

Multilocus Analysis of Nucleotide Variation of *Oryza sativa* and Its Wild Relatives: Severe Bottleneck during Domestication of Rice

Qihui Zhu,*† Xiaoming Zheng,* Jingchu Luo,† Brandon S. Gaut,‡ and Song Ge*

*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China;

†Center for Bioinformatics, State Key Laboratory of Plant Genetic Engineering and Protein Engineering and College of Life Sciences, Peking University, Beijing, China; and ‡Department of Ecology and Evolutionary Biology, University of California–Irvine

Varying degrees of reduction of genetic diversity in crops relative to their wild progenitors occurred during the process of domestication. Such information, however, has not been available for the Asian cultivated rice (*Oryza sativa*) despite its importance as a staple food and a model organism. To reveal levels and patterns of nucleotide diversity and to elucidate the genetic relationship and demographic history of *O. sativa* and its close relatives (*Oryza rufipogon* and *Oryza nivara*), we investigated nucleotide diversity data from 10 unlinked nuclear loci in species-wide samples of these species. The results indicated that *O. rufipogon* and *O. nivara* possessed comparable levels of nucleotide variation ($\theta_{\text{sil}} = 0.0077\sim 0.0095$) compared with the relatives of other crops. In contrast, nucleotide diversity of *O. sativa* was as low as $\theta_{\text{sil}} = 0.0024$ and even lower ($\theta_{\text{sil}} = 0.0021$ for *indica* and 0.0011 for *japonica*), if we consider the 2 subspecies separately. Overall, only 20–10% of the diversity in the wild species was retained in 2 subspecies of the cultivated rice (*indica* and *japonica*), respectively. Because statistic tests did not reject the assumption of neutrality for all 10 loci, we further used coalescent to simulate bottlenecks under various lengths and population sizes to better understand the domestication process. Consistent with the dramatic reduction in nucleotide diversity, we detected a severe domestication bottleneck and demonstrated that the sequence diversity currently found in the rice genome could be explained by a founding population of 1,500 individuals if the initial domestication event occurred over a 3,000-year period. Phylogenetic analyses revealed close genetic relationships and ambiguous species boundary of *O. rufipogon* and *O. nivara*, providing additional evidence to treat them as 2 ecotypes of a single species. Lowest linkage disequilibrium (LD) was found in the perennial *O. rufipogon* where the r^2 value dropped to a negligible level within 400 bp, and the highest in the japonica rice where LD extended to the entirely sequenced region (~ 900 bp), implying that LD mapping by genome scans may not be feasible in wild rice due to the high density of markers needed.

Introduction

Domestication has long been an area of great interest in study of evolution (Darwin 1859) and for understanding the process of mutual dependence between human societies and the plants and animals (Zeder et al. 2006). Varying degrees of reduction of genetic diversity in crops relative to their wild progenitors occurred during the process of domestication (Tanksley and McCouch 1997; Zeder et al. 2006). This reduction in genetic diversity has resulted partly from small initial populations in crops relative to their wild ancestors and partly from intense selection for agronomic traits (Eyre-Walker et al. 1998; Clark et al. 2004; Zeder et al. 2006). The former effect, often referred to as “domestication bottleneck” (Eyre-Walker et al. 1998), is featured by a genome-wide loss of genetic diversity, whereas selection is locus specific and occurs in regions of the genome that are tightly linked to the sites that are the targets of selection (Clark et al. 2004; Zeder et al. 2006). In addition, domestication may also have a major impact on the organization of genetic diversity within the genome because population bottlenecks and selection may elevate linkage disequilibrium (LD) throughout the genome (Nordborg 2000). Thus, knowledge of the process and dynamics of crop domestication is important, both because domestication bottlenecks limit genetic variation in crops (Eyre-Walker et al. 1998; Buckler et al. 2001) and because any approach that analyzes variation in populations such as association mapping should be based on population

genetic parameters (Eyre-Walker et al. 1998; Liu and Burke 2006).

In recent decades, molecular population genetics has been widely 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). Because natural selection generally acts on some but not all genes, it is possible to differentiate the effects of natural selection and demography using multilocus sequence data (Nordborg and Innan 2002; Wright and Gaut 2005). Although many 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 undertaken to examine the changes in nucleotide diversity and to reveal demographic history of crops and their wild relatives (Eyre-Walker et al. 1998; White and Doebley 1999; Hamblin et al. 2004; Tenaillon et al. 2004; White et al. 2004; Wright et al. 2005; Liu and Burke 2006).

Asian cultivated rice (*Oryza sativa* L.) is an economically important crop that is the staple food for more than one-half of the world’s population. It includes 2 major subspecies: *Oryza sativa* ssp. *indica* and *Oryza sativa* ssp. *japonica*. Differences between the subspecies are apparent both in a number of physiological and morphological traits and in growth habitats (Oka 1988; Garris et al. 2005). Geographically and ecologically, *indica* is primarily known as lowland rice and grown throughout tropical Asia, whereas *japonica* is typically found in temperate East Asia, upland areas of Southeast Asia, and high elevations in South Asia (Oka 1988; Khush 1997). Two Asian wild species, *Oryza*

Key words: nucleotide diversity, domestication, bottleneck, linkage disequilibrium, *Oryza sativa*, *Oryza rufipogon*.

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

Mol. Biol. Evol. 24(3):875–888. 2007

doi:10.1093/molbev/msm005

Advance Access publication January 11, 2007

rufipogon Griff. and *O. nivara* Sharma et Shastry, are most closely related to and have thus been considered as the progenitors of *O. sativa* (Vaughan 1989; Khush 1997). The perennial *O. rufipogon* is widely distributed in southern China, South and Southeast Asia, and northern Australia and inhabits wet habitats throughout its growing season. In contrast, the annual *O. nivara* is mainly found in South and Southeast Asia and occurs in seasonally dry/wet areas such as ponds, swamps, and the vicinity of rice fields with shallow water (Vaughan 1989). Because life history traits characterizing these 2 wild species vary continuously in nature and segregate in an F₂ population (Morishima 2001), they have been treated mostly as 2 different ecotypes (Oka 1988; Morishima et al. 1992; Lu et al. 2000; Cheng et al. 2003) or subspecies (Vaughan and Morishima 2003) under a single species (*O. rufipogon*). It is increasingly appreciated that *O. sativa* and its wild relatives (*O. nivara* and *O. rufipogon*) form a large species complex together with weedy races from introgression and hybridization between rice and its wild relatives (Vaughan 1989; Lu et al. 2000; Morishima 2001).

Because of the agronomic and theoretical importance of rice, extensive studies have been conducted on the genetic diversity, systematics, and evolution of rice and its relatives (see reviews in Oka 1988; Khush 1997; Ge et al. 2001; Vaughan et al. 2003; Zhu and Ge 2005). It has been well established that the Asian cultivated rice originated from its wild relative, *O. rufipogon* sensu lato (including both annual and perennial ecotypes) (Khush 1997; Vaughan et al. 2003), though the number and precise geographical location of domestications are still controversial (Morishima 2001; Zhu and Ge 2005; Londo et al. 2006). A growing number of studies have characterized nucleotide variation of cultivated rice and its wild relatives (Barbier et al. 1991; Olsen and Purugganan 2002; Garris et al. 2003; Yoshida and Miyashita 2005; Olsen et al. 2006). These investigations, however, were mainly focused on a single species and exclusively based on 1 or 2 genes or multiple linked genes. In this study, we provide the first investigation on the levels and patterns of genetic variation and relationships of the Asian cultivated rice and its close relatives based on sequence data of multiple nuclear loci. Our specific objectives were 1) to compare nucleotide variation of *O. sativa* and its wild progenitors with other major crops; 2) to elucidate the genetic relationship between *O. rufipogon* and *O. nivara*; 3) to study population demography and explore the severity of population bottleneck during rice domestication; and 4) to characterize the level of LD, which is a key issue in the design of fine-scale mapping of agronomically important genes. To address these questions not only provides important insights into the domestication process and the evolution of cultivated rice but also facilitates the effective use of the wild rice germplasm.

Materials and Methods

Plant Materials

A total of 30 accessions of cultivated rice (16 *O. sativa* ssp. *indica* and 14 *O. sativa* ssp. *japonica*) and 30 wild individuals (18 *O. rufipogon* and 12 *O. nivara*) were sampled and listed in table 1. The cultivated accessions originated

from 21 countries and included all the 5 distinct varietal groups (*indica*, *aus*, *aromatic*, *temperate japonica*, and *tropical japonica*) (Garris et al. 2005) to represent broad genetic diversity of Asian cultivated rice. The wild individuals were sampled to cover the entire distribution range of the 2 wild species. We also included one *Oryza barthii* accession from Africa as an outgroup because this species is closely related to the species complex of *O. sativa* and its relatives (Zhu and Ge 2005). Total DNA was extracted from fresh or silica gel-dried leaves, using the hexadecyltrimethylammonium bromide method as described in Ge et al. (1999).

