Genet. Res., Camb. (2005), **85**, *pp.* 1–14. *With 3 figures.* © 2005 Cambridge University Press
doi:10.1017/S0016672304007293 Printed in the United Kingdom

Microsatellite diversity within *Oryza sativa* with emphasis on *indica–japonica* divergence

LI-ZHI GAO[1]*, CHI-HONG ZHANG[1], LI-PING CHANG[2], JI-ZENG JIA[1], ZONG-EN QIU[1] AND YU-SHEN DONG[1]
[1] *Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences, Beijing 100081, P.R. China*
[2] *Guang' anmen Hospital, Chinese Academy of Traditional Chinese Medicine, Beijing 100053, P.R. China*

(*Received 12 July 2004 and in revised form 5 November 2004*)

Summary

The molecular evolution of cultivated rice *Oryza sativa* L. has long been a subject of rice evolutionists. To investigate genetic diversity within and differentiation between the *indica* and *japonica* subspecies, 22 accessions of *indica* and 35 of *japonica* rice were examined by five microsatellite loci from each chromosome totalling 60 loci. Mean gene diversity value in the *indica* rice ($H=0{\cdot}678$) was 1·18 times larger than in the *japonica* rice ($H=0{\cdot}574$). Taking the sampling effect into consideration, average allele number in the *indica* rice was 1·40 times higher than that in the *japonica* rice (14·6 vs 10·4 per variety). Chromosome-based comparisons revealed that nine chromosomes (1, 2, 3, 4, 5, 8, 9, 10 and 11) harboured higher levels of genetic diversity within the *indica* rice than the *japonica* rice. An overall estimate of F_{ST} was 0·084–0·158, indicating that the differentiation is moderate and 8·4–15·8% of the total genetic variation resided between the *indica* and *japonica* groups. Our chromosome-based comparisons further suggested that the extent of the *indica–japonica* differentiation varied substantially, ranging from 7·62% in chromosome 3 to 28·72% in chromosome 1. Cluster analyses found that most varieties formed merely two clusters for the *indica* and *japonica* varieties, in which two *japonica* varieties and five *indica* varieties were included in the counterpart clusters, respectively. The 12 chromosome-based trees further showed that 57 rice varieties cannot be clearly clustered together into either the *indica* or *japonica* groups, but displayed relatively different clustering patterns. The results suggest that the process of *indica–japonica* differentiation may have proceeded through an extensive contribution by the alleles of the majority in the rice genome.

1. Introduction

Cultivated rice (*Oryza sativa* L.) is the staple food crop for more than half of the world's populations (Chang, 1984). It is commonly recognized that the cultivated rice has differentiated into two subspecies, subsp. *indica* and subsp. *japonica*, during the domestication and selection process (Chang, 1976; Oka, 1988; Morishima *et al*., 1992). The adaptive evolution of rice varieties to different ecological environments may have led to the *indica–japonica* differentiation of Asian cultivated rice. The genetic variation within the two subspecies, which represents two partially isolated gene pools, is a major source of genetic diversity in the world rice germplasm. *Indica*/*japonica* hybrid varieties promise to be of tremendous yield potential compared with intra-subspecific rice hybrids because two parents of the former are more distantly related (Yuan *et al*., 1989). The exploration and utilization of the genes which are involved in such a differentiation and are capable of overcoming the partial reproductive isolation mechanisms between these two groups have always been pursued by rice geneticists and breeders (Ikehashi & Araki, 1986). The molecular evolution, particularly genetic basis of diversity and differentiation in the *indica* and *japonica* rice

* Corresponding author: Li-zhi Gao, University of Texas Health Science Center at Houston, School of Public Health, Human Genetics Center, 1200 Hermann Pressler, Houston, TX 77030, USA. Fax: +1 (713) 500 0900. e-mail: LZGao@sph.uth.tmc.edu

Table 1. *Oryza sativa varieties listed in this study*

Sample no.	Sample varieties	Subspecies	Sample no.	Sample varieties	Subspecies
1	Meidalu	Jp.	30	Dianchao101	Id.
2	Jinglu	Jp.	31	Jingdao1187	Jp.
3	Yingdao	Jp.	32	Yunxian36	Jp.
4	Jingxian89	Id.	33	Hebeiaiyuan	Jp.
5	Zhou903	Id.	34	Hanjing911	Jp.
6	Aijiunante	Id.	35	Hejiang19hao	Jp.
7	Qishanzhan	Id.	36	Xifeng98-12	Jp.
8	Xianghanxian3hao	Id.	37	Tainan17	Jp.
9	Honghandao	Jp.	38	Zhenfudao	Id.
10	Yuanjiugu	Jp.	39	Dianrui456	Id.
11	Dahongmang	Id.	40	Qiuguang	Jp.
12	Liuhuangzhan1	Jp.	41	Dianchao103	Id.
13	Wuyiujing3hao	Jp.	42	Chao40	Id.
14	Shengyiu2	Id.	43	Xiengshi	Jp.
15	Teqing2	Id.	44	Heidao110	Jp.
16	Duochandao	Id.	45	Yuanfu	Jp.
17	Dashengchan	Id.	46	Chunfeng	Jp.
18	Xiushui11	Jp.	47	Hunanruanmi	Id.
19	Giayu293	Jp.	48	Dianlong201	Id.
20	Tongshansanlicun	Jp.	49	Xiushui04	Jp.
21	Zipihandao	Jp.	50	Juanguang	Jp.
22	Zhenguiai	Id.	51	Jingruan1ao	Jp.
23	Ribenqing	Jp.	52	Hailin1hao	Jp.
24	SLG-1	Jp.	53	Xianghu84	Jp.
25	Guangluai4hao	Id.	54	Xiushui46	Jp.
26	Puanshendao	Id.	55	Dianrui449	Id.
27	Yinmixian	Jp.	56	Minghui63	Id.
28	Ranfen98-8	Jp.	57	Taihujing2hao	Jp.
29	Ji89-60	Jp.			

Id, indica; Jp., japonica.

genomes of the species, besides its huge economic significance, has long been an unending interest among rice evolutionists. Therefore, clarifying evolutionary relationships between the two subspecies and a clear understanding of genetic diversity within the species not only will be essential in guiding our efforts to seek beneficial genes for the hybrid rice breeding programmes, but also will form a foundation for further evolutionary studies in the genomics era.

The assessment of genetic diversity and/or differentiation of rice has been attempted in diverse rice samples with many systems of genetic markers such as morphological traits (Oka, 1964; Morishima & Oka, 1981; Cheng, 1985), biochemical markers (Nakagahra, 1978; Second, 1982; Glaszmann, 1987), DNA randomly amplified polymorphisms (RAPDs) (Mackill, 1995), DNA restriction fragment length polymorphisms (RFLPs) (Wang & Tanksley, 1989; Zhang *et al*., 1992; Sun *et al*., 2002), amplified fragment length polymorphisms (AFLPs) (Mackill *et al*., 1996; Zhu *et al*., 1998), and simple sequence repeats (SSRs) (Yang *et al*., 1994). These studies revealed different extents of genetic variation within cultivated rice and suggested that the *indica* rice is genetically more diverse than the *japonica* rice. However, the estimates of *indica–japonica* divergence of Asian cultivated rice were surprisingly high; that is, it is about 50% with morphological analysis (Morishima & Oka, 1981) and about 34% based on RFLP data (Zhang *et al*., 1992). When measured using allozymes, the variability of highly selfing species (e.g. rice) was often seen to be largely reduced (Nakagahra, 1978; Second, 1982; Glaszmann, 1987); the nature of RFLPs of low polymorphisms assayed within a self-pollinated species (e.g. *O. sativa*) (Zhang *et al*., 1992) determined that it is not suitable to be a genetic marker to give an accurate estimate of the genetic diversity and particularly population structure. Therefore, fine-scale analyses to better test the *indica–japonica* divergence and understand the population structure of Asian cultivated rice require more polymorphic markers. The most promising candidates are microsatellites, which exhibit the same desirable attributes as population genetic markers as allozymes (co-dominance, Mendelian inheritance, presumed neutrality), though they are much more variable and abundant in most eukaryotic species (e.g. Bowcock

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

et al., 1994; Estoup *et al*., 1995). The microsatellite markers have proved a powerful tool to analyse significantly higher polymorphisms in rice for evolutionary analyses (e.g. Wu & Tanksley, 1993; Yang *et al*., 1994; Nagaraju *et al*., 2002; Gao *et al*., 2002). They have made rice even more attractive as a model system for molecular evolutionary studies, since such markers could be further verified on the fully sequenced rice genome of two subspecies in the near future. By using diverse samples of rice germplasm, Yang *et al*. (1994) studied genetic variation and *indica–japonica* differentiation at the 10 polymorphic microsatellites. However, a common problem associated with the estimate in this study and most estimates in other above-mentioned studies is that these markers only represented a minority of the genome, and were not sampled from the whole genome.

