Theor Appl Genet (2008) 117:1181–1189
DOI 10.1007/s00122-008-0855-7

ORIGINAL PAPER

Contrasting population genetic structure and gene flow between *Oryza rufipogon* and *Oryza nivara*

Hai-Fei Zhou · Xiao-Ming Zheng · Ruo-Xun Wei · Gerard Second · Duncan A. Vaughan · Song Ge

Received: 10 March 2008 / Accepted: 28 July 2008 / Published online: 20 August 2008

Abstract The cross compatible wild relatives of crops have furnished valuable genes for crop improvement. Understanding the genetics of these wild species may enhance their further use in breeding. In this study, sequence variation of the nuclear *Lhs1* gene was used to investigate the population genetic structure and gene flow of *Oryza rufipogon* and *O. nivara*, two wild species most closely related to *O. sativa*. The two species diverge markedly in life history and mating system, with *O. rufipogon* being perennial and outcrossing and *O. nivara* being annual and predominantly inbreeding. Based on sequence data from 105 plants representing 11 wild populations covering the entire geographic range of these wild species, we detected significantly higher nucleotide variation in *O. rufipogon* than in *O. nivara* at both the population and species levels. At the population level the diversity in *O. rufipogon* (Hd = 0.712; θ_{sil} = 0.0017) is 2–3 folds higher than that in *O. nivara* (Hd = 0.306; θ_{sil} = 0.0005). AMOVA partitioning indicated that genetic differentiation among *O. nivara* populations (78.2%) was much higher than that among *O. rufipogon* populations (52.3%). The different level of genetic diversity and contrasting population genetic structure between *O. rufipogon* and *O. nivara* might be explained by their distinct life histories and mating systems. Our simulation using IM models demonstrated significant gene flow from *O. nivara* to *O. rufipogon*, indicating a directional introgression from the annual and selfing species into the perennial and outcrossing species. The ongoing introgression has played an important role in shaping current patterns of genetic diversity of these two wild species.

Communicated by T. Sasaki.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0855-7) contains supplementary material, which is available to authorized users.

H.-F. Zhou · X.-M. Zheng · R.-X. Wei · S. Ge (✉)
State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 100093 Beijing, China
e-mail: gesong@ibcas.ac.cn

G. Second
The Research Institute for Development (IRD), Montpellier, France

D. A. Vaughan
National Institute of Agrobiological Sciences, Tsukuba 305-8602, Japan

S. Ge
Graduate School, Chinese Academy of Sciences, 100039 Beijing, China

Introduction

Assessment of the level and distribution of genetic diversity within a species may contribute to a better understanding of its evolutionary history and potential for use in breeding. Knowledge of population structure and evolutionary history of wild progenitors of cultivated plants is particularly critical both to understand the evolutionary processes leading to domestication and to their conservation and management (Schaal et al. 1998; Wright and Gaut 2005). *Oryza rufipogon* Griff. and *O. nivara* Sharma *et* Shastry are the most closely related species to Asian cultivated rice (*O. sativa*) and are considered its progenitors (Oka 1988; Khush 1997). *O. rufipogon* is perennial, photoperiod sensitive, largely cross-fertilized, and widely distributed from southern China, South and Southeast Asia to Papua New Guinea and northern Australia. It grows in areas with year

round water, such as swamps and lakes. In contrast, *O. nivara* is an annual, photoperiod insensitive and predominantly self-fertilized species. This species is restricted to South and mainland Southeast Asia and is usually found in ponds and swamps that dry up completely in the dry season (Vaughan and Morishima 2003; Sang and Ge 2007). Because of their important role in providing beneficial genes for rice breeding (Khush 1997; Vaughan et al. 2003), *O. rufipogon* and *O. nivara* have long been the subject to extensive taxonomic, phylogenetic and population studies using a variety of approaches. These studies reveal that these species are cross compatible (Oka 1988; Lu et al. 2000) and exhibit little genetic differentiation (Lu et al. 2002; Cheng et al. 2003; Zhu and Ge 2005; Zhu et al. 2007). Consequently, *O. rufipogon* and *O. nivara* have been treated as two different ecotypes (Second 1985; Oka 1988; Lu et al. 2000; Cheng et al. 2003) or subspecies (Vaughan and Morishima 2003) under a single species (*O. rufipogon sensu lacto*) and in some cases, considered a single large gene pool (e.g., Londo et al. 2006). We here refer to *O. rufipogon* and *O. nivara* as different species in this population study because they are phenotypically distinct (Vaughan and Morishima 2003; Sang and Ge 2007).

