MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 2439–2454 doi: 10.1111/j.1365-294X.2010.04674.x

Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*)

XIAO-MING ZHENG and SONG GE

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China and Graduate University, Chinese Academy of Sciences, Beijing 100039, China

Abstract

Ecological divergence plays a prominent role in the process of speciation, but how divergence occurs in the face of gene flow is still less clear, and remains controversial among evolutionists. Here we investigated the nucleotide diversity, divergence and gene flow between *Oryza nivara* and *O. rufipogon* using sequences of seven chloroplast and nuclear loci. By analysing samples from 26 wild populations across the geographic ranges of the two species, we showed that both species were highly structured and *O. rufipogon* maintained a higher level of species-wide diversity than *O. nivara*. Notably, phylogenetic, AMOVA and F_{ST} analyses were unable to detect significant nucleotide differentiation between the two species. We estimated that the two species began to diverge at *c.* 0.16 million years ago. Our coalescent-based simulations strongly rejected the simple isolation model of zero migration between species, but rather provided unambiguous evidence of bidirectional gene flow between species, particularly from *O. rufipogon* to *O. nivara*. Our simulations also indicated that gene flow was recurrent during the divergence process rather than arising from secondary contact after allopatric divergence. In conjunction with different morphological and life-history traits and habitat preference in the two species, this study supports the hypothesis that these *Oryza* species are better treated as ecotypes that diverged quite recently and are still under the process of divergence. Importantly, we demonstrate the ecological divergence between *O. rufipogon* and *O. nivara* in the presence of significant gene flow, implying that natural selection plays a primary role in driving the divergence of the two *Oryza* species.

Keywords: divergence, ecological speciation, gene flow, nucleotide diversity, *Oryza nivara*, *Oryza rufipogon*

Received 12 February 2010; revision received 13 April 2010; accepted 16 April 2010

Introduction

Reconstructing the evolutionary process of population divergence is fundamental to understanding the mechanisms by which a population adapts to its environment and ultimately the process of speciation (Coyne & Orr 2004; Hey 2006). Ecological divergence between populations is driven by divergent natural selection between contrasting environments, resulting in ecological speciation through the evolution of reproductive isolation (Schluter 2000, 2009; Coyne & Orr 2004). Given the fact that ecological speciation is increasingly regarded as a common means by which new species arise, ecological divergence has been assumed to play a prominent role in the process of speciation (Schluter 2000, 2009; Abbott & Comes 2007). Ecological divergence or adaptation is driven by natural selection and countered by gene flow between populations, particularly at the early stages of divergence process (Hey 2006; Foster *et al*. 2007; Nosil 2008). Therefore, ecological divergence occurs when natural selection is strong enough to overcome gene flow or when gene flow is limited.

Gene flow is one of the important evolutionary forces that maintains cohesion among geographically separated populations and accelerates evolution through

Correspondence: Song Ge, Fax: +86 10 62590843;
E-mail: gesong@ibcas.ac.cn

adaptive introgression (Morjan & Rieseberg 2004; Abbott & Comes 2007; Lexer & Widmer 2008). An accurate characterization of gene flow under ecological divergence would provide significant insights into the understanding of speciation models and the strength of natural selection contributing to adaptive divergence and speciation. For example, if homogenizing gene flow was found between closely related or incipient species, sympatric or parapatric speciation rather than strict allopatric speciation must be invoked, thus implying that divergent natural selection would play a critical role in species divergence (Coyne & Orr 2004; Schluter 2009). Despite its importance, however, the role of gene flow in population divergence and, in particular, the likelihood of speciation in the face of gene flow has long been contentious among evolutionary biologists and remains poorly understood (Hey 2006; Abbott & Comes 2007; Nosil 2008).

In overall agreement with the conventional view that gene flow is absent during speciation (Mayr 1963; Coyne & Orr 2004), theoretical models demonstrated that speciation in the face of gene flow was difficult though possible under particular conditions (Bolnick & Fitzpatrick 2007). Nevertheless, owing to methodological advances in studying speciation with gene flow (Hey & Nielsen 2004; Hey 2006), a growing number of studies have convincingly demonstrated that speciation in the presence of gene flow could be common in both plants and animals (e.g. Niemiller *et al.* 2008; Nosil 2008; Stadler *et al.* 2008; Faure *et al.* 2009; Nadachowska & Babik 2009; Ross-Ibarra *et al.* 2009; for reviews see Morjan & Rieseberg 2004; Hey 2006). Although molecular population genetics of gene flow have been investigated in numerous plant species (Morjan & Rieseberg 2004; Zhang & Ge 2007; Lexer & Widmer 2008; Stadler *et al.* 2008; Strasburg & Rieseberg 2008; Ross-Ibarra *et al.* 2009), relatively few have studied the gene flow during ecological divergence and its consequence to local adaptation and speciation (Foster *et al.* 2007; Lowry *et al.* 2008).

Two wild *Oryza* species, *O. rufipogon* Griff. and *O. nivara* Sharma et Shastry, are more closely related and collectively regarded as the progenitors of *O. sativa* (Sang & Ge 2007a; Vaughan *et al.* 2008). *O. rufipogon* had always been the species name widely accepted to accommodate the wild Asian A-genome species until the annual and predominantly self-fertilized populations was treated as an independent species *O. nivara* by Sharma & Shastry (1965). Although taxonomic debate on whether *O. rufipogon* and *O. nivara* are treated as two species or ecotypes of the same species has continued (Morishima 2001; Sang & Ge 2007a; Vaughan *et al.* 2008), we refer to these two taxa as different species in the present study for convenience. *O. rufipogon* is perennial, photoperiod sensitive and largely cross-fertilized; whereas *O. nivara* is annual, photoperiod insensitive and predominantly self-fertilized. Geographically, *O. rufipogon* is widely distributed from southern China, South and Southeast Asia to Papua New Guinea and northern Australia, and inhabits areas with year-round water, such as swamps and lakes. In contrast, *O. nivara* is mainly found in South and Southeast Asia, and usually occurs in ponds and swamps that dry up completely in the dry season (Vaughan & Morishima 2003; Vaughan *et al.* 2008). Artificial hybridization (Oka 1988; Lu *et al.* 2000), molecular phylogenetic analyses (Barbier *et al.* 1991; Lu *et al.* 2002; Zhu & Ge 2005) and molecular population genetic work (Zhu *et al.* 2007) indicate that the two species are cross compatible and exhibit little interspecific genetic differentiation. On the other hand, contrasting morphology and life-history traits as well as habitat preference in the two species have been documented by a variety of experimental and field investigations (Sano *et al.* 1980; Sano & Morishima 1982; Morishima *et al.* 1984; Barbier 1989; Cai *et al.* 2004; Vaughan *et al.* 2008). Consequently, *O. rufipogon* and *O. nivara* have been regarded as different subspecies (Vaughan & Morishima 2003) or ecotypes (Oka 1988; Barbier 1989; Morishima 2001; Zhu & Ge 2005) under a single species (*O. rufipogon sensu lato*). Multiple lines of evidence suggest that the presence of annual *O. nivara* and perennial *O. rufipogon* are an adaptive response to divergent habitats and reflect the different adaptive strategies (Morishima 2001; Sang & Ge 2007a). As such, these two species are sister or incipient species in the early stage of speciation, and thus provide an ideal biological model to study the history of divergence with gene flow and ecological adaptation.

In the present study, we provide the first investigation on the divergence population genetics of *O. nivara* and *O. rufipogon* using multilocus sequences. First, we examined the nucleotide diversity and differentiation of the two *Oryza* species. All previous studies have focused either on local populations in specific regions (Morishima *et al.* 1984; Barbier 1989; Ge *et al.* 1999; Gao *et al.* 2000; Zhou *et al.* 2003; Cai *et al.* 2004; Kuroda *et al.* 2007) or on the species based on one or few individuals from disparate populations (Second 1985; Sun *et al.* 2001; Lu *et al.* 2002; Caicedo *et al.* 2007; Zhu *et al.* 2007). Moreover, *O. nivara* populations have not been sufficiently sampled at the species level or not sampled at all in all the cases. Here, by collecting samples from 26 wild populations across the entire geographic ranges of the two species, we were able to determine the patterns of nucleotide diversity both between species and among populations within species. Specifically, we asked whether *O. nivara* and *O. rufipogon* represent genetically differentiated taxa or good species because different species names have been inconsistently assigned to the annual

plant (Vaughan 1989; Sang & Ge 2007a; Vaughan *et al.* 2008). Such information also provides additional insights into the speciation history of the species and helps better characterize the wild progenitor of cultivated rice.