Here, we report a comprehensive analysis of genetic variation and differentiation by using a sample of 57 rice varieties representing considerable breadth of the species *O. sativa* and 60 microsatellite markers that provide a broad coverage of the rice genomes and are evenly distributed on the 12 chromosomes. The aims in the present study were to investigate patterns of the genome-wide genetic diversity and differentiation that occur at different loci, chromosomes and the whole genome, and gain insights into the nature of the *indica–japonica* differentiation with an evolutionary perspective.

2. Plant materials and methods

(i) *Plant materials*

The materials used in this study were 57 varieties of cultivated rice *O. sativa* including 22 of the *indica* type and 35 of the *japonica* type (Table 1). Most materials were obtained from the Institute of Crop Germplasm Resources (Beijing City), Chinese Academy of Agricultural Sciences, and others were provided by the China Rice Research Institute (Hanzhou City) and Yunnan Academy of Agricultural Sciences (Kunming City), respectively. Leaves were harvested from a single plant of each variety grown in the greenhouses of the Institute of Crop Germplasm Resources. DNA was isolated from young fresh leaf tissues according to the method of Edwards *et al*. (1991).

(ii) *Simple sequence repeat genotyping*

When this study was conducted in 1997, there were a total of 323 microsatellite loci publicly available (Wu & Tanksley, 1993; Panaud *et al*., 1996; Akagi *et al*., 1996; Chen *et al*., 1997). We selected 5 loci for each chromosome totalling 60 primer pairs that are evenly distributed throughout the rice genome (Table 2). These microsatellite loci are also displayed on the Cornell University Rice Genes web site (http://www.gramene.org/microsat/ssr.txt). Microsatellite polymorphisms were assayed by specific PCR conditions following Panaud *et al*. (1996). PCR products were run on 6% polyacrylamide denaturing gels, and marker bands were revealed using silver staining as described by Panaud *et al*. (1996). The null alleles were confirmed after several repetitions with different amplification conditions to ensure that no reaction failure existed. To determine the allele size, the samples were directly compared with band sizes from an allelic ladder prepared by amplification of an artificial mixture of DNA from all the assayed samples.

Table 2. *Sixty microsatellites and their chromosome locations used in the study*

Chromosome location	Microsatellite markers	Chromosome location	Microsatellite markers
1	RM212	7	RM47
1	OSR23	7	OSR22
1	OSR27	7	RM234
1	RM220	7	RM82
1	RM200	7	RM248
2	RM211	8	OSR35
2	OSR8	8	OSR7
2	OSR11	8	OSR30
2	RM233A	8	RM223
2	RM263	8	RM25
3	RM231	9	OSR28
3	RM55	9	OSR29
3	RM168	9	RM215
3	RM60	9	RM245
3	RM232	9	RM242
4	RM241	10	RM244
4	RM261	10	OSR33
4	OSR15	10	RM222
4	RM252	10	RM258
4	RM255	10	RM228
5	RM249	11	RM206
5	RM164	11	RM202
5	RM233B	11	RM167
5	OSR34	11	RM209
5	RM26	11	RM224
6	RM225	12	RM83
6	OSR19	12	OSR20
6	OSR21	12	OSR32
6	OSR25	12	RM235
6	RM253	12	RM247

(iii) *Evaluation of polymorphisms and population structure*

Genetic variability for each subspecies was first assessed by calculating the number of alleles per locus (A) and allelic richness (R_S) (a measure of allele number independent of sample size; see Petit *et al*., 1998). For selfing species (e.g. rice) which are

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

Table 3. *A summary of population genetic values of O. sativa at 60 microsatellite loci*

Loci/chromosomes	*H*_I	*H*_J	*Rs*_I	*Rs*_J	*Rs*_ Sativa	*A*_I	*A*_J	*A*_ Sativa	F_{IS}_I	F_{IS}_J
RM212	0·511	0·243	2·942	3·411	3·262	3	4	4	1·000	1·000
OSR27	0·766	0·482	6·785	5·147	7·109	6	5	10	1·000	1·000
OSR23	0·485	0·297	2·96	3·609	3·532	3	4	4	1·000	1·000
RM200	0·901	0·881	11·46	10·338	11·832	13	13	17	0·102	−0·005
RM220	0·732	0·805	5·952	6·798	7·411	6	7	8	1·000	1·000
All for chromosome 1	**0·679**	**0·542**	**6·020**	**5·861**	**6·629**	**31**[a] **(6·2**[b]**)**	**33 (6·6)**	**43 (8·6)**	**0·820**	**0·799**
RM211	0·541	0·165	3·808	2·577	3·17	4	3	4	1·000	1·000
OSR11	0·66	0·554	4	6·284	6·737	5	7	8	1·000	1·000
RM263	0·934	0·706	8·978	4·912	7·675	9	6	9	1·000	1·000
OSR8	0·719	0·593	4·92	5·57	5·695	4	7	7	1·000	1·000
RM233A	0·838	0·728	6·842	5·165	7·582	7	5	7	1·000	1·000
All for chromosome 2	**0·738**	**0·549**	**5·71**	**4·902**	**6·172**	**29 (5·8)**	**28 (5·6)**	**37 (7·4)**	**1·000**	**1·000**
RM168	0·788	0·316	5·896	3·376	5·457	6	3	7	1·000	1·000
RM60	0	0	1	1	1	1	1	1	–	–
RM55	0·688	0·598	4·889	3·871	4·84	5	4	6	1·000	1·000
RM232	0·779	0·819	5·882	6·594	6·71	6	7	8	1·000	1·000
RM231	0·817	0·825	6·928	7·285	7·224	6	7	8	1·000	1·000
All for chromosome 3	**0·614**	**0·512**	**4·919**	**4·425**	**5·046**	**24 (4·8)**	**22 (4·4)**	**30 (6·0)**	**1·000**	**1·000**
RM255	0·683	0·675	4·994	6·192	5·813	5	8	8	1·000	1·000
RM252	0·835	0·585	5·984	5·037	6·828	8	6	9	1·000	1·000
RM261	0·749	0·313	6·615	4·002	5·737	7	6	9	1·000	1·000
RM241	0·858	0·751	7·767	8·04	9·691	8	11	14	1·000	1·000
OSR15	0·381	0·377	2	4·309	3·694	3	6	6	1·000	1·000
All for chromosome 4	**0·701**	**0·540**	**5·472**	**5·516**	**6·353**	**31 (6·2)**	**37 (7·2)**	**46 (9·1)**	**1·000**	**1·000**
OSR34	0·634	0·545	3	2·754	3	3	3	3	0·912	1·000
RM233B	0·742	0·234	3·998	2·854	4·491	5	4	5	1·000	1·000
RM26	0·61	0·116	2·996	2·383	2·859	3	3	3	1·000	1·000
RM249	0·899	0·843	12·081	14·368	14·276	7	13	15	0·722	0·763
RM164	0·85	0·686	6·735	7·024	7·662	7	8	10	0·941	1·000
All for chromosome 5	**0·747**	**0·485**	**5·762**	**5·877**	**6·458**	**25 (5·0)**	**31 (6·2)**	**36 (7·2)**	**0·915**	**0·953**
RM225	0·9	0·872	8·696	9·619	9·756	8	11	12	1·000	1·000
OSR21	0·522	0·578	2·989	3·989	3·895	3	4	4	1·000	1·000
OSR19	0·686	0·653	3·991	5·207	4·891	4	6	6	1·000	1·000
OSR25	0	0	1	1	1	1	1	1	–	–
RM253	0·81	0·85	7·617	8·157	8·234	7	9	9	1·000	1·000
All for chromosome 6	**0·584**	**0·591**	**4·859**	**5·594**	**5·555**	**23 (4·6)**	**31 (6·2)**	**32 (6·4)**	**1·000**	**1·000**
RM234	0·662	0·563	3·923	5·265	5·805	2	7	8	1·000	1·000
RM47	0·091	0·113	1·904	2·534	2·444	2	5	6	1·000	0·746
RM248	0·905	0·724	8	6·265	9·215	9	6	11	0·926	1·000
OSR22	0·61	0·634	3	5·284	5·11	3	5	6	0·926	0·955
RM82	0·205	0·632	2·92	3	2·995	3	4	4	1·000	1·000
All for chromosome 7	**0·495**	**0·533**	**3·949**	**4·47**	**5·114**	**19 (3·8)**	**27 (3·4)**	**35 (7·0)**	**0·970**	**0·940**
OSR07	0·642	0·526	3·942	4·452	4·531	5	4	6	1·000	1·000
RM025	0·813	0·667	7·85	4·686	6·539	8	5	8	0·928	1·000
RM223	0·848	0·731	7·689	6·994	8·061	8	8	9	1·000	1·000
OSR35	0·88	0·841	7·874	6·602	8·246	7	7	9	0·940	1·000
OSR30	0·726	0·628	4	3·753	3·997	4	4	4	0·595	0·522
All for chromosome 8	**0·782**	**0·679**	**6·271**	**5·297**	**6·275**	**32 (6·4)**	**28 (5·6)**	**36 (7·2)**	**0·893**	**0·904**
RM215	0·6	0·817	5·693	6·879	7·401	6	8	9	1·000	1·000
OSR28	0·616	0·749	5·622	7·252	7·109	5	8	9	0·756	0·833
OSR29	0·526	0·631	2·96	5·461	5·244	2	6	6	1·000	1·000
RM245	0·783	0·131	4·938	2·508	4·156	5	2	5	0·920	1·000
RM242	0·695	0·603	3·995	4·931	5·174	4	5	5	1·000	1·000
All for chromosome 9	**0·644**	**0·586**	**4·642**	**5·406**	**5·817**	**22 (4·4)**	**29 (5·4)**	**35 (7·0)**	**0·935**	**0·967**

