 | 3.1 |
| SA | 51,667 | 1,653 | 3.2 | 46,884 | 811 | 1.7 | 64,100 | 1,980 | 3.1 |
| ME | 15,128 | 0 | 0.0 | 9,019 | 0 | 0.0 | 30,371 | 0 | 0.0 |
| Total | 57,630 | 3,407 | 5.9 | 59,685 | 2,344 | 3.9 | | | |

[a] N and n are the total number of pairwise LD performed and the number of LD significant at $P$<0.001

**Table 7** Partitioning of variance components of linkage disequilibrium between alleles at different SSR loci in different subdivided populations

| Subdivision | $D_{IT}^2$ | $D_{IS}^2$ | $D_{ST}^2$ | $D'_{IS}^2$ | $D'_{ST}^2$ |
|---|---|---|---|---|---|
| A: I–J subspecies | 0.0423 | 0.0107 | 0.0307 | 0.0307 | 0.0116 |
| B: geographic regions | 0.0416 | 0.0183 | 0.0222 | 0.0328 | 0.0088 |
| C: sub + regions | 0.2648 | 0.0531 | 0.2103 | 0.2369 | 0.0279 |

$D_{IT}^2$, $D_{IS}^2$, $D_{ST}^2$, $D'_{IS}^2$ and $D'_{ST}^2$ are the total variance of LD, the components due to linkage disequilibrium within subpopulations, allelic differentiation between pairwise loci, gametic differentiation between pairwise loci, and linkage disequilibrium in the total population, respectively

## LD patterns in subpopulations

The components resulting from partitioning the total variance ($D_{IT}^2$) of LDs averaged over all possible pairs are listed in Table 7. The total samples were subdivided into different subpopulations in three ways, with A for the two clustered subspecies, B for the four main geographic subpopulations of large sample sizes, and C for the cross-classified subdivisions into which the samples within the two subspecies were further delineated into geographic subpopulations. The relationship between $D'_{IS}^2$ (the component of the gametic differentiation) and $D_{ST}^2$ (the allelic differentiation) was the same for subdivision A and greater in B and C, respectively, indicating a more-pronounced differentiation at the multilocus level than at the single-locus level. The observed $D_{ST}^2$ was nearly 4- and 3-times larger than $D_{IS}^2$ in subdivisions C and A, indicating that the allele frequencies were very different among the subdivisions. The $D'_{IS}^2$ was about 9-, 4- and 2-times as large as $D'_{ST}^2$ for subdivisions C, B and A, respectively, indicating that the gametic frequency differentiation as the major contributor (72.6%, 78.8% and 89.5%) to the LD observed in the subspecies and geographic populations. Overall, the results of $D_{IS}^2 < D_{ST}^2$ and $D'_{IS}^2 > D'_{ST}^2$ indicated that migration among the subdivisions is very limited.

To evaluate the genetic structure associated with the subspecies and geographic groups, analysis of multilocus associations was applied to the SSR data. Table 8 shows the components of multilocus associations among and within the populations obtained by the method of Brown and Feldman (1981). These included the single-locus components representing the average variation among subpopulations in genic diversity and the variance among subpopulations in allele frequency, and the two-locus components, such as the mean and variance of LD, the

**Table 8** Brown's multilocus analysis of the SSR diversity and linkage disequilibrium of the subpopulations from three classified ways for the 193 accessions

| Source of variance | Abbreviated as[a] | Values[b] | | |
|---|---|---|---|---|
| | | C | B | A |
| Single-locus effect | | | | |
| Mean gene diversity | MH | 20.64 | 20.54 | 20.45 |
| Variance of diversity | VH | 2.35 | 0.42 | 0.56 |
| Wahlund's effect | WH | −3.17 | −1.01 | −1.19 |
| Two-locus effect | | | | |
| Mean disequilibrium | MD | 43.92 | 76.37 | 55.74 |
| Wahlund's effect | WC | 29.12 | 3.74 | 19.50 |
| Interaction between MD and WC | AI | 10.69 | 3.42 | 5.55 |
| Variance of disequilibrium | VD | 276.24 | 111.47 | 42.76 |
| Covariance of interactions | CI | 5.48 | 3.66 | 1.69 |
| Total variance | $\sigma^2_t$ | 103.54 | 103.48 | 100.58 |
| Average variance | $\sigma^2_k$ | 356.97 | 215.45 | 126.10 |

[a] $\sigma^2_t$=MH+VH+WH+MD+WC+AI, $\sigma^2_k$=MH+MD+AI+VD+CI

[b] A, B and C represent analyses based on the subspecific division (two subspecies), the geographically subdivision (four main subpopulations for the analysis), and the nested subdivision (the total sample was first classified into two subspecies and then each subspecies was subdivided into geographic subpopulations, forming a total of 15 population subdivisions), respectively

covariance of allele frequencies over subpopulations, and various interactions. These components were measured by their contributions to the total variance regarding the number of heterozygous loci between two randomly chosen gametes.