Although genetic diversity and population genetic structure of *O. rufipogon* and *O. nivara* have been investigated extensively, geographic patterns of genetic diversity across the entire geographic range of these species is less well understood because most studies have focused either on the local populations in specific regions (Morishima et al. 1984; Morishima and Barbier 1990; Ge et al. 1999; Gao et al. 2000; Zhou et al. 2003; Kuroda et al. 2007) or on the species based on single accessions from disparate populations (Sun et al. 2001; Lu et al. 2002; Zhu et al. 2007; Rakshit et al. 2007). Particularly, populations of the annual *O. nivara* have been insufficiently sampled in most cases. Cai et al. (2004) used quantitative traits, allozyme and RFLP markers to investigate the population genetic structure of one annual and six perennial populations of the species, which is so far the only investigation of natural populations sampled across their entire geographic range. They found that different classes of data detected different patterns of variation, with quantitative traits reflecting ecotype differentiation into annual and perennial types but allozymes and RFLPs indicating geographic differentiation among populations. Since only one population of *O. nivara* was included in this study it provided limited information about the genetic differentiation and introgression between the annual and perennial species. Previous studies have shown that intermediate populations between annual (*O. nivara*) and perennial (*O. rufipogon*) types are present in some disturbed habitats and introgression occurs between these wild species and between cultivated rice and these wild species (Oka 1988; Song et al. 2006; Kuroda et al. 2007). However, information on the patterns and extent of introgression between *O. nivara* and *O. rufipogon* remains largely unknown (Vaughan et al. 2003; Sang and Ge 2007).

Sequences of nuclear genes can provide useful information on intraspecific variation and have been successfully used to address various population and evolutionary questions in plant species (for reviews Schaal et al. 1998; Wright and Gaut 2005). In this study, we used a single-copy nuclear gene (*Lhs1*) to investigate the nucleotide variation and population differentiation of *O. rufipogon* and *O. nivara* populations sampled from their entire geographic range. The objective was to quantify the genetic diversity at both the population and species levels and compare the result with those from previous studies using different markers such as allozyme and SSRs. We were particularly interested in the population genetic structure of *O. rufipogon* and *O. nivara* because they have distinctive habitat preferences associated with different life history and mating system. Also, we investigated the gene flow between the two wild species and its potential impact on the patterns of variation in the species. These studies will not only facilitate the effective use of the wild rice germplasm but also be instructive to better understand the origin and domestication process of cultivated rice.

Materials and methods

Populations sampling and DNA extraction

A total of 105 plants representing six *O. rufipogon* and five *O. nivara* populations were sampled across the entire geographic range of these two species (Fig. 1). Two *O. rufipogon* populations (ruf-PNG and ruf-IDN) and one *O. nivara*

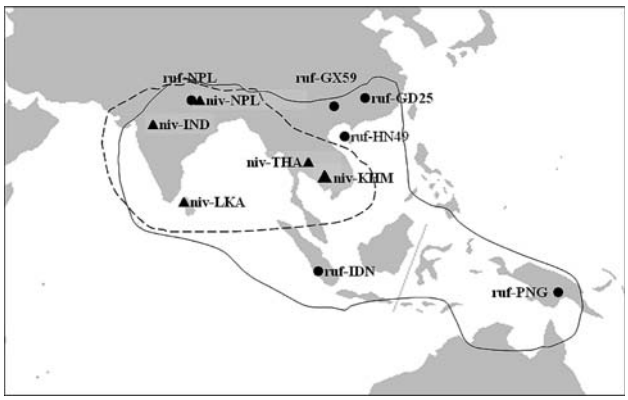

Fig. 1 Geographical localities where the populations were sampled. Solid circles and triangles represent the populations of *O. rufipogon* and *O. nivara*, respectively. Solid and broken lines indicate the geographic range of *O. rufipogon* and *O. nivara*, respectively. Detailed information on these populations is shown in Supplementary Table 1

population (niv-KHM) were provided by the International Rice Research Institute (IRRI) and the remaining eight populations were collected by the authors. Detailed information on the populations sampled is provided in the supplementary materials (Supplementary Table 1). Seed germination and seedling cultivation followed Ge et al. (1999). One *O. barthii* accession from Africa was used as an outgroup because this species is closely related to the species complex of *O. sativa* and its relatives (Zhu and Ge 2005). Total DNA was extracted from fresh or silica gel-dried leaves of one plant of each accession, using the CTAB method as described by Ge et al. (1999).

PCR amplification, cloning, and sequencing

Leafy hull sterile 1 (*Lhs1*), located in chromosome 3 of rice, is a MADS-box transcription factor and plays an essential role in determining floral meristem identity and floral organ development (Jeon et al. 2000). Based on our previous population genetics study (Zhu et al. 2007), we observed that, out of the ten nuclear loci amplified, *Lhs1* evolved neutrally with low recombination rate and the amplified portions of this gene was long enough with high proportion of introns to ensure sufficient information. Therefore, *Lhs1* was chosen as a suitable marker for this study. Polymerase chain reaction (PCR) primers (forward: 5′-agaggtaccgcagctg caac-3′ and reverse: 5′-ctccttcatgcttagtgggc-3′) were designed for amplifying a region between exons II and IV of *Lhs1*. An touch-down PCR amplification was performed in a total volume of 50 µl in thin-walled tubes using T-gradient thermocycler. The procedure includes initial denaturation of 5 min at 94°C; 10 cycles of 40 s denaturing at 94°C, 40 s annealing at 62°C (0.5°C descending each cycle), 2 min extension at 72°C; 25 cycles of 40 s at 94°C, 40 s at 52°C, 2 min at 72°C, and 10 min at 72°C for final extension. The PCR reaction mix contained approximately 50–100 ng of template DNA, 0.2 µM of each primer, 0.2 mM of each dNTP, 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 2.0 mM $MgCl_2$, 5% DMSO, and a mixture of 1.0 U Ex *Taq* DNA polymerase (TaKaRa, Japan). Amplification products were separated on 1.5% agarose gels and purified with either a Pharmacia purification kit (Amersham Pharmacia Biotech, USA) or QIAquick kit (QIAGEN, Germany). Sequencing reactions were performed by an ABI3730XL sequencer (Applied Biosystems Corp., USA) or a Megabase1000 sequencer (Amersham Pharmacia Biotech, USA).