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

Table 3. (*Cont.*)

Loci/chromosomes	H_I	H_J	Rs_I	Rs_J	Rs_ Sativa	A_I	A_J	A_ Sativa	F_{IS}_I	F_{IS}_J
RM244	0·658	0·494	4·807	3·413	4·128	5	4	5	1·000	1·000
OSR33	0·838	0·704	5·979	4·768	5·861	5	4	6	1·000	1·000
RM222	0·837	0·922	7·712	12·265	11·539	8	13	14	1·000	1·000
RM258	0·816	0·243	5·938	3·411	5·148	6	3	6	1·000	1·000
RM228	0·752	0·576	4·976	6·625	6·666	4	8	8	1·000	1·000
All for chromosome 10	**0·78**	**0·588**	**5·882**	**6·096**	**6·668**	**28 (7·6)**	**32 (6·4)**	**39 (7·8)**	**1·000**	**1·000**
RM202	0·879	0·806	6·93	7·133	8·309	7	8	10	0·943	1·000
RM206	0·904	0·874	8·947	9·92	11·225	9	11	15	1·000	1·000
RM167	0·706	0·807	5·711	6·958	6·889	6	8	8	1·000	1·000
RM224	0·84	0·579	5·903	4·728	6·685	5	7	9	1·000	1·000
RM209	0·625	0·299	3	4·379	4·644	3	6	7	0·900	0·892
All for chromosome 11	**0·791**	**0·673**	**6·098**	**6·624**	**7·55**	**30 (6·0)**	**40 (8·0)**	**49 (9·8)**	**0·969**	**0·978**
RM235	0·924	0·79	11·163	6·352	9·898	13	6	13	0·902	1·000
OSR32	0·699	0·558	4	3·685	3·934	4	4	4	0·774	0·832
OSR20	0·781	0·87	5·869	7·57	7·731	7	8	8	0·942	0·928
RM247	0·489	0·84	3·885	9·522	8·294	3	11	11	1·000	1·000
RM83	0	0	1	1	1	1	1	1	–	–
All for chromosome 12	**0·579**	**0·612**	**5·183**	**5·626**	**6·171**	**28 (5·6)**	**30 (6·0)**	**37 (7·4)**	**0·905**	**0·940**
ALL	**0·678**	**0·574**	**5·397**	**5·474**	**6·151**	**322 (5·4) 14·6[c]**	**368 (6·1) 10·5**	**455 (7·6) 8·0**		

Gene diversity within samples (H), allelic richness (R_S), Weir and Cockerham's F_{IS} (the heterozygote deficit within subspecies), allele numbers (A) at every locus, and allele numbers (A_T) across two subspecies for each subspecies are shown, respectively. Mean values for 12 chromosomes and the entire genome are also given.

[a] Total allele number observed on each chromosome in all the samples.

[b] Total allele number per locus observed for each chromosome in all the samples.

[c] Average allele number per variety observed in all the genome. The sampling effect was considered to estimate the values. Such data for each chromosome were not shown.

composed of non-random mating populations, the estimates that are usually called 'heterozygosity' may not be appropriate. Therefore, we estimated this quantity (gene diversity) within samples (following Nei, 1987) with the program FSTAT version 2.9.3 (Goudet, 2001). Genetic differentiation between the two subspecies was quantified for each locus and for all loci together. For this purpose, the program FSTAT version 2.9.3 (Goudet, 2001) was used to quantify Weir & Cockerham's (1984) estimators of F-statistics (f estimates F_{IS}, θ estimates F_{ST}). Tests for population differentiation were made using an unbiased estimated P value for a log-likelihood (G)-based exact test (Goudet *et al.*, 1996). In addition, Nei's estimation of gene diversity within and differentiation between the *indica* and *japonica* rice was estimated by H_S, H_T, D_{ST}, D_{ST}', H_T, G_{ST}, G_{ST}' and G_{IS} by using FSTAT version 2.9.3 (Goudet, 2001). For each subspecies–locus combination, departure from Hardy–Weinberg expectation was assessed by exact tests, with unbiased P values estimated through a Markov chain method (Guo & Thompson, 1992); a global test across loci and populations was constructed using Fisher's method (Raymond & Rousset, 1995). To test the hypothesis of heterozygote deficiency, the multiscore (U) test of Raymond & Rousset (1995) was employed. Tests for genotypic linkage disequilibrium among pairs of loci in each population were performed using Fisher's exact tests (Raymond & Rousset, 1995), with unbiased P values again derived by a Markov chain method. Pearson's correlation coefficient was employed to estimate relationships between different values.

(iv) *Phylogenetic analyses*

To illustrate *indica–japonica* differentiation on different rice chromosomes, 12 UPGMA trees of 57 genotypes were constructed using the chord distance of Cavalli-Sforza & Edwards (1967) based on the microsatellite variation of 5 loci from each chromosome. Then all 60 loci were pooled to construct UPGMA and neighbor-joining (Saitou & Nei, 1987) trees by means of Cavalli-Sforza & Edwards's chord distance to better investigate the differentiation between the two subspecies. Using NEIGHBOR in the PHYLIP computer package (version 3.5c; Felsenstein, 1995), UPGMA phenograms were generated. The

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

robustness of Nei's unbiased genetic distance trees was assessed by creating 999 bootstrap replicates of the data set with the SEQBOOT algorithm in PHYLIP, and then generating a majority rule consensus tree in the CONSENSE program. Two accessions of wild progenitor, *O. rufipogon*, were used as outgroups with respect to the relationships of two rice subspecies. They were collected from Shangsi (Guangxi Province) and Yuanjiang (Yunnan Province), respectively. These distance trees were viewed by the program TREEVIEW (Page, 1996).

3. Results

(i) *Evaluation of SSR polymorphisms detected in the total sample*

A total of 455 allelic variants were identified in 57 cultivars with 60 microsatellite markers (data not shown). In this study, the banding patterns resolved by each primer pair were in accordance with single locus variation, and polymorphisms detected have been subjected to Mendelian heritage. Therefore, we refer to the sequence amplified by each primer pair as a locus and a variant as an allele.

