Multilocus Analysis of Nucleotide Variation and Speciation in *Oryza officinalis* and Its Close Relatives

Lin-Bin Zhang and Song Ge*†*

*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing, China; and †The Graduate School, Chinese Academy of Sciences, Beijing, China

Nucleotide variation in 10 unlinked nuclear genes was investigated in species-wide samples of *Oryza officinalis* and its close relatives (*Oryza eichingeri* and *Oryza rhizomatis*). Average estimates of nucleotide diversity were the lowest in *O. rhizomatis* ($\theta_{sil} = 0.0038$) and the highest in *O. eichingeri* ($\theta_{sil} = 0.0057$) that is disjunctly distributed in Africa and Sri Lanka. These wild rice species appeared to harbor relatively low levels of nucleotide variation relative to other plant species because the diversity level of *O. eichingeri* is only 23–46% of those in *Zea* species and 35% of that in *Arabidopsis thaliana*. The lower nucleotide diversity in these *Oryza* species could be best explained by their smaller historic effective population sizes. The speciation model test indicated that *O. officinalis* and its close relatives might have undergone a process of population contraction since divergence from their ancestor. Incongruent topologies among 10 gene trees, particularly regarding the positions of *O. eichingeri* and *O. rhizomatis* accessions might be attributed to lineage sorting arising from ancient polymorphism and hybridization/introgression between the Sri Lankan *O. eichingeri* and *O. rhizomatis*. However, the null hypothesis of the isolation model was not rejected for any contrast between taxa, which suggested that no subsequent gene flow shaped the present patterns of nucleotide variation since their divergence and that introgression was not pervasive in this group of species. Our molecular dating provides an approximate divergence time of 0.37 Myr between 2 geographical races of *O. eichingeri*, much more recent compared with the times of other speciation events in this group (0.63–0.68 Myr). A long-distance dispersal from West Africa to Sri Lanka was more likely to play a role in the disjunct distribution of *O. eichingeri*.

Introduction

Inference of recent evolutionary history of closely related species is one of the most intricate questions for evolutionary biologists. The level and pattern of nucleotide variation in DNA sequences provide important information on the evolutionary history of a species and divergent process of closely related species. In recent decades, molecular population genetics and genealogical approaches have been successfully used to reveal the patterns of genetic diversity within and between populations and to trace the histories of divergence and speciation in plants (Eyre-Walker et al. 1998; Hilton and Gaut 1998; Savolainen et al. 2000; Tiffin and Gaut 2001; Olsen and Purugganan 2002; Ramos-Onsins et al. 2004; Wright and Gaut 2005). Genetic information recorded in multilocus DNA sequences exceptionally benefits us to explore the various forces (mutation, migration, selection, and random drift) in evolutionary history (Nordborg and Innan 2002; Wright and Gaut 2005). By analyzing multilocus nucleotide diversity in closely related species, we could detect both the polymorphisms accumulated independently in each species since their divergence and the variation segregated originally in their common ancestor (Wang et al. 1997; Kliman et al. 2000; Machado et al. 2002; Ramos-Onsins et al. 2004). Given natural selection generally acting on some but not all genes, it is possible to differentiate the effects of natural selection and demography (Nordborg and Innan 2002; Wright and Gaut 2005). The molecular population genetics approaches have been proved to be especially useful to study the history of species by revealing the polymorphism patterns at randomly selected genes, as demonstrated by many cases in *Drosophila* species (Wang et al. 1997; Kliman et al. 2000; Machado et al. 2002; Hey and Nielsen 2004), primates (Yu et al. 2004; Won and Hey 2005), and humans (Akey et al. 2004; Enard and Pääbo 2004). Although studies on nucleotide variation of plant species have been conducted mainly focusing on the model plant *Arabidopsis* and several crops (see review in Wright and Gaut 2005), relatively few investigations have been conducted on the speciation and divergence of closely related species in higher plants with molecular population genetics methods (Ramos-Onsins et al. 2004; Städler et al. 2005).

In the rice genus (*Oryza* L.), 10 genome groups (i.e., the A-, B-, C-, BC-, CD-, E-, F-, G-, HJ-, and HK-genomes) have been recognized (Ge et al. 1999; Khush and Brar 2001), including the A-genome group that the cultivated rice (*Oryza sativa*) belongs to. The C-genome group, a well-defined monophyletic clade (Ge et al. 1999), includes 3 closely related diploid species, that is, *Oryza officinalis* Wall. ex Watt., *Oryza eichingeri* Peter, and *Oryza rhizomatis* Vaughan (Tateoka 1965; Vaughan 1989). *Oryza officinalis* is the most common species and distributed widely in southern China, South and Southeast Asia, and Papua New Guinea, whereas *O. rhizomatis* has only been reported from Sri Lanka. The third species, *O. eichingeri*, is distributed in Sri Lanka and West and East Africa and is the only wild *Oryza* species reported from both Asia and Africa. It is intriguing that the Sri Lankan *O. eichinger* is sympatric to *O. rhizomatis* with their population being overlapping in both northern and southern Sri Lanka, though their habitats are distinctly different (Bautista et al. 2006). Phylogenetic and population genetic studies showed that these 3 C-genome species have diverged recently with low level of species differentiation (Ge et al. 1999; Bao and Ge 2003; Bao et al. 2006; Bautista et al. 2006). Therefore, *O. officinalis* and its close relatives provide an ideal system to explore demographic history and speciation processes in plants. In this study, we investigate the patterns of nucleotide variation in 10 unlinked nuclear loci in species-wide samples of the 3 C-genome species of

Key words: nucleotide variation, divergence, speciation, *Oryza officinalis*.

E-mail: gesong@ibcas.ac.cn.

Mol. Biol. Evol. 24(3):769–783. 2007
doi:10.1093/molbev/msl204
Advance Access publication December 20, 2006

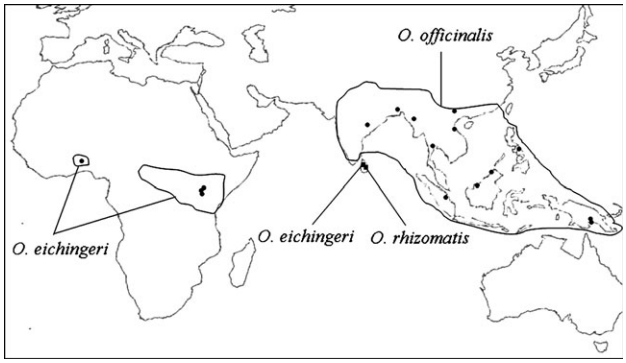

FIG. 1.—Geographical distribution of *Oryza officinalis* and its close relatives. Dots indicate the sampled accessions.

Oryza. We aim to use multiple genealogies and population parameters to explore whether they have remained isolated since their divergence and to address the demographic and geographic aspects of their speciation history.

This study also seeks to qualify species-wide levels of nucleotide diversity of *O. officinalis* and its close relatives and compare the result with those from previous studies using different markers. Although a few studies have been undertaken on the genetic diversity of the C-genome species (Aggarwal et al. 1999; Gao et al. 2001; Gao 2005; Gao and Zhang 2005; Bautista et al. 2006), inconsistent results have been obtained probably due to the different samples and molecular markers used. For example, using amplified fragment length polymorphism markers, Aggarwal et al. (1999) and Bao et al. (2006) inferred that *O. rhizomatis* harbored the lowest genetic variation among 3 C-genome species, whereas Bautista et al. (2006) found that *O. eichingeri* maintained lower genetic diversity than the other 2 species. To date, a few studies on nucleotide variation have been undertaken on the cultivated rice (*O. sativa*) and its wild relative *Oryza rufipogon* (Olsen and Purugganan 2002; Garris et al. 2003; Yoshida and Miyashita 2005; Olsen et al. 2006). These investigations, however, mainly focused on a single species and were exclusively based on 1 or 2 genes or multiple linked genes. The present study is the first attempt to investigate the nucleotide polymorphism and divergence among the wild *Oryza* species using multilocus sequence data of unlinked genes. Such information will facilitate the effective use of the wild rice germplasm because the wild species in *Oryza* possess abundant genes valuable for rice breeding and improvement, such as resistance to diseases and insects and stress tolerances (Khush and Brar 2001; Vaughan et al. 2003).

Materials and Methods

Species Sampling

The geographic distribution of 3 C-genome species is shown in figure 1. The 3 species are largely allopatric across the pantropical Old World, but *O. eichingeri* is sympatric to *O. rhizomatis* in Sri Lanka. The identity and geographic origin of the individuals sampled for each species are presented in table 1 and figure 1. Twelve *O. officinalis* individuals were collected from 11 countries, covering the entire distribution range of the species. Four *O. rhizomatis* individuals were sampled to represent the species that is found only in Sri Lanka. Because our previous studies found high level of genetic divergence between the African and Sri Lankan populations of *O. eichingeri* (Bao and Ge 2003; Bao et al. 2006), we treated this species as 2 geographic races and sampled 4 individuals from each of them. Additionally, 3 accessions of the diploid *Oryza punctata*, a B-genome species, were sampled as outgroups because previous studies showed that the B-genome species was closely related to the C-genome group (Ge et al. 1999). Seed germination and seedling cultivation followed the description in Bao and Ge (2004). Total genomic DNA was extracted from fresh young leaves or silica gel–dried leaves, using the hexadecyltrimethylammonium bromide method as previously described in Ge et al. (1999).

Table 1
Accessions and Geographic Origin of the Individuals Sampled

Species	Accession[a]	Source	Code[b]
Oryza officinalis	101412	India	off-IND
	102460	Bangladesh	off-BGD
	105081	Myanmar	off-MMR
	7904	Yunnan, China	off-CHN
	81972	Thailand	off-THA
	105080	Vietnam	off-VNM
	105093	Malaysia	off-MYS
	81796	Indonesia	off-IDN
	105100	Brunei	off-BRN
	105085	Philippines	off-PHL
	106519	Papua New Guinea	off-PNG1
	106522	Papua New Guinea	off-PNG2
Oryza eichingeri-LKA	81803	Sri Lanka	eic-LKA1
	105407	Sri Lanka	eic-LKA2
	105413	Sri Lanka	eic-LKA3
	105415	Sri Lanka	eic-LKA4
O. eichingeri-AFR[c]	101425	Uganda	eic-UGA1
	105159	Uganda	eic-UGA2
	105162	Uganda	eic-UGA3
	IP7	Cote d'Ivoire	eic-CIV
Oryza rhizomatis	103410	Sri Lanka	rhi-LKA1
	103421	Sri Lanka	rhi-LKA2
	105448	Sri Lanka	rhi-LKA3
	105950	Sri Lanka	rhi-LKA4
Oryza punctata	103903	Tanzania	pun-TZA
	104067	Chad	pun-TCD
	105984	Cameroon	pun-CMR

[a] All accessions were obtained from leaf materials or seeds provided by the Genetic Resources Center of the International Rice Research Institute at Los Banos, Philippines, except for 7904 that was collected by the authors and IP7 that was provided by Dr G. Second (France).