Polymerase Chain Reaction Amplification, Cloning, and Sequencing

The genes used in this study represent 10 unlinked nuclear loci on 9 different chromosomes in rice. Detailed information about the location and functional association as well as the primer sequences of the amplified regions is provided in table 2. Polymerase chain reaction (PCR) amplification was performed in a total volume of 25 μ l using T-gradient 96U thermocycler (Biometra, Göttingen, Germany). The reaction mix contained 5–50 ng of template DNA, 0.2 μ mol/l of each primer, 200 μ mol/l of each deoxyribonucleotide triphosphate, 10 mmol/l Tris-Cl (pH 8.3), 50 mmol/l KCl, 25 mmol/l MgCl₂, 1.5% dimethyl sulfoxide, and a mixture of 0.5 U *Pfu* polymerase (Promega, Madison, WI) and 0.5 U Ex *Taq* DNA polymerase (TaKaRa, Shiga, Japan). Amplification was carried out in a T-gradient 96 U thermocycler (Biometra, Göttingen, Germany) as follows: 2 min at 94 °C followed by 38 cycles of 30 s at 94 °C, 30–50 s at 54 °C, 90 s at 72 °C, and a final extension at 72 °C for 10 min. Because of high T_m value and short PCR product in length for *Waxy*, a 2-step PCR was used for the gene. The condition included denaturation at 95 °C for 2 min, 38 cycles composed of denaturation at 95 °C for 45 s, annealing and extension at 68 °C for 90 s, and a final extension at 72 °C for 7 min. Amplification products were separated by electrophoresis on 1.5% agarose gels stained with ethidium bromide using a 100-bp DNA ladder and gel purified with a Pharmacia purification kit (Amersham Pharmacia Biotech, Piscataway, NJ). Sequencing was done on a Megabase1000 automatic DNA sequencer (Amersham Pharmacia Biotech) after the reaction products were purified through precipitation with 95% ethanol and 3 M sodium acetate (pH 5.2).

For the cultivated rice, we sequenced PCR products directly on both strands. For wild *O. rufipogon* and *O. nivara*, in which either homozygous or heterozygous individuals exist, PCR fragments were cloned into *pGEM* T-easy vectors (Promega) with either a Pharmacia purification kit (Amersham Pharmacia Biotech) or a Dinggou purification kit (Dingguo, Beijing, China). Independent plasmid DNAs were selected randomly and at least 6 clones were sequenced individually using the DYEnamic ET Terminator kit (Amersham Pharmacia Biotech), following the manufacturer's protocol. Because *Taq* errors occur at random, it is unlikely that polymorphisms shared among more than 1 clone (sequence) are artificial (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*

Table 1
Plant Materials Used in This Study

Taxon	Accession ^a	Country	Code ^b	Variety Group ^c	
<i>Oryza sativa</i> ssp. <i>indica</i>	10594	Indonesia	ind-IDN	—	
	11713	Sri Lanka	ind-LKA1	—	
	11724	Sri Lanka	ind-LKA2	—	
	30416	Philippines	ind-PHL	<i>indica</i>	
	71503	Malaysia	ind-MYS	<i>indica</i>	
	74716	India	ind-IND	—	
	RA4870	Vietnam	ind-VNM	<i>indica</i>	
	RA4890	Iran	ind-IRN	<i>aus</i>	
	RA4921	Thailand	ind-THA	<i>indica</i>	
	RA4933	Pakistan	ind-AUS1	<i>aus</i>	
	RA4939	Bhutan	ind-AUS2	<i>aus</i>	
	RA4966	Brazil	ind-BRA	<i>indica</i>	
	RA4976	Afghanistan	ind-AFG	<i>aus</i>	
	RA5020	Japan	ind-JPN	<i>indica</i>	
	RA5321	Bangladesh	ind-AUS3	<i>aus</i>	
	ZH249	China	ind-CHN	—	
	<i>Oryza sativa</i> ssp. <i>japonica</i>	Nipponbare	Japan	jap-JPN	<i>temperate japonica</i>
		11169	Philippines	jap-PHL	<i>temperate japonica</i>
		69317	France	jap-FRA	—
		70932	Pakistan	jap-PAK	—
Au70140		Greece	jap-GRC	—	
Au8134		Vietnam	jap-VNM	—	
RA4875		United States	jap-USA2	<i>tropical japonica</i>	
RA4913		India	jap-IND	<i>aromatic</i>	
RA4950		Brazil	jap-BRA	<i>aromatic</i>	
RA4972		South Korea	jap-KOR	<i>temperate indica</i>	
RA5123		United States	jap-USA1	<i>temperate japonica</i>	
RA5363		Australia	jap-AUS	<i>temperate japonica</i>	
RA5364		Nigeria	jap-NRA	<i>tropical japonica</i>	
Taizhong 65		China	jap-CHN	—	
<i>Oryza rufipogon</i>		80542	India	ruf-IND	—
		80774	Philippines	ruf-PHL	—
	100678	Taiwan, China	ruf-CHN1	—	
	103423	Sri Lanka	ruf-LKA	—	
	105494	Myanmar	ruf-MMR	—	
	105720	Cambodia	ruf-KHM	—	
	105942	Thailand	ruf-THA	—	
	105958	Indonesia	ruf-IDN	—	
	105960	Bangladesh	ruf-BGD	—	
	106036	Malaysia	ruf-MYS	—	
	106161	Laos	ruf-LAO	—	
	106505	Papua New Guinea	ruf-PNG	—	
	0318	Jiangxi, China	ruf-CHN2	—	
	2510	Guangdong, China	ruf-CHN3	—	
	5213	Hainan, China	ruf-CHN4	—	
	6102	Guangxi, China	ruf-CHN5	—	
	6803	Yunnan, China	ruf-CHN6	—	
<i>Oryza nivara</i>	VOC4	Nepal	ruf-NPL	—	
	103407	Sri Lanka	niv-LKA1	—	
	103416	Sri Lanka	niv-LKA2	—	
	104687	India	niv-IND1	—	
	105319	India	niv-IND2	—	
	105391	Thailand	niv-THA	—	
	105431	Sri Lanka	niv-LKA3	—	
	105705	Nepal	niv-NPL	—	
	105742	Cambodia	niv-KHM	—	
	105895	Bangladesh	niv-BGD	—	
	106061	India	niv-IND	—	
	106148	Laos	niv-LAO	—	
	106345	Myanmar	niv-MMR	—	
	<i>Oryza barthii</i>	106234	Sierra Leone	barthii	—

^a Accession numbers preceded by RA were provided by Dr Susan McCouch in Cornell University; ZH249, VOC4, and the accessions with 4 digits were collected by the authors; All the other accessions were obtained from the Genetic Resources Center of the International Rice Research Institute (IRRI) at Los Banos, Philippines.

^b Individual accession is abbreviated by first 3 letters of the taxon name followed by the code of its origin of country.

^c Followed Garris et al. (2005).

Table 2
Summary of the Genes Surveyed and the Primer Sequences Used in the Study

Locus	Chromosome Location	Alignment Length (bp)			Primer Sequence	Functional Association
		Total	Coding	Noncoding		
<i>Adh1</i>	11	853	0	853	TCCCGTGTCCCTCGGATCTTC; GTCACACCTCTCCAACACTCT	Alcohol dehydrogenase I
<i>CatA</i>	2	610	0	610	TCGTTATCATCAGCGTGGAC; CATGCCACTAAGTCTCAG	Catalase gene
<i>Cbp1</i>	12	824	317	507	CATTTTACATAGCTGGGGAGTC; CTGGGGTCACTTCTTGATAG	Serine carboxypeptidase I
<i>GBSSII</i>	7	624	139	485	GTGACAGGACAGAAACAGTG; TCCAGTGTGCCAGTCATTG	Granule-bound starch synthase II
<i>Ks1</i>	4	941	481	460	TTTCCTGGTATGCTTAGCCTTGC; GCGAAAAGCCATTGGCACATGTTG	<i>ent</i> -kaurene synthase I
<i>Lhs1</i>	3	1,057	161	896	AGAGGTACCGCAGCTGCAAC; CTCCTTCATGCTDAGTGGGC	MADS box transcription factor 1
<i>Os0053</i>	2	738	149	589	AGAGAAAAGCGGTGTCAGACG; CACGTTCTCCTTGGCGAGCG	Putative aldehyde decarboxylase
<i>SSIII</i>	10	917	73	844	AAGAGCAAAGAGGTGTTGGAG; CATCGTCTGGAACCTCATCAAG	Soluble starch synthase II-1
<i>TFIIAγ-1</i>	1	971	75	896	AGGTGAAGAGCAAGGTTA; TCRCAGGCCACRATCTTCAC	Gamma subunit of transcription factor II A
<i>Waxy</i>	6	544	179	365	CTTCGAGCCCTGTGGACTC; CACGTTCTCCTTGGCGAGCG	Granule-bound starch synthase
Total		8,079	1,574	6,505		

polymerase artifact. To confirm the singletons, we repeated PCR amplification, cloning, and sequencing and finally excluded those singletons resulting from *Taq* polymerase error. By means of multicloning sequencing and reamplifying and resequencing, interallelic PCR recombinants were also verified and removed. All sequences were deposited into GenBank, with accession numbers EF069438–EF070112.

Sequence Analysis

Initial sequence data were assembled with the Contig-Express program (Informax Inc. 2000, North Bethesda, MD). Allele sequences were aligned using a combination of methods implemented in BioEdit version 7.0.1 (Hall 1999) and ClustalX 1.81 (Thompson et al. 1997), with further manual refinements. Sequences were edited with DAMBE version 4.1.19 (Xia and Xie 2001). Insertions/deletions (indels) were not included in the analysis. For each locus and taxon, we calculated the number of segregating sites (S), the number of haplotypes (h) as well as 2 parameters of nucleotide diversity: π , the expected heterozygosity per nucleotide site (Nei 1987) and θ_w , an estimate of $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per nucleotide (Watterson 1975). Estimates of nucleotide diversity were based on total sequences and silent sites separately using the DnaSP version 4.10.8 (Rozas et al. 2003). The observed heterozygous rate at each locus was estimated as $H_o = \text{number of heterozygous individuals}/\text{number of total individuals}$. An individual was considered as heterozygous when 2 alleles were different in 1 site at any locus.