Genetic diversity for each of the microsatellite loci and each of 12 chromosomes in two rice groups is summarized in Table 3. The number of alleles varied greatly among the 60 loci with an average of 7·58 alleles per locus. RM60, OSR25 and RM83 were monomorphic, whereas all other loci were polymorphic in both *indica* and *japonica* groups with the number of alleles ranging from 3 (OSR34 and RM26) to 17 (RM200). There appears to be no correlation between the number of alleles detected and the number of SSR repeats in the SSR loci. For example, 7 microsatellite loci containing the (GA) repeat motifs did not display obvious correlation with the number of alleles detected ($r=0{\cdot}164$, $P=0{\cdot}726$), while 32 microsatellite loci containing the (CT) repeat motifs showed no correlation with the detected alleles ($r=0{\cdot}140$, $P=0{\cdot}445$). The diversity values (H, R_S and A) in the total sample also varied widely from one locus to another and from one chromosome to another across the two rice groups (Table 3, Fig. 1). Although the extent of such variation in average diversity is not as large as the difference in the number of alleles per locus, genetic diversity values (H) seem to be significantly correlated with the number of alleles (A) (the *indica* rice varieties: $r=0{\cdot}774$, $P=0{\cdot}000$; the *japonica* rice varieties: $r=0{\cdot}772$, $P=0{\cdot}000$). Across the two rice groups, for example, the highest gene diversity of 0·891 was observed for RM200 with 17 alleles, whereas OSR34 and RM26 with 3 alleles each exhibited diversity values of 0·589 and 0·361, respectively (Tables 3, 4).

(ii) *Chromosome-based comparison of genetic diversity between indica and japonica rice*

More alleles were resolved in the *japonica* varieties ($Na=368$) than in the *indica* varieties ($Na=322$). Forty-six (12·5%) more alleles were observed in the *japonica* rice varieties than in the *indica* rice varieties, but most of these alleles had low frequencies (data not shown). When the sampling effect was considered to estimate average allele numbers, however, the *indica* rice had more alleles than the *japonica* rice at the 60 loci detected (14·6 vs 10·4 per variety), although the difference is not significant when all 60 loci were pooled by t-test ($P=0{\cdot}426489$). Allelic richness (R_S) allows comparison of this quantity between different sample sizes. Our results suggested that the allele number in the *indica* rice ($R_S=5{\cdot}397$) is similar to that in the *japonica* rice ($R_S=5{\cdot}474$). Chromosome-based comparison of gene diversity further showed that the mean diversity value for all chromosomes in the *indica* rice ($H=0{\cdot}678$) was 1·18 times higher than in the *japonica* rice ($H=0{\cdot}574$) (Table 3, Fig. 1). Nevertheless, these differences are not statistically significant. Of the 57 polymorphic loci, 41 showed higher H values in the *indica* group than in the *japonica* group. Moreover, the mean values (H) for each chromosome indicated that microsatellite variation is not randomly distributed among the different chromosomes of the two rice groups. Nine chromosomes (1, 2, 3, 4, 5, 8, 9, 10 and 11) possessed higher levels of genetic diversity in the *indica* rice than in the *japonica* rice, whereas only three chromosomes (6, 7 and 12) had more variation in the *japonica* rice than in the *indica* rice. However, only four chromosomes (2, 5, 10 and 11) have significant differences between two rice groups when every five loci were pooled.

(iii) *Population structure and indica–japonica differentiation within O. sativa*

Single-locus exact tests for Hardy–Weinberg equilibrium showed significant deviations in 118 out of 120 comparisons ($P<0{\cdot}05$), a number comparable to what would be expected by type I error alone. Moreover, the two rice groups showed significant heterozygosity deficits across loci ($P<0{\cdot}05$), and a global test across two groups and all loci for the species (Fisher's method) indicated a significant deviation from Hardy–Weinberg expectations (d.f. = 114, $P<0{\cdot}001$). Specifically testing for heterozygote deficiency (U-test) suggested statistically significant deficits at all loci across the two groups, all of which were significant following sequential Bonferroni corrections. Thus, there seems to be strong evidence for drastic departures from Hardy–Weinberg expectations and non-random mating. Exact tests for genotypic linkage disequilibrium further showed significant deviations

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

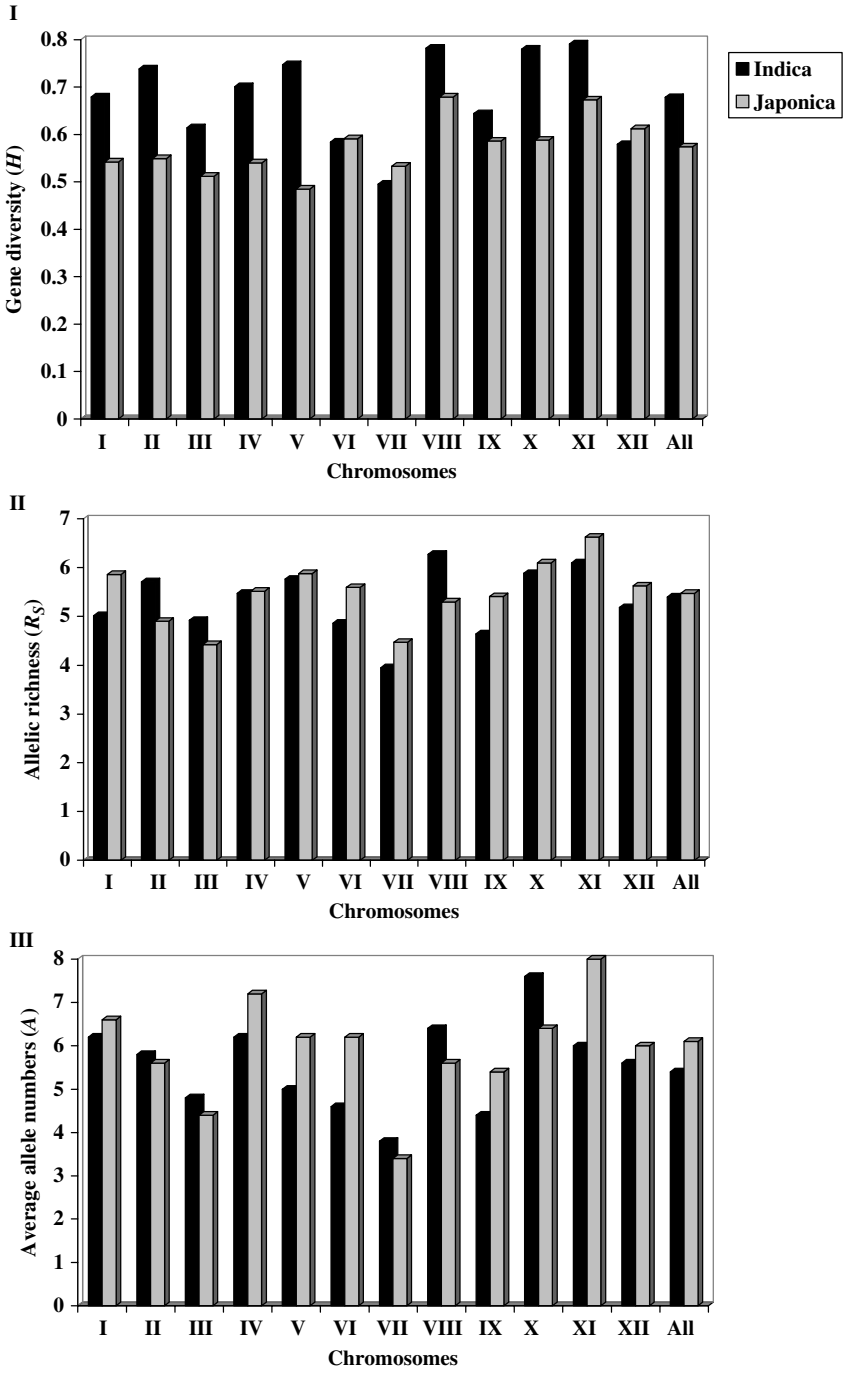

Fig. 1. Chromosome-based comparisons of genetic diversity within the *indica* and *japonica* subspecies based on genetic data of sixty microsatellites: (I) gene diversity; (II) allelic richness; and (III) average allele numbers.

across the two groups, suggesting that most of these loci are associated in very strong linkage disequilibrium. The estimates of F_{IS} for all loci ranged from 0·0351 at the locus RM200 to 1·0000 at 38 loci (Table 5), suggesting non-random mating and heterozygosity deficits within populations ($P<0·001$). At the intra-population level, F_{IS} was positive ($F_{IS}=0·9529$ across two rice groups of the species, $P<0·001$, Table 5; similar to Nei's $G_{IS}=0·946$ in Table 4), a pattern consistent with the heterozygosity deficits observed in tests of Hardy–Weinberg equilibrium. F_{IS} values showed that the two groups deviated from Hardy–Weinberg expectation within the species with a deficiency of heterozygotes in all 12 chromosomes. As a primarily selfing species with an estimated outcrossing rate of 0–5% (Oka, 1988), the above results are expected, suggesting that effective population sizes of both subspecies may be small.