For a nuclear locus, individuals of wild rice can be either homozygous or heterozygous and thus the purified PCR products were cloned into pGEM T-easy vector (Promega Corp., USA) following the manufacturers instructions. To obtain both alleles from the heterozygous individuals, we sequenced randomly six to ten clones per accession. To find alleles easier, direct sequencing was also performed in comparison with plasmid sequences. Since *Taq* errors occur at random, it is unlikely that polymorphisms shared among more than one clone (sequence) are artificial (Eyre-Walker et al. 1998). However, "singletons", i.e. polymorphisms occurred in only one sequence relative to all the remainder sequences can be either true sequence variation or *Taq* polymerase artifact. To confirm the singletons, we repeated PCR amplification, cloning, and sequencing and finally exclude those singletons resulting from *Taq* polymerase error. Using this strategy, interallelic PCR recombinants were also verified and removed (Zhu et al. 2007). All sequences were deposited in GenBank under accession numbers (EU327204–EU327254).

Data analysis

Initial sequences were assembled and edited using the program Sequencher (Gene Codes Corp., USA). Multiple sequences were aligned using ClustalX 1.81 (Thompson et al. 1997), with further manual refinements. Insertions/deletions (indels) were not included in the analysis. We calculated the number of segregating sites (S), the number of haplotypes (h), haplotype diversity (Hd) and two parameters of nucleotide diversity: π (Nei 1987) and θ_w from segregating sites (Watterson 1975) using the DnaSP version 4.10.9 (Rozas et al. 2003). The observed heterozygosity was estimated as H_o (H_o = number of heterozygous individuals/number of total individuals).

Two neutrality tests, based on either the frequency spectrum of polymorphism or the haplotype distribution, were used to detect possible deviations from the predictions of a neutral equilibrium model. Tajima's D statistic, Fu and Li's D^*/F^* were calculated at all sites and at silent sites separately for each locus, population and species. Tajima's D (Tajima 1989) was based on the discrepancy between the mean pairwise differences (π) and Watterson's estimator (θ), while Fu and Li's D^*/F^* (Fu and Li 1993) rely on the difference between the number of polymorphic sites in external branches (polymorphisms unique to an extant sequence) and number of polymorphic site in internal phylogenetic branches (polymorphisms shared by extant sequences). In these two tests, negative values indicate an excess of low frequency polymorphisms, while positive values indicate an excess of intermediate variants. *O. barthii* sequences were used as an outgroup for Fu and Li's tests.

To assess the divergence and genetic relationships among taxa, fixed difference and shared polymorphism for pairwise comparisons were computed between the two species. A fixed difference refers to the nucleotide site where all sampled sequences of one species are different from all sequences of another species, whereas a shared polymorphism occurs when two species have the same two bases

segregating at the same site. Shared polymorphisms reveal a history of polymorphism that has not yet been erased by genetic drift and thus reflect either a short divergence time between species or historically large population sizes (Hilton et al 1994). A hierarchical analysis of population subdivision and pairwise F_{ST} analyses were conducted using an analysis of molecular variance (AMOVA) as implemented in Arlequin 3.1 (Excoffier et al. 2005). Two measures of population differentiation, G_{ST} and N_{ST}, were compared by the program PermutCpSSR (Pons and Petit 1996). G_{ST} makes use of haplotype frequencies while N_{ST} takes into account differences between haplotypes. A higher N_{ST} than G_{ST} indicates the presence of phylogeographical structure with closely related haplotypes being found more often in the same area than less closely related haplotypes (Pons and Petit 1996). The Mantel test performed in Arlequin was used to examine the correlation between geographical distance and Slatkin's measure M [$M = (1/F_{ST} - 1)/2$], a measure of the extent of gene flow in an island model at equilibrium (Slatkin 1993).

To further assess the gene flow between species, we used the IM software to test the "isolation with migration" (IM) models (Hey and Nielsen 2004). This model has six demographic parameters that include bidirectional migration rates, one for each species. The posterior probability densities of migration rates are generated by Markov chain Monte Carlo (MCMC) simulations assuming the HKY model of sequence evolution. The analysis was done with ten independent chains under Metropolis coupling and performed with different seed numbers to guarantee convergence of the sample (Hey 2006). The model fitting requires that the loci under study have no recombination and selective neutrality. Therefore, we identified four different non-recombining blocks of *Lhs1* region, which were treated as different loci in the analysis (Won et al. 2005) to fit the assumption that different loci have segregated independently. Four-gametes test (Hudson and Kaplan 1985) applied by DnaSP was used for the identification of possible recombination events.