Our second goal was to quantitatively characterize the level, direction and temporal distribution of gene flow between *O. nivara* and *O. rufipogon*. It is noted that weak genetic differentiation between species could be due to recent divergence, gene flow, or a combination of these factors (Hey 2006; Nosil 2008). Moreover, gene flow might have been ongoing during the process of divergence at one extreme and occurred upon secondary contact after a period of allopatric divergence at the other extreme (Coyne & Orr 2004; Nosil 2008). It has long been difficult to assess the extent of gene exchange between closely related or incipient species during their divergence until powerful analytical methods have been developed to accurately estimate molecular demographic parameters including effective population size, divergence time, and gene flow (Nielsen & Wakeley 2001; Hey & Nielsen 2004, 2007). Weak or no genetic differentiation between the two *Oryza* species, as evidenced in previous studies, implies the likelihood of potential gene flow between them. No attempt has been made to estimate gene flow in *Oryza* species under the analytical framework of divergence population genetics until recently Zhou *et al.* (2008) revealed an asymmetrical gene flow from *O. nivara* to *O. rufipogon* using the isolation with migration model on *Lhs1* sequences. However, accurate documentation of gene flow requires multilocus studies because estimated rates of gene flow may vary widely among loci (Won & Hey 2005; Strasburg & Rieseberg 2008). Thus, we aim to use multilocus sequences to address how strong and at what point the gene flow occurred during the divergence of the two *Oryza* species. Characterization of the extent and timing of gene flow may also help reveal speciation models involving the species divergence and contribute to our understanding of how these two species evolve in contrasting environmental conditions.

Materials and methods

Population sampling and loci studied

A total of 243 individuals representing 11 *O. nivara* and 15 *O. rufipogon* populations were sampled across the entire geographical distribution of the species. The geographic distribution of the two species and localities of the populations sampled in this study are showed in Fig. 1. Detailed information on the sampled populations is provided in Table S1, Supporting information. The distribution of the annual *O. nivara* is confined to tropical continental Asia, in contrast to the distribution of perennial *O. rufipogon*, which extends to China in the north and to Oceania in the south (Fig. 1). In the region

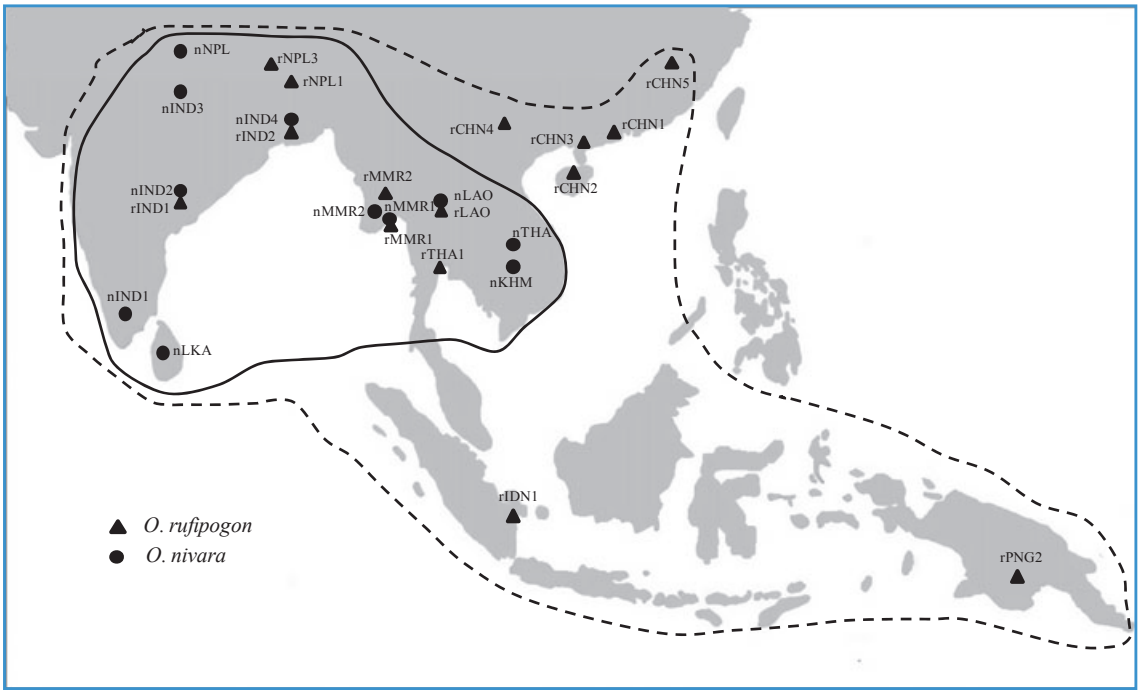

Fig. 1 Geographical distribution of two species and the localities of the populations sampled in this study. Solid and broken lines indicate the geographic range of *O. nivara* and *O. rufipogon*, respectively. Solid circles and triangles represent the populations of *O. rufipogon* and *O. nivara*, respectively. Detailed information on these populations is provided in Table S1, Supporting information.

where *O. nivara* and *O. rufipogon* overlap, the populations of the two species are largely sympatric or parapatric despite different habitats, and usually found in close proximity from several kilometres to less than 100 m (Kuroda *et al.* 2007; Vaughan *et al.* 2008). In this study, 8 out of 15 *O. rufipogon* populations were collected from localities within the distributional range of *O. nivara* (Fig. 1). Because of the potential introgression/hybridization between cultivated rice and the two wild species under study (Kuroda *et al.* 2005; Sang & Ge 2007a; Vaughan *et al.* 2008), we grew the samples from all populations in the greenhouse of Institute of Botany (CAS) in Beijing and China National Rice Research Institute in Hangzhou, Zhejiang Province. Based on a few morphological features such as grain shattering, tiller and panicle structures, and the flowering time, we were able to distinguish among the cultivated rice and the two wild species (Sang & Ge 2007a; Vaughan *et al.* 2008). Particularly, we were able to reliably identify admixed individuals because of the reduction of grain shattering that occurs exclusively in cultivated rice. Thus we excluded from this study the populations that might be admixed with cultivated rice. We also sampled two *Oryza barthii* accessions as outgroups because this species is closely related to *O. nivara/O. rufipogon* group (Zhu & Ge 2005). Total DNA was extracted from fresh or silica gel-dried leaves, using the hexadecyltrimethylammonium bromide method as described in Ge *et al.* (1999).

Two chloroplast and five nuclear loci were PCR amplified and sequenced from 7 to 15 individuals for each of 26 populations. We chose a subset of the nuclear loci used in our previous study (Zhu *et al.* 2007), including the single-copy *Ks1*, *Lhs1*, *SSII1* and *Waxy*. One additional nuclear gene, *Heading data 3a* (*Hd3a*), located in rice chromosome 6, was also sampled. This gene is a rice orthologue of *Arabidopsis FLOWERING LOCUS T* (*FT*) that contributes to diversity in flowering time (Takahashi *et al.* 2009). These genes represent five unlinked loci on four different chromosomes in rice. In addition, two chloroplast fragments (*ndhC-tRNV* and *trnP-RPL33*) were sampled because they are intergenic spacers and evolve relatively quickly in grasses (Tang *et al.* 2010). The schematic diagrams of all five nuclear loci are shown in Fig. S1, Supporting information, and the amplification primers and sequenced regions of all loci are provided in Table S2, Supporting information.

Amplification, cloning and sequencing

Polymerase chain reaction (PCR) amplification methods generally followed those of our previous studies (Zhang & Ge 2007; Zhu *et al.* 2007; Tang *et al.* 2010). Sequencing was done on an ABI3730XL automatic sequencer (Applied Biosystems Corp.). Because heterozygous individuals may exist for both *O. rufipogon* and *O. nivara* (Zhu *et al.* 2007), PCR fragments were cloned into pGEM T-easy vectors (Promega Corp.) with a Pharmacia purification kit (Amersham Pharmacia Biotech) and three cloned DNA fragments were sequenced for each individual. Since *Taq* errors occur at random, polymorphisms shared among more than two clones (sequences) are unlikely to be artificial; whereas singletons can be either true sequence variation or *Taq* polymerase error (Eyre-Walker *et al.* 1998; Zhu *et al.* 2007). To correct for PCR errors in cloned fragments, we identified individuals in the alignments that contained singletons and then resequenced at least four clones after a second round of PCR. In addition to confirm the singletons, this strategy of repeated PCR amplification, cloning and sequencing could verify and remove interallelic PCR recombinants (Zhang & Ge 2007; Zhu *et al.* 2007). All sequences have been deposited in GenBank, with the accession nos GU727015–GU727308.

Sequence analysis and tests of neutrality

Sequences were edited with BioEdit 7.0.9.0 (Hall 1999). For each locus, we calculated the number of segregating sites (S), number of haplotypes (h) and two parameters of nucleotide diversity: Nei's π, the expected heterozygosity per nucleotide site (Nei 1987) and Watterson's θ, an estimate of the population mutation parameter $4N_e\mu$ (Watterson 1975) using DnaSP version 5.0 (Rozas *et al.* 2003). We estimated the minimum number of recombination events (Rm) using the four-gamete test (Hudson & Kaplan 1985).