After jackknifing over loci, an overall estimate of F_{ST} was 0·158, indicating that 15·8% of the total

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

Table 4. *Nei's estimators of gene diversity within and differentiation between indica and japonica rice at 60 microsatellite loci**

Locus	H_S	H_T	D_{ST}	D_{ST}'	H_T'	G_{ST}	G_{ST}'	G_{IS}
RM212	0·376	0·535	0·16	0·32	0·695	0·299	0·46	1
OSR27	0·623	0·736	0·113	0·226	0·849	0·154	0·266	1
OSR23	0·39	0·562	0·172	0·343	0·733	0·306	0·468	1
RM200	0·891	0·915	0·024	0·048	0·939	0·026	0·051	0·049
RM220	0·769	0·834	0·065	0·129	0·898	0·078	0·144	1
Mean for chromosome 1	**0·61**	**0·716**	**0·107**	**0·213**	**0·823**	**0·173**	**0·278**	**0·81**
RM211	0·351	0·532	0·181	0·362	0·713	0·34	0·507	1
OSR11	0·606	0·735	0·128	0·257	0·863	0·175	0·297	1
RM263	0·819	0·848	0·03	0·059	0·878	0·035	0·068	1
OSR8	0·655	0·722	0·067	0·134	0·789	0·093	0·17	1
RM233A	0·782	0·843	0·06	0·121	0·903	0·072	0·134	1
Mean for chromosome 2	**0·643**	**0·736**	**0·093**	**0·187**	**0·829**	**0·143**	**0·235**	**1**
RM168	0·55	0·596	0·046	0·092	0·642	0·077	0·143	1
RM60	0	0	0	0	0	0	0	0
RM55	0·643	0·636	−0·007	−0·014	0·629	−0·011	−0·023	1
RM232	0·799	0·826	0·027	0·054	0·853	0·033	0·063	1
RM231	0·821	0·848	0·027	0·054	0·875	0·032	0·061	1
Mean for chromosome 3	**0·563**	**0·581**	**0·019**	**0·037**	**0·6**	**0·026**	**0·049**	**1**
RM255	0·679	0·714	0·034	0·069	0·748	0·048	0·092	1
RM252	0·709	0·796	0·087	0·174	0·883	0·109	0·197	1
RM261	0·529	0·569	0·04	0·08	0·609	0·071	0·132	1
RM241	0·804	0·825	0·022	0·043	0·847	0·026	0·051	1
OSR15	0·379	0·561	0·182	0·364	0·743	0·325	0·49	1
Mean for chromosome 4	**0·62**	**0·693**	**0·073**	**0·146**	**0·766**	**0·116**	**0·192**	**1**
OSR34	0·589	0·668	0·079	0·158	0·747	0·118	0·212	0·953
RM233B	0·486	0·626	0·14	0·28	0·766	0·224	0·366	1
RM26	0·36	0·448	0·088	0·176	0·536	0·196	0·328	1
RM249	0·87	0·866	−0·004	−0·008	0·863	−0·005	−0·009	0·742
RM164	0·767	0·781	0·014	0·027	0·794	0·017	0·034	0·967
Mean for chromosome 5	**0·614**	**0·678**	**0·063**	**0·127**	**0·741**	**0·11**	**0·186**	**0·932**
RM225	0·886	0·892	0·006	0·012	0·898	0·007	0·014	1
OSR21	0·551	0·631	0·081	0·162	0·712	0·128	0·227	1
OSR19	0·669	0·697	0·028	0·056	0·725	0·04	0·078	1
OSR25	0	0	0	0	0	0	0	0
RM253	0·83	0·863	0·033	0·066	0·896	0·038	0·074	1
Mean for chromosome 6	**0·587**	**0·617**	**0·03**	**0·059**	**0·646**	**0·043**	**0·079**	**0·8**
RM234	0·612	0·727	0·115	0·23	0·842	0·158	0·273	1
RM47	0·102	0·101	−0·001	−0·002	0·1	−0·009	−0·018	0·86
RM248	0·813	0·836	0·023	0·047	0·86	0·028	0·054	0·959
OSR22	0·623	0·687	0·064	0·128	0·751	0·093	0·171	0·941
RM82	0·421	0·467	0·046	0·092	0·513	0·099	0·179	1
Mean for chromosome 7	**0·514**	**0·564**	**0·049**	**0·099**	**0·613**	**0·074**	**0·132**	**0·952**
OSR07	0·584	0·585	0·001	0·003	0·587	0·002	0·005	1
OSR30	0·676	0·728	0·052	0·104	0·78	0·072	0·134	0·561
RM25	0·739	0·758	0·019	0·039	0·777	0·025	0·05	0·96
RM223	0·789	0·868	0·08	0·159	0·948	0·092	0·168	1
OSR35	0·86	0·886	0·025	0·05	0·911	0·028	0·055	0·969
Mean for chromosome 8	**0·73**	**0·765**	**0·035**	**0·071**	**0·8**	**0·044**	**0·082**	**0·898**
RM215	0·709	0·774	0·065	0·13	0·839	0·084	0·155	1
OSR28	0·683	0·768	0·085	0·17	0·854	0·111	0·2	0·799
OSR29	0·579	0·679	0·1	0·199	0·778	0·147	0·256	1
RM245	0·452	0·6	0·148	0·295	0·748	0·246	0·395	0·931
RM242	0·649	0·759	0·11	0·221	0·869	0·145	0·254	1
Mean for chromosome 9	**0·614**	**0·716**	**0·102**	**0·203**	**0·818**	**0·147**	**0·252**	**0·946**

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

Table 4. (*Cont.*)

Locus	H_S	H_T	D_{ST}	D_{ST}'	H_T'	G_{ST}	G_{ST}'	G_{IS}
RM244	0·576	0·572	−0·003	−0·007	0·569	−0·006	−0·012	1
OSR33	0·77	0·806	0·035	0·071	0·841	0·044	0·084	1
RM222	0·88	0·894	0·014	0·029	0·908	0·016	0·032	1
RM258	0·527	0·622	0·095	0·191	0·718	0·153	0·266	1
RM228	0·662	0·676	0·014	0·027	0·69	0·02	0·039	1
Mean for chromosome 10	**0·683**	**0·714**	**0·031**	**0·062**	**0·745**	**0·045**	**0·082**	**1**
RM202	0·842	0·863	0·02	0·041	0·883	0·024	0·046	0·97
RM206	0·889	0·898	0·009	0·018	0·907	0·01	0·02	1
RM167	0·757	0·818	0·061	0·123	0·88	0·075	0·139	1
RM224	0·709	0·772	0·063	0·126	0·835	0·082	0·151	1
RM209	0·459	0·487	0·027	0·055	0·514	0·056	0·106	0·897
Mean for chromosome 11	**0·731**	**0·768**	**0·036**	**0·073**	**0·804**	**0·049**	**0·092**	**0·973**
RM235	0·857	0·891	0·034	0·068	0·925	0·038	0·074	0·947
OSR32	0·628	0·686	0·058	0·116	0·744	0·084	0·156	0·8
OSR20	0·826	0·862	0·036	0·072	0·898	0·041	0·08	0·935
RM247	0·666	0·773	0·107	0·213	0·879	0·138	0·242	1
RM83	0	0	0	0	0	0	0	0
Mean for chromosome 12	**0·595**	**0·642**	**0·047**	**0·094**	**0·689**	**0·06**	**0·11**	**0·736**
Overall	**0·625**	**0·683**	**0·057**	**0·114**	**0·74**	**0·084**	**0·154**	**0·946**

* G_{ST} is Nei's coefficient of gene variation; H_S and H_T are the mean gene diversity within subspecies and the overall gene diversity in the entire species, respectively; D_{ST} is the amount of gene diversity between subspecies; D_{ST}', H_T', and G_{ST}' are the equivalent estimators of D_{ST}, H_T and G_{ST}, respectively, which are independent of the number of samples; G_{IS} is an estimator of F_{IS}.

genetic variation resided between the *indica* and *japonica* groups (Table 5). Our chromosome-based comparison suggested that the extent of *indica–japonica* differentiation varied substantially, ranging from 7·62% in chromosome 3 to 28·72% in chromosome 1. F_{ST} values varied widely from–0·0226 at locus RM55 to 0·5408 at locus RM211 at the 57 assayed polymorphic loci (Table 5). Similar patterns of differentiation between two groups are seen at different loci, chromosomes, and the entire genome using Nei's estimators of G_{ST} and G_{ST}' (Table 5, Fig. 2). However, G_{ST} are apparently lower than the F_{ST} estimates with an overall estimate of $G_{ST}=0·084$, whereas the G_{ST}' estimates are close to the F_{ST} estimates with a mean estimate of $G_{ST}'= 0·154$.

