J Mol Evol (2009) 68:393–402
DOI 10.1007/s00239-009-9217-6

Genetic Signature of Rice Domestication Shown by a Variety of Genes

Yuanli Zhang · Jiao Wang · Xiaohui Zhang · Jian-Qun Chen · Dacheng Tian · Sihai Yang

Received: 8 October 2008 / Accepted: 19 February 2009 / Published online: 17 March 2009
© Springer Science+Business Media, LLC 2009

Abstract Cultivated rice was domesticated from common wild rice. However, little is known about genetic adaptation under domestication. We investigated the nucleotide variation of both cultivated rice and its wild progenitors at 22 *R*-gene and 10 non–*R*-gene loci. A significant regression was observed between wild rice and rice cultivars in their polymorphic levels, particularly in their nonsynonymous substitutions (θ_a). Our data also showed that a similar proportion (approximately 60%) of nucleotide variation in wild rice was retained in cultivated rice in both *R*-genes and non–*R*-genes. Interestingly, the slope always was >1 and the intercept always >0 in linear regressions when a cultivar's polymorphism was x-axis. The slope and intercept values can provide a basis by which to estimate the founder effect and the strength of artificial direct selection. A larger founder effect than previously reported and a strong direct-selection effect were shown in rice genes. In addition, two-directional selection was commonly found in differentiated genes between *indica* and *japonica* rice subspecies. This kind of selection may explain the mosaic origins of *indica* and *japonica* rice subspecies. Furthermore, in most *R*-genes, no significant differentiation between cultivated and wild rice was detected. We found evidence for genetic introgression from wild rice, which may have played an important role during the domestication of rice *R*-genes.

Keywords Genetic variation · NBS-LRR genes · Two-directional selection

Yuanli Zhang and Jiao Wang contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00239-009-9217-6) contains supplementary material, which is available to authorized users.

Y. Zhang · J. Wang · X. Zhang · J.-Q. Chen · D. Tian (✉) · S. Yang (✉)
State Key Laboratory of Pharmaceutical Biotechnology, Department of Biology, Nanjing University, Nanjing 210093, China
e-mail: dtian@nju.edu.cn

S. Yang
e-mail: sihaiyang@nju.edu.cn

Introduction

Asian cultivated rice, including the *indica* and *japonica* subspecies, was domesticated approximately 10,000 years ago (Brar and Khush 1997; Sang and Ge 2007). The origins of Asian cultivated rice from its wild ancestors have been debated for decades, mainly regarding the question of whether it originated monophyletically or polyphyletically. The monophyletic origin hypothesis suggests that common wild rice (*Oryza rufipogon*) evolved into the *indica* and *japonica* subspecies in different locations and at different times (Vitte et al. 2004; Ma and Bennetzen 2004). In contrast, the polyphyletic origin hypothesis postulates that diversification of *indica* and *japonica* subspecies occurred before they were separately domesticated and subsequently developed into two major subspecies (Second 1982; Wang and Sun 1996; Tang et al. 2006). During the process of domestication, founder events can decrease genetic diversity, change allele frequencies, increase linkage disequilibrium (LD), and eliminate rare alleles in the resulting population (Ross-Ibarra et al. 2007). Recent research has shown that only 10 to 20% of the genetic diversity in wild rice relatives was retained in cultivated rice, indicating a severe genetic bottleneck during domestication (Zhu et al. 2007). Therefore, common wild rice serves as a vast, important genetic reservoir and germplasm

resource to improve traits of agronomic importance and broaden genetic diversity in rice breeding (Song et al. 2005). Many agronomically beneficial traits in wild rice, such as rice tungro virus resistance, bacterial leaf blight (*Xa21*) resistance, and acid sulfate soil tolerance, play important roles in rice breeding (Song et al. 2005).

