Theor Appl Genet (2000) 101:379–387

Z. Li · J.N. Rutger

Geographic distribution and multilocus organization of isozyme variation of rice (*Oryza sativa* L.)

Received: 5 August 1999 / Accepted: 13 December 1999

Abstract Genetic organization of isozyme variation in rice (*Oryza sativa L.*) was investigated based on 17 polymorphic isozyme loci using a sample of 511 accessions of worldwide origin. The genetic diversity within the species was very high (H=0.36 with 4.82 alleles per locus), as compared with most selfing plant species. Three diversity centers were detected for isozyme variation including South Asia, China and Southeast Asia. The accessions were classified into three well-differentiated cultivar groups corresponding to the *indica* and *japonica* subspecies, and a new unnamed group. Variation within the cultivar groups accounted for 80% of the total isozyme variation. Within-country variation accounted for 58% of the total variation while among-region and among-country variation within the cultivar groups accounted for only 14% and 8% of the total variation. Analyses using log-linear models revealed that pronounced non-random associations between and among alleles at many unlinked isozyme loci were organized in a non-hierarchical pattern, and subspecific and macrogeographic differentiation was much more pronounced in multilocus phenotype frequencies than in allelic frequencies at individual loci. These results suggest that selection on multilocus gene complexes was largely responsible for the maintenance of the extensive isozyme variation within the species and the *indica-japonica* differentiation. Our results further suggest the independent domestication of *indica* and *japonica*, the dual origins of the *indica* rice from China and South Asia (India), and the differentiation of the ecotypes '*javanica*' and the 'temperate *japonica*' within the *japonica* subspecies.

Key words Rice · Isozyme variation · Multilocus organization · Evolution

Communicated by A.L. Kahler

Z. Li (✉)
Department of Agronomy and Range Sciences,
University of California and USDA-ARS,
Davis, CA 95616, USA
e-mail: zli@cgiar.org

J.N. Rutger
National Rice Germplasm Evaluation and Enhancement Center,
USDA-ARS, P.O. Box 287, Stuttgart, AR 72160, USA

Present address:
Z. Li, The International Rice Research Institute,
P.O. Box 3127, 1271 Makati City, Philippines

Introduction

Genetic diversity is a ubiquitous property of all species in nature. The distribution and organization of genetic variation within and among populations of a species are the consequences of its evolution. As an important aspect of genetic diversity, allozyme variation in many crop plants and their wild relatives has been extensively surveyed regarding its geographic distribution and multilocus organization (Brown 1978; Brown and Jain 1979; Nevo and Ben-Shlomi 1984; Hamrick and Godt 1989; Allard et al. 1993; Allard 1996). Results from those studies have provided important knowledge and strategies for efficient exploration, conservation and utilization of genetic resources in crop plants.

Rice, *Oryza sativa* L., is one of the most important grain crops in the world, and the staple food for more than half the world population. As a self-pollinated crop species with extensive intra-specific variation, differentiation of *O. sativa* into two major subspecies, '*Xian*' (*indica*) and '*Geng*' (*japonica*), has been well documented (Kato et al. 1928; Terao and Mizushima 1939; Matsuo 1952; Ting 1957; Oka 1958, 1988). Differentiation of *O. sativa* into different cultivar groups at the isozyme level has also been reported (Nakagahra 1977, 1978, 1984; Second 1982; Glaszmann 1987), but the geographic distribution and the multilocus organization of the isozyme variation within *O. sativa* have not been fully characterized. In the present study, we attempt to answer two important questions: (1) what are the geographic distribution patterns of isozyme variation within *O. sativa* and the subspecies, and (2) how and to what extent are the alleles at multiple isozyme loci associated

with one another? Information about the detailed distribution and organization of isozyme variation should provide insights into the origin and evolution of the cultivated rice.

Materials and methods

Materials and electrophoresis of isozymes

Five hundred and eleven rice cultivars originating from ten geographical regions and 31 countries representing the entire growing areas of *O. sativa* in the world, provided by the USDA National Small Grains Germplasm Collection (Table 1), were investigated for isozyme variation revealed by starch gel-electrophoresis. The seeds were germinated in plastic Petri dishes at room temperature. At 4 days after germination, crude composite extracts of water-soluble proteins were prepared from 8 to 15 whole young seedlings of each cultivar by homogenization in distilled water, then absorbed in filter paper wicks. The wicks were inserted into a cut located 4.0-cm from the cathodal end of a horizontal starch gel. Each gel accommodated 30 samples (the standard cultivar IR36 was placed in the center). The gels were prepared using three buffer systems to detect polymorphisms in nine different enzymes. System 1, modified from Glaszmann (1987), (gel buffer: 0.005 M tris-histidine buffer, tray buffer: 0.4 M tris-citrate buffer, pH=7.0) allowed the detection of eight polymorphic loci for the enzyme systems leucine aminopeptidase (LAP, two loci), isocitrate dehydrogenase (IDH, one locus), shikimate dehydrogenase (SDH, one locus), 6-phosphogluconate dehydrogenase (PGD, two loci) and phosphoglucose isomerase (PGI, two loci); system 2 (Poulik buffer system with pH=7.2 in gel buffer and pH =8.6 in tray buffer) permitted the detection of six more polymorphic loci in esterase (EST, four loci), malic enzyme (ME, one locus) and glutamate oxalate transaminase (GOT, one locus); while system 3 (borate buffer system with pH=8.2 in gel buffer and pH=8.6 in tray buffer) allowed the survey of two more polymorphic loci in acid phosphatase (ACP, two loci). The phenol-oxidizing enzyme (*Ph*, one locus) was surveyed by soaking ten grains and ten kernels (dehulled grains) from each of the accessions in 1% phenol solution for 48 h. The status of grain and kernel color after soaking indicated a positive reaction (*indica* type, black color) or a negative reaction (*japonica* type, unchanged color).