(iv) *Genetic relationships*

Dendrograms were constructed based on every 5 loci for each chromosome (data not shown) and 60 loci for the entire genome (Fig. 3). Phylogenetic analyses showed that the 57 rice varieties can not be clustered together into either monophyletic *indica* or monophyletic *japonica* groups in the 12 chromosome-based trees, which displayed relatively different clustering patterns. The varieties formed two major clusters of *indica* and *japonica* for most chromosomes (chromosomes 1, 2, 3, 4, 9 and 10). In chromosome 1, for example, all the *indica* varieties except for five (Dahongmang, Dashenchan, Puanshendao, Zhenguiai and Zhenfudao) were included in one cluster, whereas another cluster had all the *japonica* varieties except Liuhuangzhan1 and Giayu293. Among these six chromosome-based trees, there were obvious differences in the total number of varieties as well as specific varieties that were included in the counterpart group. For example, we observed that, in addition to two major clusters of the *indica* and *japonica* groups with most varieties, small portions appeared grouped into either single or mixed subclusters for the three chromosomes 5, 6 and 11. Interestingly, the last three chromosomes (7, 8 and 12) showed relatively different clustering patterns from those described above. Several subclusters at a larger or smaller scale were formed mainly by *indica* or *japonica* varieties, but they randomly grouped together regardless of subspecies boundary. Our cluster analyses, therefore, suggest that the process of *indica–japonica* differentiation may have proceeded through an extensive contribution by the alleles of the majority of the rice genome. As a result of non-random mutation or substitution, however, the differentiation may not have occurred evenly in all 12 chromosomes under natural or artificial selection during the evolution of cultivated rice. In the genome-based trees constructed by UPGMA and neighbor-joining methods (Fig. 3), on the other hand, the majority of varieties formed merely two clusters for the *indica* and *japonica*

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

Table 5. *Genetic differentiation within and between* indica *and* japonica *rice and both subspecies combined*

Loci/chromosomes	F_{IS}	F_{IT}	F_{ST}
RM212	1·0000***	1·0000	0·4806***
OSR27	1·0000***	1·0000	0·2780***
OSR23	1·0000***	1·0000	0·4837***
RM200	0·0351	0·0848	0·0515**
RM220	1·0000***	1·0000	0·1422***
Mean for chromosome 1	**0·8070*****	**0·8170**	**0·2872*****
RM211	1·0000***	1·0000	0·5408***
OSR11	1·0000***	1·0000	0·3023***
RM263	1·0000***	1·0000	0·0713**
OSR08	1·0000***	1·0000	0·1727***
RM233A	1·0000***	1·0000	0·1359***
Mean for chromosome 2	**1·0000*****	**1·0000**	**0·2446*****
RM168	1·0000***	1·0000	0·1583***
RM60	0	0	0
RM55	1·0000***	1·0000	–0·0226
RM232	1·0000***	1·0000	0·0626***
RM231	1·0000***	1·0000	0·0613***
Mean for chromosome 3	**1·0000*****	**1·0000**	**0·0762*****
RM255	1·0000***	1·0000***	0·0921**
RM252	1·0000***	1·0000***	0·2032***
RM261	1·0000***	1·0000***	0·1462***
RM241	1·0000***	1·0000***	0·0521***
OSR15	1·0000***	1·0000***	0·4904***
Mean for chromosome 4	**1·0000*****	**1·0000*****	**0·1968*****
OSR34	0·9640***	0·9717	0·2152***
RM233B	1·0000***	1·0000	0·3982***
RM26	1·0000***	1·0000	0·3711***
RM249	0·7456***	0·7433	–0·0091
RM164	0·9756***	0·9765	0·0362***
Mean for chromosome 5	**0·9370*****	**0·9383**	**0·2060*****
RM225	1·0000***	1·0000	0·0141**
OSR21	1·0000***	1·0000	0·2249***
OSR19	1·0000***	1·0000	0·0780***
OSR25	0	0	0
RM253	1·0000***	1·0000	0·0731***
Mean for chromosome 6	**1·0000*****	**1·0000**	**0·0975*****
RM234	1·0000***	1·0000	0·2770***
RM47	0·8317***	0·8287	–0·0178
RM248	0·9710***	0·9727	0·0574***
OSR22	0·9439***	0·9534	0·1699***
RM82	1·0000***	1·0000	0·1599***
Mean for chromosome 7	**0·9493*****	**0·9510**	**0·1364*****
OSR07	1·0000***	1·0000	0·0059*
RM25	0·9704***	0·9719	0·0520***
RM223	1·0000***	1·0000	0·1707***
OSR35	0·9775***	0·9788	0·0556***
OSR30	0·5504***	0·6119	0·1366***
Mean for chromosome 8	**0·8997*****	**0·9141**	**0·0841*****
RM215	1·0000***	1·0000	0·1501***
OSR28	0·8073***	0·8449	0·1955***
OSR29	1·0000***	1·0000	0·2509***
RM245	0·9385***	0·9667	0·4590***
RM242	1·0000***	1·0000	0·2574***
Mean for chromosome 9	**0·9491*****	**0·9623**	**0·2626*****
RM244	1·0000***	1·0000	–0·0109
OSR33	1·0000***	1·0000	0·0865***
RM222	1·0000***	1·0000	0·0308***
RM258	1·0000***	1·0000	0·2938***
RM228	1·0000***	1·0000	0·0425**
Mean for chromosome 10	**1·0000*****	**1·0000**	**0·0929*****
RM202	0·9765***	0·9776	0·0466***
RM206	1·0000***	1·0000	0·0204***
RM167	1·0000***	1·0000	0·1374***
RM224	0·9723***	0·9763	0·1556***
RM209	0·8956***	0·9084	0·1226***
Mean for chromosome 11	**0·9689*****	**0·9725**	**0·0965*****
RM235	0·9561***	0·9594	0·0753***
OSR32	0·8072***	0·8381	0·1602***
OSR20	0·9334***	0·9387	0·0786***
RM247	1·0000***	1·0000	0·2328***
RM83	0	0	0
Mean for chromosome 12	**0·9241*****	**0·9341**	**0·1367*****
Total[b]	**0·947 (0·025)**	**0·955 (0·021)*****	**0·158 (0·018)*****
95%[c]	**0·890**	**0·907*****	**0·125*****
99%[d]	**0·868**	**0·888*****	**0·115*****

[a] Values of F_{IS} (the heterozygote deficit within subspecies) and F_{ST} (the fixation of different alleles between two subspecies) for each locus studied are given. Weir and Cockerham's *F*-statistic overall values of F_{ST} and F_{IS} for each chromosome and the entire genome are also shown. Statistically significant deviations from Hardy–Weinberg expectations are indicated by: $*P<0{\cdot}05$, $**P<0{\cdot}01$ and $***P<0{\cdot}001$.

[b] Total mean values with standard errors after jackknifing over loci.

[c] Values obtained by bootstrapping over loci; number of replicates = 999; nominal confidence interval = 95%.

[d] Values obtained by bootstrapping over loci; number of replicates = 999; nominal confidence interval = 99%.

varieties (with <70 bootstrap support), in which two *japonica* varieties (Giayu293 and Liuhuangzhan1) and five *indica* varieties (Dahongmang, Dashenchan, Puanshendao, Zhenfudao and Zhenguiai) were included in the counterpart clusters, respectively.

4. Discussion

Our analyses, in contrast to the very great differentiation ($F_{ST}>0{\cdot}25$) reported in previous studies, showed only moderate differentiation ($F_{ST}=0{\cdot}5$–$0{\cdot}15$) between the two subspecies, although the extent of differentiation varied with the rice materials assayed and the genetic markers used. For example, Morishima & Oka (1981) suggested that the

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

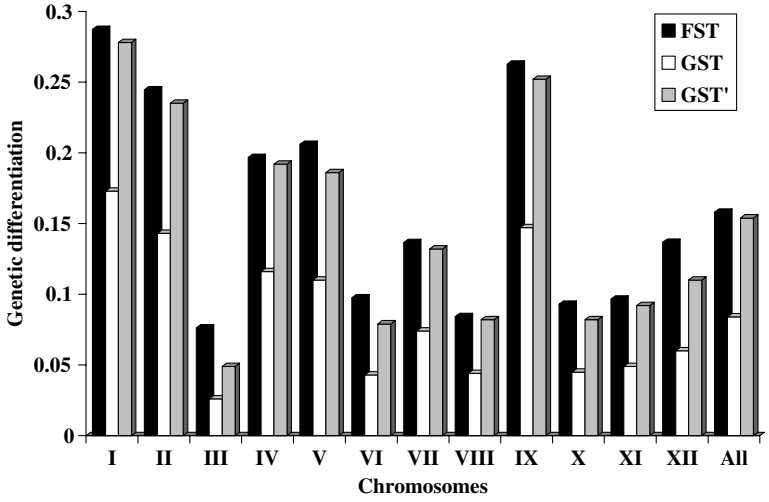

Fig. 2. Chromosome-based comparisons of genetic differentiation between the *indica* and *japonica* groups based on 60 microsatellite markers.