[b] Individual accession is abbreviated by the first 3 letters of the species name followed by the code of its origin of country, such as LKA referring to Sri Lanka.

[c] AFR, Africa.

Sampled Loci

DNA sequences were obtained for 10 nuclear loci that are located on 10 different chromosomes in rice (*O. sativa*) (table S1, Supplementary Material online). *Adh1* gene encodes alcohol dehydrogenase I (alcohol nicotinamide adenine dinucleotide^{+}: oxidoreductase, EC 1.1.1.1), an important protein in the process of anaerobic metabolism. It is a single copy in the *Oryza* species and located in the short arm of chromosome 11 in rice (Tarchini et al. 2000). As a single copy located in chromosome 5 in rice, *GPA1* encodes a G protein α subunit that functions in various systems

of signal transduction in diverse tissues or cells in flowering plants (Seo et al. 1995). *Leafy hull sterile 1* (*Lhs1*), located in chromosome 3 in rice, is a MADS-box transcription factor and plays an essential role in determining floral meristem identity and in floral organ development (Jeon et al. 2000). Serine carboxypeptidase I (*CBP1*) plays an important role during development and following germination of cereal grains (Washio and Ishikawa 1994). Both *ent*-copalyl diphosphate synthases I (*CPS1*) and *ent*-kaurene synthase I (*Ks1*) are central catalyzing enzymes in the early steps of the gibberellin biosynthetic pathway (Sakamoto et al. 2004). Starch synthase II1 (*SSII1*) and granule-bound starch synthase (*Waxy-GBSS*) are genes related to starch biosynthesis, whereas granule-binding starch synthase II (*GBSSII*; ADP-glucose–starch glucosyltransferase) is a gene involved in synthesis of amylose in rice leaves (Dian et al. 2003). Gamma subunit of transcription factor II A (*TIFFA*γ-1) is one of the recessively inherited resistance genes that provide race-specific resistance to bacterial blight (Blair et al. 2003). Amplifying primers for these loci were designed based on the sequences from rice, maize, sorghum, and/or oat, and the amplified regions spanned 0.7–1.4 kb in length. The amplification primers and the sequenced regions of the 10 loci are shown in table S1 and figure S2, Supplementary Material online.

Polymerase Chain Reaction Amplification, Cloning, and Sequencing

All polymerase chain reaction (PCR) amplifications were performed in a total volume of 25 μl on a Tpersonal thermocycler (Biometra, Germany), using 10–30 ng genomic DNA. The reaction mixture was supplemented with 0.2 μM of each primer, 200 μM of each deoxyribonucleotide triphosphate, 10 mM Tris-Cl (pH = 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, and 0.75 U ex*Taq* DNA polymerase (TaKaRa). To reduce recombinant molecule during PCR (Judo et al. 1998; Shammas et al. 2001), long extension time during PCR reactions was used. For instance, a 3-min extension was used during each amplification cycling for ~1.4-kb *Adh1*, a 2.5-min extension for ~1.3-kb *GPA1*, and a 2-min extension for ~1.0-kb *Lhs1*. Amplified products were ligated into *pGEM* T-easy vectors (Promega, Madison, WI) after being purified from agrose gel with either a Pharmacia purification kit (Amersham Pharmacia Biotech, Piscataway, NJ) or a Dingguo purification kit (Dingguo, Beijing, China). Independent plasmid DNAs were selected randomly and isolated by the method of alkaline lysis plasmid miniprep as described (Ausubel 1992). Sequencing reactions were performed by a MegaBACE 1000 automated sequencer (Amersham Pharmacia Biotech) or an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA).

Although both outcrossing and inbreeding species have been recorded for the wild *Oryza* species (Vaughan 1989; Dally and Second 1990), mating system of the C-genome species is largely unclear. Therefore, individuals in these species can be either homozygous or heterozygous at nuclear loci. Thus all amplification products of 10 loci from each individual were cloned, and multiple clones were sequenced. To obtain both alleles from the heterozygous samples, we adopted the partial sequencing strategy (Tiffin and Gaut 2001) to sequence 8 to 10 clones at each locus. The advantage of the multiclone sequencing is that pseudo-polymorphisms induced by *Taq* polymerase or recombination during PCR amplification could be avoided in the final sequence data set, although this method is relatively laborious and high costing (Palumbi and Baker 1994; Tiffin and Gaut 2001; Clark et al. 2004).

Because *Taq* errors occur at random, it is unlikely that polymorphisms shared among more than one clones (sequences) are artifactual (Palumbi and Baker 1994; Eyre-Walker et al. 1998; Hilton and Gaut 1998). However, "singletons," that is, polymorphisms occurred in only 1 sequence relative to all the remainder sequences, can represent either true sequence variation or *Taq* polymerase artifact. Previous studies found that the percentage of singletons resulting from *Taq* polymerase error ranged from 29% (Hilton and Gaut 1998) to 100% (Small et al. 1999; White and Doebley 1999), depending on the genes and taxa. Because we sequenced more than 8 clones for each individual, most alleles could be easily determined and some of the artificial singletons were removed from the original data set. To confirm the remaining singletons, we performed repeated PCR amplification, cloning, and sequencing and found that 63% (51 out of 81) of the singletons resulted from *Taq* polymerase error and the corrected sequences were used in the analyses. By means of multiclone sequencing and reamplifying and resequencing, interallelic PCR recombinants were also verified and removed. Therefore, accuracy and reliability of nucleotide polymorphisms in this study are sufficiently guaranteed for subsequent analyses.

All allele sequences have been deposited in GenBank, and their accession numbers are DQ223326–DQ223418, DQ901744–DQ901953, and DQ911245–DQ911249.

Sequence Analysis

Sequence data were edited and assembled with the ContigExpress program from the Vector NTI Suite 6.0 (Informax Inc., North Bethesda, MD). Allele sequences for each locus were aligned using a combination of methods implemented in DAMBE version 4.1.19 (Xia and Xie 2001) and ClustalX version 1.81 (Thompson et al. 1997), with additional manual refinements. Levels of intraspecific genetic variation were calculated with estimates of average pairwise differences per basepair between sequences (π) (Nei and Li 1979) and Watterson's estimates (θ_w) from S (Watterson 1975) using both DnaSP version 4.10 (Rozas et al. 2003) and SITES (Hey and Wakeley 1997), where S is the number of segregating sites. With the assumption of the standard neutral model of a random-mating population of constant size, the statistic estimate of nucleotide variation θ_w in an autosomal gene is equal to $4N_e\mu$, where N_e is the effective population size and μ, the mutation rate per generation per site. The minimum number of recombination events was assessed using the algorithm of Hudson and Kaplan (1985) in the SITES program.

Under the null assumption that molecular variation is evolving neutrally, a number of statistical tests have been used to assess whether selective forces exert influences on patterns of genetic variation. In this study, deviation from

standard neutral equilibrium model was tested based on both the frequency spectrum of polymorphisms or the haplotype distribution and the relationship between intraspecific and interspecific diversity. If the hypothesis of the neutrality is not rejected based on single statistical test, we still could not determine if the locus is evolving neutrally because failure of rejection could be simply due to the fact that the test is not sensitive enough to detect certain type of selective force (Wayne and Simonsen 1998). Therefore, multiple statistic tests for individual locus (Tajima 1989; Fu and Li 1993), as well as multilocus tests (Hudson et al. 1987), were performed to determine the departure from the neutrality hypothesis and make inferences on the species history using the program DnaSP (Rozas et al. 2003).

Tajima's *D* (Tajima 1989) was based on the discrepancy between the mean pairwise differences (π) and Watterson's estimator (θ_w), whereas *D** and *F** of 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). Because selective force is generally considered to affect a particular locus in evolutionary history, the multilocus HKA test across unlinked or loosely linked loci was performed using the program HKA to discriminate between selection forces and population demography during the speciation process. For the HKA tests, *O. punctata* sequences were used as outgroups. The SITES, HKA, and WH (mentioned below) software packages were distributed kindly by Jody Hey (http://lifesci.rutgers.edu/~heylab). Insertion/deletion polymorphisms were excluded from the analyses.

The genealogical trees of 10 nuclear loci were constructed using the parsimony and distance methods as implemented in PAUP* version 4.0b10 (Swofford 2002). The Neighbor-Joining (NJ) method (Saitou and Nei 1987) was performed with Kimura's 2-parameter distances (Kimura 1980). Maximum parsimony (MP) analyses were performed using heuristic search with MULPARS, Tree Bisection-Reconnection branch swapping, and RANDOM stepwise addition with 1,000 replicates. Topological confidence was assessed by bootstrap analysis with 1,000 replicates. Furthermore, we estimated the divergence time of the C-genome species based on molecular clock hypothesis. To examine rate heterogeneity among lineages, we used the program MEGA version 3.0 (Kumar et al. 2004) to assess the constancy of molecular evolution across individual lineages.

Ancestral Parameter Estimates

The simple speciation model was fitted based on the different classes of mutations from the multilocus sequence comparisons between 2 species using the program WH (Wakeley and Hey 1997; Wang et al. 1997; Kliman et al. 2000). Three classes of nucleotide mutations for the WH program, including polymorphisms that are exclusive to 1 species, shared polymorphisms between 2 species, and fixed differences between 2 species, could be obtained from the SITES program. Under the assumption of 2 descendent populations (species) separated from an ancestral population (species) with constant population sizes and no gene flow between the populations (species) after their separation, the model presented estimates of population parameters in the ancestral and descendant populations (species) (θ_A, θ_1, θ_2) as well as the time since separation T (scaled in $2N_1$ generations). A rejection of the model may indicate that gene flow has occurred between a species pair after the time of divergence. Both a simple measure (the difference between the highest and lowest numbers of shared polymorphisms plus the difference between highest and lowest numbers of fixed differences among multiple loci) (Wang et al. 1997; Machado et al. 2002) and a χ^2 statistic (Kliman et al. 2000) have been used to assess the overall fit to the simple speciation model.