Results

Nucleotide variation and divergence between species

Total length of aligned sequence is 1,055 bp, including 161 bp of coding sequence and 894 bp of noncoding sequence. A total of 126 sequences from *O. rufipogon* and 84 from *O. nivara* were determined, with two sequences per individual. Twenty-six polymorphic sites (*S*) were observed in total, with 28 and seven haplotypes found in *O. rufipogon* and *O. nivara*, respectively (Table 1). Different numbers of heterozygotes were observed for the two species (Table 1). All *O. rufipogon* populations consisted of heterozygous individuals ranging from 20.0 (ruf-IDN) to 84.6% (ruf-NPL), while only one out of 42 *O. nivara* individuals was heterozygote (niv-THA). The average observed heterozygosity in *O. rufipogon* populations was 49.2% and much higher than that in *O. nivara* populations (4.0%). These results were consistent with their mating system

Table 1 Summary of nucleotide polymorphism and neutrality tests at *Lhs1*

Species	Population	*N*	H_o %	*S*	*h*	*Hd*	π_{sil}	θ_{sil}	*D*	*D**	*F**
O. rufipogon	ruf-GD25	26	61.5	9	9	0.748	0.0021	0.0026	−0.610	0.253	−0.006
	ruf-GX59	22	81.8	1	2	0.519	0.0006	0.0003	1.554	0.635	1.009
	ruf-HN49	24	25.0	8	8	0.768	0.0031	0.0023	1.046	1.334	1.452
	ruf-NPL	26	84.6	5	6	0.732	0.0023	0.0014	1.811	1.158	1.563
	ruf-PNG	18	22.2	7	7	0.791	0.0018	0.0025	−1.044	0.253	−0.126
	ruf-IDN	10	20.0	3	3	0.711	0.0015	0.0012	1.227	1.154	1.310
	Average	21	49.2	5.5	5.8	0.712	0.0019	0.0017	0.664	0.798	0.867
	Total	126	54.0	20	28	0.919	0.0030	0.0042	−0.850	0.867	0.247
O. nivara	niv-IND	16	0	0	1	0	0	0	–	–	–
	niv-LKA	12	0	2	3	0.667	0.0009	0.0007	0.554	0.973	0.979
	niv-NPL	26	0	1	2	0.148	0.0002	0.0003	−0.714	0.612	0.288
	niv-KHM	20	0	1	2	0.358	0.0004	0.0006	−0.769	0.866	0.487
	niv-THA	10	20.0	2	2	0.356	0.0008	0.0008	0.019	1.026	0.873
	Average	16.8	4.0	1.2	2	0.306	0.0004	0.0005	−0.228	0.869	0.657
	Total	84	2.4	6	7	0.675	0.0020	0.0015	0.827	1.206	1.275

N number of sequences sampled, H_o observed heterozygosity, *S* number of segregating sites, *h* number of haplotypes, *Hd*: haplotype diversity, π_{sil} and θ_{sil} nucleotide diversity and Watterson's parameter for silent sites, respectively; *D* Tajima's *D* (Tajima 1989); *D** and *F** Fu and Li's *D** and Fu and Li's *F** (Fu and Li 1993), respectively

Table 2 Divergence between *O. nivara* and *O. rufipogon*

S_{niv}	2
S_{ruf}	16
Dxy	0.0032 (0.0001)
Da	0.0008 (0.0001)
S_S	7
S_F	0

S_{niv} mutations polymorphic in *O. nivara* but monomorphic in *O. rufipogon*

S_{ruf} mutations polymorphic in *O. rufipogon* but monomorphic in *O. nivara*

Dxy (Nei 1987) the average number of nucleotide substitution per site between the two species. Those in parentheses are standard errors

Da (Nei 1987) the number of net nucleotide substitution per site

S_S the number of polymorphic sites shared by the two species

S_F the number of fixed differences between the two species

because the outcrossing rates of *O. rufipogon* and *O. nivara* were estimated to be 30–56% and 5%–20%, respectively (Vaughan and Morishima 2003).

A higher level of variation was found in *O. rufipogon* than in *O. nivara*, as measured by all parameters (Table 1). In particular, the number of segregating sites (S) and nucleotide diversity at silent sites (θ_{sil}) in *O. rufipogon* (20 and 0.0042) were about three times that in *O. nivara* (6 and 0.0015). At the population level, nucleotide diversity varied greatly among populations with the average being much higher in *O. rufipogon* than in *O. nivara*. For example, on average, haplotype diversity of *O. rufipogon* populations ($Hd = 0.712$) is twice that of *O. nivara* populations ($Hd = 0.306$). Average nucleotide diversity of *O. rufipogon* populations ($\pi_{sil} = 0.0019$; $\theta_{sil} = 0.0017$) is much higher than those of *O. nivara* populations ($\pi_{sil} = 0.0004$; $\theta_{sil} = 0.0005$) (Table 1). Estimates for DNA divergence between the two species is shown in Table 2. More polymorphic sites were observed in *O. rufipogon* ($S_{ruf} = 16$) than in *O. nivara* ($S_{niv} = 2$). Seven shared polymorphisms and no fixed differences were detected between *O. rufipogon* and *O. nivara*, suggesting the persistence of either ancestral polymorphisms or interspecific gene flow.

To test if the *Lhs1* locus evolves neutrally, a neutrality test was performed for the species and population datasets separately. Neither Tajima's D nor Fu and Li's D^* and F^* rejected the null hypothesis of neutral evolution at the species and population levels, indicating that there was no evidence of departure from the neutrality model at *Lhs1*. Therefore, the pattern of variation at this locus might be used to infer the population structure and history of the two species.