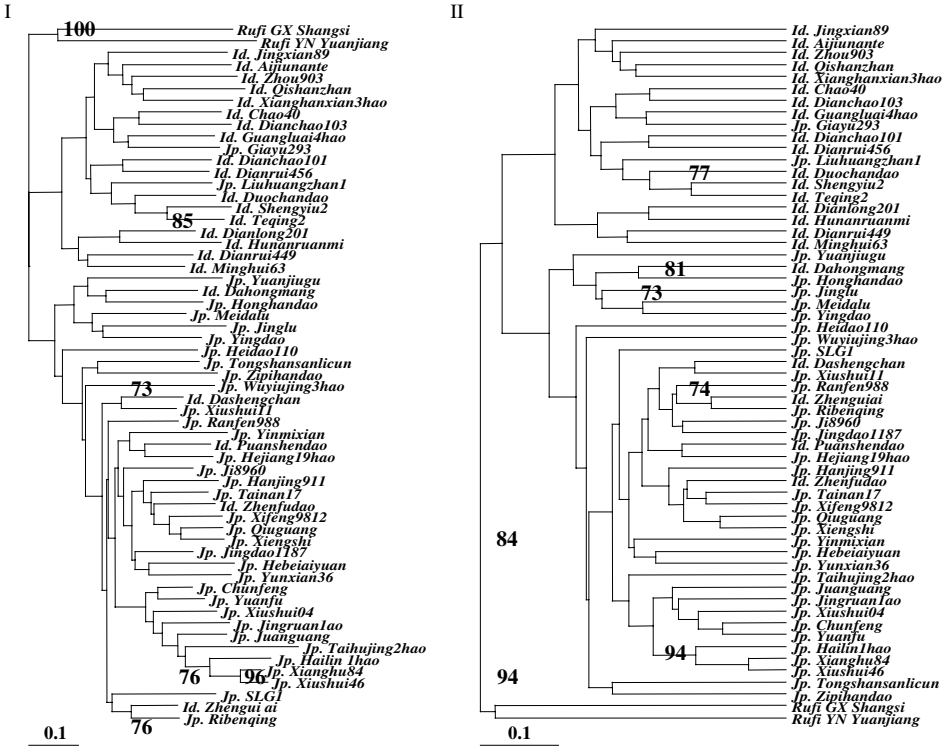

Fig. 3. Phenograms based on genetic distances among 57 rice varieties. I and II are UPGMA and neighbor-joining trees using the chord distance of Cavalli-Sforza & Edwards (1967) based on the microsatellite variation of 60 loci of the entire genome. Two accessions of wild rice, *O. rufipogon*, were added as outgroup. Numbers on branches indicate bootstrap values of >70 (999 replications).

component representing *indica–japonica* differentiation accounted for about 50% of the total genetic diversity; Zhang *et al.* (1992) later found that 33 of 49 probes demonstrated significant differentiation and suggested that about 34% (equal to G_{ST}) of the total variation could be explained by this component. One likely explanation for the difference is that a low polymorphism detected within this selfing species using a method such as RFLPs (Zhang *et al.*, 1992) may tend to overestimate the differentiation, because G_{ST} (an estimator equivalent to F_{ST} used by Zhang *et al.*, 1992) is a relative value to estimate genetic differentiation among populations. Obviously, the *indica–japonica* differentiation that accounted for approximately 10% (an estimator equal to G_{ST} was used) of the total genetic diversity at the

https://doi.org/10.1017/S0016672304007293 Published online by Cambridge University Press

10 polymorphic microsatellites (Yang *et al*., 1994) is fairly consistent with our estimation of 8·4–15·8%.

In agreement with the previous analyses of genetic diversity within these two rice types (Morishima & Oka, 1981; Second, 1982; Glaszmann, 1987; Oka, 1988; Zhang *et al*., 1992; Yang *et al*., 1994), we provide further evidence that *indica* rice is genetically more diverse than *japonica* rice as indicated by the estimators of genetic diversity (H and A) (Table 3, Fig. 1). Taking the sampling effects into consideration (e.g. R_S), we show that *japonica* rice generally has more alleles per locus than *indica* rice, but the difference is not statistically significant. The result agrees with a former study, in which Yang *et al*. (1994) reported that *indica* rice has a larger number of allele per locus than *japonica* rice using 238 accessions at 10 microsatellites. A recent comparative genomics study, for instance, revealed that about 1000 more SNPs were located in the intergenic regions of *indica* GLA4 than in the same regions of the *japonica* Nipponbare sequence, although a reverse trend was observed in the coding regions (Han & Xue, 2003). The comparatively high diversity displayed by *indica* rice may be the result of a diffuse origin from various wild populations in the lowland tropics (Glaszmann, 1987), whereas *japonica* rice was most likely domesticated somewhere in the northern parts of Southeast Asia or South China and then moved north out of the range of its wild progenitor. This means that the effective size of initial populations may be one of the important contributors to genetic diversity that differ between the two subspecies. More likely, the frequent introgression from sympatric wild and weedy rice in tropical or subtropical lowlands and subsequent fixation by self-crossing may also have historically contributed to the detected increase in genetic diversity within *indica* rice. It will be worthwhile further verifying the roles of the above-mentioned factors and other forces (such as natural and artificial selection) that may act on the two rice genomes.

In this study, we observed non-random distribution of microsatellite variation across different loci and chromosomes in the rice genome. Further chromosome-based comparisons revealed that nine chromosomes (1, 2, 3, 4, 5, 8, 9, 10 and 11) harboured higher levels of genetic diversity within *indica* rice than *japonica* rice. Similarly, differences in genetic diversity were detected in different loci and chromosomes in the wheat genome by 472 loci of the RFLPs genetic map (Jia *et al*., 2001). Our results show that rice varieties possess a great depth of gene diversity that is not randomly distributed at different loci of the two rice genomes, suggesting great potential to utilize inter-subspecific heterosis hybrid rice breeding for the future. Wright's F-statistics, Nei's estimations and phylogenetic analyses further demonstrated that *indica–japonica* differentiation occurred on all 12 chromosomes and almost all the detected loci with a variable range. This suggested that the process of *indica–japonica* differentiation may have proceeded through almost the entire rice genome and was genetically contributed to by almost all 12 chromosomes. The effect of selection for the agronomic traits that distinguish crops from their ancestors has been well recognized (Clark *et al*., 2004), but it may be too early to state that a bottleneck effect and selection has acted on rice genomic regions of different chromosomes during the domestication process. Therefore, more studies regarding the evolutionary forces that result in genetic differences between the two rice types could be of great value in understanding what shaped the population structure and how the two rice genomes of current cultivated rice evolved.

To date, the molecular nature of *indica–japonica* differentiation has been an interesting but uncertain area. The molecular diversity and hybrid sterility in *indica–japonica* rice crosses suggested that the genetic basis of *indica–japonica* differentiation may be complex (Zhang *et al*., 1997). The majority of *indica–japonica* differentiation resolved in the RFLP analysis is attributable to insertion/deletion or other rearrangements of genomic sequences that occurred in one group but not the other (McCouch *et al*., 1988; Wang & Tanksley, 1989; Zhang *et al*., 1992). All the microsatellite loci revealing the largest amounts of *indica–japonica* differentiation are those whose allelic differences are caused by variable numbers of di-nucleotide repeats such as (GA)n or (GT)n (Yang *et al*., 1994). Compared with the 20 *indica–japonica* differentiating RFLP markers, SSLP markers were shared by the same regions on chromosomes 2, 6, 7 and 9 with the *indica–japonica* differentiating RFLP markers, but others appeared on other chromosomes, where no *indica–japonica* differentiating RFLP markers had been detected (Fan *et al*., 1999). In addition to insertion/deletion or other rearrangements of genomic sequences discussed above, it is likely that gene inactivation, subspecific gene duplication, amino acid substitution, and/or transposable elements have contributed to the process of *indica–japonica* differentiation. For example, most recently, sequence alignments between 2·3 Mb of three contiguous segments of chromosome 4 from *indica* Guangluai 4 and its collinear sequences from *japonica* Nipponbare suggested frequent deviations from collinearity by insertions or deletions (Feng *et al*., 2002). An expansion by insertions of transposable elements might lead to the observation that chromosome 4 of *japonica* is probably larger than that of *indica*. These insertions occur not only in the intergenic regions, but also in some of the coding regions. When the finished sequences of two rice genomes become available in the

near future (Normile & Pennisi, 2002), these can be mined to better understand how Asian cultivated rice varieties evolved and differentiated into *indica* and *japonica* subspecies through comparing their genome organization, characterizing the allelic variability, expression, distribution of the potentially important candidate genes, and exploring other evolutionary contributors. From the viewpoint of evolution and domestication, however, a better understanding of rice evolution and *indica–japonica* differentiation will rely on our knowledge of comparative genomics, evolutionary dynamics and, in particular, the molecular evolution and regulation of associated agronomic traits.