Results

Nucleotide Variation

Forty-eight sequences were obtained for each of the 10 loci, with 2 sequences per individual. Total length of the aligned sequences for the 10 genes is 9,916 bp, including 3,321 bp of coding sequence and 6,585 bp of noncoding sequence (table S1, Supplementary Material online). The number of insertion–deletion polymorphisms ranged from 0 to 12 across loci, with a total of 37 indel polymorphisms for the 4 taxa. A detailed examination of these indel polymorphisms showed that all of them occurred in noncoding regions, with 26 (~70%) being 1-bp indels. The remaining indels included four 2-bp, five 3-bp, one 15-bp, and one 39-bp polymorphisms. All indels were not considered in subsequent analyses. The schematic diagrams and the nucleotide polymorphisms in the sequenced regions of 10 genes are provided in Supplementary Materials online (figs. S1 and S2, Supplementary Material online).

Standard statistics of sequence variation for each locus are summarized in table 2, including the estimates of nucleotide variation in different regions at individual loci. As expected, due to strong functional constraint, the levels of nucleotide variation at coding regions were lower than those at noncoding regions at all 10 loci except for 6 cases involving 3 loci (*Lhs1*, *Ks1*, and *SSIII*), where no silent substitution was observed. Levels of polymorphisms varied across loci, with *CBP1*, *GPA1*, and *Ks1* being the least variable genes in *O. eichingeri*, *O. officinalis*, and *O. rhizomatis*, respectively. At the species level, the average estimates of variation over 10 loci were comparable for all the taxa although *O. officinalis* has much wider distribution than the other 3. The π_{sil} ranged from 0.0033 (Sri Lankan *O. eichingeri*) to 0.0044 (*O. officinalis*) and θ_{sil} ranged from 0.0038 (*O. rhizomatis*) to 0.0042 (Sri Lankan *O. eichingeri*) (table 2). For *O. eichingeri*, the Sri Lankan race (π_{sil} = 0.0033; θ_{sil} = 0.0042) possessed almost similar level of nucleotide diversity with the African race (π_{sil} = 0.0040; θ_{sil} = 0.0039), but the value in the Sri Lankan race was much lower (θ_{sil} = 0.0013; π_{sil} = 0.0011) if one introgressed individual was excluded (see Discussion). It is noted that the diversity values of *O. eichingeri* would be increased (θ_{sil} = 0.0057; π_{sil} = 0.0052) if both races were combined into a single data set, indicative of the impact of population subdivision on genetic diversity.

Table 2
Estimates of Nucleotide Variation

Species	Locus	N^a	L_T^b	S^c	π_T^d	θ_T^e	π_{sil}^f	θ_{sil}^g	K_{sil}^h	D^i	$D^{j,*}$	$F^{k,*}$	R_M^l
Oryza officinalis	*Adh1*	24	1,427.1	12	0.0026	0.0023	0.0048	0.0042	0.1028	0.4839	1.4645*	1.3650	0
	CBP1	24	816.3	22	0.0117	0.0072	0.0149	0.0094	0.0137	2.2761*	1.6244**	2.1332**	6
	CPS1	24	717.9	7	0.0026	0.0026	0.0027	0.0027	0.0503	−0.0324	1.2873	1.0477	0
	GBSSII	24	798.3	8	0.0031	0.0027	0.0040	0.0035	0.0246	0.4814	1.3336	1.2592	1
	GPA1	24	1,373.3	2	0.0003	0.0004	0.0005	0.0006	0.0431	−0.3543	0.8373	0.5856	0
	Ks1	24	922.0	17	0.0044	0.0049	0.0063	0.0066	0.1428	−0.4094	0.5756	0.3232	4
	Lhs1	24	1,179.2	4	0.0010	0.0009	0.0012	0.0012	0.0707	0.3734	1.0844	1.0206	0
	SSII1	24	937.0	10	0.0023	0.0029	0.0025	0.0032	0.0376	−0.7092	0.9039	0.4951	1
	TFIIAγ-1	24	946.0	12	0.0034	0.0034	0.0036	0.0036	0.0658	0.0197	0.1549	0.1329	0
	Waxy	24	701.0	10	0.0028	0.0038	0.0038	0.0048	0.0039	−0.9123	−0.1036	−0.4000	0
	Average				0.0034	0.0031	0.0044	0.0040	0.0555				
Oryza eichingeri-LKA	*Adh1*	8	1,426.1	17	0.0039	0.0046	0.0071	0.0085	0.0963	−0.8312	−0.5257	−0.6636	0
	CBP1	8	817.0	0	0.0000	0.0000	0.0000	0.0000	0.0018	—	—	—	0
	CPS1	8	718.0	2	0.0012	0.0011	0.0012	0.0011	0.0486	0.4142	1.1112	1.0379	0
	GBSSII	8	800.1	7	0.0028	0.0034	0.0036	0.0044	0.0241	−0.7927	−0.8937	−0.9616	0
	GPA1	8	1,372.5	5	0.0013	0.0014	0.0020	0.0021	0.0464	−0.3355	0.1265	0.0214	0
	Ks1	8	922.0	9	0.0024	0.0038	0.0031	0.0048	0.1372	−1.7232*	−1.8850*	−2.0454**	0
	Lhs1	8	1,175.6	15	0.0032	0.0050	0.0040	0.0061	0.0703	−1.8000*	−1.9760**	−2.1494**	0
	SSII1	8	937.0	10	0.0046	0.0041	0.0046	0.0042	0.0323	0.5506	0.9039	1.4194	0
	TFIIAγ-1	8	947.0	9	0.0026	0.0037	0.0027	0.0039	0.0660	−1.4712	−1.5105	−1.6623	0
	Waxy	8	700.9	8	0.0029	0.0044	0.0045	0.0069	0.0186	−1.7012*	−1.8590*	−2.0159*	0
	Average				0.0025	0.0032	0.0033	0.0042	0.0542				
O. eichingeri-AFR	*Adh1*	8	1,427.1	31	0.0078	0.0084	0.0133	0.0140	0.0972	−0.3768	−0.3957	−0.4360	0
	CBP1	8	816.8	1	0.0005	0.0005	0.0008	0.0007	0.0004	0.3335	0.8878	0.8253	0
	CPS1	8	739.0	5	0.0030	0.0027	0.0032	0.0028	0.0472	0.5884	0.1265	0.2753	0
	GBSSII	8	800.8	9	0.0049	0.0043	0.0062	0.0056	0.0237	0.5448	1.4857*	1.4026	0
	GPA1	8	1,372.5	6	0.0020	0.0017	0.0030	0.0026	0.0443	0.8062	1.4086	1.4002	0
	Ks1	8	922.0	8	0.0035	0.0034	0.0038	0.0034	0.1332	0.2580	1.0497	0.9575	0
	Lhs1	8	1,178.0	3	0.0012	0.0010	0.0009	0.0009	0.0681	0.8387	0.3007	0.4656	0
	SSII1	8	739.3	6	0.0032	0.0025	0.0029	0.0023	0.0308	1.3818	1.4087	1.5476	0
	TFIIAγ-1	8	947.0	4	0.0018	0.0016	0.0019	0.0017	0.0635	0.4852	1.3125	1.2325	0
	Waxy	8	701.0	6	0.0027	0.0033	0.0041	0.0051	0.0180	−0.9203	−0.7219	−0.8475	0
	Average				0.0031	0.0029	0.0040	0.0039	0.0526				
Oryza rhizomatis	*Adh1*	8	1,429.0	24	0.0052	0.0065	0.0092	0.0115	0.1062	−1.0777	−0.9729	−1.1103	0
	CBP1	8	815.6	7	0.0024	0.0033	0.0022	0.0034	0.0135	−1.3592	−1.3604	−1.5030	0
	CPS1	8	718.0	10	0.0066	0.0054	0.0069	0.0056	0.0524	1.1466	0.8209	0.9936	0
	GBSSII	8	800.7	6	0.0021	0.0029	0.0027	0.0037	0.0225	−1.2801	−1.2545	−1.3910	0
	GPA1	8	1,371.5	5	0.0012	0.0014	0.0018	0.0021	0.0442	−0.7554	−0.4941	−0.6110	0
	Ks1	8	922.0	4	0.0014	0.0017	0.0004	0.0007	0.1426	−0.6257	−0.9208	−0.9372	0
	Lhs1	8	1,160.8	7	0.0034	0.0023	0.0045	0.0031	0.0700	2.2290*	1.4401*	1.7939**	0
	SSII1	8	937.0	4	0.0016	0.0017	0.0018	0.0018	0.0337	−0.0198	0.5681	0.4750	0
	TFIIAγ-1	8	946.4	8	0.0043	0.0033	0.0045	0.0034	0.0667	1.5454	1.4652*	1.6415*	0
	Waxy	8	700.0	3	0.0018	0.0017	0.0029	0.0026	0.0243	0.4577	1.2338	1.1560	0
	Average				0.0030	0.0030	0.0037	0.0038	0.0576				

[a] Total number of sequences.
[b] Average length (bp) of the sequences from each species.
[c] Total number of polymorphic sites.
[d] Average number of pairwise nucleotide differences per site (Nei 1987) calculated on the total number of polymorphic sites.
[e] Watterson's estimator of θ per basepair (Watterson 1975) calculated on the total number of polymorphic sites.
[f] Average number of pairwise nucleotide differences per site calculated on the silent sites.
[g] Watterson's estimator of θ per basepair calculated on the silent sites.
[h] Average divergence per silent site (with the Jukes and Cantor [1969] correction) between the sample sequences and the alleles of *Oryza punctata*.
[i] Tajima's *D* (Tajima 1989).
[j] *D** of Fu and Li (1993).
[k] *F** of Fu and Li (1993).
[l] The minimum number of recombination events (Hudson and Kaplan 1985).
* $0.01 < P \leq 0.05$; ** $0.001 < P \leq 0.01$; *** $P \leq 0.001$.

Because the amount of recombination influences the coalescent simulations of isolation species model and the phylogenetic inferences (Wang et al. 1997; Kliman et al. 2000), we estimated the recombination rate at all loci. Based on the algorithm of Hudson and Kaplan (1985), the minimum number of historical recombination events (*R*m) at each locus was estimated with 4-gamete test for each species. Recombination events in our data sets were low and observed only for *O. officinalis*, with a minimum 6 at *CBP1* locus, 4 at *Ks1*, and 1 each at *GBSSII* and *SSII1* (table 2).