Data analysis

Nei's (1972) unbiased genetic distance (D) was used as a relative measure of difference between all pairs of the 511 accessions for 16 polymorphic isozyme loci and the *Ph* locus (phenol reaction) to create a 465×465 (we detected 465 distinct phenotypes) distance matrix. Then, a two-step classification based on this distance matrix was performed based on the method of average distance using SAS PROC CLUSTER (SAS Institute 1990). Nei's (1973) heterozygosity index was used as a measure of isozyme variation, which is given by:

$$\hat{H}_k = 1 - \sum_{i=1}^{n} \hat{p}_i^2,$$

where $\hat{p}_i$ is the frequency of ith allele at locus k, and n is the number of alleles at this locus. For l loci, the average gene diversity index was estimated by:

$$\hat{\bar{H}}_T = \frac{1}{l} \sum_{k=1}^{l} \hat{H}_k.$$

The total genetic diversity within a species was partitioned into its components in a nested manner (Nei 1973) as follows:

$$\hat{\bar{H}}_T = \hat{\bar{H}}_{WC} + \hat{D}_{AC} + \hat{D}_{AR} + \hat{D}_{AG},$$

where $\hat{\bar{H}}_T$ is the total gene diversity in the sample, $\hat{\bar{H}}_{WC}$ is the within-country gene diversity within-regions within cultivar groups, $\hat{D}_{Ac}$ is the among-country diversity within-regions within groups, $\hat{D}_{AR}$ is the among-region diversity within cultivar groups, and $\hat{D}_{AG}$ is the among-cultivar group variation. The corresponding relative gene diversity estimates for the within country, the inter-countries within regions, the inter-geographic regions within cultivar groups, and the inter-groups within the species, were calculated as $\hat{G}_{WC}=\hat{H}_{WC}/\hat{\bar{H}}_T$, $\hat{G}_{AC}=\hat{D}_{AC}/\hat{\bar{H}}_T$, $\hat{G}_{AR}=\hat{D}_{AR}/\hat{\bar{H}}_T$, and $\hat{G}_{AG}=\hat{D}_{AG}/\hat{\bar{H}}_T$, respectively.

The multilocus organization of isozyme variation within *O. sativa* was analyzed with log-linear models (Bishop et al. 1975; Feinberg 1980) using SAS PROC CATMOD (SAS Institute 1990). The most-parsimonious models with best goodness-of-fit to the frequency tables of the observed data sets, were obtained by model selection procedures based on the partitioning of likelihood-ratio statistics (G^2) (Bishop et al. 1975; Feinberg 1980).

Results

Diversity, subspecific differentiation, and geographic distribution

Overall variation

The ten enzymes allowed the detection of 82 alleles (including six null alleles) at 17 polymorphic loci (*Ph*, *Idh1*, *Sdh1*, *Pgd1*, *Pgd2*, *Pgi1*, *Pgi2*, *Lap1*, *Lap2*, *Est1*, *Est2*, *Est3*, *Est4*, *Acp1*, *Acp2*, *Me2* and *Got1*) among the 511 accessions assayed (Table 1). The average diversity index of the sample was 0.36±0.04 with an average of 4.82 alleles per locus. Gene diversity indices at individual isozyme loci varied considerably, ranging from 0.02 for *Got1* to 0.56 for *Idh1*. Rare alleles with a frequency

Table 1 Polymorphic loci of ten enzyme systems and their diversity indices ($\hat{H}$) estimated from a worldwide sample of 511 rice accessions

Enzyme	Loci	Number of alleles[a]	$\hat{H}$
Acid phosphatase	*Acp2*	4	0.40
	Acp3	6*	0.53
Esterase	*Est1*	4*	0.51
	Est2	5*	0.07
	Est3	4*	0.39
	Est4	6	0.18
Glutamate oxaloacetate fransaminase	*Got1*	2	0.02
Isocitrate dehydrogenase	*Idh1*	4	0.56
Leucine aminopeptidase	*Lap1*	6*	0.22
	Lap2	6	0.45
Malic enzyme	*Me2*	6*	0.54
Phosphogluconate dehydrogenase	*Pgd1*	3	0.05
	Pgd2	5	0.45
Phosphoglucose isomerase	*Pgi1*	6	0.42
	Pgi2	7	0.50
Phenol oxidizing enzyme	*Ph*	2	0.46
Shikimate dehydrogenase	*Sdh1*	6	0.29
Average		4.82	0.36

[a] * Indicates a null allele detected

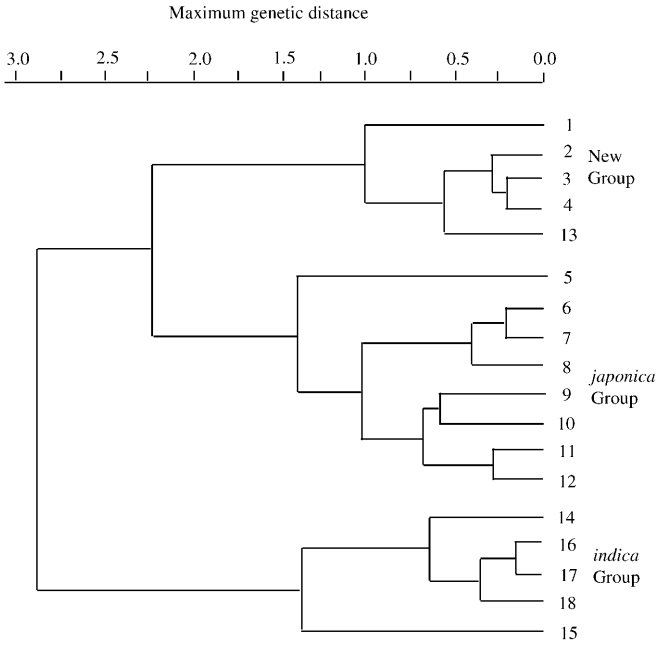

Fig. 1 Phylogenetic relationships of 18 clusters (465 genotypes) of rice cultivars sampled worldwide

<0.01 accounted for a significant proportion (30.5%) of the 82 alleles detected at these enzyme loci. Among the 17 polymorphic loci, three loci (*Pgd1*, *Est2* and *Got1*) had a single predominant allele with a frequency >0.95, six loci (*Ph*, *Sdh1*, *Est3*, *Est4*, *Lap2* and *Me2*) had two predominant alleles, and the remaining eight (*Idh1*, *Pgd2*, *Pgi1*, *Pgi2*, *Lap1*, *Est1*, *Acp1* and *Acp2*) had three or more alleles with frequencies ≥0.05.