For the subspecific division (A), the single-locus components accounted for approximately 20% of the total variance ($\sigma^2_t$). The variance of mean disequilibrium accounted for 55.7% of the total variance in the two-locus components, indicating that the disequilibria were not only high within subspecies, but also highly different between subspecies. The large proportion of the differentiation was due to the joint responses involving two or more loci, indicating that the indica-japonica differentiation is much-more pronounced in multilocus structure. For geographic (B) and cross-classified (C) subpopulations, similar patterns were observed, indicating that the multilocus associations were predominant in both geographic and cross-classified subdivisions. The multilocus structure varied across the geographic regions, as suggested by the large portion of the disequilibrium variance (VD) in the total variance in all three subdivisions. Wahlund's effect was small in all population subdivisions, indicating that selection, rather than population admixture or founder effect, was largely responsible for the observed multilocus structure.

## Discussion

This study was the first step in characterizing the molecular diversity of a large set of parental lines used in the International Rice Molecular Breeding Program coordinated by IRRI. The 193 accessions included 65 elite commercially grown cultivars, each of which is currently used as the recipient in a large-scale backcross-breeding program. All 193 accessions are also used as donors in different participating institutions. This large backcross-breeding program was designed to exploit the diversity of complex phenotypes within the rice primary gene pool by transferring useful genetic diversity from the primary gene pool into the elite rice cultivars grown across Asia and integrating DNA markers with breeding for the discovery of useful genes/QTLs underlying important traits such as grain quality, yield and abiotic stress tolerances during the process of cultivar development. Thus, information regarding genetic diversity, SSR polymorphism and relationships among the materials is essential for achieving our objectives.

The sampling of the 193 accessions in this study was not random and tentative, because it aimed to cover as much diversity as possible based on our current knowledge on the genetic diversity within *O. sativa*. The observed level of genetic diversity from the SSRs and its distribution pattern were generally consistent with those of most previous studies based on much larger samples (Glaszmann 1987; Li and Rutger 2000). Thus, it can be concluded that the 193 accessions sampled indeed represent a significant breadth of the total genetic diversity within *O. sativa*. SSR markers for this assay were chosen because of their obvious advantages such as abundance in the genome, high level of polymorphism, co-dominance and cost-effectiveness (Yang et al. 1994; McCouch et al. 1997; Ni et al. 2002), which were clearly demonstrated in this study. The high level of SSR polymorphism detected among the japonica accessions, however, was surprising since japonica accessions tended to show a low level of polymorphism for other types of markers such as isozyme and RFLP loci (Glaszmann 1987; Zhang et al. 1992; Li and Rutger 2000). This could be attributed primarily to the diverse origins of the japonica accessions sampled in our materials.

Our results revealed several unique features of the SSR diversity in *O. sativa*. The first unique feature of the SSRs in rice is its high level of diversity within both japonica and indica groups. This was different from the result from

isozymes of the same set of materials, which revealed a much higher diversity in the indica accessions than the japonica ones (data not shown). Second, the SSRs clearly show the well-documented indica-japonica differentiation that was reflected as subspecies-specific alleles at many loci, but this subspecific differentiation accounted for only 6.5% of the total SSR diversity in the samples and was much-less pronounced than the isozyme variation (data not shown). While this is expected, based on the genetic neutrality of most SSR loci and easier generation of SSR diversity by mutation, unequal crossover and other mechanisms (McCouch et al. 1997), it does suggest the independent evolution of japonica from indica. Third, the japonica group exhibited larger geographic differentiation at both single-locus and multi-locus levels than the indica group, and the japonica group was clearly differentiated into two subgroups, providing evidence supporting the presence of tropical and temperate japonica ecotypes first proposed by Chang (1976). The single solitary genotype for the deepwater rice (Jalmagna) in our grouping indicated that novel genotypes often arise from the extreme selection pressure, though what was responsible for another solitary genotype, Baobaomi from Yunnan of China, remains unknown. The fourth feature of SSRs in rice was the prevalence of the non-random associations between or among unlinked SSR loci resulting from the subspecific and geographic differentiation within *O. sativa*. The multilocus combinations tended to show a greater level of differentiation than alleles at single SSR loci within various subdivisions. This result is much the same as that observed for the isozyme loci in *O. sativa* (Li and Rutger 2000).