Geographic pattern and population differentiation

AMOVA partitioning was conducted to investigate the overall distribution of genetic diversity between species, among population within species and within populations. Table 3 shows the AMOVA results with different variation partitioned. Significant genetic differentiation exists between species ($P < 0.05$), among populations within species ($P < 0.001$) and within populations ($P < 0.001$). Of the total genetic diversity, about 16% was attributable to divergence between species, ~52% to population differentiation within species, and ~32% resided within populations. When considered for two species separately, a significantly larger proportion ($P < 0.001$) of the total variation was distributed among populations for both species (Table 3). Genetic differentiation among *O. nivara* populations (78.19%) was much higher than that among *O. rufipogon* populations (52.53%). A high level of genetic differentiation among populations for both species was also reflected by pairwise comparisons of F_{ST} between populations (Table 4). The F_{ST} values ranged from 0.0134 (niv-IND vs. niv-NPL) to 0.9486 (niv-IND vs. niv-KHM) and, except for one contrast (niv-IND vs. niv-NPL), they were all significant, including the comparisons of populations between and within species.

Tests for phylogeographical structure of haplotype variation indicated that the N_{ST} was significantly higher than G_{ST} for both *O. nivara* ($P < 0.001$) and *O. rufipogon* ($P < 0.01$) as well as for the combined populations of the two species ($P < 0.01$), supporting the presence of phylogeographical structure. Mantel's test detected significant association

Table 3 AMOVA results for *O. rufipogon* and *O. nivara* populations

Source of variation		d. f.	Sum of squares	Variance component	% variance
Among species		1	81.392	0.48473*	16.09*
Among populations within species		9	271.577	1.55778***	51.70***
Within populations		199	193.098	0.97034***	32.21***
O. rufipogon	Among populations	5	183.282	1.69690***	52.53***
	Within populations	120	184.004	1.53336 (na)	47.47 (na)
O. nivara	Among populations	4	28.802	0.43407***	78.19***
	Within populations	79	9.567	0.12110 (na)	21.81 (na)

d. f. degree of freedom, na significance not applicable, *** $P < 0.001$, * $P < 0.05$

Table 4 Pairwise population differentiation as measured by F_{ST}

Population	niv-IND	niv-LKA	niv-NPL	niv-KHM	niv-THA	ruf-GD25	ruf-GX59	ruf-HN49	ruf-NPL	ruf-PNG
niv-LKA	0.8725***									
niv-NPL	0.0134	0.8616***								
niv-KHM	0.9486***	0.8879***	0.9383***							
niv-THA	0.8605***	0.7613***	0.8633***	0.0940*						
ruf-GD25	0.4615***	0.4419***	0.4963***	0.4861***	0.3113***					
ruf-GX59	0.9417***	0.9116***	0.9448***	0.9364***	0.8959***	0.7039***				
ruf-HN49	0.5573***	0.5266***	0.5993***	0.5448***	0.4023***	0.2980***	0.6896***			
ruf-NPL	0.6285***	0.5957***	0.6582***	0.7154***	0.5790***	0.3894***	0.7640***	0.3803***		
ruf-PNG	0.7746***	0.6243***	0.7950***	0.7964***	0.6619***	0.4219***	0.8273***	0.4524***	0.3459***	
ruf-IDN	0.6708***	0.5876***	0.7083***	0.6036***	0.3478***	0.2946***	0.8093***	0.3379***	0.2371*	0.3206***

*** P < 0.001, ** P < 0.01, * P < 0.05

between geographical distance and F_{ST} values for *O. nivara* populations (r = 0.713, P = 0.029) but not for *O. rufipogon* populations (r = −0.022, P = 0.515), suggesting that isolation by distance model applied only for *O. nivara* populations.

Gene flow and introgression between species

Recent gene flow between *O. rufipogon* and *O. nivara* was examined by the IM model and the marginal posterior density distribution for migration rates is shown in Fig. 2. Each of the three repeats was run for 20 million steps of burn-in lengths under the HKY model and actual run durations after burn-in were 50 million steps. From three different runs, we got the same distribution and the effective sample size (ESS) of every parameter is larger than 50, indicating a well mixing chain. The curves generated by migration analyses have clear peaks, reflecting well the occurrence of migration in both directions. Estimates of migration parameters obtained

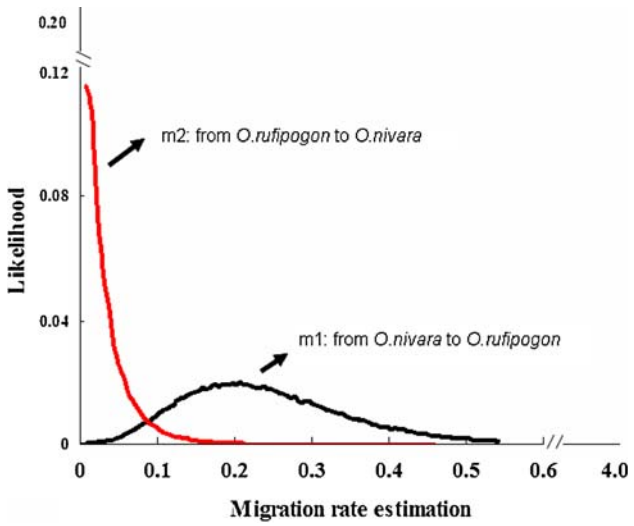

Fig. 2 Likelihood distributions for migration rate estimates under the "Isolation with Migration" model