We performed Tajima's D (1989) and D^* and F^* of Fu & Li (1993) using the program DnaSP to test for the neutral equilibrium model of evolution. These tests were conducted for both the species-wide pooled samples and the local populations at each locus. We obtained the associated one-tailed P-valued for Tajima's D and Fu and Li's D^* and F^* by computing 1000 coalescent simulations, taking into account estimates of the recombination per gene in DnaSP. Additionally, multilocus tests for Tajima's D and Fu and Li's D^* and F^* as well as the multilocus HKA test (Hudson *et al.* 1987) across loci were performed at the species level using the HKA program (http://lifesci.rutgers.edu/;heylab/heylabsoftware htm#HKA). We used *O. barthii* sequences as outgroups to conduct the HKA tests for the two species separately.

Phylogenetic analyses

The genealogical trees of alleles/haplotypes were constructed by PAUP* version 4.0b10 (Swofford 2002), using

the Neighbour-Joining (NJ) method with Kimura's 2-parameter distances (Kimura 1980). Maximum parsimony (MP) and maximum likelihood (ML) methods were also performed and a heuristic search with tree-bisection–reconnection, ACCTRAN and 1000 random-taxon-addition replicates was implemented in these analyses. The optimal model of sequence evolution for each data set was determined using Akaike's information criterion (AIC) implemented in MODELEST 3.7 (Posada & Crandall 1998). ML analyses were performed under the model of evolution selected for each data set. In all phylogenetic analyses, bootstrap analyses were performed to assess the confidence of internal nodes with 1000 replicates for NJ and MP and 300 replicates for ML. The hypothesis of monophyly of the two species was evaluated with the Shimodaira–Hassegawa (SH) test (Shimodaira & Hassegawa 1999) by comparing an unconstrained tree and the trees that are constrained by reciprocal monophyly of the two species.

The overall distribution of nucleotide diversity among populations was investigated using an Analysis of Molecular Variance (AMOVA) as implemented in Arlequin 3.01 program. Sequence variation was hierarchically partitioned between the two species, among populations within species and within populations. The significance of all estimated fixation indices was tested using 10,000 permutations as described in Excoffier *et al.* (1992). Pairwise F_{ST}, generally expressed as the proportion of genetic diversity due to allele frequency differences among populations (Holsinger & Weir 2009), was used to measure population differentiation within and between species, as implemented in Arlequin 3.01 program (Excoffier *et al.* 1992).

Fitting an isolation with migration model

We used the isolation with migration (IM) model (Nielsen & Wakeley 2001; Hey & Nielsen 2004) to estimate population demographic parameters, with which the relative effects of divergence and gene flow can be distinguished with sufficient accuracy and precision (Won & Hey 2005; Hey 2006; Strasburg & Rieseberg 2010). This model uses Markov chain Monte Carlo (MCMC) to estimate the posterior probability densities of six demographic parameters scaled by the neutral mutation rate (μ), including the effective population sizes of the ancestral (θ_A) and the descendant populations (θ_1 and θ_2), the divergence time (t) and bidirectional migration rates (m_1 and m_2). The IM model involves several simplifying assumptions such as neutrality and nonrecombination of genetic loci, random mating in ancestral and descendent populations, and no genetic contribution from unsampled populations (Nielsen & Wakeley 2001; Hey & Nielsen 2004). Although these assumptions might be violated in empirical data, recent studies has shown that parameter estimates of the IM model were robust to even high levels of population structure and to recombination as long as loci were divided into nonrecombining blocks (Carstens *et al.* 2007; Strasburg & Rieseberg 2010). Strasburg & Rieseberg (2010) further indicated that IMa estimates are generally quite robust to small to moderate violations of the IM model assumptions, but model selection of mutation rate had significant impact on estimates of almost all parameters. To avoid these biases, therefore, we first chose the largest nonrecombining blocks of each locus in our analyses. The length of the longest nonrecombining block ranged from 165 bp (*Waxy*) to 326 bp (*SSII1*), with the average length of about 250 bp. Then, we use MODELEST 3.7 (Posada & Crandall 1998) to obtain the optimal model of sequence evolution to assess the potential error due to improper model use. Of the nuclear loci, *Waxy* fit the GTR model and the remaining four loci to HKY model.

We estimated demographic parameters using the IMa program (Hey & Nielsen 2007). After a few preliminary runs were made to optimize prior boundaries for six parameters, we conducted the final simulation for a burn-in of one million generations and five million steps. The analysis was done with 10 independent chains under Metropolis coupling and performed with different seed numbers to guarantee convergence of the sample (Hey & Nielsen 2004; Won & Hey 2005). To convert divergent parameter estimates to easily interpreted units, we followed the modified method of Tenaillon *et al.* (2004) to estimate the mutation rate (μ) at silent sites for each locus, i.e. $\mu = \mu_{adh1} \times K_{sil}/K_{sadh1} \times L$, where L is the length of the locus and K_{sil} and K_{sadh1} are silent distance for that locus and synonymous distance at the *Adh1* locus (Zhang & Ge 2007), respectively. μ_{adh1} is estimated to be 7.0×10^{-9} substitution per synonymous site per year, a fossil-calibrated synonymous rate of *Adh1* divergence in grasses (Gaut *et al.* 1996). The geometric average mutation rate of the five loci ($\mu = 9.0 \times 10^{-6}$) was used to rescale the IMa parameter estimates from the combined analysis. The estimated divergence time and effective population sizes were then converted to years (t) and effective number of individuals (N_1, N_2 and N_A), respectively. The migration parameter ($2Nm$) is the product of the effective number of gene copies and the gene copy migration rate ($2Nm = \theta \times m/2$) (Hey & Nielsen 2004).

Results

Nucleotide diversity and neutrality tests

We sequenced five unlinked nuclear loci and two chloroplast fragments from 243 individuals representing 26

populations of the two species, with a total concatenated length of 4087 bp and 1320 bp for nuclear and chloroplast loci, respectively. The length of the aligned sequence for each locus ranged from 565 to 945 bp. The number of insertion–deletion (indel) polymorphisms ranged from 1 to 53 across loci, with a total of 27 indel polymorphisms for the two taxa. Indels were not considered for subsequent analyses.

Standard statistics of sequence polymorphism for each locus are shown in Table 1. For the nuclear loci, the species-wide levels of silent nucleotide variation (θ_{sil}) vary across loci from 0.0064 (*SSII1*) to 0.0214 (*Hd3a1*) in *O. rufipogon* and from 0.0039 (*Ks1*) to 0.0202 (*Hd3a1*) in *O. nivara*. Average silent nucleotide variation across loci is significantly higher in *O. rufipogon* ($\pi_{sil} = 0.0080$; $\theta_{sil} = 0.0131$) than in *O. nivara* ($\pi_{sil} = 0.0063$; $\theta_{sil} = 0.0101$) ($P < 0.05$, for both π_{sil} and θ_{sil}), in accordance with the fact that *O. rufipogon* has a wider distribution and is largely outcrossing while *O. nivara* is primarily inbreeding (Vaughan 1989; Morishima 2001). As expected, sequence polymorphism of chloroplast fragments is on average lower than that of nuclear loci. An interesting pattern is that θ_{sil} values are consistently higher than those of π_{sil} at all loci for both species except for the chloroplast loci where the two values are close. The difference between θ_{sil} and π_{sil} might reflect strong population structure within species, as found in other species such as *Arabidopsis* (Schmid *et al.* 2005), *Zea mays* (Moeller *et al.* 2007) and wild tomato species (Arunyawat *et al.* 2007). Four-gamete tests indicated that the minimum number of recombination events (*Rm*) ranged from 0 to 3 in *O. nivara* and from 0 to 13 in *O. rufipogon*, with the estimates being higher in *O. rufipogon* than *O. nivara* at all nuclear loci (Table 1), likely a reflection of high heterozygosity due to high levels of outcrossing rate in *O. rufipogon*.