Classification of cultivar groups

Four hundred and sixty five phenotypes were detected among the 511 accessions and were classified into 18 clusters (with minimal within-cluster variation) based on the 465×465 genetic-distance matrix. Three distinct cultivar groups were revealed by the cluster analysis (Fig. 1). The first group corresponded to the traditional *indica* subspecies, containing 132 accessions (25.8% of the 511 accessions) from 19 countries in nine geographical regions (all except the Middle East and Africa). The second group corresponded to the *japonica* subspecies, which contained 350 (68.5%) accessions from 27 countries in ten geographical regions (all except the Middle East). A new group was also identified in the present study. This group had genetic distances of 0.45 to *indica* and 0.29 to *japonica* and contained only 29 accessions distributed primarily along the Himalayas (India, Pakistan and Iran).

Gene diversity and its geographical distribution patterns

Table 2 shows the geographic distribution of the absolute gene diversity indices ($\hat{H}$) of the 17 isozyme loci, revealing three diversity centers for isozyme variation in *O. sativa*. The first diversity center was South Asia, primarily India and Burma. This region had the highest diversity ($\hat{H}$=0.38 and 3.35 alleles per locus). The second diversity center was East Asia, primarily China, which had an $\hat{H}$ value of 0.34 and 3.53 alleles per locus. The third diversity center was Southeast Asia, primarily Philippines, which had an $\hat{H}$ value of 0.34 and 3.00 alleles per locus.

Table 3 shows the average relative components of gene diversity among these groups of accessions. The among-cultivar group variation component, $\hat{D}_{AG}$, was 0.07 and accounted for 20% of the total isozyme variation. The weighted average within group variation, $\hat{H}_{WG}$, was 0.29, accounting for 80% of the total variation. The new group and the *indica* group were more diverse ($\hat{H}_{WG}$ was 0.36) than the *japonica* group ($\hat{H}_{WG}$=0.25). The average among-population variation, including among-region variation within groups, $\hat{D}_{AR}$, and among-country variation within regions, $\hat{D}_{AC}$, was 0.05 and 0.03, accounting for 14% ($\hat{G}_{AR}$) and 8% ($\hat{G}_{AC}$) of the total variation (Table 3). The *indica* group and the new group showed a greater geographic differentiation than the *japonica* group ($\hat{D}_{AR}$ and $\hat{G}_{AR}$ were 0.07 and 0.188 for *indica*, 0.13 and 0.37 for the new group, and 0.04 and 0.14 for *japonica*). The average relative within-population variation, $\hat{G}_{WR}$ and $\hat{G}_{WC}$, accounted for 66% and 58% of the total variation, respectively (Table 3). $\hat{G}_{WR}$ was greater in the *japonica* group than in the other two groups. In the *indica* group, the absolute within-population variation, $\hat{H}_{WC}$, was high in India (0.33), the Philippines (0.31), and China (0.28). Within the *japonica* group, $\hat{H}_{WC}$ was high in China (0.25) and USA (0.25).

Multilocus organization of *O. sativa*

Multi-locus organization due to indica-japonica differentiation

Eight loci, *Ph*(**A**), *Sdh1*(**B**), *Pgd2*(**C**), *Pgi2*(**D**), *Est4*(**E**), *Est1*(**F**), *Acp3*(**G**) and *Me2*(**H**), where allelic frequencies differentiated significantly between *indica* and *japonica*, were used to construct three four-dimension contingency tables. When the full model (which contains all possible effects) was examined, three 4-locus associations in the three data subsets were all significant but some of the lower-order associations (2-locus and 3-locus effects) were insignificant, indicating that the non-random associations among alleles at these loci had a general non-hierarchical pattern. Table 4 shows the selected most parsimonious models which fitted the data subsets. Strong pairwise associations occurred between 11 of 18 possible gene pairs. Five (55.6%) 3-locus associations were also significant, indicating that the presence of most pairwise associations between alleles at those loci depended on the presence of alleles at a third locus.

To consider how the alleles at the eight loci were associated with one another, quadruplex (4-locus) phenotypes, **ABCD** (*Ph*, *Sdh1*, *Pgd2* and *Pgi2*), **EFGH** (*Est4*, *Est1*, *Acp3* and *Me2*), **ABEF** and **CDGH**, within the *indica* and *japonica* groups were examined (Table 5). They

Table 2 Geographical pattern of $\hat{H}$ over 17 isozyme loci of a worldwide sample of 511 rice accessions