Tests of Neutrality

To test the standard neutral equilibrium model, we first performed the tests of Tajima's *D* (Tajima 1989), and *D**

Table 3
Summary of Multilocus Neutrality Tests

Species	N[a]	Tajima's D^{b}_{all}	Tajima's D^{c}_{sil}	D_{all} of Fu and Li[d]	F^{*}_{all} of Fu and Li[e]	HKA^{f}_{all}	HKA^{g}_{sil}
Oryza officinalis	24	−0.068	−0.070	−0.064	−0.064	85.966***	81.851***
Oryza eichingeri-LKA	8	−0.050**	−0.050**	−0.084**	−0.086	32.291***	33.079*
O. eichingeri-AFR	8	−0.050	−0.048	−0.064	−0.086	39.493**	42.210***
Oryza rhizomatis	8	−0.049	−0.045	−0.076	−0.076	34.219*	33.388*

[a] Number of alleles for each species across 10 loci.
[b] Average simulated mean values of Tajima's D across 10 loci.
[c] Average simulated mean values of Tajima's D on silent sites across 10 loci.
[d] Average simulated mean values of D across 10 loci of Fu and Li.
[e] Average simulated mean values of F^* across 10 loci of Fu and Li.
[f] The probability from chi-square distribution of the HKA test across 10 loci.
[g] The probability from chi-square distribution of the HKA test on silent sites across 10 loci.
* $0.01 < P \leq 0.05$; ** $0.001 < P \leq 0.01$; *** $P \leq 0.001$.

and F^* of Fu and Li (1993) to address whether the data show evidence that natural selection has shaped levels of variation for individual loci. The values of Tajima's D and D^* and F^* of Fu and Li varied vastly across 10 loci and most of them were not significant (table 2). It is noteworthy that both Tajima's D and D^* and F^* of Fu and Li indicated significant positive values for the locus *CBP1* in *O. officinalis* and *Lhs1* in *O. rhizomatis*, suggesting the presence of balancing selection at these loci. This explanation is consistent with the finding that the elevated diversity was observed at the loci, in particular for *CBP1* in *O. officinalis* (table 2). For *O. eichingeri*-AFR, no locus was significantly different from 0. For *O. eichingeri*-LKA, however, 7 values were negative with 3 being significantly less than 0, and 2 (*CBP1* and *SSII1*) were slightly but not significantly positive (one could not be calculated). D^* and F^* of Fu and Li gave similar patterns (table 2). To determine whether the average values of these tests within taxa significantly deviate from zero, we used a multilocus test based on coalescent simulations to compare the observed Tajima's D and D^* and F^* values of Fu and Li across all loci against the neutral expectation. Significantly negative mean values of Tajima's D and D^* of Fu and Li were observed only for the Sri Lankan *O. eichingeri* (table 3), consistent with the tests at individual loci (table 2). The overall negative patterns of the tests from *O. eichingeri*-LKA indicated an excess of low-frequency polymorphisms in sequence data, with the simplest explanation being a recent demographic expansion (Tajima 1989, and see below) because demographic forces affect all loci simultaneously. It is interesting that level of nucleotide variation at *CBP1* is significantly lower in 2 races of *O. eichingeri* (θ_{sil} = 0.0–0.0007) than in the other species (θ_{sil} = 0.0034–0.0094), suggesting that natural selection might remove variation from this species at this locus.

We further performed the multilocus HKA test (Hudson et al. 1987) to examine whether levels of polymorphism and divergence across loci would be correlated, as expected under neutral model of molecular evolution. The multilocus HKA tests using all sampled sequences were carried out for any pair of 4 taxa, and no signature of departure from the neutral model was detected (*O. officinalis*/*O. eichingeri*-LKA, $\chi^2 = 15.66$, $P = 0.616$; *O. officinalis*/*O. eichingeri*-AFR, $\chi^2 = 23.49$, $P = 0.173$; *O. officinalis*/*O. rhizomatis*, $\chi^2 = 14.52$, $P = 0.695$; *O. eichingeri*-LKA/*O. rhizomats*, $\chi^2 = 11.61$, $P = 0.867$; *O. eichingeri*-AFR/*O. rhizomats*, $\chi^2 = 11.87$, $P = 0.853$; and *O. eichingeri*-LKA/*O. eichingeri*-AFR, $\chi^2 = 7.267$, $P = 0.988$). Because the HKA test statistic for closely related species is not expected to follow the χ^2 distribution (Machado et al. 2002), we compared the test statistic with a distribution generated from 10,000 coalescent simulations (Hilton et al. 1994). Using 2 sequences of *O. punctata* as outgroups, HKA tests across loci were applied to each of the 4 taxa. Figure 2 shows, for each taxon, the contribution from each locus to the overall test statistic, indicating whether or not the observed values of polymorphism and divergence are higher or lower than expected. In each case, the overall test statistic indicated a rejection of the neutral model (table 3). Of the loci with the largest contribution to the overall test statistic, *CBP1* and *Waxy* contributed greatly to the significant multilocus HKA statistics for 3 taxa, with *CBP1* to *O. officinalis*, *O. eichingeri*-AFR, and *O. rhizomatis*, whereas *Waxy* to *O. officinalis*, *O. eichingeri*-LKA, and *O. eichingeri*-AFR. Additionally, *GBSSII* and *Lhs1* also contributed to significant multilocus HKA statistics for *O. eichingeri*-AFR (fig. 2). When the HKA test was repeated with the exclusion of those loci that showed the strongest departures from expectations, the value of the overall test statistics dropped but those for *O. officinalis*, *O. rhizomatis*, and *O. eichingeri*-AFR were still significant (data not shown). As indicated by Ramos-Onsins et al. (2004), significant values can be explained by a larger variance in the polymorphism/divergence ratio than that expected under a neutral equilibrium model. This large variance might be attributed to selection on some of the loci as mentioned above, such as *CBP1* and *Lhs1*.

Shared/Fixed Polymorphism and Divergence

As incipient species diverge from each other, shared polymorphisms are expected to lose whereas fixed differences gradually accumulate (Wakeley and Hey 1997). Thus, closely related species are expected to harbor a relative higher level of shared polymorphisms because the divergence event has not lasted long enough to erase all ancestral polymorphisms. The numbers of shared polymorphisms and fixed differences between 4 taxa pairs are presented in table 4. The number of shared polymorphisms and

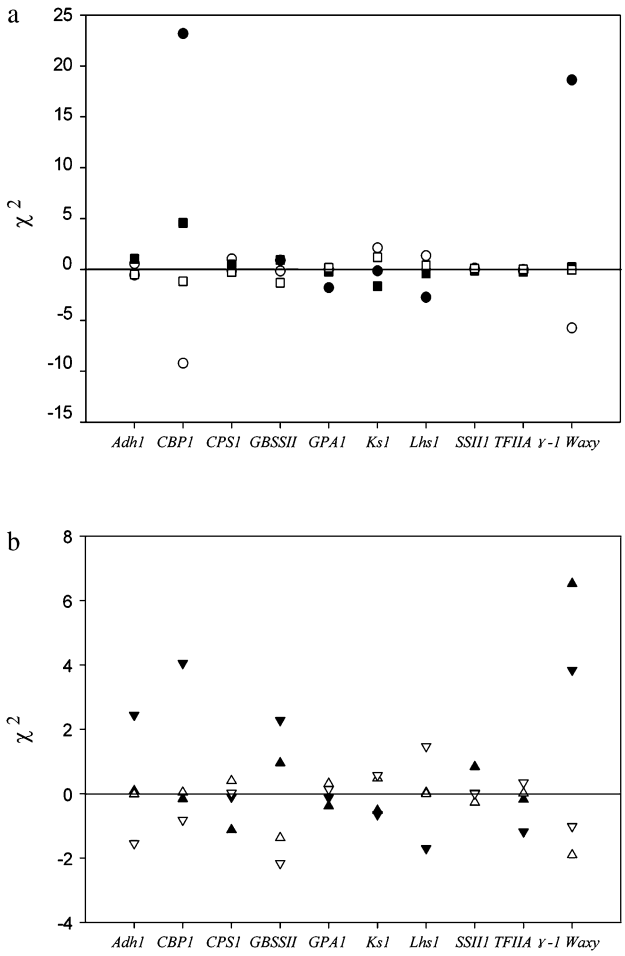

FIG. 2.—Results of 4 multilocus HKA tests (Hudson et al. 1987). In each test, polymorphisms within an individual species are compared with divergence from multiple *Oryza punctata* sequences. The contributions to the overall χ^2 statistic that resulted from polymorphisms are shown by solid symbols and from divergence by open symbols. (*a*) ● and ○, *Oryza officinalis*; ■ and □, *Oryza rhizomatis*. (b) ▲ and △, *Oryza eichingeri*-LKA; ▼ and ▽, *O. eichingeri*-AFR. Symbols above the line indicate that the observed values are greater than the simulated. Similarly, symbols below the line indicate that the observed values are less than the simulated.

fixed differences at each locus was different, which might reveal unique evolutionary histories of individual genes.

A larger number of fixed differences but no shared polymorphism were observed at *Lhs1* locus between all contrasts except for the *O. eichingeri*-LKA/*O. rhizomatis* comparison (table 4). This result might imply that this locus has experienced directional selection, in agreement with the positive Tajima's *D* in 3 taxa except for *eichingeri*-LKA (table 2). Much more shared polymorphisms than fixed differences (19 vs. 2) were observed between the 2 races of *O. eichingeri* relative to more fixed differences than shared polymorphisms between the other contrasts. Moreover, the shared polymorphisms between them involved in 6 loci though there was no polymorphism observed at 2 loci (table 4). It is interesting that the number of shared polymorphisms was more than twice that of fixed differences (25 vs. 13) between *O. eichingeri*-LKA and *O. rhizomatis*, in contrast to the numbers (10 vs. 34) between *O. eichingeri*-AFR and *O. rhizomatis* (table 4). More shared polymorphism and less fixed differences between *O. eichingeri*-LKA and *O. rhizomatis* might indicate closer genetic affinity or hybridization/introgression between them. Because shared polymorphisms can also be generated by parallel mutations, we calculated the amount of shared polymorphisms under the assumption that mutations occur randomly and independently with equal probability at all sites to assess whether the shared polymorphisms could arise just by recurrent mutation (table 4). In all contrast pairs, the expected values of shared polymorphisms were very low (generally near zero) and comprised a small fraction of the observed number (table 4). Therefore, the probability of recurrent mutation was rather low in our data, indicating that a significant fraction of shared polymorphism could not be explained by parallel mutation (Clark 1997; Kliman et al. 2000).

The level of net divergence, the average pairwise divergence between species minus the average intraspecific pairwise variation (Nei 1987), was also used to measure interspecific difference. Over all the 10 loci, levels of net pairwise divergence (*D*) among 4 taxa and between each of them and *O. punctata*, an outgroup used in this study, were calculated (table S2, Supplementary Material online). Estimates of net divergence were similar between each of the 4 taxa and *O. punctata*, with the average values over the 10 loci ranging from 0.0394 to 0.0413. The net divergence between the C-genome species and *O. punctata* was obviously higher than those of the 4 taxa pairs (0.0022–0.0072). Note that the average net divergence between African and Sri Lankan races of *O. eichingeri* was 3-fold lower than the estimations of the other pairs, indicating close genetic relationship between the 2 geographic races (table S2, Supplementary Material online).