The tests of Tajima's *D* and Fu and Li's *D** and *F** for each locus and for multiple loci are shown in Table 1. All the values of multilocus tests were negative but not significant. For single locus tests, most populations of *O. nivara* had negative but not significant values except for *Waxy* ($D = -1.8176$, $P = 0.043$). In *O. rufipogon*, significant negative Tajima's *D* and/or Fu and Li's *D** and *F** were found in a few cases, including the *Lhs1* for *D* and *D** and *F**, *Ks1* for *D* and *Waxy* for *D** and *F**. A significant departure from neutrality at a specific locus may not necessarily indicate the signature of selection because violation of nonrecombination within loci and population structure could lead to significant values of Tajima's *D* and Fu and Li's *D** and *F** tests (Ramos-Onsins & Rozas 2002; Moeller *et al.* 2007). Therefore, we calculated these statistics using the largest nonrecombining sequence blocks for each locus (Strasburg & Rieseberg 2008) to test whether recombination within loci had a significant effect on *D*, *D** and *F**. The results are qualitatively similar to those in Table 1 (data not shown). To examine whether a significant excess of low frequency polymorphism in the pooled

Table 1 Summary of nucleotide polymorphism and neutrality tests

Taxon	Locus	*S*	*h*	π_{sil}	θ_{sil}	*D*	*D**	*F**	*Rm*
O. nivara (*n* = 102)	*Hd3a*	50	19	0.0149	0.0202	−0.8365	1.5033	0.6722	3
	Ks1	18	16	0.0027	0.0039	−0.8585	−0.7227	−0.9232	1
	Lhs1	28	16	0.0044	0.0057	−0.7297	−0.5479	−0.7406	2
	SSII1	28	27	0.0035	0.0061	−1.2818	−0.5479	−0.9973	3
	Waxy	53	33	0.0062	0.0144	−1.8176*	−0.4802	−1.2345	2
	Average[a]	**35**	**22**	**0.0063**	**0.0101**	**−0.1691**	**−0.1203**	**−0.5324**	**2**
	ndhC-trnV	9	6	0.0025	0.0027	−0.1428	0.5730	0.3866	0
	trnP-rpl33	4	3	0.0025	0.0012	2.0802*	0.9434	1.5391	0
O. rufipogon (*n* = 141)	*Hd3a*	56	36	0.0147	0.0214	−0.9676	−0.0930	−0.5624	8
	Ks1	36	37	0.0026	0.0074	−1.9394*	−0.7726	−1.5040	5
	Lhs1	39	32	0.0038	0.0096	−1.8362*	−2.3589*	−2.5735*	4
	SSII1	31	33	0.0028	0.0064	−1.6694	−0.1168	−0.8939	4
	Waxy	83	53	0.0159	0.0208	−0.7467	−3.5281*	−2.7636*	13
	Average[a]	**49**	**38**	**0.0080**	**0.0131**	**−0.2259**	**−0.1963**	**−0.1258**	**7**
	ndhC-trnV	7	7	0.0025	0.0018	0.7650	−0.7429	−0.2595	0
	trnP-rpl33	5	5	0.0018	0.0014	0.5509	0.9806	0.9962	0

S, number of segregating sites.π_{sil}, average number of pairwise nucleotide differences per site calculated on the silent sites.
θ_{sil}, Watterson's estimator of θ per base pair calculated on the silent sites.
D and *D**/*F**, Tajima's *D* (Tajima 1989) and Fu & Li's *D** and *F** (Fu & Li 1993).
Rm, estimate of the minimum number of recombination events (Hudson & Kaplan 1985).
D and *D**/*F** values are based on multilocus tests using the HKA program.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

samples is due to population structure, we performed the neutrality tests for each population of the two species separately. We found that only 34 out of total 546 tests (6.2%) obtained significant departures from neutral equilibrium expectation, including those loci at which species-wide tests were significant (Table S3, Supporting information). These observations suggest that the significant Tajima's *D* and Fu and Li's *D** and *F** may be attributed partly to the interplay of population structure and demography.

Because Tajima's *D*-statistic is sensitive to population structure and demography (Tajima 1989), we further conducted a multilocus HKA test that is robust to population structure and demography. We did not detect significant departure from the equilibrium model for the *O. nivara*–*O. rufipogon* contrast ($\chi^2 = 2.74$, $P = 0.94$). Given that the HKA test statistic for closely related species is not expected to follow the χ^2 distribution (Machado *et al.* 2002), we repeated HKA tests for *O. nivara* and *O. rufipogon* separately using one sequence of *O. barthii* as an outgroup. We compared the test statistics with a distribution generated from 10 000 coalescent simulations (Hilton *et al.* 1994) and did not detect significant values for the *O. rufipogon*/*O. barthii* contrast ($\chi^2 = 3.63$, $P = 0.85$). However, the null hypothesis of proportionality between polymorphism and divergence was rejected in the *O. nivara*/*O. barthii* contrast ($\chi^2 = 35.02$, $P < 0.001$), with a major contribution of *Waxy* to the significant statistics. Removal of *Waxy* from the multilocus HKA test caused the statistic to drop below the critical value. Taken together, no unequivocal evidence of selection was found for the loci sampled in this study except for *Waxy* at which the variation pattern might be affected by selection to some extent.

Genealogical analysis and species divergence

We used NJ, MP and ML methods to construct the genealogies for the five nuclear loci and the combined chloroplast fragment separately. Figure 2 shows the NJ trees for each locus, and the MP and ML analyses obtained similar phylogenies with slightly different statistical support for some clades (data not shown). It should be cautioned that these phylogenetic trees cannot be treated as true genealogies but rather an approximation of genealogy based on overall similarity (Broughton & Harrison 2003; Niemiller *et al.* 2008). As shown in Fig. 2, the number of clades supported by nonparametric bootstrap over 50% varied among loci, from 4 (*Lhs1*) to 11 (*Ks1*). The alleles or haplotypes sampled from the two species have a polyphyletic relationships. Of the 8 haplotypes in the combined chloroplast locus, 5 (62%) are shared by both species. The shared alleles on the nuclear trees are from 3 (*Ks1*) to 10 (*Lhs1*) (6–26%), and distributed at both deep branches and tips of the trees. Phylogenetic analyses indicated that at most loci, numerous haplotypes from *O. nivara* are more closely related to those sampled from *O. rufipogon*. To test for the monophyly of the two species, we compared the tree topologies using the SH test to compare trees constrained to be reciprocally monophyletic to unconstrained trees. We found a significant reduction in likelihood between the constrained and unconstrained trees for all loci (Table S4, Supporting information), rejecting the monophyly of either of the two species.

Divergence between and within species

Hierarchical AMOVAs were used to investigate the overall distribution of genetic diversity between the two species in the entire and overlapping regions (Table 2). Variance components attributed to between species were significant at only 1 (*Waxy*) and 2 (*Hd3a* and *Waxy*) loci for the populations in the overlapping and entire regions, respectively. Less than 10% of the total variation is attributed to variation between species for all loci except for *Waxy* (32.8%). Strikingly, genetic differentiation between populations within species and within populations was significant at all seven loci. Of the total variation, components among populations within species ranged from 40.82% (*Hd3a*) to 68.25% (*trnP-rpl33*), while components within populations ranged from 20.84% (*Waxy*) to 54.32% (*ndhC-trnV*). The average of variance components among populations within species (49.56%) and that of within populations (42.58%) are much higher than that found between species (7.85%) (Table 2).

Pairwise F_{ST} values revealed strong genetic differentiation at nuclear loci between population contrasts both within species (0.4651–0.5267 for *O. nivara*; 0.3173–0.5748 for *O. rufipogon*) and between species (0.4473–0.7196 for the entire region; 0.3860–0.6039 for the overlapping region) (Table 3), with almost all pairwise values being significant (not shown). However, the average F_{ST} values for the population contrasts between species, whether based on the entire region (0.4924) or on the overlapping region (0.4391) were not significantly higher than that of those contrasts within species (0.5008 for *O. nivara* and 03827 for *O. rufipogon*). These results were consistent with AMOVA analyses, reflecting highly structured populations in both species and little differentiation between species.

Isolation with migration model

We repeated runs of simulations with the IMa program and revealed unambiguous marginal posterior