Plant disease resistance (*R*) genes have been used in resistance-breeding programs for decades, which have been proven to be the most economic and effective strategy by which to control crop diseases (Moffat 2001). Compared with rice domesticates, more genetic diversity was located in its wild relatives, and these were valuable sources for resistance to disease or insect pests (Moffat 2001). Some of these genes have successfully been transferred or introgressed into varieties of cultivated rice. For example, bacterial leaf blight *R*-genes *Xa21*, *Xa23*, and *Xa27* have been introgressed into rice cultivars from *Oryza longistaminata*, *O. rufipogon,* and *O. minuta*, respectively (Meyers et al. 2005). *Pib*, a NBS-LRR class *R*-gene that confers a high level of resistance to most types of rice blast, was introgressed independently from two Indonesian and two Malaysian rice cultivars into *japonica* rice cultivars (Liu et al. 2002). *Pi-9*, another resistance gene to rice blast, was introduced from *O. minuta*, a tetraploid wild species of the *Oryza* genus (Amante-Bordeos et al. 1992).

Abundant genetic diversity in *R*-genes is important as a necessary condition for response-to-recognition specificity of pathogens, not only in wild relatives of rice but also in rice-breeding programs. Polymorphic studies have demonstrated that extremely high levels of diversity were detected among different populations or even within populations in some *R*-loci, e.g., in *Rpp13*, *Rpp8*, and *L* genes (Allen et al. 2004; Ding et al. 2007a, b; Sun et al. 2008; Tian et al. 2008; Yang et al. 2008). The absence of genetic diversity in *R*-genes can have dire consequences in terms of disease epidemics in excessively uniform crops. The most famous example of this was the potato famine that struck Ireland in the middle of the nineteenth century. Excessive reliance on potato as a staple food led to a widespread famine when the crop was devastated by a fungal disease, the potato late blight.

The largest class of *R*-genes encodes nucleotide binding site–leucine-rich repeat proteins (NBS-LRRs). The only known function of these proteins is to condition disease resistance (Nimchuk et al. 2003). A total of 480 NBS-LRRs have been identified in the rice genome (Zhou et al. 2004; Yang et al. 2006). A genome-wide survey of *R*-gene polymorphisms showed that four variation types—the conserved, the diversified, the intermediate-diversified, and the present/absent patterns—were detected in rice lines (Yang et al. 2006, 2008). However, previous studies did not explore the fundamental role of plant domestication in human history and the critical importance of a relatively small number of crop plants to modern societies. To understand which factors influenced genetic diversification and genomic evolution in *R*-genes during domestication, this study investigated the quantitative (genetic variation) and qualitative (genetic differentiation) differences in NBS-LRR *R*-genes and other functional genes between cultivated rice and its wild relatives. This is a first step in understanding the evolutionary forces acting on rice *R*-genes between domestic and progenitor rice species.

Materials and Methods

Plant Material and DNA Isolation

The variation patterns of *R*-genes were investigated in >17 worldwide rice cultivars and 14 wild rice individuals. Rice cultivars were obtained from the United States Department of Agriculture, from the United States National Plant Germplasm System, and from Cailin Wang at the Institute of Food Crops, Jiangsu Academy of Agricultural Sciences, China. The wild rice accessions were provided by Dajian Pan at the National Guangzhou Wild-Rice Conservation, China, and the National Institute of Genetics, Japan. Genomic DNA was extracted from fresh leaves using the cetyltrimethyl ammonium bromide (CTAB) method.

PCR Amplification and DNA Sequencing

A total of 22 NBS-LRR genes (Table 1), except for 2 functional *R*-genes (*Pib* and *Pita*), were randomly selected from the conserved, the highly diversified, and the intermediate-diversified *R*-genes, which were defined by Yang et al. (2006, 2008). Nine of 10 non–*R*-genes in Table 1 were chosen on the basis of their functions confirmed experimentally (Table S1), and 1 (*Vatp*) was selected as a neutrally evolved gene (Londo et al. 2006). Each of these genes (excluding *Vatp*) was assumed to be functional, at least in some alleles, because their sequences were identical (or almost identical) to the full-length cDNA and expressed sequence tags (EST) sequences recorded in GenBank. In addition, other 10 non–*R*-genes from Tang et al. (2006), representing highly divergent genes, were employed as references in this study (Table S2).