Region		Cultivar group						**Totals**		
Country		***indica***		***japonica***		**New group**		N	$\hat{H}$	Number of alleles
		N[a]	$\hat{H}$[b]	N	$\hat{H}$	N	$\hat{H}$			
Northeast Asia (NEA)		**3**	**0.18**	**66**	**0.24**	**2**	**0.15**	**71**	**0.27**	**2.82**
Japan		1	0.00	49	0.18	2	0.15	52	0.22	2.71
Korea		2	0.06	17	0.18	–	–	19	0.22	1.71
East Asia (EA)		**71**	**0.28**	**67**	**0.25**	**1**	**0.00**	**139**	**0.34**	**3.53**
China		67	0.28	52	0.24	1	0.00	120	0.34	3.35
Taiwan		4	0.30	15	0.25	–	–	19	0.31	2.35
Southeast Asia (SEA)		**17**	**0.36**	**39**	**0.20**	–	–	**56**	**0.34**	**3.00**
Philippines		11	0.31	23	0.21	–	–	34	0.36	2.65
Indonesia		1	0.00	11	0.18	–	–	12	0.22	1.82
Malaysia		3	0.25	3	0.11	–	–	6	0.32	1.88
Vietnam		2	0.18	2	0.03	–	–	4	0.23	1.53
South Asia (SA)		**34**	**0.32**	**18**	**0.20**	**15**	**0.31**	**67**	**0.38**	**3.35**
India		17	0.33	11	0.18	8	0.26	36	0.39	3.06
Burma		4	0.27	4	0.08	1	0.00	9	0.36	2.29
Thailand		9	0.26	–	–	1	0.00	10	0.27	1.94
Pakistan		–	–	3	0.09	4	0.28	7	0.31	1.94
Bangladesh		2	0.00	–	–	1	0.00	3	0.18	1.41
Ceylon		1	0.00	–	–	–	–	1	0.00	1.00
Sri Lanka		1	0.00	–	–	–	–	1	0.00	1.00
Middle East	Iran	–	–	–	–	**5**	**0.29**	**5**	**0.29**	**2.47**
USSR		**2**	**0.09**	**12**	**0.15**	**3**	**0.13**	**17**	**0.17**	**1.53**
Europe		**1**	**0.00**	**46**	**0.16**	**1**	**0.00**	**48**	**0.18**	**3.18**
France		1	0.00	7	0.13	–	–	8	0.18	1.71
Italy		–	–	12	0.17	–	–	13	0.22	2.00
Hungary		–	–	4	0.12	–	–	4	0.12	1.29
Bulgaria		–	–	5	0.12	–	–	5	0.11	1.35
Turkey		–	–	10	0.09	–	–	10	0.09	1.41
Spain		–	–	7	0.12	–	–	7	0.12	1.41
New Caledonia		–	–	1	0.00	–	–	1	0.00	0.00
North America (NA)	USA	**1**	**0.00**	**68**	**0.25**	–	–	**69**	**0.26**	**2.82**
South America		**1**	**0.00**	**19**	**0.17**	**1**	**0.00**	**21**	**0.22**	**2.24**
Brazil		–	–	9	0.16	1	0.00	10	0.21	1.71
Argentina		–	–	7	0.13	–	–	7	0.13	1.35
Peru		–	–	3	0.11	–	–	3	0.11	1.24
El Salvador		1	0.00	–	–	–	–	1	0.00	1.00
Australia		**3**	**0.15**	**8**	**0.18**	–	–	**11**	**0.34**	**2.18**
Africa	Egypt	–	–	**7**	**0.04**	–	–	**7**	**0.03**	**1.12**
Total		**132**	**0.36**	**350**	**0.25**	**29**	**0.36**	**511**	**0.36**	**4.82**

[a] N, the number of accessions sampled in specific regions or countries
[b] $\hat{H}$ is the mean diversity index averaged from 17 isozyme loci

Table 3 Patterns of $\hat{H}$ for isozyme variation in rice

Group	$\hat{D}_{AG}$[a]	$\hat{G}_{AG}$	$\hat{H}_{WG}$	$\hat{G}_{WG}$	$\hat{D}_{AR}$	$\hat{G}_{AR}$	$\hat{H}_{WR}$	$\hat{G}_{WR}$	$\hat{D}_{AC}$	$\hat{G}_{AC}$	$\hat{H}_{WC}$	$\hat{G}_{WC}$
indica			0.36	1.00	0.07	0.19	0.29	0.81	0.03	0.09	0.26	0.72
japonica			0.25	1.00	0.04	0.14	0.21	0.86	0.02	0.09	0.19	0.76
New group			0.36	1.00	0.13	0.37	0.23	0.62	0.04	0.12	0.18	0.51
Average	0.07	0.20	0.29	0.80	0.05	0.14	0.24	0.66	0.03	0.08	0.21	0.58

[a] $\hat{D}$ and $\hat{G}$ are absolute and relative components of the mean gene diversity $\hat{H}$.
The subscripts AG, AR and AC represent the among-group, among-region and among-country variation: WG, WR and WC represent the within-group, within-region, and within-country variation

Table 4 Non-random associations of alleles at loci *Ph* (A), *Sdh1* (B), *Pgd2* (C), and *Pgi2* (D), and *Est4* (E), *Est1* (F), *Acp3* (G), and *Me2* (H), due to the *indica-japonica* differentiation, detected by estimated U-terms and their standardized values in the selected log-linear models 1 (1-D, 2-C, 3-B and 4-A), 2 (1-H, 2-G, 3-F and 4-E), and 3 (1-D, 2-B, 3-F and 4-G)

U terms[a]	Model 1		Model 2		Model 3	
	Estimated value	Standard. value	Estimated value	Standard. value	Estimated value	Standard. value
U_{12}	0.41	3.50***	0.05	0.31	0.06	0.72
U_{13}	0.26	2.67**	0.34	5.58****	–	–
U_{23}	0.05	0.47	0.01	0.17	0.71	7.48****
U_{123}	0.27	2.73**	0.18	2.85**	–	–
U_{14}	0.88	10.17****	0.40	2.69**	0.31	3.67***
U_{24}	0.30	2.73**	0.37	2.54**	0.38	3.89***
U_{124}	–	–	0.47	3.15**	0.30	3.53***
U_{34}	–	–	0.63	5.23****	0.13	1.32
U_{134}	–	–	–	–	0.25	2.59**

[a] '–' indicates the absence of the corresponding U terms in the selected model. The standardized U terms are standardized normal variables with zero mean and unit variance under the null hypothesis (completely independence). Thus, when an observed U term has a standardized value >1.96, this term is said to be significant at P=0.05 (the association is non-random)

[b] *, **, *** and **** indicate that the observed genotypic frequencies deviated from independence (based on frequencies at individual loci) at significance levels of P=0.05, 0.01, 0.001, and 0.0001, respectively