This research was supported by the International Foundation for Sciences (IFS) (grant no. C/2738-1, 2) to L.G. The rice materials were provided by Mr Yi-ping Wang at the China Rice Research Institute and Professor Qingrui He at YunnanAAS. Finally we thank Dr M. David Webb, Sara Barton and an anonymous reviewer for their careful reading and useful comments to improve the manuscript.

References

Akagi, H., Yokozeki, Y., Inagaki, A. & Fujimura, T. (1996). Microsatellite DNA markers for rice chromosomes. *Theoretical and Applied Genetics* **93**, 1071–1077.

Bowcock, A., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. & Cavalli-Sforza, L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457.

Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967). Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* **19**, 233–257.

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

Chang, T. T. (1984). Conservation of rice genetic resources: luxury or necessity? *Science* **224**, 251–256.

Chen, X., Temnykh, S., Xu, Y., Cho, Y. G. & McCouch, S. R. (1997). Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **95**, 553–567.

Cheng, K. S. (1985). A statistical evaluation of the classification of rice cultivars into hsien and keng subspecies. *Rice Genetics Newsletter* **2**, 46–48.

Clark, R. M., Linton, E., Messing, J. & Doebley, J. F. (2004). Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proceedings of the National Academy of Sciences of the USA* **101**, 700–707.

Edwards, K., Johnstone, C. & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research* **19**, 1349.

Estoup, A., Taillez, C., Cornuet, J.-M. & Solignac, M. (1995). Size homoplasy and mutational processes of interrupted microsatellites in Apidae species, *Apis mellifera* and *Bombus terrestris*. *Molecular Biology and Evolution* **12**, 1074–084.

Fan, Y. Y., Zhuang, J. Y., Wu, J. L. & Zhang, K. L. (1999). SSLP-based subspecies identification in *Oryza sativa* L. *Rice Genetics Newsletter* **16**, 19.

Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., *et al*. (2002). Sequence and analysis of rice chromosome 4. *Nature* **420**, 316–320.

Felsenstein, J. (1995). *PHYLIP* (*Phylogeny Inference Package*) *Version 3.57c*. Washington, D.C.: University of Washington, Seattle.

Gao, L. Z., Schaal, B. A., Zhang, C. H., Jia, J. Z. & Dong, Y. S. (2002). Assessment of population genetic structure in common wild rice *Oryza rufipogon* Griff. using microsatellite and allozyme markers. *Theoretical and Applied Genetics* **106**, 173–180.

Glaszmann, J. C. (1987). Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics* **74**, 21–30.

Guo, S. W. & Thompson, E. A. (1992). Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48**, 361–372.

Goudet, J., Raymond, M., Meeüs, T. de & Rousset, F. (1996). Testing differentiation in diploid populations. *Genetics* **144**, 1933–1940.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available via http://www.unil.ch/izea/softwares/fstat.html

Han, B. & Xue, Y. B. (2003). Genome-wide intraspecific DNA-sequence variations in rice. *Current Opinion in Plant Biology* **6**, 134–138.

Ikehashi, H. & Araki, H. (1986). Genetics of F1 sterility in remote crosses of rice. In: *Rice Genetics*, pp. 119–130. Manila, The Philippines: IRRI.

Jia, J. Z., Zhang, Z. B., Devos, K. & Gale, M. D. (2001). Analyses of genetic diversity of the 21 wheat chromosomes as revealed by RFLP mapping loci. *Science in China* (*Series C*) **31**, 13–21.

Mackill, D. J. (1995). Classifying *japonica* rice cultivars with RAPD markers. *Crop Science* **35**, 889–894.

Mackill, D., Zhang, Z., Redona, E. & Colowit, P. (1996). Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome* **39**, 969–977.

McCouch, S. R., Kochert, G., Tu, Z. H., Wang, Z. Y., Khush, G. S., Coffman, W. R. & Tanksley, S. D. (1988). Molecular mapping of rice chromosome. *Theoretical and Applied Genetics* **76**, 815–829.

Morishima, H. & Oka, H. I. (1981). Phylogenetic differentiation of cultivated rice: 21 Numerical evaluations of the *indica–japonica* differentiation. *Japanese Journal of Breeding* **31**, 402–413.

Morishima, H., Sano, Y. & Oka, H. I. (1992). Evolutionary studies in cultivated rice and its wild relatives. *Oxford Surveys in Evolutionary Biology* **8**, 135–184.

Nakagahra, M. (1978). The differentiation, classification and center of genetic diversity of cultivated rice (*Oryza sativa* L.) by isozyme analysis. *Tropical Agriculture Research Service* **11**, 77–82.

Nagaraju, J., Kathirvel, M., Kumar, R. R., Siddiq, E. A. & Hasnain, S. E. (2002).Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *Proceedings of the National Academy of Sciences of the USA* **99**, 5836–5841.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Normile, D. & Pennisi, E. (2002). Rice: boiled down to bare essentials. *Science* **296**, 32–36.

Oka, H. I. (1964). Considerations on the genetic basis of intervarietal sterility in *Oryza sativa*. In: *Rice Genetics and Cytogenetics*, pp. 159–174. Symposium on Rice Genetics and Cytogenetics. Manila, The Philippines: IRRI / Amsterdam : Elsevier.

Oka, H. I. (1988). *Origin of Cultivated Rice*. Tokyo: Japan Scientific Society Press/Elsevier.

Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Computational Molecular Biology* **12**, 357–358.

Panaud, O., Chen, X. & McCouch, S. R. (1996). Development of microsatellite markers and characterization of simple sequence length polymorphism (SSR) in rice (*Oryza sativa* L.). *Molecular Genetics and Genomics* **252**, 597–607.

Petit, R. J., El Mousadik, A. & Pons, O. (1998). Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844–855.

Raymond, M. & Rousset, F. (1995). GENEPOP: population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–249.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Second, G. (1982). Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Japanese Journal of Genetics* **57**, 25–27.

Sun, C. Q., Wang, X. K. & Yoshimura, A. (2002). Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **104**, 1335–1345.

Wang, Z. Y. & Tanksley, S. D. (1989). Restriction fragment length polymorphism in *Oryza sativa* L. *Genome* **32**, 1113–1118.

Weir, B. S. & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Wu, K. S. & Tanksley, S. D. (1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular and General Genetics* **241**, 225–235.

Yang, G. P., Maroof Saghai, M. A., Xu, C. G., Zhang, Q. F. & Biyashev, R. M. (1994). Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Molecular Genetics and Genomics* **245**, 187–194.

Yuan, L. P., Virmani, S. S. & Mao, C. X. (1989). Hybrid rice: achievement and outlook. In: *Progress in Irrigated Rice Research*, pp. 219–235. Manila, The Philippines: pp. 219–235.

Zhang, Q. F., Maroof Saghai, M. A., Lu, T. Y. & Shen, B. Z. (1992). Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. *Theoretical and Applied Genetics* **83**, 494–499.

Zhang, Q. F., Liu, K. D., Yang, G. P., Maroof Saghai, M. A., Xu, C. G. & Zhou, Z. Q. (1997). Molecular marker diversity and hybrid sterility in *indica–japonica* rice crosses. *Theoretical and Applied Genetics* **95**, 112–118.

Zhu, J., Gale, M. D., Quarrie, S., Jackson, M. T. & Bryan, G. J. (1998). AFLP markers for the study of rice biodiversity. *Theoretical and Applied Genetics* **96**, 602–611.