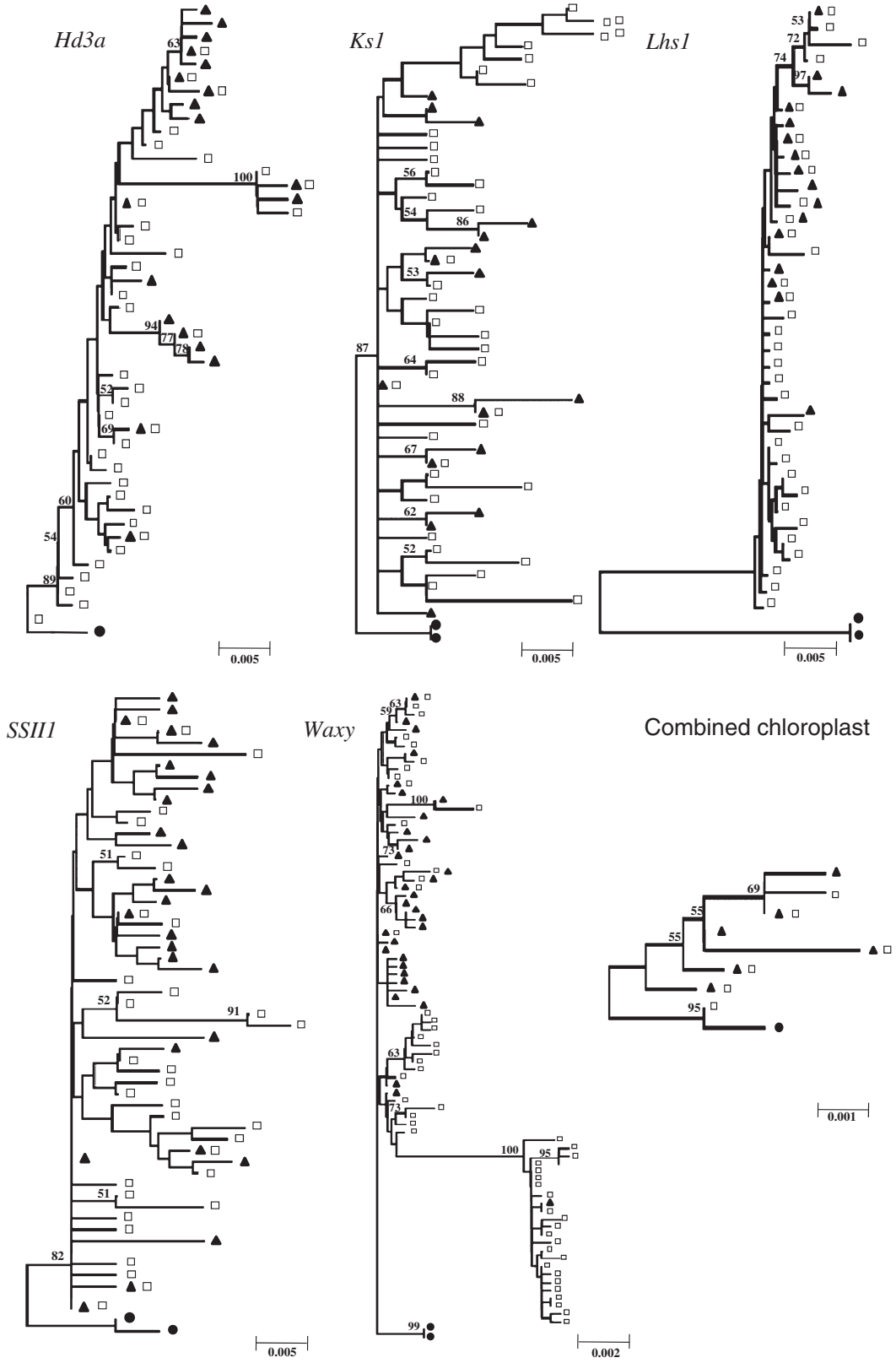

Fig. 2 Phylogenetic trees of five nuclear genes and the combined chloroplast fragment used in this study. Bootstrap support values above 50% are shown on the trees. *Oryza nivara* alleles are indicated by filled triangles and *O. rufipogon* alleles by open squares. Trees were rooted with *O. barthii* sequences indicated by solid circles.

Table 2 Hierarchical analysis of molecular variance for 26 populations of two species (AMOVA)

Source of variation		Nuclear					Chloroplast		Average
		Hd3a	*Ks1*	*Lhs1*	*SSII1*	*Waxy*	*ndhC-trnV*	*trnP-rpl33*	
Two species (the entire region)	Between species	9.76*	2.80	4.34	4.27	32.77***	2.52	−1.50	7.85
	Among populations within species	40.82***	52.33***	44.63***	51.38***	46.38***	43.16***	68.25	49.56
	Within populations	49.42***	44.88***	51.03***	44.35***	20.84***	54.32***	33.25	42.58
Two species (the overlapping region)	Between species	7.79	2.47	−1.48	4.02	29.00***	3.26	−3.40	5.95
	Among populations within species	32.72***	52.71***	48.12***	53.71***	43.03***	31.96***	67.92***	47.16
	Within populations	58.49***	44.82***	53.36***	47.67***	27.97***	65.68***	35.57***	47.65

*$P < 0.05$; ***$P < 0.001$.

Table 3 Summary statistic of pairwise divergence (F_{ST}) between populations within species and between species

Locus		Within *O. nivara*	Within *O. rufipogon*	*O. nivara* vs *O. rufipogon* (the entire region)	*O. nivara* vs *O. rufipogon* (the overlapping region)
Hd3a	Average	0.4651	0.3268	0.4948	0.4298
	Min/max	−0.0973/0.8642	−0.0499/0.9381	−0.0599/0.93608	−0.0599/0.9361
Ks1	Average	0.5267	0.4421	0.5039	0.4593
	Min/max	0.0311/0.8776	0.0350/0.5498	0.0175/0.9415	0.0175/0.8287
Lhs1	Average	0.4857	0.3173	0.4473	0.3860
	Min/max	0.0011/0.7427	−0.0831/0.6978	−0.0129/0.9852	−0.0129/0.8766
SSII1	Average	0.5256	0.4446	0.5234	0.4811
	Min/max	0.1166/0.8621	0.0128/0.9232	0.0548/0.9668	0.0548/0.8438
Waxy	Average	0.4924	0.5748	0.7196	0.6039
	Min/max	−0.0357/0.9942	−0.0463/0.9765	−0.0349/0.9955	−0.0349/0.9955
Average*		**0.5008**	**0.3827**	**0.4924**	**0.4391**
ndhC-trnV	Average	0.3649	0.3733	0.3663	0.3088
	Min/max	−0.1321/1.0000	−0.1376/1.0000	−0.2000/1.0000	−0.1265/0.8907
trnP-rpl33	Average	0.4514	0.5272	0.4695	0.4215
	Min/max	−0.2500/1.0000	−0.1531/1.0000	−0.1546/1.0000	−0.1546/1.0000

**Waxy* was excluded from the average calculation because of its violation of neutral evolution.

probability distributions of the demographic parameters for the two species. The effective sample size (ESS) values for the time parameter ranged from approximately 150 to 500 in different analyses, well over the recommended cut-off. The maximum-likelihood estimates (MLE) and the 90% highest posterior density (HPD) are shown in Table 4. To make the parameter estimates more easily interpreted, we convert them to a scale of effective individuals or years based on the average mutation rate across five loci (9.0×10^{-6}). Effective population sizes for each species are about 0.2 million (90% HPD interval: 0.17–0.27 million) for *O. nivara* and 0.49 million (90% HPD interval: 0.44–0.63 million) for *O. rufipogon*. These effective sizes of descendant populations are 2–5 times larger than the effective size of the ancestral population, indicating that both *O. nivara* and *O. rufipogon* have undergone dramatic population expansion. The marginal posterior probability distribution of the divergence parameter, *t*, showed a sharp peak at 1.455, which was converted into a divergence time of approximately 0.16 million years with 90% HPD ranged from 0.12 to 0.20 million years ago (Table 4).

Long-term estimates of effective migration between *O. nivara* and *O. rufipogon* are exceptionally high and asymmetric (Table 4). MLEs of $2Nm$ from *O. nivara* into *O. rufipogon* and from *O. rufipogon* into *O. nivara* are 0.70 and 2.69, respectively. Because effective gene flow higher than 1 is often regarded as high enough to prevent population differentiation due to genetic drift (Moeller *et al.* 2007), migration from *O. nivara* into *O. rufipogon* could be considered to be moderate while that from *O. rufipogon* into *O. nivara* reflects a high level of effective gene flow. Given the possibility of selection on *Waxy*, we repeated the above simulations by

Table 4 Maximum-likelihood estimates (MLE) and the 90% highest posterior density (HPD) intervals of demographic parameters from IMa multilocus analyses

Comparison	θ_1	θ_2	θ_A	m_1	m_2	t	N_1	N_2	N_A	$2N_1m_1$	$2N_2m_2$	t (years × 1000)
O. nivara vs *O. rufipogon* (the entire region)												
MLE	7.859	19.020	4.045	0.125	0.265	1.455	218.3	528.3	112.3	0.49	2.52	161
Lower 90% HPD	6.278	15.950	2.371	0.025	0.155	1.075	174.3	443.0	65.8			119
Upper 90% HPD	9.812	22.647	6.185	0.315	0.425	1.725	272.5	629.0	171.8			191
O. nivara vs *O. rufipogon* (the overlapping region)												
MLE	7.531	17.631	3.787	0.185	0.303	1.480	209.1	489.7	105.1	0.70	2.69	163
Lower 90% HPD	5.876	14.230	2.307	0.035	0.125	1.135	163.2	395.2	64.0			126
Upper 90% HPD	9.707	21.896	5.876	0.455	0.495	1.775	269.6	608.2	163.0			197

θ_1, θ_2, θ_A, m_1, m_2, t are scaled by the mutation rate, while N_1, N_2, N_A, $2N_1m_1$, $2N_2m_2$ and t are scaled by individuals or years × 10^3.
θ_1: effective population size of *O. nivara*.
θ_2: effective population size of *O. rufipogon*.
θ_A: effective population size of ancestral population.
m_1: population migration rate from *O. nivara* to *O. rufipogon*.
m_2: population migration rate from *O. rufipogon* to *O. nivara*.
t: time since species divergence.

excluding this locus and obtained same conclusions, though some of the estimated values are slightly different (Table S5, Supporting information).