Table 5 Quadriplex phenotypes with appreciable frequencies in the *indica* and *japonica* groups and their geographic distribution at loci *Ph* (**A**), *Sdh1* (**B**), *Pgd2* (**C**), *Pgi2* (**D**), *Est4* (**E**), *Est1* (**F**), *Acp3* (**G**), *Me2* (**H**), *Idh1* (**I**), *Lap1* (**J**), *Lap2* (**K**) and *Est3* (**L**)

indica						***japonica***								
Genotype	Frequency					Genotype	Frequency							
	Total	EA[a]	SA	SEA	Other		Total	EA	NEA	EUR	NA	SEA	SA	Other
ABCD														
2133	0.24	**0.41**	0.06	–	–	1121	0.67	0.45	0.53	0.65	0.91	0.66	0.72	0.50
2134	0.05	**0.08**	–	–	–	2121	0.06	**0.10**	**0.61**	**0.13**	–	–	0.06	0.04
2132	0.06	**0.08**	–	0.07	0.09	1122	0.05	**0.12**	**0.05**	**0.04**	–	–	–	0.02
2233	0.07	**0.07**	0.03	0.07	0.18	1111	0.05	0.03	0.05	0.04	0.01	**0.08**	**0.11**	**0.09**
2223	0.11	0.01	**0.11**	0.40	0.27	2111	0.04	0.03	**0.18**	–	–	0.03	–	–
2123	0.13	0.07	**0.26**	0.07	0.09	1131	0.03	0.04	–	–	–	**0.03**	–	0.04
2213	0.05	0.01	**0.06**	0.20	–									
2222	0.07	–	**0.06**	0.20	0.46									
	0.76						**0.87**							
EFGH														
2221	0.24	**0.37**	–	–	0.46	2012	0.45	0.43	0.61	0.57	0.44	0.31	0.22	0.39
2021	0.20	**0.35**	–	0.07	–	2011	0.12	**0.21**	0.11	0.09	0.10	–	**0.28**	0.11
1222	0.11	0.03	**0.20**	**0.27**	0.09	2022	0.12	0.13	0.09	0.11	0.16	0.10	–	0.15
2121	0.10	0.06	**0.17**	–	0.27	2212	0.09	0.01	–	–	0.12	**0.41**	**0.11**	0.11
2211	0.07	0.06	0.03	0.27	–	2021	0.05	**0.10**	**0.09**	–	0.04	–	–	–
1211	0.06	–	**0.17**	0.13	–	2032	0.05	**0.09**	0.03	0.11	–	–	–	0.07
1111	0.03	–	**0.11**	–	–	2312	0.03	–	0.01	–	**0.12**	–	–	–
	0.80						**0.90**							
IJKL														
2221	0.23	**0.41**	–	0.06	–	**2221**	0.29	**0.49**	**0.50**	**0.22**	**0.25**	0.03	–	0.17
2220	0.15	**0.21**	–	0.18	0.09	**1221**	0.29	0.24	0.32	0.63	0.01	0.31	0.28	0.39
1211	0.14	0.04	**0.38**	0.12	–	**1211**	0.04	0.01	–	–	–	**0.10**	**0.28**	0.11
1221	0.14	0.08	**0.24**	0.24	–	3210	0.03	–	–	–	**0.15**	–	–	–
1220	0.11	0.06	**0.21**	0.12	0.18	**1220**	0.05	0.02	0.03		0.06	0.03	0.06	0.15
2211	0.10	**0.10**	–	0.12	0.37	**2211**	0.03	–	0.03	0.02	**0.06**	–	–	0.02
1210	0.05	–	**0.15**	0.12	–	**1210**	0.07	0.02	0.01	0.02	–	**0.36**	**0.17**	–
2021	0.03	**0.06**	–	–	–	3211	0.02	–	–	–	**0.12**	–	–	–
	0.81						**0.82**							

[a] EA (East Asia), SA (South Asia), SEA (Southeast Asia), NEA (Northeast Asia), EUR (Europe), and NA (North America)

revealed two important features of the multilocus organization in *O. sativa*. First, in both *indica* and *japonica* groups, alleles at these isozyme loci were present in only a few predominant phenotypes. For example, in the *indica* group, 109 (23%) of 480 possible quadruplex phenotypes (based on the numbers of alleles at individual loci) were detected, of which only 28 (6%) had frequencies≥0.03. In the *japonica* group, 93 (12%) of 769 possible quadruplex phenotypes were detected, of which only 23 (3%) had frequencies≥0.03. There was a single predominant quadruplex phenotype (>0.45) detected in each of the three data subsets, from which the basic *japonica* phenotype ('11332021') at the eight loci (*Ph*, *Sdh1*, *Pgd2*, *Pgi2*, *Est4*, *Est1*, *Acp3* and *Me2*) could be determined. Second, all but one ('2120' at *Ph*, *Sdh1*, *Est4* and *Est1*) of the four-locus predominant phenotypes in the *indica* group were not present in the *japonica* group, and vice versa, indicating that *indica-japonica* differentiation is much more pronounced in multi-locus phenotypic frequencies than in allelic frequencies at individual loci.

Table 6 Selected log-linear models of the polymorphic isozyme loci *Me2* (H), *Acp3* (G), *Lap2* (K), *Idh1* (I), *Pgd2* (C) and *Pgi1* (N), within both *indica* and *japonica* groups for four data sets – 1 (G, H, I and K), 2 (C, I and N), 3 (H, K, C and N) and 4 (C, H, K and N)

Data subset	Group	Selected model	G^2	*df*	*P*
1	*indica*	[HI][GI][IK]	6.75	8	0.563
	japonica	[HI][GI][IK]	5.21	14	0.982
2	*indica*	[CI][IN][CN]	5.61	6	0.469
	japonica	[CI][IN][CN]	5.04	4	0.283
3	*indica*	[H][CK][CN]	19.96	23	0.644
	japonica	[CH][KN][CN]	13.91	21	0.873
4	*indica*	[CG][CK][CN]	16.82	21	0.722
	japonica	[G][CN][KN]	20.68	23	0.600

Multilocus organization within indica *and* japonica

Six loci, *Pgd2*, *Acp3*, *Me2*, *Idh1*, *Lap2* and *Pgi1*, which were polymorphic within both *indica* and *japonica*, were organized to form four data subsets, and subjected to log-linear model analysis. The selected models for data subsets 1 and 2 (Table 6) were identical for both groups, i.e. strong pairwise associations occurred between *Idh1* and all other loci, as well as between *Pgi1* and *Pgd2*. There were no three or higher order associations among the five loci. However, the pairwise associations in the *indica* and *japonica* groups were in opposite directions (Table 7). For example, in the *indica* group, significant positive associations were detected for the gamete type '12' for *Me2* and *Idh1* (allele 1 of *Me2* and allele 2 of *Idh1*), '11' for *Acp3* and *Idh1* and for *Lap2* and *Idh1*, '32' for *Pgd2* and *Idh1*, and '21' for *Pgd2* and *Pgi1*, while the significant negative associations for those gamete types were found in the *japonica* group. In other words, the gamete types favored (in excess under the hypothesis of independence) in the *indica* group were not favored (in deficiency) in the *japonica* group. Data in subsets 3 and 4 (Table 6) show that associations between *Lap2* and *Pgd2*, and *Acp3* and *Pgd2*, were significant in the *indica* group but not in the *japonica* group where strong associations were detected between *Lap2* and *Pgi1*, *Me2* and *Pgd2*.