Genealogical Analyses

Genealogical trees of the 10 loci were constructed for all samples using both NJ and MP methods. Several characteristics were observed from the genealogical trees as showed in the NJ trees (fig. S3, Supplementary Material online). First, sequences from the 4 taxa sufficiently diverged from the B-genome species at 8 out of 10 loci, implying that the C-genome species started to diverge relatively recently compared with its divergence from the B-genome species. The 2 exceptions involved the loci *CBP1* and *Waxy*, in which zero-length branches were found involving the outgroup (*O. punctata*) samples and ingroup accessions (fig. S3b and S3j, Supplementary Material online), suggestive of the persistence of ancestral alleles together with their descendants and the derived lineages evolved from single ancestral alleles (Posada and Crandall 2001). This deep coalescence of alleles at *CBP1* and *Waxy* might reflect maintenance by balancing selection, in agreement with the results of neutral tests. Second, on most trees, *O. officinalis* sequences formed a monophyletic clade, whereas sequences of the remaining 3 taxa did not cluster by taxon despite a tendency for sequences to cluster by the taxonomic designation. This pattern suggests that coalescence for alleles at most loci occurs after divergence of *O. officinalis* and the common ancestor of *eichingeri* and *O. rhizomatis*. Third, alleles from *eichingeri* and

Table 4
Numbers of Shared Polymorphisms and Fixed Differences

	officinalis/ *rhizomatis*		*officinalis*/ *eichingeri*-LKA		*officinalis*/ *eichingeri*-AFR		*eichingeri*-LKA/ *eichingeri*-AFR		*eichingeri*-LKA/ *rhizomatis*		*eichingeri*-AFR/ *rhizomatis*	
Locus	Shared	Fixed	Shared	Fixed	Shared	Fixed	Shared	Fixed	Shared	Fixed	Shared	Fixed
Adh1	0 (0.20)	1	0 (0.14)	3	0 (0.25)	1	8 (0.36)	1	8 (0.28)	1	8 (0.51)	0
CBP1	0 (0.19)	2	0 (0.00)	0	1 (0.03)	0	0 (0.00)	0	0 (0.00)	10	0 (0.01)	9
CPS1	0 (0.02)	3	0 (0.02)	3	0 (0.05)	2	0 (0.03)	1	0 (0.03)	2	1 (0.07)	1
GBSSII	3 (0.06)	0	0 (0.07)	0	0 (0.09)	0	4 (0.03)	0	2 (0.05)	0	0 (0.07)	0
GPA1	0 (0.09)	1	0 (0.01)	1	0 (0.01)	1	1 (0.01)	0	3 (0.02)	0	0 (0.02)	0
Ks1	3 (0.07)	0	8 (0.14)	0	1 (0.13)	0	1 (0.07)	0	3 (0.03)	0	0 (0.03)	4
Lhs1	0 (0.02)	11	0 (0.04)	11	0 (0.01)	17	0 (0.03)	0	6 (0.08)	0	0 (0.02)	7
SSII1	0 (0.04)	4	0 (0.10)	1	0 (0.06)	5	2 (0.06)	0	0 (0.04)	0	0 (0.03)	7
TFIIAγ-1	0 (0.10)	1	0 (0.10)	1	0 (0.05)	1	3 (0.04)	0	3 (0.07)	0	1 (0.04)	0
Waxy	0 (0.04)	7	0 (0.11)	2	0 (0.09)	3	0 (0.07)	0	0 (0.03)	0	0 (0.03)	7
Total	6	30	8	22	2	30	19	2	25	13	10	34

NOTE.—The numbers in parentheses are the expected shared polymorphisms that arose by recurrent mutations (Clark 1997).

O. rhizomatis were intermixed entirely at 9 loci, whereas accessions from 2 *O. eichingeri* races were basically separated into 2 groups, corresponding to the African and Sri Lankan races. It should be noted that alleles from the African *O. eichingeri* in Cote d'Ivoire (eic-CIV) were clustered with alleles of the Sri Lankan *O. eichingeri* at 8 loci (fig. S3, Supplementary Material online), implying their closer relationship. In brief, phylogenetic analyses indicate that for a particular locus some alleles in 1 taxon are more closely related to those sampled from another taxon, which occurs at different hierarchical levels. Such genealogical patterns of lack of concordance among the 10 gene trees at different taxonomical levels might result from lineage sorting and gene flow because the 4 taxa were closely related and diverged very recently (see below) and gene flow has been documented previously, at least for that between the Sri Lankan *eichingeri* and *O. rhizomatis* (Bautista et al. 2006).

To explore the phylogenetic relationship among species, we reconstructed the phylogeny among the 4 taxa based on a combined data set of the 10 genes. Because there were 2 alleles for a heterozygote, we chose 1 allele randomly in each locus for the heterozygous individuals. Both NJ and MP analyses demonstrated essentially the same topology except for slightly different bootstrap supports for some clades (data not shown). The combined phylogeny (fig. 3) revealed that accessions from the same species formed well-supported, monophyletic clades, including a highly supported monophyly of 2 races of *O. eichingeri*. It is worthwhile mentioning that 2 accessions had particular positions in figure 3. One accession of the Sri Lankan *O. eichingeri* (eic-LKA4) was not clustered with the other *O. eichingeri* accessions and instead formed a clade with the *O. rhizomatis* accessions. Taking into consideration at individual loci, eic-LKA4 was heterozygous at 8 loci, and importantly, 1 of the alleles at all the 8 loci clustered with the *O. rhizomatis* alleles (fig. S3, Supplementary Material online). This phylogenetic pattern could be best explained by hybridization/introgression between the Sri Lankan *O. eichingeri* and *O. rhizomatis* (see Discussion). Within the *O. eichingeri* clade, an African accession (eic-CIV) formed a clade with the Sri Lankan accession before clustering with the other African accessions. Similarly, alleles of this accession were clustered with alleles of Sri Lankan *O. eichingeri* at 8 loci (fig. S3, Supplementary Material online), suggestive of a long-distance dispersal of the African *O. eichingeri* into Sri Lanka (see Discussion).

Testing Speciation Model and Estimating Divergence Time

Upon the assumption of constant population size in history and no gene flow between 2 descendent species since separation from an ancestral species, the simple speciation model (WH) was used to estimate the relative sizes

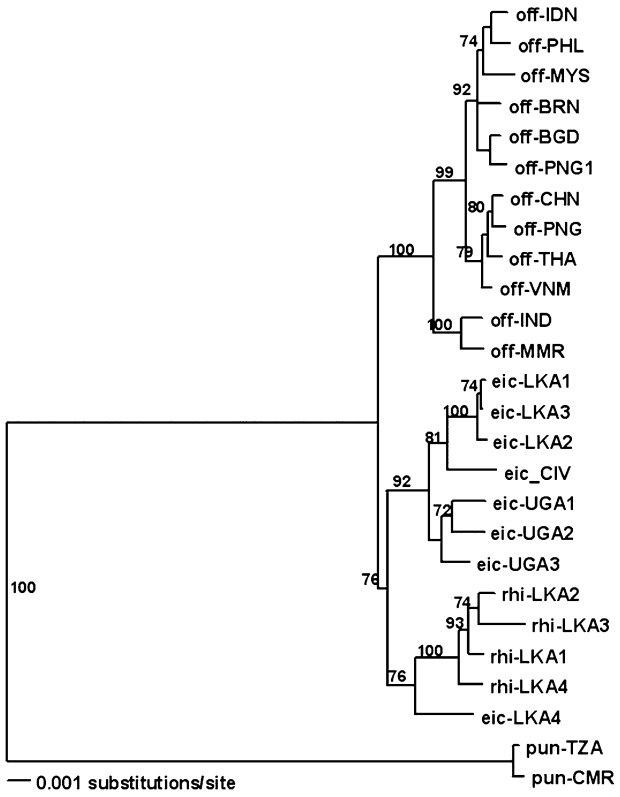

FIG. 3.—NJ tree (Saitou and Nei 1987) using Kimura 2-parameter distances (Kimura 1980) based on combined sequence data. The numbers above nodes are bootstrap values when greater than 50%, and the scale bar indicates level of sequence divergence.

Table 5
Isolation Model Fitting

Species 1	Species 2	θ_1	θ_2	θ_A	T	χ^2	$P_{\chi 2}$	WWH	P_{WWH}
officinalis	*rhizomatis*	23.230	24.092	52.153	0.7438	111.377	0.4887	14	0.5458
officinalis	*eichingeri*-LKA	23.412	25.783	47.599	0.6675	150.810	0.1687	19	0.2824
officinalis	*eichingeri*-AFR	27.579	30.122	32.213	0.9428	130.848	0.0705	18	0.1426
eichingeri-LKA	*rhizomatis*	6.372	5.962	66.131	0.379 1	108.646	0.8404	18	0.5224
eichingeri-AFR	*rhizomatis*	16.365	16.159	72.888	0.6415	78.624	0.9531	17	0.6106
eichingeri-LKA	*eichingeri*-UGA	36.575	33.569	27.874	0.2344	46.351	0.8498	9	0.5347

NOTE.—θ_1, θ_2, and θ_A are the estimates of the population mutation parameters for species 1, species 2, and the ancestral species, respectively. T is the estimated speciation time in units of $2N_1$ generations, where N_1 is the estimate of the effective population size of species 1. The P values, for both χ^2 and the WWH test (Wang et al. 1997), are the proportion of simulated values greater than or equal to the observed values.

of the ancestral and descendant populations (Wakeley and Hey 1997). Based on 3 classes of polymorphisms (exclusive polymorphisms for 1 species, shared polymorphisms, and fixed differences) across all loci, we have applied the simple isolation model to taxon comparisons (table 5). The null hypothesis of the isolation model was not rejected for any of the 2 comparisons when both χ^2 (Kliman et al. 2000) and WWH statistics (Wang et al. 1997) were used, suggesting that no subsequent gene flow shaped the present patterns of nucleotide variation since their divergence. It is noteworthy that in 4 out of the 6 contrasts between *O. officinalis*, *O. rhizomatis*, and *O. eichingeri*, the ancestral population sizes (θ_A) were approximately 2-fold to 10-fold larger than those of either descendents (θ_1, θ_2) (table 5), indicating that the population sizes of descendent taxa might have experienced a contracting process since the time of their separation. This implies a speciation scenario in which a larger ancestral population becomes subdivided and the resulting daughters each occupies part of the ancestral species range. The exception is the contrast between the 2 races of *O. eichingeri* (*O. eichingeri*-LKA/*O. eichingeri*-AFR), in which the ancestral population sizes (θ_A = 27.874) was estimated slightly smaller than those of the 2 descendents (θ_1 = 36.575, θ_2 = 33.569). This result did not support a speciation event where a large ancestral population becomes subdivided, as would be expected by a vicariance event, but rather suggests that the formation of 2 races (*O. eichingeri*-LKA and *O. eichingeri*-AFR) might involve long-distance dispersal (see Discussion) and subsequently slight population expansion, consistent with the average negative Tajima's *D* value and *D** value of Fu and Li (table 3).