Despite the advantage of statistical power when using multilocus estimates, the single locus analysis of population differentiation might show more detailed demographic history (Strasburg & Rieseberg 2008; Nadachowska & Babik 2009). Thus we further estimate the posterior probability density of the mean time of migration events for each locus in both directions to obtain the temporal distributions of migration rates across loci. It is clear from Fig. 3, that all the loci showed evidence of migration for both directions, with the peaks far from zero. Although the mean time of migration events varied across loci from 0.355 (*Waxy*) to 0.545 (*Ks1*), their distributions are broad for all loci, indicating continuous or recurrent gene flow from the time of species divergence to present.

Discussion

Molecular diversity and differentiation between O. rufipogon *and* O. nivara

Species-wide patterns of nucleotide variation of *O. rufipogon* and *O. nivara* have been studied using sequences of multiple genes (e.g. Caicedo *et al.* 2007; Rakshit *et al.* 2007; Zhu *et al.* 2007). However, a major limitation of these studies was that none of them has sampled multiple populations covering the geographic range of the species. Here, based on sequences from wild populations covering the entire geographical range of the two species, we found higher level of species-wide nucleotide diversity in *O. rufipogon* than in *O. nivara*, and particularly, we showed that both wild species were highly structured with a large proportion of total genetic diversity residing among populations within species. In addition, pairwise F_{ST} analysis indicated that genetic differentiation among populations was significantly higher in *O. nivara* (0.5008) than in *O. rufipogon* (0.3827), consistent with previous studies (Kuroda *et al.* 2007; Zhu *et al.* 2007; Zhou *et al.* 2008).

Notably, we were unable to find significant nucleotide differentiation between the two *Oryza* species at almost all neutral loci sampled. The evidence included (1) phylogenetic analyses showed no support for two major clades corresponding to the two species and that many alleles/haplotypes were shared between species at all the loci; (2) AMOVA partitioning did not reveal significant interspecific genetic differentiation, with a small proportion of total genetic diversity attributed to divergence between species; (3) the interspecific F_{ST} values were not significantly higher than those of the intraspecific F_{ST}. In fact, the lack of interspecific differentiation for the two species was documented at both the species and the regional levels in many previous studies based on allozymes, RFLPs and multilocus sequences (Second 1985; Barbier 1989; Barbier *et al.* 1991; Lu *et al.* 2002; Cai *et al.* 2004; Zhu & Ge 2005; Zhu *et al.* 2007). This lack of genetic differentiation between species could be explained by a combination of lineage sorting of ancestral polymorphism due to recent divergence and historical gene flow (see next section).

It should be mentioned that previous studies of morphological and life-history traits all found significant or clear differentiation between *O. rufipogon* and *O. nivara* (Sano *et al.* 1980; Morishima *et al.* 1984; Barbier 1989; Cai *et al.* 2004). For instance, Barbier (1989) studied the

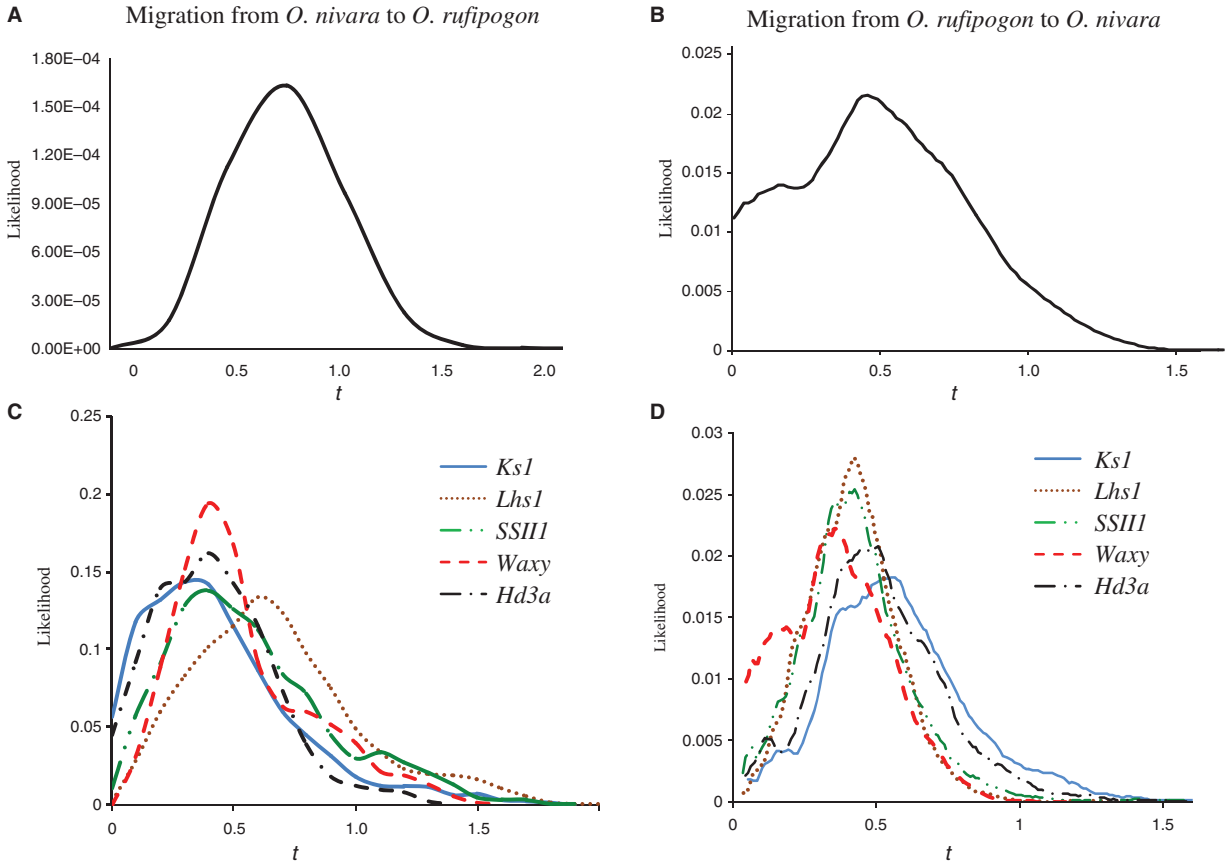

Fig. 3 Distributions of mean time of migration events. Summary distributions for mean time of migration events across all loci are shown in (A) for *O. nivara* to *O. rufipogon* and (B) for *O. rufipogon* to *O. nivara*. Marginal posterior probability distributions by locus are shown in (C) for *O. nivara* to *O. rufipogon* and (D) for *O. rufipogon* to *O. nivara*.

genetic variation and differentiation of four annual (*O. nivara*) and four perennial (*O. rufipogon*) populations in Thailand and found no appreciable interspecific differentiation at enzyme loci but a clear differentiation between the annuals and the perennials in life-history traits such as flowering date, reproductive effort, regenerative ability, number of panicle per plant, etc. Similarly, Cai *et al.* (2004) revealed significant differences for quantitative traits but no differentiation for allozyme and RFLP markers between the perennial (*O. rufipogon*) and annual (*O. nivara*) populations.

Taken together, the present study supports previous arguments (Oka 1988; Barbier 1989; Cheng *et al.* 2003; Zhu *et al.* 2007) for the existence of two ecotypes rather than two distinct species because distinction between *O. rufipogon* and *O. nivara* involves only morphological (mainly life history) traits associated with divergent habitats (Sang & Ge 2007a; Vaughan *et al.* 2008). The contrasting differentiation patterns at life-history traits and genetic markers between the two *Oryza* species raises an interesting question regarding the role of gene flow and divergent selection in the maintenance of the annual and perennial ecotypes in nature.

Recent divergence in face of gene flow

The observation of no molecular divergence but clear differences in life-history traits and habitat preferences between *O. rufipogon* and *O. nivara* is in contrast to a few studies on plant species, in which both substantial morphological and significant molecular divergence were found between ecotypes or species from different habitats (e.g. Remington & Robichaux 2007; Lowry *et al.* 2008). As indicated by some authors (Barbier 1989; Sang & Ge 2007b; Vaughan *et al.* 2008), the lack of genetic differentiation between *O. rufipogon* and *O. nivara* might be a consequence of recent divergence and/or gene flow between species that prevents divergence at the molecular level. In recent years, coalescent-based analyses such as the 'isolation with migration model (IM)' have been successfully used to distinguish between the effects of time since divergence and gene flow on levels

of molecular genetic differentiation (see reviews in Hey 2006; Strasburg & Rieseberg 2010). Our multilocus data sets enable us to estimate important demographic parameters including the scaled effective population sizes, divergence time and migration rate for the two species. In addition to apparent population expansion of the two species after separation from an ancestral population, we estimated that the divergence of the two *Oryza* species occurred approximately 0.12–0.19 million years ago. Therefore, the two *Oryza* species diverged quite recently because the gene pool of *O. rufipogon* started diversifying at ~0.4 million years ago (Zhu & Ge 2005), much earlier than the estimated divergence time of the two species. This result is in accordance with evidence from previous studies based on isozymes and DNA markers (Second 1985; Oka 1988; Barbier *et al.* 1991).