Tests of Neutrality

To test for deviations from the neutral equilibrium model of evolution, we performed several tests using the programs DnaSP and HKA (Hudson–Kreitman–Aguade

(<http://lifesci.rutgers.edu/~heylab/>). In each locus, Tajima's D (Tajima 1989) and D^* and F^* of Fu and Li (1993) were calculated at all sites and at silent sites separately. Tajima's D 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 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). For both tests, negative values indicate an excess of low-frequency polymorphisms, whereas positive values indicate an excess of intermediate polymorphisms. To assess the neutral prediction of the ratio of polymorphism to divergence across loci, a multilocus HKA test (Hudson et al. 1987) was applied to all 10 loci. Because selective force generally effect on single locus in the history of species evolution, the multilocus HKA test across unlinked or loosely linked loci can discriminate between selection forces and population demography during the speciation process. *Oryza barthii* sequences were used as outgroups for the HKA and tests of Fu and Li.

Genealogy and Divergence of Taxa

The genealogical trees of 10 nuclear loci were constructed by PAUP* version 4.0b10 (Swofford 2001), using the Neighbor-Joining (NJ) and maximum parsimony (MP) methods. The NJ analysis was conducted with Kimura's 2-parameter distances (Kimura 1980), and the MP was performed using heuristic searches (Tree Bisection-Reconnection branch swapping; MulTrees option in effect) with 10 random additions of taxa. The stability of internal nodes was assessed by bootstrap analysis with 1,000 replicates. Because of the close relationship of the 4 taxa (Morishima et al. 1992; Lu et al. 2000), we computed fixed differences and shared polymorphisms for pairwise comparisons

among 4 taxa. In addition, to test for monophyly of taxa, constraint trees were made to hold monophyly for the 4 taxa separately and evaluated with a Kishino–Hasegawa test (Kishino and Hasegawa 1989), as implemented in PAUP*.

Linkage Disequilibrium

The decay of LD with physical distance was estimated using a nonlinear regression analysis of LD between polymorphic sites versus the distance between sites in basepairs (Remington et al. 2001). The expected value of r^2 at drift–recombination equilibrium is $E(r^2) = 1/(1 + \rho)$, where ρ is $4Nc$ and c is the recombination rate in Morgans between the 2 markers and N is the effective population size (Hill and Weir 1988). Under the assumption of a low level of mutation and finite sample size, the expectation (Hill and Weir 1988) becomes

$$E(r^2) = \left[\frac{10 + \rho}{(2 + \rho)(11 + \rho)} \right] \left[1 + \frac{(3 + \rho)(12 + 12\rho + \rho^2)}{n(2 + \rho)(11 + \rho)} \right],$$

where n is the number of sequences sampled. We pooled our sequences across loci and fit this model individually for the 4 taxa as well as separately for the wild and cultivated samples using PROC NLIN in SAS Ver. 6.12 (SAS Institute, Cary, NC) and SigmaPlot 10.0 (SPSS Inc., Chicago, IL). Although a least-square estimate of ρ per basepair by the nonlinear regression might be biased because of the nonindependence among linked sites and nonequilibrium populations (Remington et al. 2001), such analysis is still useful for investigating the overall rate of decay of LD (Ingvarsson 2005; Liu and Burke 2006). DnaSP was also used to estimate the minimum number of recombination events in the history of each gene sample (R_M) (Hudson and Kaplan 1985). Squared allele-frequency correlations for LD (r^2) were also calculated with program SITES (Hey and Wakeley 1997).

Coalescent Simulations

We used coalescent simulations to model the impact of a bottleneck on sequence diversity using a modified version of Hudson's *ms* program (Hudson 2002). We modeled the divergence of 2 populations representing the cultivated (*O. sativa*) and wild rice (*O. rufipogon*/*O. nivara*), with a population bottleneck in cultivated rice. The bottleneck model was previously described in domesticated maize (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005) and was modified slightly in this study as follows.

To model the bottleneck, we combined *O. rufipogon* and *O. nivara* as a single population as the progenitor of *O. sativa* because previous and present studies supported treating them as a large gene pool (Morishima 2001; Londo et al. 2006). Due to 2 alternative hypotheses regarding origin of the Asian cultivated rice, we modeled the bottleneck under 2 conditions: 1) to treat the Asian cultivated rice as a whole, assuming that it originated monophyletically (Oka 1988; Lu et al. 2002) and 2) to perform the simulations on 2 subspecies of the cultivated rice, *indica* and *japonica*, separately, assuming that they had separate domestication origins (Second 1982; Yamanaka et al. 2003; Zhu and Ge 2005).

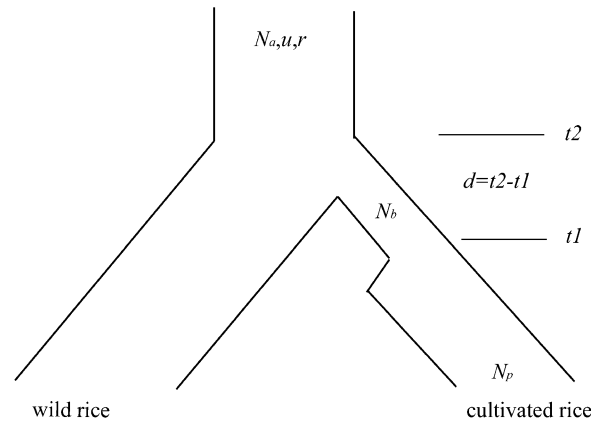


FIG. 1.—Diagrammatic representation of the coalescent model used in simulation. See text for details.

The goal of these simulations was to estimate the bottleneck severity that best explained the cultivated rice data.

The assumptions and parameters used for simulation are illustrated in figure 1. It was supposed that at time t_2 generations ago, a single ancestral population of size N_a experienced an instantaneous shift in size to the bottlenecked population size N_b , and at time t_1 generations ago, the bottlenecked population expanded instantaneously to the present population size N_p . The bottleneck was characterized by 2 parameters: d , the duration of the bottleneck in generations (measured as $t_2 - t_1$) and N_b , the population size during the bottleneck. We used the bottlenecked stringency K (Wright et al. 2005), which is the ratio of N_b and d , to describe the severity of bottleneck in the domestication because an obvious positive correlation between N_b and d were detected in previous maize studies (Tenaillon et al. 2004; Wright et al. 2005). In this study, a range of the parameter d (200, 500, 1,000, 1,500, 2,000, and 3,000) was explored. We used 3,000 generations to be the maximum duration of a bottleneck associated with rice domestication because archeological evidence suggested that the duration of the domestication bottleneck in rice seems less than 3,000 years (see Discussion). For a given d value, K value ranged from 0.001 to 7 and were used to assess the fit of different models. A total of 150 scenarios were examined for each locus with 10,000 simulations.

For each simulation at each locus, we started with silent θ_w , recombination rate, and Tajima's D calculated for *O. rufipogon*/*O. nivara*, and the cultivated rice data were used to assess fit of statistics. Then we calculated values of S , θ_w , π , and Tajima's D for wild rice and cultivated rice, respectively. The simulation was accepted if the number of segregating sites (S) fell within 20% of the observed wild rice data (Weiss and von Haeseler 1998). To assess fit, we defined levels of acceptance corresponding to a range of 20% for each of 4 summary statistics (S_{rice} , $\theta_{w\text{-rice}}$, π_{rice} , and D_{rice}). For each of the 150 explored bottleneck scenarios, we calculated the approximate likelihood for locus i as the proportion of simulations among 10,000 that fit the data (Weiss and von Haeseler 1998). Finally, after calculating likelihoods on a per-locus basis, we calculated the multilocus likelihood by multiplying across loci. This approach

made the implicit assumption that loci are independent (Tenaillon et al. 2004). Furthermore, a recent study on human population size change indicated that combining multiple statistics might result in a significant reduction of the compatible parameter space and thus effectively improve the power to detect bottlenecks (Voight et al. 2005). Therefore, we combined the 4 statistics, S , θ_w , π , and Tajima's D to better measure the fit to coalescent simulations.

Results

Nucleotide Diversity

Sequences of the 10 unlinked loci were obtained from 60 accessions, representing 2 subspecies of *O. sativa* and its wild relatives. The length of aligned sequence for each locus varied between 544 bp and 1,057 bp, with a total of 8,079 bp in length. Of the 10 loci, 8 contained both coding and noncoding sites and the remaining 2 were intron (*Adh1*) and 5'-flanking (*CatA*) regions. The schematic diagrams of the 10 genes are provided in supplementary figure S1 (Supplementary Material online). Both indel and single nucleotide polymorphisms (SNPs) were present in the sequenced regions. A total of 69 indel polymorphisms were found exclusively in introns or flanking regions and excluded from all further analyses. SNP polymorphisms were substantially higher than indel polymorphisms. A total of 357 SNPs and thus an average of 1 SNP every 23 nucleotides were found across the 4 taxa, with the number of SNP being 53 in *indica*, 30 in *japonica*, 263 in *O. rufipogon*, and 199 in *O. nivara*.