To understand speciation history of *O. officinalis* and its close relatives, we used the molecular clock approach to estimate the time of divergence. Three loci (*CBP1*, *Lhs1*, and *Waxy*) were excluded from the divergence analyses because of their heterogeneous mutation rates detected by the program MEGA (data not shown) and significant deviation from neutrality in the data sets (table 2). For the remaining 7 loci, the relative-rate test (Tajima 1993) showed no evidence of rate heterogeneity among the lineages between the 3 C-genome species. Therefore, nucleotide divergences at the 7 loci should be appropriate for determining the divergence time among the species. Because the divergence rates of silent sites of the 7 loci varied greatly, we used a modified method to that of Tenaillon et al. (2004) to estimate the mutation rate (μ) at silent sites for each locus by $\mu = \mu_{adh1} * K_{sil}/K_{sadh1}$, where K_{sil} and K_{sadh1} are silent distance for that locus and synonymous distance at *Adh1* locus, 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 at grass *Adh1* locus (Gaut et al. 1996). As a result, the mutation rates between *O. officinalis* and the outgroup (*O. punctata*) at silent site varied fourfold from 3.99×10^{-9} for *GBSSII1* to 16.82×10^{-9} for *Ks1*. Using these estimates of absolute rates and sequence divergences at the 7 loci, we calculated approximate divergence times between *O. officinalis* and *O. punctata* at 3.8 Myr. Within the C-genome species, we obtain the average divergence times of 0.63–0.68 Myr for 2 consecutive speciation events involving that between *O. officinalis* and the *O. eichingeri*/*O. rhizomatis* lineage and that between *O. eichingeri* and *O. rhizomatis*, indicative of rapid succession of speciation in this group. The mean sequence divergence between *O. eichingeri*-AFR and *O. rhizomatis*-LKA was 0.55%, which translated into a divergence of 0.37 Myr. Despite a number of limitations to the use of clocks based on sequence data, particularly the potential error of the presumed maize–rice divergence date of 50 Myr (Gaut et al. 1996; White and Doebley 1999), the estimates of divergence events provide a rough and relative time frame for understanding of the speciation tempos and biogeographic history of this group of species.

Discussion

Low Level of Nucleotide Variation in the C-Genome Species in *Oryza*

Although nucleotide variation varied 30-fold among the 10 loci, the 4 taxa in the present study possessed similar levels of nucleotide diversity (θ_{sil} = 0.0038–0.0042; π_{sil} = 0.0033–0.0044). When the 2 races of *O. eichingeri* were considered together, diversity value for this species would be elevated slightly (θ_{sil} = 0.0057; π_{sil} = 0.0052) because of population subdivision. A literature survey on the multiple gene studies of plant species has shown a wide range of nucleotide variation across species, even among closely related species (table 6). However, the *Oryza* species in the present study maintain apparently lower nucleotide diversity compared with the estimates of other angiosperm species based on estimates of multiple loci. As shown in table 6, the nucleotide diversity from a majority of plant species is 2–6 times that of the C-genome species, except for the cultivated sorghum (*Sorghum bicolor*) that is largely self-pollinating and has a smaller effective population size (Hamblin et al. 2004) and for *Arabidopsis lyrata* ssp. *lyrata*

Table 6
Estimates of Nucleotide Diversity of Angiosperm Species Based on Multiple Loci for Species-Wide Samples

Species	Number of Genes	θ_{sil}	π_{sil}	Predominant Self-Fertilization	References
Oryza officinalis	10	0.0040	0.0044	No	This study
Oryza eichingeri	10	0.0057	0.0052	No	This study
O. eichingeri-LKA	10	0.0042	0.0033	No	This study
O. eichingeri-AFR	10	0.0039	0.0040	No	This study
Oryza rhizomatis	10	0.0038	0.0037	No	This study
Oryza nivara	10	0.0077	0.0063	Yes	Zhu et al. (2007)
Oryza rufipogon	10	0.0095	0.0072	No	Zhu et al. (2007)
Oryza sativa	10	0.0024	0.0037	Yes	Zhu et al. (2007)
Arabidopsis thaliana	33	0.0161	—	Yes	Wright and Gaut (2005)
Arabidopsis halleri	8	—	0.0150	No	Ramos-Onsins et al. (2004)
Arabidopsis lyrata ssp. *lyrata*	8	0.0040	—	No	Wright and Gaut (2005)
Arabidopsis lyrata ssp. *petraea*	13	0.0247	0.0230	No	Ramos-Onsins et al. (2004); Wright and Gaut (2005)
Zea mays ssp. *mays*	29	0.0149	—	No	Wright and Gaut (2005)
Zea mays ssp. *parviglumis*	18	0.0247	—	No	Wright and Gaut (2005)
Zea diploperennis	4	0.0123	0.0113	No	Tiffin and Gaut (2001)
Zea perennis	4	0.0130	0.0140	No	Tiffin and Gaut (2001)
Sorghum bicolor	52	0.0034	0.0039	Yes	Hamblin et al. (2004)
Hordeum vulgare	9	0.0109	—	Yes	Morrell et al. (2003)
Hordeum vulgare ssp. *spontaneum*	18	0.0081[a]	0.0075[b]	Yes	Morrell et al. (2005)
Helianthus annuus (cultivated)	9	0.0072[a]	0.0096	No	Liu and Burke (2006)
H. annuus (wild)	9	0.0144[a]	0.0234	NA	Liu and Burke (2006)
Quercus petraea	7	—	0.0072[b]	No	Pot et al. (2005)

NOTE.—NA, mating system unclear in this species.
[a] θ_T from entire region.
[b] π_T from entire region.

that has comparable diversity to the 4 taxa investigated. Although *O. eichingeri* maintained the highest level of nucleotide diversity among the 3 species under study, its mean estimate of nucleotide diversity at silent sites ($\theta_{sil} = 0.0057$) was only 23–46% of those observed in *Zea* species ($\theta_{sil} = 0.0123$–0.0247) and 35% of that in *Arabidopsis thaliana* ($\theta_{sil} = 0.0161$), a highly self-fertilizing species (table 6). Low level of genetic variation in *O. officinalis* and its closely related species was previously reported using molecular markers, such as allozyme, restriction fragment length polymorphisms, and simple sequence repeats (Gao et al. 2001; Federici et al. 2002; Gao and Zhang 2005; Bautista et al. 2006). Based on more than 20 allozyme loci, Gao et al. (2000, 2001) studied the genetic diversity of the Chinese populations of *O. officinalis* and *O. rufipogon* and found that *O. officinalis* maintained significantly lower variation ($P = 12.5$; $H_e = 0.0266$) than *O. rufipogon* ($P = 22.7$; $H_e = 0.068$). In a recent study, Gao and Zhang (2005) used SSR markers to comparatively investigate the population genetics of the Chinese *O. officinalis* and *O. rufipogon* and found that *O. rufipogon* had higher levels of genetic diversity but lower differentiation among populations ($P = 100.0$; $H_e = 0.580$; $F_{st} = 0.271$) than *O. officinalis* ($P = 57.1$; $H_e = 0.283$; $F_{st} = 0.554$).

As pointed out by Wright and Gaut (2005), a number of different factors may contribute to nucleotide variation across plant species, including sample representative, mutation rate, demography, and selection. The former one can be ruled out in this study because we used species-wide samples for all taxa. Mutation rate is also not relevant because we found much higher diversity ($\theta_{sil} = 0.0095$; $\pi_{sil} = 0.0072$) in a related A-genome species (*O. rufipogon*) based on sequences of same set of loci (Zhu et al., 2007, table 6). It is well established that nucleotide diversity can be affected by a consequence of selection (either long-term balancing selection or a recent selective sweep). However, it is unlikely that the low diversity in the species under study was attributable to selection because statistic tests did not find overall significant deviation from neutrality except for a few of loci where both balancing selection (elevated the diversity) and selective sweep (reduced the diversity) were present (table 2).

The effects of demography on diversity and speciation have been well appreciated in plants and animals (see Machado et al. 2002; Llopart et al. 2005; Wright and Gaut 2005 for reviews). Ramos-Onsins et al. (2004) conducted a comprehensive investigation on the divergence and speciation of the closely related outcrossing *Arabidopsis halleri* and *A. lyrata* based on sequence data of 3–8 nuclear loci and found that levels of nucleotide variation in different species reflected the differences among species in effective population size. In the present study, the speciation model test indicates that the estimated effective population sizes of all 4 taxa (θ_1, θ_2) are smaller than those of their ancestors (θ_A) at all comparison pairs except for the contrast between the 2 races of *O. eichingeri* (table 5). This result suggests that *O. officinalis* and its close relatives might have undergone a process of population contraction since divergence from their ancestor. This phenomenon has been evidenced from other organisms such as *Drosophila*, *Arabidopsis*, and crop species because of either a reduction of effective population size or a founder effect and bottleneck during domestication (Kliman et al. 2000; Machado et al. 2002; Hamblin et al. 2004; Ramos-Onsins et al. 2004; Städler et al. 2005; Wright and Gaut 2005). Consequently, the low level of nucleotide diversity in the C-genome species

is most likely explained by the demographic factor, that is, a smaller historic effective population size. Recent population reduction and extinction because of habitat fragmentation and deterioration may also have led to genetic reduction for these species. For example, several studies (Gao et al. 2001; Gao 2005) showed that the natural populations of *O. officinalis* have been isolated due to habitat deterioration and human destruction, which in turn caused the spatial distribution of this species to be fragmented. Under the circumstances of the fragmented habitats and isolated populations associated with low level of migration, individual local populations are apt to extinction from stochastic processes (Amos and Harwood 1998).

Evolutionary History and Introgression between Species

Traditionally, *O. officinalis* and its close relatives have been delineated on the basis of morphological characters, distinct habitats, and different geographical distributions (Tateoka 1965; Vaughan 1989, 1990). *Oryza officinalis* and *O. rhizomatis* can be differentiated from *O. eichingeri* in that the former 2 have rhizomes (Vaughan 1990). Compared with *O. rhizomatis*, *O. officinalis* has smaller spikelets, shorter palea tip, and more approximately equal branches from the lowest panicle node. Morphologically, *O. officinalis* is more similar to *O. rhizomatis* than to *O. eichingeri* (Vaughan 1990), which was supported by phylogenetic analysis based on multiple gene sequences (Bao and Ge 2003). In contrast, recent AFLP (Bautista et al. 2006) and SSR (Bao et al. 2006) analyses suggested that *O. rhizomatis* was more genetically similar to *O. eichingeri* than to *O. officinalis*.