When the quadruplex phenotypic frequencies within both *indica* and *japonica* groups were examined (Table 5), geographic differentiation of isozyme variation was again more pronounced in multilocus phenotype frequencies than in allelic frequencies at individual loci. In the *indica* group, two distinct subgroups, accessions from China and those from South Asia, could be clearly distinguished from each other based on the quadruplex phenotypic frequencies at 12 polymorphic loci. For instance, accessions from East Asia (China) were characterized with high frequencies of phenotypes '2133', '2134' and '2132' at *Ph*, *Sdh1*, *Pgd2* and *Pgi2* (**ABCD**), '2221' and '2021' at *Est4*,

Table 7 Standardized association effects (U terms) estimated from the selected models of the polymorphic isozyme loci *Pgd2* (C), *Idh1* (I), *Acp3* (G), *Me2* (H), *Lap2* (K) and *Pgi1* (N) of the *indica* and *japonica* groups

Data subset 1 (I, G, H and K)				Data subset 2 (C, I and N)				
Association	Gamete type	*indica*	*japonica*	Association	*indica*		*japonica*	
		U term	U term		Gamete type	U term	Gamete type	U term
HI	21	1.6	–2.1*	CI	31	–3.2*	31	1.7
	11	–1.6	2.1*		32	3.2*	3	–1.7
					21	1.4	21	–3.1***
					22	–1.4	22	3.1***
GH	11	2.9***	3.0***	CN	35	1.6	22	2.8**
	21	–2.9***	3.0***		21	2.1*	24	–3.4***
					31	–4.1***	34	2.4**
IK	11	2.5**	–1.8	IN	12	1.6	12	3.5***
	12	–2.5**	1.8		22	–1.6	22	–3.5***
					11	–0.9	11	–2.0*
					21	0.9	21	2.0*

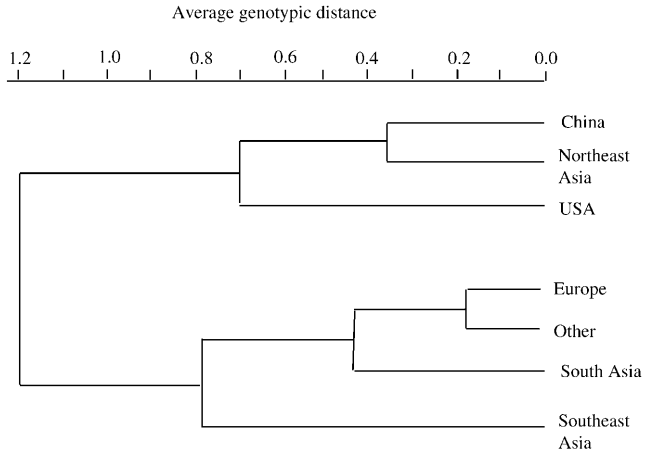

Fig. 2 Geographic differentiation in multilocus phenotypic frequencies within the *japonica* group

Est1, *Acp3* and *Me2* (**EFGH**), and '2221', '2220' and '2211' for *Idh1*, *Lap1*, *Lap2* and *East3* (**IJKL**). In contrast, accessions from South Asia had high frequencies of the phenotypes '2123', '2223' and '2222' for **ABCD**, '1222', '2121', '1211' and '1111' for **EFGH**, and '1211', '1221', '1220' and '1210' for **IJKL**. In all these cases, the multilocus phenotypes abundant in China were either absent or had very low frequencies in South Asia, and vise versa.

For the *japonica* group, East Asia (China) was the region with all predominant *japonica* phenotypes, which supported the common view that China was the center of origin for *japonica* rice (Oka 1988; Wang 1990). Cluster analysis using the genetic distances computed from the multi locus phenotypic frequencies revealed two subgroups with clear geographic distributions (Fig. 2). The major group covered accessions primarily from East Asia, Northeast Asia, and USA (primarily from California). The other subgroup included accessions from South Asia, Europe, Southeast Asia, and other regions. It is interesting to note that many accessions from the Southern US shared some unique phenotypes such as '3210' and '3211' at **IJKL** (Table 5) which were absent at all other regions, while the phenotype '1210' was present in all geographic regions of the *japonica* group but was absent in the Southern US cultivars.