Standard statistics of sequence variation for each locus are summarized in table 3, 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 all substantially lower than those at noncoding regions (data not shown). Levels of nucleotide diversity are heterogeneous among loci, with *GBSSII* being the least variable loci (mean $\theta_{\text{sil}} = 0.0036$) and *Waxy* the highest (mean $\theta_{\text{sil}} = 0.0161$). Diversity pattern of the entire region was similar to that of the silent sites at each locus. To determine whether silent diversity and interspecific divergence are correlated with each other as expected under neutral evolution, we calculated the correlation between θ_{sil} and K_{sil} and found significant positive correlation for *O. rufipogon* ($P < 0.01$) and *O. nivara* ($P < 0.05$), but no significant correlation for 2 subspecies of the cultivated rice ($P > 0.55$). The positive correlations between θ_{sil} and K_{sil} in 2 wild species conformed to the pattern expected under neutrality, suggesting that neutral mutation rates (and/or selective constraint) varied among loci. No significant correlation between θ_{sil} and K_{sil} in *indica* and *japonica*, on contrast, suggested the deviation from a strict neutral equilibrium model, probably due in part to demographic processes associated with domestication.

At the taxon level, polymorphisms in the wild species ($\theta_{\text{sil}} = 0.0077$ for *O. nivara* and $\theta_{\text{sil}} = 0.0095$ for *O. rufipogon*) were significantly higher than those in the cultivated rice ($\theta_{\text{sil}} = 0.0021$ for *indica* and $\theta_{\text{sil}} = 0.0011$ for *japonica*). At both the entire region and silent sites, nucleotide diversity in *indica* rice was roughly 2 times that of the

japonica rice, whereas the 2 wild species maintained comparable levels of diversity that were 3.7-fold to 8.6-fold higher than those in the cultivated rice (tables 3 and 4). It is obvious from table 3 that diversity in wild species is much higher than that in the cultivated rice at all loci. When the data of *indica* and *japonica* were pooled, the cultivars maintained 5.7% (*Chp1*) to 50% (*GBSSII*) as much diversity across loci compared with the wild species (table 4). Over all 10 loci, domesticated rice has about $\sim 22\%$ (θ_{sil}) of the variability found in its progenitor (table 4).

Although no heterozygote was detected in the cultivated rice sample, different numbers of heterozygotes were observed for the 2 wild species. Across 10 loci, 12 out of 18 (66.7%) *O. rufipogon* individuals were heterozygous, whereas 3 of 13 (23.1%) *O. nivara* individuals were heterozygous. The frequency of heterozygotes in the 2 species was consistent with their mating system because the outcrossing rate of *O. rufipogon* was estimated to be 30–50%, whereas that of *O. nivara* was from 5% to 25% (Morishima et al. 1984; Barbier 1989).

Neutrality Tests

To examine the fit of nucleotide polymorphism data to the neutral equilibrium model, we first performed the tests of Tajima's D (Tajima 1989) and D^* and F^* of Fu and Li (1993) for each gene. No significant Tajima's D value and D^* and F^* values of Fu and Li were observed at any locus except for *O. rufipogon* at *Adh1*, where the neutral model was rejected at $P < 0.05$ (table 3). It is interesting to note that 9 of 10 loci for *O. rufipogon* and 8 of 10 loci for *O. nivara* had negative Tajima's D values, whereas 2 cultivated subspecies had positive Tajima's D at most loci (tables 3 and 4), perhaps as a consequence of population expansion in the wild species or the skewed distribution toward intermediate frequency in the cultivated species. Theoretically, Tajima's D should be higher in populations that have experienced a recent bottleneck because of the preferential loss of low-frequency variant (Tenaillon et al. 2001). This is obviously true for cultivated rice where the D values were higher than those in the wild species at all 10 loci (tables 3 and 4).

We also 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 the neutral model of molecular evolution. No significant departure from the equilibrium model was detected between any taxon pair (*japonica-indica*, $X^2 = 1.87$, $P = 0.993$; *japonica-nivara*, $X^2 = 1.93$, $P = 0.993$; *japonica-rufipogon*, $X^2 = 3.18$, $P = 0.957$; *indica-nivara*, $X^2 = 3.16$, $P = 0.958$; *indica-rufipogon*, $X^2 = 3.13$, $P = 0.959$; *rufipogon-nivara*, $X^2 = 0.79$, $P = 0.999$). Thus, there appears to be no evidence for the action of past selection at the 10 loci as a whole. 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 1 sequence of *O. barthii* as an outgroup at each locus, HKA tests across loci were applied to each of the 4 taxa. The multilocus test was not significant for the 2 wild species

Table 3
Summary of Nucleotide Polymorphisms and Neutrality Tests

Taxon	Locus	$H_o\%$	S	π_T	θ_T	π_{sil}	θ_{sil}	D	D^*	F^*	r^2	R_M	
<i>Oryza sativa</i> ssp. <i>indica</i> ($n = 32$)	<i>Adh1</i>	0	4	0.0018	0.0012	0.0018	0.0012	1.165	1.051	1.260	0.258	0	
	<i>CatA</i>	0	20	0.0122	0.0084	0.0122	0.0090	-0.138	0.587	0.440	0.198	0	
	<i>Cbp1</i>	0	1	0.0002	0.0003	0.0002	0.0004	-0.783	2.061	0.026	/	0	
	<i>GBSSII</i>	0	4	0.0015	0.0016	0.0013	0.0015	-0.229	1.363	0.481	1.000	0	
	<i>Ks1</i>	0	5	0.0018	0.0013	0.0029	0.0022	0.673	1.744	0.418	0.355	0	
	<i>Lhs1</i>	0	5	0.0018	0.0012	0.0020	0.0013	1.361	1.889	0.512	0.623	0	
	<i>Os0053</i>	0	2	0.0006	0.0007	0.0007	0.0008	-1.424	2.220	0.270	1.000	0	
	<i>SSIII</i>	0	3	0.0015	0.0008	0.0015	0.0009	2.043	1.867	0.395	0.619	0	
	<i>TFIIAγ-1</i>	0	3	0.0011	0.0008	0.0012	0.0008	1.017	2.088	0.218	0.233	0	
	<i>Waxy</i>	0	6	0.0044	0.0028	0.0048	0.0032	1.308	2.016	0.421	0.177	0	
	Average	0	5.3	0.0027	0.0020	0.0029	0.0021				0.496	0	
	<i>Oryza sativa</i> ssp. <i>japonica</i> ($n = 28$)	<i>Adh1</i>	0	1	0.0002	0.0003	0.0002	0.0003	-0.741	0.603	0.269	0.006	0
		<i>CatA</i>	0	0	0	0	0	0	—	—	—	—	0
		<i>Cbp1</i>	0	0	0	0	0	0	—	—	—	—	0
<i>GBSSII</i>		0	5	0.0019	0.0021	0.0017	0.0020	0.539	1.363	0.817	0.846	0	
<i>Ks1</i>		0	4	0.0016	0.0011	0.0027	0.0018	1.026	1.744	0.366	0.708	0	
<i>Lhs1</i>		0	6	0.0012	0.0015	0.0014	0.0017	-0.512	1.889	0.348	0.399	0	
<i>Os0053</i>		0	10	0.0035	0.0035	0.0057	0.0042	1.155	2.220	0.568	1.000	0	
<i>SSII</i>		0	2	0.0019	0.0014	0.0012	0.0006	0.356	1.867	0.409	1.000	0	
<i>TFIIAγ-1</i>		0	2	0.0003	0.0005	0.0003	0.0006	-0.972	2.088	0.056	1.000	0	
<i>Waxy</i>		0	0	0	0	0	0	—	—	—	—	0	
Average		0	3	0.001	0.0010	0.0013	0.0011				0.708	0	
<i>Oryza rufipogon</i> ($n = 36$)		<i>Adh1</i>	22.2	37	0.0047	0.0118	0.0047	0.0118	-2.136*	-3.056*	-3.251*	0.467	1
		<i>CatA</i>	38.9	40	0.0151	0.0175	0.0175	0.0181	-0.931	0.394	-0.012	0.086	3
		<i>Cbp1</i>	27.8	18	0.0025	0.0056	0.0034	0.0073	-1.777	-0.069	-1.073	0.252	7
	<i>GBSSII</i>	27.8	7	0.0023	0.0027	0.0017	0.0024	-0.589	0.638	0.076	0.263	2	
	<i>Ks1</i>	33.3	12	0.0022	0.0031	0.0028	0.0040	-0.393	0.817	-0.176	0.093	0	
	<i>Lhs1</i>	44.4	16	0.0025	0.0037	0.0026	0.0039	-1.127	-0.814	-1.253	0.061	2	
	<i>Os0053</i>	44.4	47	0.0093	0.0154	0.0108	0.0181	-1.422	-2.029	-1.861	0.063	0	
	<i>SSIII</i>	22.2	22	0.0034	0.0058	0.0038	0.0059	-1.342	-3.691	-3.051	0.122	0	
	<i>TFIIAγ-1</i>	33.3	30	0.0039	0.0072	0.0042	0.0076	-1.591	1.796	0.545	0.664	1	
	<i>Waxy</i>	55.6	34	0.0183	0.0153	0.0203	0.0162	0.904	0.068	0.310	0.243	0	
	Average	66.7 ^a	26.3	0.0064	0.0088	0.0072	0.0095				0.231		
	<i>Oryza nivara</i> ($n = 24$)	<i>Adh1</i>	23.1	33	0.0096	0.0117	0.0096	0.0117	-0.776	1.125	0.618	0.588	4
		<i>CatA</i>	0	38	0.0164	0.0172	0.0170	0.0176	-0.971	1.158	0.627	0.143	3
		<i>Cbp1</i>	0	7	0.0025	0.0024	0.0002	0.0004	-0.845	2.061	1.100	0.868	3
<i>GBSSII</i>		15.4	10	0.0045	0.0044	0.0037	0.0031	0.6117	1.363	1.358	0.236	0	
<i>Ks1</i>		7.7	19	0.0037	0.0054	0.0040	0.0052	-1.240	1.281	0.600	0.116	0	
<i>Lhs1</i>		7.7	18	0.0036	0.0046	0.0040	0.0049	-0.715	1.116	0.455	0.139	0	
<i>Os0053</i>		0	22	0.0071	0.0080	0.0085	0.0099	-0.528	2.220	0.745	0.157	0	
<i>SSIII</i>		15.4	9	0.0035	0.0026	0.0036	0.0025	1.344	1.470	0.577	0.220	0	
<i>TFIIAγ-1</i>		0	12	0.0023	0.0033	0.0024	0.0035	-1.110	2.088	0.443	0.068	0	
<i>Waxy</i>		23.1	31	0.0084	0.0156	0.0101	0.0183	-1.392	0.720	-0.069	0.224	0	
Average		23.1 ^a	19.9	0.0062	0.0075	0.0063	0.0077				0.276		