Our IM simulations strongly rejected the simple isolation model of zero migration between species but instead provided unambiguous evidence of bidirectional gene flow between the two species. Note that the estimated migration rate is much higher from *O. rufipogon* to *O. nivara* ($2Nm = 2.52$) than in the opposite direction ($2Nm = 0.49$) (Table 4). This is striking because earlier authors (e.g. Morishima *et al.* 1984) proposed that the major direction of pollen flow was from annual to perennial due to the higher outcrossing rate in the perennial species. This speculation was supported by Zhou *et al.* (2008) who found significant gene flow from *O. nivara* to *O. rufipogon* based on nuclear *Lhs1* sequences. The unexpectedly high level of gene flow from *O. rufipogon* to *O. nivara* might be partly explained by the difference of reproductive strategy in the two species because typical *O. rufipogon* populations reproduce mainly by vegetative propagation (Sang & Ge 2007a; Vaughan *et al.* 2008) and are thus less likely to receive gene flow from *O. nivara*. Another explanation appears to be much higher effective population size for *O. rufipogon* relative to that for *O. nivara*. Nevertheless, the mechanism underlying the direction of gene flow needs further investigation utilizing natural populations and through designed experiments.

The high level of bidirectional gene flow between species might represent the gene flow that has been ongoing or recurrent during the process of species divergence, or alternatively, could be a signature of gene flow that occurred upon secondary contact after a period of allopatric divergence. Our coalescent-based simulations showed that the timing of reconstructed migration events between the species was widely distributed throughout time, suggesting ongoing gene flow during divergence rather than secondary contact following allopatric divergence (Niemiller *et al.* 2008; Faure *et al.* 2009). This argument is also supported by the gene trees on which shared alleles/hyplotypes were spread throughout the tree (Fig. 2). Consequently, divergence with gene flow can be inferred to explain the origin and divergence of *O. rufipogon* and *O. nivara*.

Implications for ecological adaptation

Ecological speciation or adaptation is the process of genetic change by which a population increases its fitness to its environment and thus leads to divergent populations associated with distinct habitats because of local adaptation. It is widely appreciated that ecological divergence due to habitat difference plays an important role in the formation of new species (Schluter 2000; Foster *et al.* 2007; Lowry *et al.* 2008). Regardless of their taxonomical status, *O. nivara* and *O. rufipogon* are ecologically distinct and readily distinguished in nature. The perennial *O. rufipogon* exhibited vigorous vegetative growth, low seed productivity, late flowering and a high outcrossing rate; whereas the annual *O. nivara* showed the opposite characteristics.

Using multiple gene sequences, Zhu & Ge (2005) constructed a phylogeny of the A-genome species, on which samples of *O. rufipogon*, *O. nivara* and cultivated rice (*O. sativa*) formed a strongly supported monophyletic group. Interestingly, *O. rufipogon* accessions were divided into two subclades but all *O. nivara* accessions nested within one of them; this finding was further supported by Grillo *et al.* (2009). Phylogenetic studies based on sequences of the phytochrome gene (Barbier *et al.* 1991) and the interspersion pattern of SINEs (Cheng *et al.* 2003) also found that annual populations (*O. nivara*) clustered with some of the perennial populations (*O. rufipogon*). Therefore, it is most likely that the annual *O. nivara* was recently derived from the perennial *O. rufipogon* and the origin of *O. nivara* was associated with an ecological shift from persistently wet to seasonally dry habitats (Barbier *et al.* 1991; Morishima 2001; Sang & Ge 2007a; Vaughan *et al.* 2008). This hypothesis is supported by multiple lines of evidence including (1) *O. rufipogon* has a much wider geographic distribution than *O. nivara* (Vaughan *et al.* 2008; Fig. 1); (2) numerous life-history traits associated with typical *O. nivara* populations such as annual habit, self-fertilization and photoperiod insensitivity are derived states and these morphological alternations might ensure a high reproductive allocation and maximize the transfer of photosynthetic product to seed production (Sang & Ge 2007a; b; Grillo *et al.* 2009); (3) low levels of species-wide genetic diversity and a small effective population size were found in *O. nivara* relative to *O. rufipogon* (Zhu *et al.* 2007; this study); and (4) the divergence time between the two species is much smaller than the time when *O. rufipogon* populations started to diversify

(Barbier *et al.* 1991; this study). Interestingly, the estimated divergence time between the two species (0.12–0.19 million years ago) coincided largely with recent glaciations when many annual plants evolved from perennials in monsoonal Asia to response to a dry climate (Whyte 1972).

The recent divergence between *O. nivara* and *O. rufipogon* in the presence of significant gene exchange strongly suggests that natural selection has played an active role in the process of speciation. Some authors assumed that spatial isolation and difference in flowering time were the main factors that drove and maintained divergence between the two species (Barbier 1989; Kuroda *et al.* 2005; Vaughan *et al.* 2008). However, the two species could be found at localities a few kilometres or even less than 100 m apart despite difference in their habitats, such as in Thailand, Laos and Nepal (Barbier 1989; Morishima *et al.* 1984; Kuroda *et al.* 2005; S. Ge, personal observation). Therefore, the difference in flowering time between species would be more important, because at least 1-month difference in peak of flowering time was observed in many areas (e.g. Barbier 1989; Kuroda *et al.* 2005; Grillo *et al.* 2009). In plants, the ability to control flowering time by responding to photoperiod is an important adaptive strategy to maximize reproductive success by optimization of the timing of pollination and seed dispersal (Lexer & Widmer 2008; Grillo *et al.* 2009). It is well documented that sister species or even ecotypes in plants often evolve differences in flowering time to adapt to different habitats (e.g. Foster *et al.* 2007; Lexer & Widmer 2008; Lowry *et al.* 2008; Schluter 2009), thus causing temporal isolation between species. Divergent flowering time can promote assortative mating and act as a prezygotic barrier to gene flow (Weis & Kossler 2004; Lowry *et al.* 2008; Grillo *et al.* 2009) and thus might serve as a major mechanism of reproductive isolation between the two *Oryza* species. Recently, Grillo *et al.* (2009) reported a quantitative trait locus (QTL) analysis of phenotypic differentiation between *O. nivara* and *O. rufipogon* and found that two QTL with the largest effect (DF1 and CF1) involved the traits 'days to flowering' and 'continued flowering' (with effect sizes of 36.5 and 30.6, respectively), further supporting the importance of flowering time in the differentiation of the two species. However, the exact number and effects of loci actually responsible for the interspecific differentiation between *O. nivara* and *O. rufipogon* remains to be explored based on rigorous experimental tests. A growing number of empirical investigations demonstrated that speciation driven by ecological separation might require a few loci or even allelic variation in a single gene (Colosimo *et al.* 2005; Remington & Robichaux 2007; Lexer & Widmer 2008; Schluter 2009). It would be particularly interesting, therefore, to determine whether the phenotypic (mainly life history traits) differentiation between the two *Oryza* species is due to variation at a small number of genes with large effect or variation with a more complex genetic basis.

Acknowledgements

We thank Lei Huang, Zhe Cai, Rong Liu, Qi-Hui Zhu, Lin-Bin Zhang, Hai-Fei Zhou, Jie Guo, Bin Ai and other members of Ge's group for technical assistances. We are grateful to Tao Sang (Michigan State University, USA) for valuable comments on this study; to Duncan A. Vaughan (National Institute of Agrobiological Sciences, Japan) for providing the nLKA population; to Jeffrey Ross-Ibarra (University of California at Davis, USA) for valuable comments and smoothing the manuscript; to Chun-Hai Shi and Jian-Guo Wu (Zhejiang University, China) and Xin-Hua Wei, Ping Yu, Cai-Hong Wang and Qun Xu (China National Rice Research Institute, China) for assistance in growing samples in greenhouse. We also thank the International Rice Research Institute (Los Banos, Philippines) for providing seed samples. This work was supported by the National Natural Science Foundation of China (30990242 and 30121003) and the National Basic Research Program of China (2007CB815704).

References

Abbott RJ, Comes HP (2007) Blowin' in the wind—the transition from ecotypeto species. *New Phytologist*, **175**, 197–200.

Arunyawat U, Stephan W, Stadler T (2007) Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. *Molecular Biology and Evolution*, **24**, 2310–2322.

Barbier P (1989) Genetic variation and ecotypic differentiation in the wild rice specis *Oryza rufipogon*. I. population differentiation in life-history traits and isozymic loci. *Japanese Journal of Genetics*, **64**, 259–271.

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

Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation, models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 459–487.

Broughton RE, Harrison RG (2003) Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics*, **163**, 1389–1401.

Cai HW, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theoretical and Applied Genetics*, **104**, 1217–1228.

Cai HW, Wang XK, 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.

Caicedo AL, Williamson SH, Hernandez RD *et al.* (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics*, **3**, e163.