non-zero values for both species, with $m_1 = 0.23$ (from *O. nivara* to *O. rufipogon*) and $m_2 = 0.003$ (from *O. rufipogon* to *O. nivara*). Importantly, the probability of the migration rate from *O. nivara* to *O. rufipogon* was significantly larger than zero (P = 0.02), whereas the migration rate in the opposite direction was not significantly larger than zero (P = 0.13). Consequently, we accept that there is gene flow in the direction from *O. nivara* to *O. rufipogon*.

Discussion

Contrasting population genetic structure between *O. rufipogon* and *O. nivara*

A few of studies have investigated the species-wide pattern of nucleotide variation of *O. rufipogon* sensu lacto (Barbier et al. 1991; Yoshida et al. 2004; Yoshida and Miyashita 2005; Londo et al. 2006; Rakshit et al. 2007; Zhu et al. 2007). These studies, however, were unable to elucidate the variation pattern within and between populations because either single accessions from disparate populations were used or annual individuals were not sufficiently sampled. The present study sampled sequences from 11 populations and shows that *O. rufipogon* possesses much higher level of nucleotide variation than *O. nivara* at both population and species levels, particularly at the population level where the diversity level in *O. rufipogon* is 2–3 folds higher than that in *O. nivara* (Table 1). Moreover, markedly different population differentiation was detected between the two species. For *O. rufipogon*, about 52% of total nucleotide diversity resides among populations; whereas 78.2% of total diversity was attributable to among-populations in *O. nivara* (Table 3). The contrasting population genetic structure between the annual (*O. nivara*) and perennial (*O. rufipogon*) populations was also found in previous studies using morphological and molecular data (Morishima et al. 1984;

Morishima and Barbier 1990; Kuroda et al. 2007). For example, based on allozyme data, Morishima and Barbier (1990) found that total genetic diversity (H_T) and within-population diversity (H_S) in the perennial *O. rufipogon* ($H_T = 0.39$, $H_S = 0.24$) was much higher than that in the annual *O. rufipgon* (*O. nivara*) ($H_T = 0.14$, $H_S = 0.06$) and population differentiation in *O. rufipogon* ($G_{ST} = 0.396$) was significantly lower than in *O. nivara* ($G_{ST} = 0.600$). Recently, based on seven SSR loci, Kuroda et al. (2007) comparatively studied the population genetic structure of ten *O. rufipogon* populations and ten *O. nivara* populations sampled from Laos. In addition to higher genetic diversity in *O. rufipogon* ($H_T = 0.77$) than in *O. nivara* ($H_T = 0.64$), they also found much higher genetic differentiation among *O. nivara* populations ($G_{ST} = 0.77$, $R_{ST} = 0.71$) than that among *O. rufipogon* populations ($G_{ST} = 0.29$, $R_{ST} = 0.28$). These studies, in conjunction with the present findings, suggest that contrasting pattern of population genetic structure exists for the two species at both species and regional scales irrespective of molecular markers used .

Different level of genetic diversity and contrasting population genetic structure between *O. rufipogon* and *O. nivara* might be explained by factors such as life history, mating system, and habitat that are diverged markedly between the two species (Vaughan and Morishima 2003; Sang and Ge 2007). Population genetics theory predicts that genetic variation in a selfing species would be reduced by decreasing effective population size and eliminating the effective rate of recombination (Charlesworth 2003) and the effect of mating system on variation at the population level is stronger than that at the species level (Glemin et al. 2006). Empirical studies indicated that inbreeding species have generally lower genetic diversity than outcrossing species and their genetic variability is more structured (see reviews in Hamrick and Godt 1990 and Glemin et al. 2006). Therefore, it is not unexpected that the outcrossing and perennial *O. rufipogon* possess higher level of genetic diversity and lower genetic differentiation among populations than the predominantly inbreeding and annual *O. nivara*. A similar pattern of nucleotide variation was also found in a number of comparative studies involving closely related outcrossing and inbreeding species (e.g., Liu et al. 1998; Baudry et al. 2001; Chiang et al 2003; Sweigart and Willis 2003).

Consistent with the findings based on allozymes, RFLPs, and SSRs (Zhou et al. 2003; Cai et al. 2004), significant genetic differentiation among almost all pairwise populations of two species was found in this study (Table 4). Moreover, we detected signficant higher N_{ST} than G_{ST} for both *O. nivara* ($P < 0.001$) and *O. rufipogon* ($P < 0.01$) populations, indicating that there is a phylogeographical structure with closely related haplotypes being found more often in the same area. However, significant association between genetic differentiation and geographical distance was observed only for *O. nivara* populations when Mantel's test was performed, suggesting that isolation by distance pattern might be more profoundly influenced by the reproductive and mating strategy. In contrast to *O. rufipogon* that is propagated by seeds and vegetatively (Vaughan 1989; Morishima and Barbier 1990), *O. nivara* has abundant seeds and high ability of seed dispersal (Vaughan and Morishima 2003) and thus may lead to populations of geographical proximity being similar genetically. In a previous study of Chinese *O. rufipogon* populations, a significant correlation was found between genetic differentiation and geographical distance (Zhou et al. 2003). Lu et al. (2002) and Cai et al. (2004) also indicated that population differentiation in *O. rufipogon* reflected geographic variation due to genetic isolation. Lack of significant association between genetic differentiation and geographical distance for *O. rufipogon* populations by Mantel's test may be affected by the scale and density of the populations sampled. It should be noted that only *Lhs1* has been used in this study and such a strategy may lead to bias in parameter estimation, particularly for genetic diversity. Therefore, genetic diversity of the two species and the effect of geographic isolation on genetic differentiation between populations needs to be investigated further using more populations and additional molecular markers.