NOTE.— $H_o\%$, the proportion of heterozygous samples; n , number of sequences sampled (chromosomes); S , number of segregating sites; π_T , average number of nucleotide differences per site between 2 sequences (Nei 1987) calculated on the total number of polymorphic sites; π_{sil} , average number of pairwise nucleotide differences per site calculated on the silent sites; θ_T , the Watterson estimator of θ per basepair (Watterson 1975) calculated on the total number of polymorphic sites; θ_{sil} , Watterson's estimator of θ per basepair calculated on the silent sites; D , Tajima's D (Tajima 1989); D^* and F^* , D^* and F^* of Fu and Li (1993); r^2 , squared allele-frequency correlations (Hill and Weir 1988); R_M , estimates of the minimum number of recombination events (Hudson and Kaplan 1985).

^a Represents the sum instead of average of individual loci.

* $P < 0.05$.

(*O. rufipogon*–*O. barthii*, $X^2 = 0.61$, $P = 0.999$; *O. nivara*–*O. barthii*, $X^2 = 1.70$, $P = 0.995$). In contrast, the HKA test rejected the null hypothesis of proportionality between polymorphism and divergence in *japonica*/*O. barthii* contrast ($X^2 = 17.86$, $P = 0.037$) and the test value was close to significance in *indica*/*O. barthii* contrast ($X^2 = 15.44$, $P = 0.080$). In both cases, *CatA* contributed greatly to the significant multilocus HKA statistics for cultivated rice, whereas *Os0053* to the indica rice and *O. nivara* (fig. 2). The rejection of a multilocus HKA test can be explained by both selection and demographic history (Wright and Gaut 2005). Removal of these single loci from the multilocus HKA test causes the test statistic to drop below the critical value in all taxa.

Divergence of Species and Genealogical Patterns

To assess the divergence and genetic relationships among taxa we first calculated the numbers of shared polymorphisms and fixed differences between taxon pairs. A fixed difference refers to the nucleotide site where all sampled sequences from one taxon are different from all sequences from another taxon, whereas a shared polymorphism occurs when 2 taxa have the same 2 bases segregating at the same site (Hilton et al. 1994). Shared polymorphisms reveal a history of polymorphism that has not yet been erased by genetic drift and thus reflect either a short divergence time between taxa or historically large population sizes (Hilton et al. 1994). As shown in

Table 4
Summary of Nucleotide Polymorphisms and Neutrality Tests on the Cultivated and Wild Rice

Locus	Cultivated rice (<i>indica</i> + <i>japonica</i>)							Wild species (<i>Oryza rufipogon</i> + <i>Oryza nivara</i>)						
	$H_o\%$	S	π_{sil}	θ_{sil}	D	r^2	R_M	$H_o\%$	S	π_{sil}	θ_{sil}	D	r^2	R_M
<i>Adh1</i>	0	5	0.0014	0.0013	0.2094	0.086	0	22.6	44	0.0068	0.0121	-1.489	0.337	4
<i>CatA</i>	0	20	0.0093	0.0078	-0.526	0.224	0	22.6	61	0.0167	0.0234	-1.3347	0.058	7
<i>Cbp1</i>	0	1	0.0001	0.0004	-0.8912	/	0	19.4	20	0.0035	0.007	-1.9216	0.224	12
<i>GBSSII</i>	0	5	0.0031	0.0017	2.2176	0.667	0	22.6	12	0.0027	0.0034	-0.5206	0.152	3
<i>Ks1</i>	0	5	0.0036	0.0019	2.0911	0.424	0	22.6	26	0.0033	0.007	-1.6748	0.055	0
<i>Lhs1</i>	0	6	0.0026	0.0014	2.1142	0.407	0	29.0	27	0.0032	0.0063	-1.4591	0.064	2
<i>Os0053</i>	0	12	0.0088	0.0042	1.4172	0.705	0	25.8	51	0.0096	0.0176	-1.6949	0.048	0
<i>SSII</i>	0	5	0.0025	0.0013	2.3062	0.436	0	25.8	26	0.0036	0.006	-1.3735	0.086	0
<i>TFIIAγ-1</i>	0	3	0.0009	0.0007	0.4484	0.198	0	19.4	40	0.0037	0.0092	-2.0611	0.394	1
<i>Waxy</i>	0	5	0.0048	0.003	1.2843	0.163	0	41.9	45	0.0182	0.0181	-1.7542	0.137	0
Average	0 ^a	6.7	0.0037	0.0024		0.368		54.8 ^a	35.2	0.0071	0.0110		0.155	

NOTE.— $H_o\%$, the proportion of heterozygous samples; 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 basepair calculated on the silent sites; D , Tajima's D (Tajima 1989); r^2 , squared allele-frequency correlations (Hill and Weir 1988); R_M , estimates of the minimum number of recombination events (Hudson and Kaplan 1985).

^a Represents the sum instead of average of individual loci.

table 5, no fixed difference was detected between any pair of comparisons at all 10 loci except for the *japonica/nivara* contrast where 3 and 1 fixed differences were observed at loci *CatA* and *Adh1*, respectively. A large

amount of shared polymorphisms were detected between contrasts over all 10 loci (table 5), indicating a close affinity or hybridization/introgression among these 4 taxa. It is not surprising that no fixed difference was observed between cultivated and wild rice because of the recent domestication history of rice, probably around 10,000 years (Normile 1997). Similarly, no fixed differences were observed between wild and cultivated sorghum (White et al. 2004). Note that the highest number of shared polymorphisms (104) was found between 2 wild species (*O. rufipogon* and *O. nivara*), whereas the lowest number of shared polymorphisms (28) occurred between 2 subspecies of the cultivated rice (*indica* and *japonica*). These observations are in agreement with F_{st} analyses where the lowest genetic differentiation was found between *O. rufipogon* and *O. nivara* ($F_{st} = 0.033$), but the highest between *indica* and *japonica* ($F_{st} = 0.190$).

We also constructed the genealogical trees for each locus using both NJ and MP methods. Results showed that on almost all the trees, sequences from the same taxon did not form a monophyletic group despite a tendency for sequences to cluster by the taxon designations in the combined trees (supplementary fig. S2, Supplementary Material online). Some accessions appeared to be in different positions on single trees (data not shown), suggesting that coalescence for alleles at most loci occurred in the common ancestor of these taxa. Such genealogical patterns of lack of concordance among the 10 gene trees might result from lineage sorting because the 4 taxa were closely related and diverged very recently from a common ancestor. Extensive gene flow through hybridization and introgression might be also factors leading to inconsistent phylogenetic relationships among taxa. To test for the monophyly of the 4 taxa, we compared the tree topologies statistically by the Kishino–Hasegawa test with or without taxon constraints. Constraining samples by treating each taxon as a monophyletic group resulted in a significant reduction in likelihood ($P < 0.001$) of differences between the constraint and nonconstraint trees, rejecting the monophyly of each of the 4 taxa (supplementary table S1, Supplementary Material online).

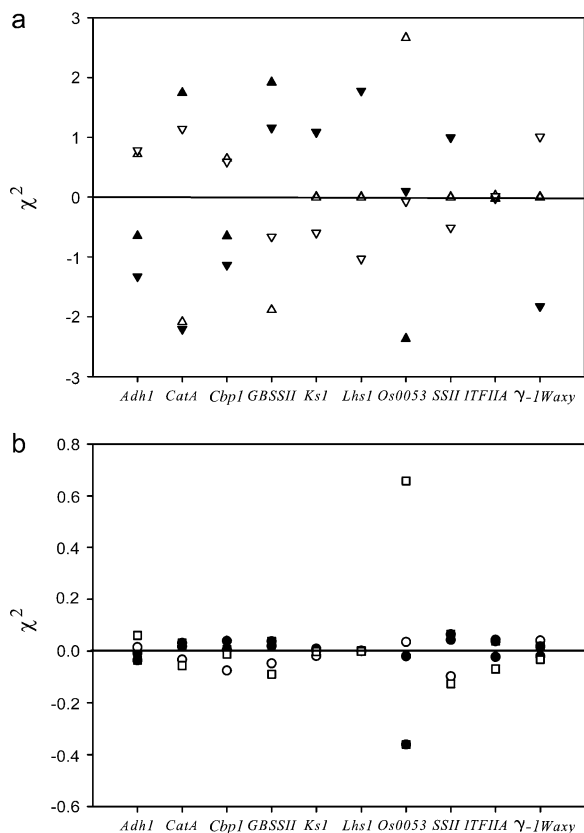


FIG. 2.—Summary of multilocus HKA tests (Hudson et al. 1987). In each test, polymorphism within individual species is compared with divergence from *Oryza barthii* sequences. Solid symbols represent the contributions to the overall χ^2 statistic that resulted from polymorphisms, and open symbols show contributions caused by divergence. (a) (▲) and (△), *indica*; (▼) and (▽), *japonica*. (b) (●) and (○), *Oryza rufipogon*; (■) and (□), *Oryza nivara*. Symbols above the horizontal line indicate that the observed deviations are greater than the simulated, and those below the line indicate that the observed deviations are less than the simulated.