In molecular phylogenetics of closely related taxa, it has been increasingly appreciated that the time back to the common ancestor of 2 DNA sequences may be longer than the time back to the common ancestor of 2 taxa (Nei 1987; Pamilo and Nei 1988). This phenomenon that gene divergence precedes species divergence originates from ancestral polymorphisms and will cause a high probability that gene trees disagree with species trees (Wu 1991; Wendel and Doyle 1998). Therefore, it is not unexpected that the phylogenetic trees based on 10 nuclear genes showed different topological relationships among species in the present study (fig. S3, Supplementary Material online). Such lineage sorting because of ancient polymorphism is more likely a source of incongruence among gene trees at lower taxonomic ranks (Wendel and Doyle 1998). As pointed out by Klein et al. (1998), at most loci, differential fixation of ancestral polymorphism influences phylogenies in which divergences occur within a time interval of less than 1–2 Myr. Our approximate estimation of divergence times for 3 species indicates that the 2 speciation events within the C-genome group happened at such a short time interval (~0.63–0.68 Myr) that the polymorphisms in the ancestral population of all 3 species could persist easily from the first divergence to the second. Such molecular phenomenon associated with speciation radiations is the main reason causing incongruent topologies by different genes in this study (fig. S3, Supplementary Material online; Bao and Ge 2003) and has been reported in other species such as human and its relatives (Enard and Pääbo 2004), *Drosophila* and field cricket species complex (Wang et al. 1997; Kliman et al. 2000; Broughton and Harrison 2003; Hey and Nielsen 2004), and plant crops (Small and Wendel 2000; Tiffin and Gaut 2001; Clark et al. 2004).

Therefore, a combined analysis based on multiple loci is generally needed to overcome the noises of ancient polymorphisms to accurately reconstruct a phylogeny of closely related species (Pamilo and Nei 1988; Wu 1991; de Queiroz et al. 1995). In this study, the combined tree indicates clearly that *O. rhizomatis* and *O. eichingeri* form a monophyletic clade, which is sister to the group containing all *O. officinalis* accessions (fig. 3). This phylogenetic relationship is also supported by the analyses based on shared polymorphisms and net pairwise divergence (table 4 and table S2, Supplementary Material online), consistent with recent population-based studies using AFLP and SSR markers (Bao et al. 2006; Bautista et al. 2006).

In the process of speciation, shared polymorphisms in newly formed species may result both from recent divergence from a common ancestor and from gene flow or introgression between species (Machado et al. 2002; Broughton and Harrison 2003; Ramos-Onsins et al. 2004). The existence of gene flow between species would alter the pattern of both within- and between-species variation (Ramos-Onsins et al. 2004). In the present study, the null hypothesis of the isolation model was not rejected for any of 6 contrasts (table 5), suggesting that no subsequent gene flow shaped the present patterns of nucleotide variation since their divergence. In another word, gene flow or introgression among species have been limited, although the divergence of *O. officinalis* and its close relatives was relatively recent and extensive polymorphism has been maintained in their common ancestor. However, a hybridization event between the Sri Lankan *O. eichingeri* and *O. rhizomatis* is plausible, for interspecific cross between them has been reported by Bautista et al. (2006). In this study, 1 of 2 alleles from 1 Sri Lankan *O. eichingeri* accession (eic-LKA4) clustered with the *O. rhizomatis* alleles at all 8 heterozygous loci (fig. S3, Supplementary Material online). This observation strongly supports the introgression or hybridization between these 2 species because most introgressed alleles tend to co-occur in the same individuals, whereas lineage sorting makes alleles in question more randomly distributed among individuals (Wendel and Doyle 1998). Nevertheless, there is little evidence of substantial introgression between the 2 species even though the Sri Lankan *O. eichingeri* and *O. rhizomatis* hybridize where they come into contact (Bautista et al. 2006). To evaluate whether or not the introgressed individual (eic-LKA4) has affected the test of isolation speciation model, we recalculated the population parameters by excluding the accession eic-LKA4. Similar estimations of the population parameters were obtained and the null hypothesis of the isolation model was not rejected (data not shown), suggesting that introgression is not pervasive between these 2 species.

Genetic Differentiation between Geographic Races and Long-Distance Dispersal of *O. eichingeri*

Oryza eichingeri is a particularly interesting species, and there has been a considerable debate regarding its taxonomic treatment because of its remarkably disjunct

distribution in Africa and Sri Lanka (fig. 1) (Nayar 1973; Biswal and Sharma 1987; Vaughan 1989; Vaughan et al. 2003). In the present study, based on sequences of 10 nuclear loci, a divergence between the 2 races of *O. eichingeri* was observed. To further explore the genetic differentiation of the 2 races, we performed a test of geographic subdivision (Hudson et al. 1992) and found a significant differentiation between the 2 races for nine loci out of 10, with the F_{st} values ranging from 0.143 to 0.714 (average F_{st} = 0.442, $P < 0.001$). Sufficient differentiation between 2 *O. eichingeri* races has also been detected using molecular markers (Shcherban et al. 2001; Federici et al. 2002; Bao and Ge 2003; Bao et al. 2006). As demonstrated previously, population subdivision can elevate the level of genetic variation in species if there is no presence of gene flow (Nei and Takahata 1993; Cherry 2004). Because no gene flow between the geographical races was detected by speciation model test, significant genetic differentiation between the 2 races might have contributed to the highest diversity of *O. eichingeri* among the diploid C-genome species (table 6). High degree of intraspecific variation in *O. eichingeri* has also been detected at the genome level (Dally and Second 1990).

Based on 1 sample collected from Sri Lanka, Sharma and Shastry (1965) named a new species, *Oryza collina*. This treatment was subsequently followed by some authors (Nayar 1973; Wang et al. 1992), but retracted by some others (Biswal and Sharma 1987; Vaughan 1990) who considered that the Sri Lankan form was within the variation range of *O. eichingeri*. This study indicates clearly that 2 geographical races of *O. eichingeri* shares a more recent common ancestor compared with either of them to the other C-genome species based on both the shared polymorphisms and the multiple gene phylogenies (table 4 and fig. S3, Supplementary Material online). Our molecular dating provides an approximate divergence time of 0.37 Myr between the 2 races, much more recent compared with the times of other speciation events in this group (0.63–0.68 Myr). These results do not support treating the Sri Lankan *O. eichingeri* as an independent species.

Oryza eichingeri is the only wild *Oryza* species reported from both Asia and Africa and thus has attracted interests regarding its geographic pattern (Vaughan et al. 2003, 2005; Bautista et al. 2006). This distribution pattern could result either from a vicariance where a large ancestral population becomes subdivided or from long-distance dispersal between 2 continents. The present study supports the long-distance dispersal hypothesis. First, our isolation model test revealed a smaller ancestral population size relative to those of 2 geographic races of *O. eichingeri* (table 5), which makes the vicariance scenario unlikely. In addition, the consistency of the negative values of the average Tajima's *D* and *D** and *F** of Fu and Li across loci (tables 2 and 3) does not corroborate the vicariance hypothesis. The long dispersal hypothesis is further supported by the observation that *O. eichingeri* has a similar morphology and is found in similar habitats in Africa and Sri Lanka (Vaughan et al. 2003).

Vaughan et al. (2005) suggested that the African *O. eichingeri* could be dispersed to Sri Lanka by the birds that migrated across the Indian Ocean from Africa. Bautista et al. (2006) comparatively studied genetic diversity of the *Oryza* species with the A- and C-genomes in southern South Asia and speculated that the Sri Lankan *O. eichingeri* might have been introduced from Africa a very long time ago. The present phylogenetic analyses found a high level of divergence between the western and eastern African accessions and that alleles from the Cote d'Ivoire accession (eic_CIV) were clustered with alleles of the Sri Lankan accessions rather than with those from the Ugandan accessions on the individual trees of 8 loci and the combined tree (fig. 3 and fig. S3, Supplementary Material online). These results suggest a closer relationship between West African and the Sri Lankan *O. eichingeri* and imply that the Sri Lankan *O. eichingeri* might be originated from West Africa.

Population genetics predicts that the derived populations would harbor much reduced genetic diversity relative to the ancient population (population bottleneck). In this study, no bottleneck effect was detected for the Sri Lankan *O. eichingeri* because the 2 geographic races had comparable levels of average nucleotide diversity (θ_{sil} = 0.0042, π_{sil} = 0.0033 for *O. eichingeri*-LKA; θ_{sil} = 0.0039, π_{sil} = 0.0040 for *O. eichingeri*-AFR; table 2). This result seems odd to the long-distance dispersal from Africa to Sri Lanka unless the genetic diversity in the African *O. eichingeri* has been significantly decreased for some reasons and/or there have been frequent dispersal events from Africa to Sri Lanka. It should be noted, however, that the diversity level of *O. eichingeri*-LKA was substantially reduced (θ_{sil} = 0.0013; π_{sil} = 0.0011) when the introgressed *O. eichingeri* accession (eic-LKA4) was excluded, whereas the diversity of *O. eichingeri*-LKA remained high if any one of the other accessions was excluded (θ_{sil} = 0.0047–0.0048; π_{sil} = 0.0039–0.0043). This suggests that much less nucleotide diversity would be expected in the Sri Lankan *O. eichingeri* if the introgression between species was precluded, consistent with a long-distance dispersal from Africa to Sri Lanka. However, it is still premature to make a conclusion about the geographic history of this species before a phylogeographic study with extensive sampling is made.

Supplementary Material

Supplementary Tables S1 and S2 and Figures S1–S3 are available at *Molecular biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Qi-Hui Zhu and Zhi-Yong Zhang for helpful suggestions on the manuscript, Jody Hey for guidance on data analysis, and Xian-Zhao Kan and the members of Ge's group for technical assistances. We are also grateful to the International Rice Research Institute (Los Banos, Philippines) for providing leaf and seed samples and Gerard Second for providing the IP7 sample. This work was supported by the National Natural Science Foundation of China (30430030 and 30121003) and the Chinese Academy of Sciences Innovation Grant.

Literature Cited

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS. 1999. Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor Appl Genet. 98:1320–1328.