Divergence and gene flow between species

It has long been a question whether to treat *O. rufipogon* and *O. nivara* as distinct species or taxa within a single species (Vaughan 1989; Vaughan and Morishima 2003; Sang and Ge 2007) because studies based on morphological analyses and artificial hybridization (Oka 1988; Lu et al. 2000) and molecular data (Lu et al. 2002; Zhu and Ge 2005; Zhu et al. 2007) did not reveal either apparent reproductive isolation or significant genetic differentiation between them. These previous results were verified in the present study where no single fixed difference was observed between the two species (Table 2) and only 16.09% of total nucleotide diversity was attributable to species divergence in the AMOVA partitioning (Table 3). Our simulation to fit the "Isolation with Migration" model to the data demonstrated significant gene flow from *O. nivara* to *O. rufipogon* though the opposite direction of introgression was near zero (Fig. 2). Such asymmetrical gene flow has been detected in other plant species (e.g., Sweigart and Willis 2003). Based on studies on wild populations of India and Thailand, Morishima et al. (1984) argued that the major direction of pollen flow was from annual to perennial because the outcrossing rates were much higher in perennial than in annual types. Gene flow was also observed across species in other closely related *Oryza* species (Vaughan and Morishima 2003) and may explain the lack of phylogenetic resolution between *O. nivara* to *O. rufipogon* (Sang and Ge 2007). The direc-

tional introgression from selfing species into their outcrossing relatives has been reported in two *Minulus* species (Sweigart and Willis 2003). Directional introgression may be partly responsible for the higher genetic diversity in *O. rufipogon* than in *O. nivara*. Therefore, gene flow or introgression has played important roles in shaping current patterns of genetic diversity and in integrating the two wild species.

It is worthwhile to note that *O. rufipogon* and *O. nivara* have distinctive phenotypes associated with their different life cycles and reproductive system in their natural habitats (Morishima et al. 1984; Vaughan and Morishima 2003). More importantly, the phenotypic distinction has a genetic basis and occurrs both regionally and over very restricted areas. For example, Cai et al. (2004) evaluated 11 morphological and physiological traits in one *O. nivara* and six *O. rufipogon* populations by growing them in a common environment. They found that major difference in these quantitative traits was ecotype differentiation towards either perennial or annual types, reflecting their genetic basis. Such adaptive differentiation in life-history traits in response to different habitats was also reported in the populations from restricted areas (Morishima et al. 1984; Kuroda et al. 2007).

These observations on *O. rufipogon* and *O. nivara* populations indicate that population pattern reflected by quantitative traits or phenotypes might be independent of that revealed by relatively neutral markers, as demonstrated in many other species (Schaal et al. 1998; Gonzalez-Rodriguez et al. 2004). This also raises an interesting question why distinctive phenotypes are persistent in the wild given gene flow between the two species. Although more studies and evidence are needed, it is reasonable to argue that strong selection resulting in ecological adaptation occurs in *O. rufipogon*/*O. nivara* populations and limited gene flow may not be strong enough to counteract the phenotypic differentiation due to adaptation. Further studies are needed to focus on population genetic structure of both neutral markers and functional genes involving adaptive phenotypes, which will provide new insights into the speciation and evolutionary dynamics of these relatives of rice.

Acknowledgments We thank Dr. Akito Kaga, National Institute of Agrobiological Sciences of Japan, for his technological assistance in sequencing. Hai-Fei Zhou thanks JIRCAS of Japan for supporting his stay in National Institute of Agrobiological Sciences of Japan. This study was supported by the National Basic Research Program of China (2007CB815704) and National Natural Science Foundation of China (30430030, 30121003) and JIRCAS of Japan.

References

Barbier P, Morishima H, Ishihama A (1991) Phylogenetic relationships of annual and perennial wild rice: probing by direct DNA sequencing. Theor Appl Genet 81:693–702

Baudry E, Kerdelhue C, Innan H, Stephan W (2001) Species and recombination effects on DNA variability in the tomato genus. Genetics 158:1725–1735

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

Charlesworth D (2003) Effects of inbreeding on the genetic diversity of populations. Phil Trans R Soc Lond B Biol Sci 358:1051–1070

Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohstubo E (2003) Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol Biol Evol 20:67–75

Chiang Y-C, Schaal BA, Chou C-H, Huang S, Chiang T-Y (2003) Contrasting selection modes at the Adh1 locus in outcrossing *Miscanthus sinensis* vs. inbreeding *Miscanthus condensatus* (*Poaceae*). Am J Bot 90:561–570