Table 5
Fixed Differences and Shared Polymorphisms at the 10 Loci^a

Locus	Cultivated/Wild	<i>indica</i> / <i>japonica</i>	<i>rufipogon</i> / <i>nivara</i>	<i>indica</i> / <i>nivara</i>	<i>indica</i> / <i>rufipogon</i>	<i>japonica</i> / <i>nivara</i>	<i>Japonica</i> / <i>rufipogon</i>
<i>Adh1</i>	3	0	26	2	3	0 (1)	0
<i>CatA</i>	21	0	18	16	19	0 (3)	0
<i>Cbp1</i>	0	0	6	0	0	0	0
<i>GBSSII</i>	5	4	4	4	4	4	4
<i>Ks1</i>	4	4	5	4	4	4	4
<i>Lhs1</i>	6	5	6	5	5	6	5
<i>Os0053</i>	12	10	15	10	12	10	10
<i>SSIII</i>	4	3	5	3	3	5	4
<i>TFIIAγ-1</i>	3	2	2	3	2	2	2
<i>Waxy</i>	4	0	16	4	4	0	0
Total	62	28	104	51	56	31 (4)	29

^a The number of shared polymorphisms is given and that of the fixed differences is given in parentheses only when higher than 0.

Coalescent Simulations and Severity of Bottleneck during Rice Domestication

Because different tests did not reject neutrality for all 10 loci (table 3), it is appropriate, as a first approximation, to use simulations of a neutral coalescent process to estimate population parameters during rice domestication. In the simulation, we assume a reduction in population size associated with the initial domestication of rice and a population increase in size after it was widely cultivated. We also assume that a domestication bottleneck first occurred 10,000 generations in the past because the earliest archaeological evidence indicated that rice domestication occurred around 10,000 BP (Khush 1997; Normile 1997). As indicated above, no significant genetic differentiation was observed between 2 wild species (*O. rufipogon* and *O. nivara*) with these data, and thus their data sets were combined into a single, large ancestral gene pool to domesticate rice.

The simulation results revealed an obvious positive correlation between N_b and d at each locus (supplementary fig. S3, Supplementary Material online), as was expected, so we estimated the parameter $K = N_b/d$. Our estimate of K values varied among summary statistics (S , θ_w , π , and Tajima's D) (supplementary fig. S4, Supplementary Material online). The summary statistics S , θ_w , and π produced an appreciable likelihood peak, but for several loci the curves based on Tajima's D did not reach a maximum within the parameter space (supplementary fig. S4, Supplementary Material online). This result might reflect different evolutionary pattern for different regions, or the polymor-

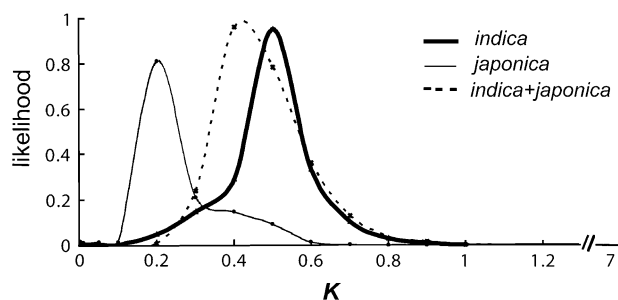


FIG. 3.—Diagram of maximum likelihood estimates for population bottleneck during rice domestication. The x axis indicates the severity parameter K , and slashes represent the omitted region with K value from 2 to 7, where the likelihood values are close to 0. The y axis represents the likelihood value based on the combined $D-S-\theta_w-\pi$.

phism distribution of wild rice might not be mimicked adequately in the simulation, or single summary statistics may be inadequate. To take fuller use of multiple aspects of the data and to narrow the confidence regions (Voight et al. 2005), we used the combinations of $D-S-\theta_w-\pi$ in the subsequent analyses of bottleneck models. Based on this analysis, the K values were lower in the japonica rice (0.2) than in the indica rice (0.5), suggestive of more severe bottleneck effect on the japonica rice (fig. 3). When the 2 subspecies of *O. sativa* were considered as a whole, the severity parameter K was 0.4 (fig. 3). Because the multilocus HKA statistics suggested that 2 loci (*CatA* and *Os0053*) may not evolve neutrally in the cultivated rice, we repeated the

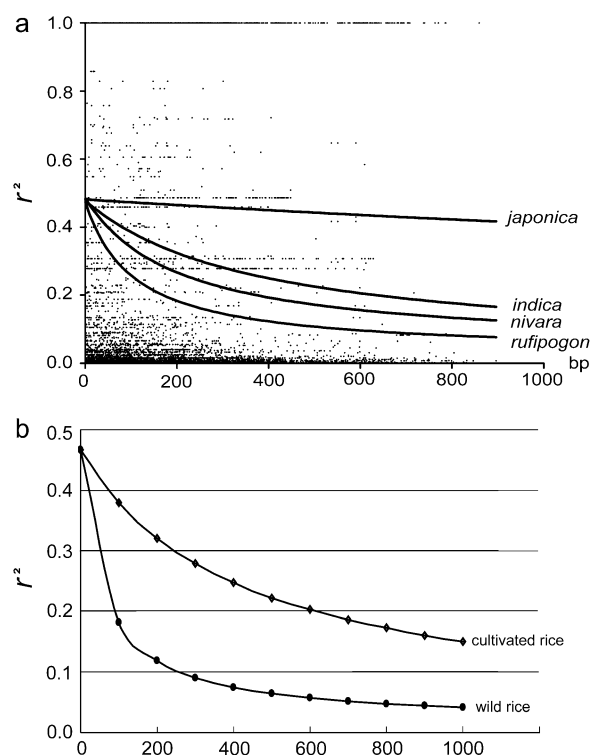


FIG. 4.—Plots of the squared allele-frequency correlations r^2 (y axis) as a function of physical distance (x axis) between sites for 10 genes in the wild and cultivated rice. The curved lines depict the fitted nonlinear regressions of mutation–recombination–drift model based on the equation in Materials and Methods. (a) Four taxa separately; (b) *indica* and *japonica* are combined as the cultivated rice, and *Oryza rufipogon* and *Oryza nivara* are combined as the wild rice.

above simulations by removing the 2 loci. Interestingly, the simulation results changed very little (data not shown). With estimated K values, we may calculate the number of individuals that have been involved in the domestication event. For example, if the domestication bottleneck was 200 generations in length, we estimated that the domestication of rice was based on 40~100 individuals of wild rice. If a bottleneck of 2,000 and 3,000 generations, we estimated that the cultivated rice was based on a wild rice population of 400~1000 individuals and 600~1,500 individuals, respectively (supplementary table S2, Supplementary Material online).

Linkage Disequilibrium

Figure 4 shows r^2 , the measure of LD, as a function of distance for comparisons within loci, pooled over the entire data set. It is obvious that both regression curves fitting to the cultivated *indica* and *japonica* were above those of 2 wild species. The observed LD in the cultivated rice extended on average much further than those in the 2 wild species where the expected values of r^2 dropped to negligible levels (i.e., <0.1) within 400 bp. In particular, the japonica rice showed extensive LD over distances that approach the length of the sequenced region (fig. 4a). Estimation of population recombination parameters showed that the minimum number of recombination events ranged from 0 to 7 across loci in the wild species, but no recombination was found in the 2 cultivated rice (table 3).

When data were pooled for the cultivated and wild samples separately, the average r^2 was significantly different between the cultivated and wild species ($P < 10^{-12}$ with a t -test), with the values being 0.368 in the cultivated rice and 0.155 in the wild species (table 4). As can be seen from figure 4b, the average r^2 in the wild species drops quickly for the first 200 bp to a value of 0.18 within 100 bp and a value of 0.12 within 200 bp; whereas the value in the cultivated rice declines much slowly to a value of only 0.38 within 100 bp and remains a value of >0.15 by 1,000 bp. The higher level of LD in the cultivated rice is consistent with an increase in LD due to domestication.