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol. 2:e286.
Amos W, Harwood J. 1998. Factors affecting levels of genetic diversity in natural populations. Phil Trans R Soc Lond B. 353:177–186.
Ausubel FM. 1992. Short protocols in molecular biology. New York: John Wiley & Sons.
Bao Y, Ge S. 2003. Phylogenetic relationships among diploid species of *Oryza officinalis* complex revealed by multiple gene sequences. Acta Phytotaxon Sin. 41:497–508.
Bao Y, Ge S. 2004. Origin and phylogeny of *Oryza* species with the CD genome based on multiple-gene sequence data. Plant Syst Evol. 249:55–66.
Bao Y, Zhou HF, Hong DY, Ge S. 2006. Genetic diversity and evolutionary relationships of *Oryza* species with the B- and C-genomes as revealed by SSR markers. J Plant Biol. 49:339–347.
Bautista N, Vaughan D, Jayasuriya A, Liyanage A, Kaga A, Tomooka N. 2006. Genetic diversity in AA and CC genome *Oryza* species in southern South Asia. Genet Resour Crop Evol. 53:631–640.
Biswal J, Sharma SD. 1987. Taxonomy and phylogeny of *Oryza collina*. Oryza. 24:24–29.
Blair MW, Garris AJ, Iyer AS, Chapman B, Kresovich S, McCouch SR. 2003. High resolution genetic mapping and candidate gene identification at the *xa5* locus for bacterial blight resistance in rice (*Oryza sativa* L.). Theor Appl Genet. 107:62–73.
Broughton RE, Harrison RG. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics. 163:1389–1401.
Cherry JL. 2004. Selection, subdivision and extinction and recolonization. Genetics. 166:1105–1114.
Clark AG. 1997. Neutral behavior of shared polymorphism. Proc Natl Acad Sci USA. 94:7730–7734.
Clark RM, Linton E, Messing J, Doebley JF. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proc Natl Acad Sci USA. 101:700–707.
Dally AM, Second G. 1990. Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section *Oryza*). Cladistic-mutation and genetic-distance analysis. Theor Appl Genet. 80:209–222.
de Queiroz A, Donoghue MJ, Kim J. 1995. Separate versus combined analysis of phylogenetic evidence. Annu Rev Ecol Syst. 26:657–681.
Dian W, Jiang H, Chen Q, Liu F, Wu P. 2003. Cloning and characterization of the *granule-bound starch synthase II* gene in rice: gene expression is regulated by the nitrogen level, sugar and circadian rhythm. Planta. 218:261–268.
Enard W, Pääbo S. 2004. Comparative primate genomics. Annu Rev Genomics Hum Genet. 5:351–378.
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.
Federici MT, Shcherban AB, Capdevielle F, Francis M, Vaughan D. 2002. Analysis of genetic diversity in the *Oryza officinalis* complex. J Biotechnol. 5:173–181.
Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics. 133:693–709.
Gao LZ. 2005. Microsatellite variation within and among populations of *Oryza officinalis* (Poaceae), an endangered wild rice from China. Mol Ecol. 14:4287–4297.
Gao LZ, Ge S, Hong DY. 2000. Allozyme variation and population genetic structure of common wild rice, *Oryza rufipogon* Griff. in China. Theor Appl Genet. 101:494–502.
Gao LZ, Ge S, Hong DY. 2001. High levels of genetic differentiation of *Oryza officinalis* Wall. et Watt. from China. J Hered. 92:511–516.
Gao LZ, Zhang CH. 2005. Comparisons of microsatellite variability and population genetic structure of two endangered wild rice species, *Oryza rufipogon* and *O. officinalis*, and their conservation implications. Biodivers Conserv. 14:1663–1679.
Garris AJ, McCouch SR, Kresovich S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics. 165:759–769.
Gaut BS, Morton BR, McCaig BC, Clegg MT. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA. 93:10274–10279.
Ge S, Sang T, Lu BR, Hong DY. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc Natl Acad Sci USA. 96:14400–14405.
Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH, Kresovich S. 2004. Comparative population genetics of the Panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. Genetics. 167:471–483.
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, Wakeley J. 1997. A coalescent estimator of the population recombination rate. Genetics. 145:833–846.
Hilton H, Gaut BS. 1998. Speciation and domestication in maize and its wild relatives: evidence from the *Globulin-1* gene. Genetics. 150:863–872.
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.
Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116:153–159.
Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics. 132:583–589.
Jeon J, Jang S, Lee S, et al. (11 co-authors). 2000. *Leafy hull sterile1* is a homeotic mutation in a rice MADS box gene affecting rice flower development. Plant Cell. 12:871–884.
Judo MSB, Wedel A, Wilson C. 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. 26:1819–1825.
Khush GS, Brar DS. 2001. Rice genetics from Mendel to functional genomics. In: Khush GS, Brar DS, Hardy B, editors. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium. Los Banos (Phillipines): IRRI. p. 3–25.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16:111–134.
Klein J, Sato A, Nagl S, O'hUigin C. 1998. Molecular trans-species polymorphism. Annu Rev Ecol Syst. 29:1–21.
Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics. 156:1913–1931.
Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinformatics. 5:150–163.

Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics. 173:321–330.
Llopart A, Lachaise D, Coyne JA. 2005. Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. Genetics. 171:197–210.
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. Mol Biol Evol. 19:472–488.
Morrell PL, Lundy KE, Clegg MT. 2003. Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. Proc Natl Acad Sci USA. 100:10812–10817.
Morrell PL, Toleno DM, Lundy KE, Clegg MT. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. Proc Natl Acad Sci USA. 102:2442–2447.
Nayar NM. 1973. Origin and cytogenetics of rice. Adv Genet. 17:153–292.
Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
Nei M, Li W-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA. 76:5269–5273.
Nei M, Takahata N. 1993. Effective population size, genetic diversity, and coalescent times in subdivided populations. J Mol Evol. 37:240–244.
Nordborg M, Innan H. 2002. Molecular population genetics. Curr Opin Plant Biol. 5:69–73.
Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD. 2006. Selection under domestication: evidence for a sweep in the rice *Waxy* genomic region. Genetics. 173:975–983.
Olsen KM, Purugganan MD. 2002. Molecular evidence on the origin and evolution of glutinous rice. Genetics. 162:941–950.
Palumbi S, Baker C. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Biol Evol. 11:426–435.
Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. Mol Biol Evol. 5:568–583.
Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol. 16:37–45.
Pot D, McMillan L, Echt C, Le Provost G, Garnier-Gere P, Cato S, Plomion C. 2005. Nucleotide variation in genes involved in wood formation in two pine species. New Phytol. 167:101–112.
Ramos-Onsins SE, Stranger BE, Mitchell-Olds T, Aguade M. 2004. Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. Genetics. 166:373–388.
Rozas J, Sánchez-Delbarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstruction phylogenetic trees. Mol Biol Evol. 4:406–425.
Sakamoto T, Miura K, Itoh H, et al. (15 co-authors). 2004. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiol. 134:1642–1653.
Savolainen O, Langley CH, Lazzaro BP, Freville H. 2000. Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. Mol Biol Evol. 17:645–655.
Seo HS, Kim HY, Jeong JY, Lee SY, Cho MJ, Bahk JD. 1995. Molecular cloning and characterization of *RGA1* encoding a G protein subunit from rice (O*ryza sativa L.* IR-36). Plant Mol Biol. 27:1119–1131.
Shammas FV, Heikkila R, Osland A. 2001. Fluorescence-based method for measuring and determining the mechanisms of recombination in quantitative PCR. Clin Chim Acta. 304:19–28.
Sharma SD, Shastry SVS. 1965. Taxonomic studies in the genus *Oryza*. VI. A modified classification of genus. Indian J Genet. 25:173–178.
Shcherban AB, Vaughan DA, Tomooka N, Kaga A. 2001. Diversity in the integrase coding domain of a *gypsy*-like retrotransposon among wild relatives of rice in the *Oryza officinalis* complex. Genetica. 110:43–53.
Small RL, Ryburn JA, Wendel JF. 1999. Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). Mol Biol Evol. 16:491–501.
Small RL, Wendel JF. 2000. Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium* L., Malvaceae). Mol Phylogenet Evol. 16:73–84.
Städler T, Roselius K, Stephan W. 2005. Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. Evolution. 59:1268–1279.
Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), Version 4.0. Sunderland (MA): Sinauer Associates.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123: 585–595.
Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics. 135:599–607.
Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A. 2000. The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell. 12:381–391.
Tateoka T. 1965. A taxonomy study of *Oryza eichingeri* and *O. punctata*. Bot Mag Tokyo. 78:156–163.
Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol. 21:1214–1225.
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.
Tiffin P, Gaut BS. 2001. Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. Genetics. 158:401–412.
Vaughan D, Kadowaki K, Kaga A, Tomooka N. 2005. On the phylogeny and biogeography of the genus *Oryza*. Breeding Sci. 55:113–122.
Vaughan D, Morishima H, Kadowaki K. 2003. Diversity in the *Oryza* genus. Curr Opin Plant Biol. 6:139–146.
Vaughan DA. 1989. The genus *Oryza* L.: current status of taxonomy. Los Banos (Phillipines): International Rice Research Institute.
Vaughan DA. 1990. A new rhizomatious *Oryza* species (Poaceae) from Sri Lanka. Bot J Linn Soc. 103:159–163.
Wakeley J, Hey J. 1997. Estimating ancestral population parameters. Genetics. 145:847–855.
Wang RL, Wakeley J, Hey J. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics. 147:1091–1106.
Wang ZY, Second G, Tanksley SD. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor Appl Genet. 83:565–581.
Washio K, Ishikawa K. 1994. Organ-specific and hormone-dependent expression of genes for serine carboxypeptidases during development and following germination of rice grains. Plant Physiol. 105:1275–1280.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombinations. Theor Popul Biol. 7:256–276.
Wayne ML, Simonsen KL. 1998. Statistical tests of neutrality in the age of weak selection. Trends Ecol Evol. 13:236–240.
Wendel JF, Doyle JJ. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis DE, Soltis PS, Doyle JJ, editors. Molecular systematics of plants, II: DNA sequencing. Boston (MA): Kluwer Press. p. 265–296.
White SE, Doebley JF. 1999. The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. Genetics. 153: 1455–1462.
Won YJ, Hey J. 2005. Divergence population genetics of chimpanzees. Mol Biol Evol. 22:297–307.
Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol. 22:506–519.
Wu CI. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. Genetics. 127:429–435.
Xia X, Xie Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. J Hered. 92:371–373.
Yoshida K, Miyashita N. 2005. Nucleotide polymorphism in the *Adh2* region of the wild rice *Oryza rufipogon*. Theor Appl Genet. 111:1215–1228.
Yu N, Jensen-Seaman MI, Chemnick L, Ryder O, Li W-H. 2004. Nucleotide diversity in gorillas. Genetics. 166:1375–1383.
Zhu Q, Zheng X-M, Luo J-C, Gaut BS, Ge S. Forthcoming 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol.

Pekka Pamilo, Associate Editor

Accepted December 18, 2006