We further found that the gametic frequency differentiation was the major contribution to the observed LD, and the relationship of $D_{IS}^2 < D_{ST}^2$ and $D'_{IS}{}^2 > D'_{ST}{}^2$ held true for all subdivisions. According to Ohta (1982), this indicates that migration was limited between accessions of the subspecies or among the geographic regions. While this interpretation appears to be the case for the landraces in our sample, it is in conflict with the fact that many elite cultivars in our sample were derived from complex pedigrees involving parents from different geographic regions. Therefore, the observed multilocus structure of the SSR diversity more likely resulted from genetic hitchhiking of those loci under the disruptive selection in the two subspecies isolated by partial reproductive barriers and/or under varied ecologically adaptive selection in different geographic regions (Allard 1996; Li et al. 1999, 2000; Li and Rutger 2000). Evidence supporting this interpretation came also from the observed genomic region-specific diversity of the SSRs, which suggests that genes involved in subspecific and geographic differentiations were not randomly distributed in the rice genome.

**Acknowledgements** We are grateful to valuable comments from Dr. Q. F. Zhang and two anonymous reviewers on the early version of this paper. We thank Dr. McCouch from Cornell University for the primer information on the SSR markers used in the study. We also thank B. Hardy for the editorial help. Financial support from the Rockefeller Foundation, BMZ of the German government, and the National Natural Science Foundation of China is greatly appreciated.

## References

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 93:1071–1077

Allard RW (1996) Genetic basis of the evolution of adaptedness in plants. Euphytica 92:1–11

Brown AHD, Feldman MW (1981) Population structure of multilocus associations. Proc Natl Acad Sci USA 78:5913–5915

Chang TT (1976) The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica 25:435–441

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and genbank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Evernson RE, Gollin D, Santaniello V (1998) Introduction and overview: agricultural values of plant genetic resources. In: Evernson RE, Gollin D, Santaniello V (eds) Agricultural values of plant genetic resources. CABI Publishing, Wallingford, UK, pp 1–29

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 74:21–30

Li YC, Fahima T, Beiles A, Korol AB, Nevo E (1999) Microclimatic stress and adaptive DNA differentiation in wild emmer wheat, *Triticum dicoccoides*. Theor Appl Genet 98:873–883

Li YC, Roder MS, Fahima T, Kirzhner VM, Beiles A, Korol AB, Nevo E (2000) Natural selection causing microsatellites divergence in wild emmer wheat at the ecologically variable microsite at Ammiad, Israel. Theor Appl Genet 100:985–999

Li Z, Rutger JN (2000) Geographic distribution and multilocus organization of isozyme variation of rice (*Oryza sativa* L.). Theor Appl Genet 101:379-387

Liu KD, Yang GP, Zhu SH, Zhang Q, Wang XM, Saghai Maroof MA (1996) Extraordinarily polymorphic ribosomal DNA in wild and cultivated rice. Genome 39:1109–1116

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair M (1997) Microsatellite marker development, mapping and application in rice genetics and breeding. Plant Mol Breed 35:89–99

Nei M (1973) The analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Ni J, Colowit PM, Mackill D (2002) Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci 42: 601–607

Ohta T (1982) Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc Natl Acad Sci USA 79:1940–1944

Saghai Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of polymorphisms scored at 40 isozyme loci. Jpn J Genet 57:25–75

StatSoft (1991) Statistica. StatSoft Inc, Tulsa, Oklahoma

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch S (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Vaughan DA, Jackson MT (1995) The core as a guide to the whole collection. In: Hodgkin T, Brown AHD, van Hintum TJL,

Morales EAV (eds) Core collections of plant genetic resources. John Wiley and Sons Inc, Chichester–New York–Brisbane–Toronto–Singapore, pp 229–239
Weir BS (1990) Genetic data analysis. Sinauer Associations, Sunderland, Massachusetts
Yang GP, Saghai Maroof MA, Xu CG, Zhang Q, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 245:187–194
Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Center, University of Alberta, Canada
Zhang Q, Saghai Maroof MA, Liu TY, Shen BZ (1992) Genetic diversity and differentiation of indica and japonica rice detected by RFLP analysis. Theor Appl Genet 83:495–499
Zhu J, Gale MD, Quarrie S, Jachson MT, Bryan GJ (1998) AFLP markers for the study of rice biodiversity. Theor Appl Genet 96:602–611