Discussion

Diversity and Differentiation between *O. rufipogon* and *O. nivara*

Although *O. rufipogon* and *O. nivara* have long been the subject of extensive studies, no investigation has been conducted on nucleotide variation of species-wide samples of these 2 species using sequence data of multiple nuclear loci despite a few of studies focusing on 1 or 2 genes (Barbier et al. 1991; Yoshida et al. 2004; Yoshida and Miyashita 2005). A literature survey on the multiple gene studies of wild plant species has revealed a wide range of nucleotide variation, from $\theta_{\text{sil}} = 0.0247$ (*Arabidopsis lyrata* ssp. *petraea*, Ramos-Onsins et al. 2004; *Zea mays* ssp. *parviglumis*, Wright and Gaut 2005) to $\theta_{\text{sil}} = 0.0040$ (*Arabidopsis lyrata* ssp. *lyrata*, Wright and Gaut 2005). The nucleotide diversity of the wild relatives of many crops fell within this range, such as wild relative of maize (*Zea mays* ssp. *mays*, $\theta_{\text{sil}} = 0.0149$; *Zea perennis*, $\theta_{\text{sil}} = 0.0130$) (Tiffin and Gaut 2001; Wright and Gaut 2005), wild barley

(*Hordeum vulgare*, $\theta_{\text{sil}} = 0.0109$) (Morrell et al. 2003), wild sunflower (*Helianthus annuus*, $\theta_{\text{sil}} = 0.0144$) (Liu and Burke 2006), and *Populus tremula* ($\theta_{\text{sil}} = 0.0167$) (Ingvarsson 2005). The present study showed that *O. rufipogon* and *O. nivara* possessed medium to low levels of nucleotide variation ($\theta_{\text{sil}} = 0.0095$ for *O. rufipogon* and $\theta_{\text{sil}} = 0.0077$ for *O. nivara*) compared with the relatives of other crop plants. However, relative to other wild species in *Oryza*, nucleotide diversity in these 2 species is much higher because our recent multiple gene study found that the θ_{sil} values ranged from 0.0038 to 0.0057 in 3 *Oryza* species with the C-genome (*Oryza officinalis*, *Oryza eichingeri*, and *Oryza rhizomatis*) (Zhang and Ge 2007). Higher level of genetic variation in *O. rufipogon* than in other *Oryza* species was also reported using allozyme, RFLPs, and SSRs (Gao et al. 2001; Gao and Zhang 2005; Bautista et al. 2006).

It is not unexpected that slightly higher nucleotide diversity is observed in *O. rufipogon* than in *O. nivara* because the former has wider distribution and is largely outcrossing species, whereas *O. nivara* is primarily inbreeding (Vaughan 1989). Over all loci, we found that 66.7% of *O. rufipogon* individuals are heterozygous, whereas 23.1% of *O. nivara* individuals are heterozygous (table 3), in general agreement with previous estimates of outcrossing rates of 5–56% for the 2 species (Morishima et al. 1984; Barbier 1989). Theoretically, genetic variation in self-pollinating species would be reduced by decreasing effective population size and eliminating the effective rate of recombination (Charlesworth 2003). The low nucleotide variation in *O. nivara* might be facilitated by self-fertilization. However, empirical studies showed that the effects of selfing on population genetics relied critically on whether the levels of variation were compared species wide or within populations (Savolainen et al. 2000; Wright et al. 2003). Because our samples were collected from disparate populations, it is still difficult to evaluate the impact of mating system on genetic variation in these species.

Taxonomic treatment and genetic relationships between *O. rufipogon* and *O. nivara* have been controversial for decades (Vaughan 1989; Morishima 2001; Zhu and Ge 2005). Although 2 separate species names have still been used in many cases (e.g., Vaughan 1989; Ge et al. 2001; Yamanaka et al. 2003), studies based on morphological analyses and artificial hybridization (Oka 1988; Lu et al. 2000) as well as recent molecular data (Lu et al. 2002; Ren et al. 2003; Zhu and Ge 2005) did not reveal either a reproductive barrier or significant genetic differentiation between them. It is a tendency, therefore, to treat these 2 species as 2 different ecotypes or subspecies under a single species (*O. rufipogon* sensu lato) (Morishima 2001; Cheng et al. 2003; Vaughan et al. 2003) and, in some cases, consider them as a single large gene pool (i.e., Londo et al. 2006). The present phylogenetic analyses based on 10 genes did not show apparent grouping of the 4 taxa, in agreement with previous results based on multiple gene phylogeny (Zhu and Ge 2005). Moreover, the Kishino–Hasegawa test demonstrated that the constraint species tree was significantly worse than the nonconstraint tree (supplementary table S1, Supplementary Material online), rejecting the hypothesis that these taxa are monophyletic groups. This result is further corroborated by abundant

shared polymorphisms, no fixed differences among them, and low differentiation, as measured by F_{st} . In brief, the multiple gene sequence data support treating *O. rufipogon* and *O. nivara* as 2 ecotypes of a large species rather than as 2 separate species (Oka 1988; Barbier et al. 1991; Morishima et al. 1992; Lu et al. 2002; Zhu and Ge 2005).

Severe Bottleneck during Domestication of *O. sativa*

Considering that the wild relatives of *O. sativa* harbor average levels of nucleotide variation, it is particularly noteworthy that the nucleotide diversity of *O. sativa* is much lower than that of other major crops including both inbreeding and outbreeding species. For example, the species-wide silent diversity (θ_{sil}) based on multilocus sequences was 0.0149 in maize (Wright and Gaut 2005), 0.0109 in barley (Morrell et al. 2003), and 0.0034 in sorghum (Hamblin et al. 2004). In this study, we found that the nucleotide variation of *O. sativa* was as low as $\theta_{sil} = 0.0024$, approximately 70% of that of sorghum (Hamblin et al. 2004), one-third of that of cultivated sunflower ($\theta_{sil} = 0.0072$, Liu and Burke 2006), and one-sixth of that of maize (Wright and Gaut 2005). The value is even lower if we consider the 2 subspecies separately (table 3).

Although previous studies using different marker systems found reduced diversity in cultivated rice relative to its wild progenitors (Oka 1988; Sun et al. 2001; Londo et al. 2006), the present study indicated that only 20% and 10% of the diversity in the wild species were retained in the subspecies *indica* and *japonica* of the cultivated rice, respectively. This is in sharp contrast to several studies of other crops where the genome-wide levels of diversity were slightly lower in the crops than in their wild progenitors. For instance, maize maintained approximately 80% of diversity found in its wild ancestor as measured by nonselected genes (Wright and Gaut 2005), and the cultivated sunflower has retained 40–50% of the diversity present in the wild (Liu and Burke 2006). A previous literature survey also indicated that approximately two-thirds of genetic diversity have been maintained in major crops relative to their wild progenitors despite the difficulty of comparisons between species due to differences in experimental systems (Buckler et al. 2001).

Why does rice differ from other domesticated plants with respect to the loss of diversity during domestication? There are 4 possible reasons that rice appears to retain an aberrantly low level of genetic diversity relative to its wild ancestors. The first is that the domestication bottleneck may have been more severe. Our simulations support this conjecture. For example, Wright et al. (2005) obtained a K value of 2.45 in maize. In our case, the K value was estimated to be 0.2 (*japonica*) and 0.5 (*indica*) assuming separate domestications or 0.4 (*japonica* + *indica*) assuming a monophyletic origin of the 2 cultivated subspecies. The founding population would have contained 200–500 individuals for rice (supplementary table S2, Supplementary Material online), but would have up to 4,650 individuals for maize (Eyre-Walker et al. 1998) if the initial domestication event occurred over a 1,000-year period. Because archaeological evidence suggested that the domestication of rice occurred at least 10,000 years ago and spread to

other regions in South and Southeast Asia around 7,000 years ago (Khush 1997), the maximum duration of the domestication bottleneck in rice is at maximum 3,000 years. Therefore, the sequence diversity currently found in rice genome can be explained by a founding population of at most 1,500 individuals (supplementary table S2, Supplementary Material online).

A second explanation for low retention of genetic diversity in domesticated rice could be sampling. If the domestication of rice was a geographically local event, based on a distinct and genetically limited subpopulation of the wild ancestor, then our use of a global sample of *O. rufipogon* and *O. nivara* could be misleading. Unfortunately, it is difficult to assess this potential problem fully, partly because there is not yet enough information about population structure in the wild relatives of rice and partly because extensive gene flow occurs between these taxa (Oka 1988). A third explanation is that the shift of mating system, from commonly outbreeding wild rice to almost exclusively inbreeding cultivated rice, may have facilitated small population sizes in the incipient domesticate. However, empirical studies showed that the mating system primarily affects diversity at population rather than species levels (Savolainen et al. 2000; Wright et al. 2003), making the selfing explanation implausible. At least, selfing does not appear to be a complete explanation for the huge diversity reduction in cultivated rice because sorghum is also selfing, yet cultivated sorghum retains 60–70% of diversity of its wild progenitor (Hamblin et al. 2004). Finally, selection reduces diversity, and the effects of selection can extend long distances along chromosomes in selfing taxa (Nordborg and Innan 2002; Olsen et al. 2006). However, various tests in this study did not detect signature of selection (tables 3 and 4) at all 10 loci despite generally low statistical power of most methods (Wright and Gaut 2005). In addition, selection may lead to both reduced diversity (e.g., directional selection) and elevated diversity (e.g., balancing selection) (Hamblin et al. 2004; Wright and Gaut 2005). Consequently, selection might not be the main reason account for the severe reduction of diversity in cultivated rice although effects of selection cannot be excluded entirely.

Linkage Disequilibrium

The extent of LD differs both among taxa and across loci (table 3). Of the 4 taxa, the lowest LD was observed in the perennial *O. rufipogon* where the r^2 value drops to a negligible level within 400 bp, and the highest in the *japonica* rice where LD extends to the entirely sequenced region (~900 bp) (fig. 4a). The overall pattern of higher LD is observed in cultivated rice, in particular the *japonica* rice (fig. 4a and b). Extensive LD has also been observed previously in both the Asian and African cultivated rice (Garris et al. 2003; Semon et al. 2005). Theoretically, LD is affected by various evolutionary and demographic factors, including selection, recombination, population admixture, inbreeding, and bottleneck (Nordborg and Tavare 2002). For instance, an investigation on a disease resistance gene (*Xa5*) in the Asian cultivated rice suggested that LD could extend to ~100 kb (Garris et al. 2003). In the *tb1* locus of maize and *Wx* locus of *O. sativa*, extended LD was

found across the entire genomic regions upstream and downstream of the genes due to selective sweeps (Clark et al. 2004; Olsen et al. 2006). However, selection and population admixture appear not to be the main reasons for the higher LD observed in the cultivated rice because neutrality tests were not significant for all loci and both subspecies *indica* and *japonica* maintained high levels of LD (fig. 4a).