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci USA 95:4441–4446

Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

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

Ge S, Oliveira GCX, Schaal BA, Gao L-Z, Hong D-Y (1999) RAPD variation within and between natural populations of the wild rice *Oryza rufipogon* from China and Brazil. Heredity 82:638–644

Glemin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc R Soc Biol Sci 273:3011–3019

Gonzalez-Rodriguez A, Bain JF, Golden JL, Oyama K (2004) Chloroplast DNA variation in the *Quercus affinis-Q. laurina* complex in Mexico: geographical structure and associations with nuclear and morphological variation. Mol Ecol 13:3467–3476

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant Population Genetics, Breeding and Genetic Resources. Sinauer Associates Inc., Sunderland, pp 43–63

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. Curr Opin Genet Dev 16:592–596

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760

Hilton H, Kliman RM, Hey J (1994) Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution 48:1900–1913

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Jeon J-S, Jang S, Lee S, Nam J, Kim C, Lee S-H, Chung Y-Y, Kim S-R, Lee YH, Cho Y-G, An G (2000) *leafy hull sterile1* is a homeotic mutation in a rice MADS box gene affecting rice flower development. Plant Cell 12:871–884

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35:25–34

Kuroda Y, Sato Y-I, Bounphanousay C, Kono Y, Tanaka K (2007) Genetic structure of three *Oryza* AA genome species (*O. rufipogon*, *O. nivara* and *O. sativa*) as assessed by SSR analysis on the Vientiane Plain of Laos. Conserv Genet 8:149–158

Liu F, Zhang L, Charlesworth D (1998) Genetic diversity in *Leavenworthia* populations with different inbreeding levels. Proc Biol Sci 265:293–301

Londo JP, Chiang Y-C, Hung K-H, Chiang T-Y, Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc Natl Acad Sci USA 103(25):9578–9583

Lu B-R, Naredo MEB, Juliano AB, Jackson MT (2000) Preliminary studies on taxonomy and biosystematics of the AA genome *Oryza* species (Poaceae). In: Jacobs SWL, Everett J (eds) Grasses: systematics and evolution. CSIRO, Melbourne, pp 51–58

Lu B-R, Zheng KL, Qian HR, Zhuang JY (2002) Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theor Appl Genet 106:101–106

Morishima H, Barbier P (1990) Mating system and genetic structure of natural populations in wild rice *Oryza rufipogon*. Plant Species Biol 5:31–39

Morishima H, Sano Y, Oka HI (1984) Differentiation of perennial and annual types due to habitat conditions in the wild rice *Oryza perennis*. Plant Syst Evol 144:119–135

Nei M (1987) Molecular evolutionary genetics. Columbia Universtiy Press, New York

Oka HI (1988) Origin of cultivated rice. Japan Scientific Societies Press, Tokyo

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144:1237–1245

Rakshit S, Rakshit A, Matsumura H, Takahashi Y, Hasegawa Y, Ito A, Ishii T, Miyashita NT, Terauchi R (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. Theor Appl Genet 114:731–743

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Sang T, Ge S (2007) Genetics and phylogenetics of rice domestication. Curr Opin Genet Dev 17:533–538

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998) Phylogeographic studies in plants: problems and prospects. Mol Ecol 7:465–474

Second G (1985) Evolutionary relationships in the sativa group of *Oryza* based on isozyme data. Genet Sel Evol 17:89–114

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279

Song Z-P, Zhu W-Y, Rong J, Xu X, Chen J-K, Lu B-R (2006) Evidences of introgression from cultivated rice to *Oryza rufipogon* (Poaceae) populations based on SSR fingerprinting: implications for wild rice differentiation and conservation. Theor Appl Genet 114:731–743

Sun CQ, Wang XK, Li ZC, Yoshimura A, Iwata N (2001) Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. Theor Appl Genet 102:157–162

Sweigart AL, Willis JH (2003) Patterns of nucleotide diversity in two species of Mimulus are affected by mating system and asymmetric introgression. Evolution 57:2490–2506

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Vaughan DA (1989) The genus *Oryza* L.: current status of taxonomy. International Rice Research Institute, Manila

Vaughan DA, Morishima H (2003) Biosystematics of the genus *Oryza*. In: Smith CW, Dilday RH (eds) Rice: origin, history, technology, and production. Wiley, London, pp 27–65

Vaughan D, Morishima H, Kadowaki K (2003) Diversity in the *Oryza* genus. Curr Opin Plant Biol 6:139–146

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Won Y-J, Sivasundar A, Wang Y, Hey J (2005) On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. Proc Natl Acad Sci USA 102:6581–6586

Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol 22:506–519

Yoshida K, Miyashita N (2005) Nucleotide polymorphism in the Adh2 region of the wild rice *Oryza rufipogon*. Theor Appl Genet 111:1215–1228

Yoshida K, Miyashita NT, Ishii T (2004) Nucleotide polymorphism in the Adh1 locus region of the wild rice *Oryza rufipogon*. Theor Appl Genet 109:1406–1416

Zhou H-F, Xie Z-W, Ge S (2003) Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon* Griff.) in China. Theor Appl Genet 107:332–339

Zhu Q, Ge S (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytol 167:249–265

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol 24:875–888