Discussion

Indica-japonica differentiation and maintenance of isozyme variation in rice

While it is generally agreed that the *indica-japonica* differentiation in *O.sativa* at the phenotypic level was largely attributable to their adaptation to very different environments such as temperature and water conditions, it is not clear what caused such differentiation at the isozyme level (Second 1982; Glaszmann 1987; Oka 1988). We found that the *indica-japonica* differentiation was associated with large allelic-frequency differentiation at many isozyme loci, but more significantly with the formation of completely different multilocus phenotypes. Furthermore, considerable non-random associations between or among alleles at many isozyme loci were organized in a non-hierarchical pattern. According to Bishop et al. (1975), a non-hierarchical pattern of non-random associations among alleles at multiple loci could be related to the concept of 'synergism', which may suggest possible epistatic relationships among the alleles at some isozyme loci. Linkage does not explain the frequency distributions of the multilocus complexes involving the eight loci (*Ph*, *Sdh1*, *Pgd2*, *Pgi2*, *Est4*, *Est1*, *Acp3* and *Me2*) because they are unlinked (Wu et al. 1988; Kinoshita and Takahashi 1991). Furthermore, we note that the sampled *indica* and *japonica* accessions consist of approximately equal numbers of land races and modern semidwarf cultivars which share the same predominant multilocus isozyme phenotypes within the same group, and the predominant multilocus *indica* and *japonica* isozyme phenotypes were positively correlated fitness characters in the same environments (Li 1989). The tendency of predominant multilocus *indica* and *japonica* isozyme phenotypes to recover quickly under no apparent selection after being broken completely by recombination has been experimentally demonstrated in progenies from *indica*×*japonica* crosses (Oka 1988; Sato 1990). These results suggest that epistasis is the intrinsic force in maintaining the integrity of these predominant *indica* and *japonica* multilocus gene complexes and that the *indica* and *japonica* differentiation at the isozyme level was due primarily to disruptive selection on the multilocus isozyme phenotypes.

Pronounced multilocus frequency differentiation at isozyme loci has been reported in several other plant species, including barley and corn (Allard et al. 1993; Allard 1996). The observation that natural and/or artificial selection operates primarily on allele combinations rather than on individual alleles at isozyme loci, suggests the presence of epistatic relationships between or among loci which jointly influence fitness traits (Allard 1996). Direct evidence has been obtained recently in a QTL mapping study in which fitness traits of the progeny from an *indica-japonica* cross were found to be determined by large numbers of complementary (epistatic) loci (Li et al. 1997a, b). Thus, epistasis may have played a key role in the evolution of predominantly selfing plant species such as rice, where the shifting balance process operates in its most extreme form for more-pronounced epistasis at multiple loci and the subdivision of populations by inbreeding (Wright 1931).

Our results indicate that the among-population variation, including the subspecific differentiation and the macro-geographic differentiation within cultivar groups, accounts for 42% of the total isozyme variation in *O. sativa*, which is consistent with that in most other self-pollinated plant species (Hamrick and Godt 1989). This proportion of isozyme variation in rice appeared to be

maintained primarily by epistasis and selection as discussed above. However, the within-population (country) variation still accounts for a greater portion of the total isozyme variation (58%). We noted that absolute within-country gene diversity in each of the groups was correlated with ecological variability and environmental heterogeneity rather than geography. For instance, maximum isozyme diversity areas (the new group) and South Asia of the *indica* group were along the Himalayas where the environments were most heterogeneous. Also, the highest gene diversity corresponded with remarkable ecotype differentiation in *indica*, the new group, and *japonica* (Chinese accessions). The observation that maximum isozyme diversity parallels the maximum variability for many morphological and life-history characters in the accessions from 'diversity centers' has been reported previously in rice (Chang 1976; Oka 1988) and in many other selfing plant species (Brown and Burdon 1987; Hamrick and Godt 1989). Direct correlations between allozyme polymorphisms with environmental factors have also been detected in many plant species (Clegg and Allard 1972; Kahler et al. 1980; Nevo and Ben-Shlomi 1984; Nevo and Krugman 1988; Allard et al. 1993; Allard 1996). Mutations apparently contribute little to the maintenance of isozyme variation in *O. sativa* since isozyme alleles present in *O. sativa* were also present in its presumed direct ancestor, *O. rufipogon* populations (Oka 1988). Natural and artificial selection, inbreeding and genetic drift during domestication have not eliminated most wild-species alleles at these isozyme loci. These results suggest that the majority, or at least part, of the within-population variation at the isozyme loci in rice has been maintained by both spatial and temporal variation with selection.

Origin of *indica* and *japonica* subspecies

Several points can be made concerning the origin of *O. sativa*. First, the observation that *indica* and *japonica* subspecies consisted of distinct multilocus phenotypes, suggesting that the origin of the *indica* and *japonica* rice accessions studied were parallel in the sense that *japonica* did not evolve from *indica*, as presumed previously (Ting 1957). Second, significant differences in allelic frequency at some isozyme loci and distinctive multilocus phenotypes of Chinese *indica* accessions from those of South Asia (India), suggest the dual origins of *indica* rice. Third, our results support the common view that China was the primary center of origin for *japonica* rice (Oka 1988; Wang 1990). Fourth, the differentiation of most accessions from Southeast Asia and South Asia from the primary *japonica* group in their multilocus phenotypes plus their unique morphology and ecological niches (Chang 1976), strongly suggest the presence of two ecotypes, 'temperate *japonica*' and 'tropical *japonica*' (or '*javanica*'), within the *japonica* subspecies, and possible multiple origins of *japonica* rice. It is noted that most accessions from the Southern US have some multilocus phenotypes which are absent or in low frequency in other regions and constitute a unique subgroup within the *japonica* group. Similar results based on RAPD markers have been reported (Mackill 1995). However, the difference at the isozyme level between the Southern US cultivars and the temperate *japonica* accessions from China and Northeast Asia appears to be less pronounced than that between the traditional temperate *japonica* and *javanica* ecotypes. Finally, the maximum differentiation at the isozyme level between *indica* and *japonica* rices from China detected here, as well as several recent archeological discoveries which revealed 7000–9000 years of rice culture history and the occurrence of mixed (*indica*-like and *japonica*-like types) ancient rice grains in China (Wang 1990; Yiu 1990), suggest that China was the place where rice (both *indica* and *japonica*) was first domesticated.

We speculate that *O. sativa* could have originated from some populations of the presumed direct ancestor, annual Asian *Oryza rufipogon*, which might have become spread in most tropical and subtropical areas in Asia. Local *O. rufipogon* populations had become differentiated more or less depending upon the environments they were subjected to and therefore were mixtures of numerous more or less *indica*-like and *japonica*-like phenotypes before domestication, as suggested by several archaeological discoveries (Yiu 1990). Considerable genetic variation at all levels within *O. rufipogon* populations could have been maintained jointly by varied selection, due to variable epistasis among alleles at multiple loci and to both spatial and temporal variation in environments, interacting with genetic drift and limited outcrossing. When the seeds of those populations were harvested by the earliest people and carried to different and more-extreme environments, they may have become more differentiated under different selection pressures. In other words, the formation of *indica* and *japonica* rices was the direct product of natural selection and domestication. This hypothesis is much the same as that proposed by Oka (1988) and Wang (1990), except that original *indica*-like and *japonica*-like allelic or character associations in the *O. rufipogon* populations were assumed to have been established primarily by natural selection rather than by founder events. Initial allele or character associations in those *O. rufipogon* populations were at a lower level (predominant pairwise associations) and maintained by weak selection varying both spatially and temporally, as suggested by a large body of evidence from previous studies (see Oka 1988).