The expectation of higher LD in selfers relative to outcrossers has been observed in many species. For example, significant LD persisted for 250 kb in selfing *Arabidopsis thaliana* (Nordborg and Innan 2002), whereas LD declined to negligible levels in <1 kb in the outcrossing maize (Remington et al. 2001) and extended only a few hundred basepairs in European aspen, an outbreeding species (Ingvarsson 2005). Recently, Liu and Burke (2006) found that LD decayed to negligible levels ($r^2 < 0.10$) within 200 bp in the obligate outcrossing wild sunflower but to the same levels within ~1,100 bp in the self-compatible cultivated sunflower. Mating system can explain partly the higher level of LD in cultivated rice because *O. sativa* is predominantly inbreeding species. Higher LD in the annual *O. nivara* than in the perennial *O. rufipogon* is also consistent with the higher outcrossing rate in *O. rufipogon* than in *O. nivara* (Morishima et al. 1984; Barbier 1989).

However, the effects of mating system and population bottleneck on LD are intermingled in many cases, particularly for crop plants given the fact that transition from outcrossing to self-fertilization and population bottleneck often take place simultaneously during crop domestications (i.e., Hamblin et al. 2004; Liu and Burke 2006). Small effective population size is most likely to be the main force responsible for the higher LD and lower recombination in cultivated rice because only 10–20% of genetic diversity in the wild rice gene pool was retained in the cultivated rice. This suggestion is also supported by our computer simulation, in which very small effective population size due to severe bottleneck was estimated in *O. sativa*. Higher LD in japonica rice is consistent with its more severe population bottleneck during its domestication. The findings in this study imply that LD mapping by genome scans may not be feasible in wild rice due to the high density of markers needed. These levels of LD, however, are sufficient to establish whether polymorphisms within candidate genes are associated with specific phenotypic variants in the cultivated rice.

Supplementary Material

Supplementary figures S1–S4 and tables S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Lin-Bin Zhang for helpful discussion on the manuscript and Zhe Li, Shu-Qi Zhao, Ge Gao, An-Yuan Guo, and other members in Ge's group for technical assistances. We are grateful to Duncan Vaughan and Tao Sang and 2 anonymous reviewers for their valuable comments. We also thank the International Rice Research Institute (Los Banos, Philippines) for providing leaf and seed

samples and Susan McCouch (Cornell University) for providing seeds of some rice cultivars. This work was supported by the National Natural Science Foundation of China (30121003, 30430030, and 90408015), the CAS Innovation Grant, the National Key Basic Research Program of China (2003CB715900), and Postdoctor Science Foundation of China (20060390012).

Literature Cited

- Barbier P. 1989. Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations. *Jpn J Genet.* 64:273–285.
- Barbier P, Morishima H, Ishihama A. 1991. Phylogenetic relationships of annual and perennial wild rice: probing by direct DNA sequencing. *Theor Appl Genet.* 81:693–702.
- 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.
- Buckler ESI, Thornsberry JM, Kresovich S. 2001. Molecular diversity, structure and domestication of grasses. *Genet Res.* 77:213–218.
- Charlesworth D. 2003. Effects of inbreeding on the genetic diversity of populations. *Phil Trans R Soc Lond B Biol Sci.* 358:1051–1070.
- Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohstubo E. 2003. Polyphyletic origin of cultivated rice: based on the interspersed pattern of SINEs. *Mol Biol Evol.* 20:67–75.
- 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.
- Darwin C. 1859. On the origin of species by means of natural selection. London: John Murray.
- Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA.* 95:4441–4446.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics.* 133:693–709.
- 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.
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics.* 169:1631–1638.
- 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.
- Ge S, Sang T, Lu BR, Hong DY. 2001. Phylogeny of the genus *Oryza* as revealed by molecular approaches. In: Khush GS, Brar DS, Hardy B, editors. Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium. Los Banos (Philippines): IRRI. p. 89–105.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.

- 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, Wakeley J. 1997. A coalescent estimator of the population recombination rate. *Genetics*. 145:833–846.
- Hill WG, Weir BS. 1988. Variances and covariances of squared linkage disequilibria in finite populations. *Theor Popul Biol*. 33:54–78.
- 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 M, 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. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*. 18:337–338.
- 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.
- Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol*. 35:25–34.
- 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.
- Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol*. 29:170–179.
- Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics*. 173:321–330.
- Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc Natl Acad Sci USA*. 103:9578–9583.
- Lu BR, Naredo MEB, Juliano AB, Jackson MT. 2000. Preliminary studies on taxonomy and biosystematics of the AA genome *Oryza* species (Poaceae). In: Jacobs SW, Everett LJ, editors. *Grasses: systematics and evolution*. Melbourne (Australia): CSIRO. p. 51–58.
- Lu BR, Zheng KL, Qian HR, Zhuang JY. 2002. Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. *Theor Appl Genet*. 106:101–106.
- 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.
- Morishima H. 2001. Evolution and domestication of rice. Manila (Philippines): International Rice Research Institute.
- Morishima H, Sano Y, Oka HI. 1984. Differentiation of perennial and annual types due to habitat conditions in the wild rice *Oryza perennis*. *Plant Syst Evol*. 144:119–135.
- Morishima H, Sano Y, Oka HI. 1992. Evolutionary studies in cultivated rice and its wild relatives. *Oxf Surv Evol Biol*. 8:135–184.
- 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.
- Nayar NM. 1973. Origin and cytogenetics of rice. *Adv Genet*. 17:153–292.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Nordborg M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics*. 154:923–929.
- Nordborg M, Innan H. 2002. Molecular population genetics. *Curr Opin Plant Biol*. 5:69–73.
- Nordborg M, Tavare S. 2002. Linkage disequilibrium: what history has to tell us. *Trends Genet*. 18:83–90.
- Normile D. 1997. Archaeology—Yangtze seen as earliest rice site. *Science*. 275:309.
- Oka HI. 1988. *Origin of cultivated rice*. Tokyo (Japan): Japan Scientific Societies Press.
- 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.
- 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.
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA*. 98:11479–11484.
- Ren F, Lu BR, Li S, Huang J, Zhu Y. 2003. A comparative study of genetic relationships among the AA-genome *Oryza* species using RAPD and SSR markers. *Theor Appl Genet*. 108:113–120.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 19:2496–2497.
- 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.
- Second G. 1982. Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Jpn J Genet*. 57:25–57.
- Semon M, Nielsen R, Jones MP, McCouch SR. 2005. The population structure of African cultivated rice *Oryza glaberrima* (Steud.): evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. *Genetics*. 169:1639–1647.
- Sun CQ, Wang XK, Li ZC, Yoshimura A, Iwata N. 2001. Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. *Theor Appl Genet*. 102:157–162.
- Swofford D. 2001. PAUP: phylogenetic analysis using parsimony (and other methods), version 4.0b10. Sunderland (MA): Sinauer Associates.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123:585–595.
- Tanksley SD, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*. 277:1063–1066.
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA*. 98:9161–9166.

- 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, 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, Philippines. Manila (Philippines): International Rice Research Institute.
- Vaughan DA, Morishima H. 2003. Biosystematics of the genus *Oryza*. In: Smith CW, editor. Rice: origin, history, technology, and production. Hoboken, NJ: John Wiley & Sons, Inc. p. 27–65.
- Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Ricnzo AD. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci USA.* 102:10508–18513.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7:256–276.
- Weiss G, von Haeseler A. 1998. Inference of population history using a likelihood approach. *Genetics.* 149:1539–1546.
- White GM, Hamblin MT, Kresovich S. 2004. Molecular evolution of the phytochrome gene family in sorghum: changing rates of synonymous and replacement evolution. *Mol Biol Evol.* 21:716–723.
- White SE, Doebley JF. 1999. The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. *Genetics.* 153:1455–1462.
- Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. *Science.* 308:1308–1310.
- Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol.* 22:506–519.
- Wright SI, Lauga B, Charlesworth D. 2003. Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Mol Ecol.* 12:1247–1263.
- Xia X, Xie Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J Hered.* 92:371–373.
- Yamanaka S, Nakamura I, Nakai H, Sato YI. 2003. Dual origin of the cultivated rice based on molecular markers of newly collected annual and perennial strains of wild rice species, *Oryza nivara* and *O. rufipogon*. *Genet Resour Crop Evol.* 50:529–538.
- Yoshida K, Miyashita N. 2005. Nucleotide polymorphism in the *Adh2* region of the wild rice *Oryza rufipogon*. *Theor Appl Genet.* 111:1215–1228.
- Yoshida K, Miyashita NT, Ishii T. 2004. Nucleotide polymorphism in the *Adh1* locus region of the wild rice *Oryza rufipogon*. *Theor Appl Genet.* 109:1406–1416.
- Zeder MA, Emswiller E, Smith BD, Bradley DG. 2006. Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* 22:139–155.
- Zhang L-B, Ge S. 2007. Multilocus analysis of nucleotide variation and speciation in *Oryza officinalis* and its close relatives. *Mol Biol Evol.* 10.1093/molbev/msl204.
- Zhu Q, Ge S. 2005. Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytol.* 167:249–265.

William Martin, Associate Editor

Accepted January 4, 2007