Carstens BC, Knowles L (2007) Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, **16**, 619–627.

Cheng C, Motohashi R, Tsuchimoto S *et al.* (2003) Polyphyletic origin of cultivated rice, based on the interspersion pattern of SINEs. *Molecular Biology and Evolution*, **20**, 67–75.

Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*, **307**, 1928–1933.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 4441–4446.

Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N (2009) Speciation in the deep sea, multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent Mussels. *PLoS ONE*, **4**, e6485.

Foster SA, McKinnon GE, Steane DA, Potts BM, Vaillancourt RE (2007) Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. *New Phytologist*, **175**, 370–380.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Gao LZ, Ge S, Hong DY (2000) Allozyme variation and population genetic structure of common wild rice, *Oryza rufipogon* Griff. in China. *Theoretical and Applied Genetics*, **101**, 494–502.

Gaut BS, Morton BR, McCaig BC, Morton BR (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proceedings of the National Academy of Sciences, USA*, **93**, 10274–10279.

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

Grillo MA, Li C, Fowlkes AM *et al.* (2009) Genetic architecture for the adaptive origin of annual wild rice, *Oryza nivara*. *Evolution*, **63**, 870–883.

Hall TA (1999) Biodit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics and Development*, **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.

Hey J, Nielsen R (2007) Integratoin within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences, USA*, **104**, 2785–2790.

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.

Holsinger KE, Weir BS. (2009) Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature Reviews Genetics*, **10**, 639–650.

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.

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.

Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics*, **169**, 945–953.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–134.

Kuroda Y, Sato YI, Bounphanousay C, Kono Y, Tanaka K (2005) Gene flow from cultivated rice (*Oryza sativa* L.) to wild *Oryza* species (*O. rufipogon* Griff. and *O. nivara* Sharma and Shastry) on the Vientiane plain of Laos. *Euphytica*, **142**, 75–83.

Kuroda Y, Sato YI, 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. *Conservation Genetics*, **8**, 149–158.

Lexer C, Widmer A (2008) The genic view of plant speciation, recent progress and emerging questions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **363**, 3023–3036.

Lowry DB, Rockwood CR, Willis JH (2008) Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution*, **62**, 2196–2214.

Lu BR, Naredo MEB, Juliano AB *et al.* (2000) Preliminary studies on taxonomy and biosystematics of the AA genome *Oryza* species (Poaceae). In: *Grasses, Systematics and Evolution* (eds Jacobs SWL, Everett J), pp. 51–58. CSIRO, Melbourne.

Lu BR, Zheng KL, Qian HR, Zhuang JY (2002) Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. *Theoretical and Applied Genetics*, **106**, 101–106.

Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*, **19**, 472–488.

Mayr E (1963) Animal Species and Evolution. *Belknap Press*, Cambridge, MA.

Moeller DA, Tenaillon MI, Tiffin P (2007) Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. parviglumis). *Genetics*, **176**, 1799–1809.

Morishima H (2001) Evolution and domestication of rice. In: *Rice Genetics IV. Proceedings of the Fourth International Rice Genetics Symposium* (eds Khush GS, Brar DS, Hardy B), pp. 63–77. IRRI, Los Banos (Phillipines).

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

Morjan CL, Rieseberg LH (2004) How species evolve collectively, implications of gene flow and selection for the

spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.

Nadachowska K, Babik W (2009) Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). *Molecular Biology and Evolution*, **26**, 829–841.

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

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae, Gyrinophilus) inferred from gene genealogies. *Molecular Ecology*, **17**, 2258–2275.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

Oka HI (1988) *Origin of Cultivated Rice*. Japan Scientific Societies Press, Tokyo.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Rakshit S, Rakshit A, Matsumura H *et al*. (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. *Theoretical and Applied Genetics*, **114**, 731–743.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Remington DL, Robichaux RH (2007) Influences of gene flow on adaptive speciation in the *Dubautia arborea–D. ciliolata* complex. *Molecular Ecology*, **16**, 4014–4027.

Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics*, **181**, 1399–1413.

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 (2007a) The puzzle of rice domestication. *Journal of Integrative Plant Biology*, **49**, 760–768.

Sang T, Ge S (2007b) Genetics and phylogenetics of rice domestication. *Current Opinion in Genetics and Developments*, **17**, 533–538.

Sano Y, Morishima H (1982) Variation in resource allocation and adaptive strategy of a wild rice, *Oryza perennis* Moench. *Botanical Gazette*, **143**, 518–523.

Sano Y, Morishima H, Oka HI (1980) Intermediate perennial-annual populations of *Oryza perennis* found in Thailand and their evolutionary significance. *Botanical Magazine Tokyo*, **93**, 291–305.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press Inc., New York.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Schmid KJ, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B, Mitchell-Olds T (2005) A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics*, **169**, 1601–1615.

Second G (1985) Evolutionary relationships in the Sativa group of *Oryza* based on isozyme data. *Genetics Selection Evolution*, **17**, 89–114.

Sharma SD, Shastry SVS (1965) Taxonomic studies in genus *Oryza* L. III. O. rufipogon Griff. sensu stricto and O. nivara Sharma et Shastry nom. nov. *Indian Journal of Genetics and Plant Breeding*, **25**, 157–167.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Stadler T, Arunyawat U, Stephan W (2008) Population genetics of speciation in two closely related wild tomatoes (Solanum Section Lycopersicon). *Genetics*, **178**, 339–350.

Strasburg JL, Rieseberg LH (2008) Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*—large effective population sizes and rates of long-term gene flow. *Evolution*, **62**, 1936–1950.

Strasburg JL, Rieseberg LH (2010) How robust are 'Isolation with Migration' analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution*, **27**, 297–310.

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. *Theoretical and Applied Genetics*, **102**, 157–162.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0*. Sinauer Associates, Sunderland, MA.

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

Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto k (2009) Variations in *Hd1* proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. *Proceedings of the National Academy of Sciences, USA*, **106**, 4555–4560.

Tang L, Zou X, Achoundong G *et al*. (2010) Phylogeny and biogeography of the rice tribe (Oryzeae): evidence from combined analysis of 20 chloroplast fragments. *Molecular Phylogenetics and Evolution*, **54**, 266–277.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. *Molecular Biology and Evolution*, **21**, 1214–1225.

Vaughan DA (1989) *The Genus Oryza L., Current Status of Taxonomy*. International Rice Research Institute, Manila, Philippines.

Vaughan DA, Morishima H (2003) Biosystematics of the genus *Oryza*. In: *Rice, Origin, History, Technology, and Production* (eds Smith CW, Dilday RH), pp 27–65. John Wiley & Sons, Inc., New Jersey).

Vaughan DA, Lu BR, Tomooka N (2008) The evolving story of rice evolution. *Plant Science*, **174**, 394–408.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Weis A, Kossler T (2004) Genetic variation in flowering time induces phonological assortative mating: quantitative genetic methods applied to *Brassica rapa*. *American Journal of Botany*, **91**, 825–836.

Whyte RO (1972) The gramineae, wild and cultivated, of monsoonal and equatorial Asia. *Asian Perspectives*, **15**, 127–151.

Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.

Zhang LB, Ge S (2007) Multilocus analysis of nucleotide variation and speciation in *Oryza officinalis* and its close relatives. *Molecular Biology and Evolution*, **24**, 769–783.

Zhou HF, Xie ZW, Ge S (2003) Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon* Griff.) in China. *Theoretical and Applied Genetics*, **107**, 332–339.

Zhou HF, Zheng XM, Wei RX *et al.* (2008) Contrasting population genetic structure and gene flow between *Oryza rufipogon* and *Oryza nivara*. *Theoretical and Applied Genetics*, **117**, 1181–1189.

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

Zhu QH, Zheng XM, Luo JC, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives, severe bottleneck during domestication of rice. *Molecular Biology and Evolution*, **24**, 875–888.

X.-M. Zheng is a PhD student working on population genetics, molecular adaptation and speciation of wild rice species. This study represents part of her PhD dissertation. S. Ge is a research professor at the State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences. He has diverse interests in plant phylogenetic reconstruction, evolutionary genomics and conservation genetics. He is particularly interested in phylogeny, speciation and molecular adaptation in rice genus and its relatives.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 List of the populations sampled in this study

Table S2 Summary of the genes sequenced and the primer sequences used in this study

Table S3 Neutrality tests for each population of the two species

Table S4 Summary of the Shimodaira–Hasegawa (SH) tests comparing relaxed trees with constraint trees

Table S5 Maximum-Likelihood Estimates (MLE) and the 90% Highest Posterior Density (HPD) intervals of demographic parameters from IMa analyses except for *Waxy*

Fig. S1 Schematic diagrams of five nuclear loci and locations of the regions sequenced. Exons are shown as open boxes, and exon numbers are labelled with capital roman numbers. Thin lines between open boxes refer to introns. Locations of primers for each fragment are sketched above the diagrams.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.