Acknowledgements We are grateful for many valuable suggestions and cricsisms from Dr. A.L. Kahler. We also thank Drs. David Mackill and Ken Gravois for their valuable comments and suggestions. This research was supported in part by the Rockefeller Foundation Grant 86058 #50, by USDA-ARS, and by funds from the California Rice Research Board.

References

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

Allard RW, García P, Sáenz de Miera LE, Pérez de la Vega M (1993) Evolution of multilocus genetic organization in *Avena hirtula* and *Avena barbata*. Genetics 135:1125–1139

Bishop YMM, Feinberg SE, Holland PW (1975) Discrete multivariate analysis: theory and practice. The MIT Press, Cambridge, Massachusetts

Brown AHD (1978) Isozymes, plant population genetic organization and genetic conservation. Theor Appl Genet 52:145–157

Brown AHD, Burdon JJ (1987) Mating system and colonizing success in plants. In: Gray AJ, Crawley MJ, Edwards PJ (eds) Colonization, succession, and stability. Blackwell Science Publishers, Oxford, pp 115–131

Brown CR, Jain SK (1979) Reproductive system and pattern of genetic variation in two Limnanthes species. Theor Appl Genet 54:181–190

Chang Te-Tzu (1976) The origin, evolution, cultivation, and diversification of Asian and African rices. Euphytica 25:425–441

Clegg MT, Allard RW (1972) Patterns of genetic differentiations in the slender wild oat species *Avena barbata.* Proc Nat Acad Sci USA 69:1820 –1824

Feinberg SE (1980) The analysis of cross-classified categorical data. The MIT Press, Cambridge, Massachusetts, USA and London, England

Hamrick JL, Godt MJ (1989) Allozyme diversity in plants. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer Association Incorporated, Sunderland, Massachusetts, pp. 43–63

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

Kahler AL, Allard RW, Krzakowa M, Wehrhahn CF, Nevo E (1980) Associations between isozyme phenotypes and environment in the slender wild oat (*Avena barbata*) in Israel. Theor Appl Genet 56:31–47

Kato S, Kosaka H, Hara S (1928) On the affinity of rice varieties as shown by the fertility of rice plants (*in Japanese* with *English* summary). Centr Agric Inst Kyushu Imp Univ 2: 241–276

Kinoshita T, Takahashi M (1991) The one hundredth report of genetical studies on rice plants – linkage studies and future prospects. J Fac Agric Hokkaido Univ 65:1–61

Li Z (1989) Genetic structure of cultivated rice (*Oryza sativa* L.): isozymes and morphology. PhD dissertation, University of California, Davis

Li Z, Pinson SRM, Park WD, Paterson AH, Stansel JW (1997a) Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 145:453–465

Li Z, Pinson SRM, Park WD, Paterson AH, Stansel JW (1997b) Genetics of hybrid sterility and hybrid breakdown in rice (*Oryza sativa* L.). Genetics 145:1139–1148

Mackill DJ (1995) Classifying *japonica* rice cultivars with RAPD markers. Crop Sci 35:889–894

Matsuo T (1952) Geoecological studies on cultivated rice (in *Japanese* with *English* summary). Bull Natl Inst Agric Sci Jpn D3:1–111

Nakagahra M (1977) Genetic analysis for esterase isozymes in rice cultivars. Jpn J Breed 27:141–148

Nakagahra M (1978) The differentiation, classification and center of genetic diversity of cultivated rice (*Oryza sativa* L.) by isozyme analysis. Trop Agr Res Ser 11:77–82

Nakagahra M (1984) Geographical distribution of esterase genotypes of rice in Asia. Rice Genet Newslett 1:118–120

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nei M (1973) The theory and estimation of genetic distance. In: Morton NE (ed) Genetic structure of populations. University of Hawaii Press, Honolulu, pp 45–54

Nevo EBA, Ben-Shlomi B (1984) The evolutionary significance of genetic diversity: ecological, demographic and life history correlates. In: Mani GS (ed) Evolutionary dynamics of genetic diversity. Lecture notes in biomathematics, vol **53**. Springer, New York, pp 13–213

Nevo EBA, Krugman T (1988) Natural selection of allozyme polymorphisms: a microgeographic climatic differentiation in wild emmer wheat (*Triticum dicoccoides*). Theor Appl Genet 75:529–538

Oka HI (1958) Intervarietal variation and classification of cultivated rice. Genet Plant Breed 18:79–89

Oka HI (1988) Origin of cultivated rice. Japan Sci Soc Press, Tokyo

Sato YI (1990) How was rice differentiated into indica and japonica? In: IRRI (ed) Rice genetics II. Proc 2nd Int Rice Genet Symp, Manila, The Philippines, pp 45–53

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

Terao H, Mizushima U (1939) Some considerations on the classification of *O. sativa* L. Jpn J Bot 10:213–258

Ting Y (1957) The origin and evolution of cultivated rice in China (*in Chinese*). Acta Agron Sinica 8:242–260

Wang XK (1990) Origin, evolution and classification of chinese rice (*Oryza sativa* L.) (in *Chinese*). In: Ying CS (ed) Rice germplasm resources in China, pp 1–16

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159

Wu KS, Glaszmann JC, Khush GS (1988) Locations of isozyme loci in rice (*Oryza sativa* L.) through trisomic analysis. Biochem Genet 26:303–320

Yiu XL (1990) On the origin of rice in China based on ancient rice grains in Homudu archaeological discoveries (in Chinese with *English* summary). Acta Sinica 5:1–10