Several studies have shown that LRR domains of *R*-gene are the major determinants of recognition specificity for *Avr* factors (Elli et al. 2000) and that these domains have the highest nucleotide polymorphism (Jiang et al. 2007). Therefore, the LRR region was chosen to detect *R*-gene variation pattern. Polymerase chain reaction (PCR) primers were designed based on conserved sites between Nipponbare and 93-11 (Yang et al. 2006). If no PCR products were obtained in individuals, then PCR was repeated for two

Table 1 Nucleotide variation at *R*-genes and non–*R*-loci

Loci	π (%)			θ (%)		θ_{sil} (%)		θ_a (%)	
	C	W	D_{xy}	C	W	C	W	C	W
R-loci									
Os06g06400	0.00	0.35	0.20	0.00	0.45	0.00	0.97	0.00	0.25
Os06g06380	0.00	0.55	0.42	0.00	0.63	0.00	1.87	0.00	0.27
Os01g33690	0.00	0.58	0.33	0.00	0.63	0.00	1.18	0.00	0.35
Os06g06390	0.00	0.61	0.56	0.00	0.72	0.00	1.79	0.00	0.10
Os12g33740	0.00	0.71	0.45	0.00	0.76	0.00	1.57	0.00	0.53
AL006613	0.08	0.02	0.20	0.05	0.42	0.00	0.44	0.06	0.36
Pita	0.09	0.15	0.12	0.31	0.35	0.45	0.36	0.27	0.35
Os02g25900	0.22	0.25	0.27	0.16	0.40	0.18	0.36	0.16	0.37
Os07g29820	0.24	0.51	0.46	0.16	0.70	0.27	1.29	0.12	0.52
Os01g72390	0.26	0.35	0.32	0.28	0.40	1.08	0.46	0.07	0.44
Os01g23380	0.35	0.64	0.56	0.70	0.71	1.54	1.23	0.44	0.55
Os01g16400	0.37	1.00	0.75	0.45	1.01	1.19	2.53	0.22	0.54
Os01g16390	0.42	1.76	1.39	0.38	1.76	0.42	2.51	0.38	1.20
Os06g48520	0.47	0.45	0.56	0.35	0.38	0.57	0.29	0.25	0.34
Os08g09430	0.48	0.72	0.64	0.36	0.61	0.00	0.31	0.47	0.70
Os04g02110	0.55	0.62	0.56	0.40	0.73	0.17	0.55	0.47	0.78
Os02g18510	0.70	0.70	0.74	0.87	1.16	1.23	1.30	0.76	1.11
Os07g40810	0.72	0.82	0.86	0.41	0.59	1.16	1.51	0.19	0.32
Os07g08890	0.72	1.62	1.66	0.74	1.56	1.49	1.94	0.52	1.44
Os10g22484	1.30	1.04	1.25	0.78	0.99	0.92	1.30	0.74	0.86
Os12g28250	3.15	4.67	4.13	3.05	2.50	3.04	3.85	2.34	2.83
Pib	6.51	7.73	6.51	4.83	5.99	6.21	7.81	3.55	4.44
Average	0.76	1.04	1.01	0.65	0.99	0.91	1.60	0.50	0.85
Non–*R*-loci									
Sh4	0.00	0.11	0.16	0.00	0.19	0.00	0.35	0.00	0.13
Sp17	0.01	0.23	0.13	0.03	0.50	0.41	1.51	0.04	0.20
MSP1	0.06	0.44	0.29	0.06	0.70	0.00	1.49	0.07	0.05
sbe1	0.07	0.06	0.08	0.04	0.08	0.17	0.35	0.00	0.00
Gigantea	0.13	0.18	0.16	0.08	0.31	0.11	0.81	0.07	0.15
qSH-1	0.14	0.22	0.20	0.09	0.50	0.21	1.32	0.06	0.25
EMF1	0.15	0.37	0.35	0.15	0.66	0.25	1.17	0.12	0.51
HKT1	0.20	0.20	0.23	0.14	0.28	0.42	0.58	0.04	0.19
COC1	0.20	0.27	0.26	0.11	0.45	0.25	1.12	0.07	0.25
Vatp	0.39	0.40	0.46	0.45	1.16	0.37	1.21	–	–
Average	0.14	0.25	0.22	0.12	0.48	0.22	0.99	0.05	0.19

C rice cultivars, *W* wild rice, π nucleotide diversity with Jukes and Cantor correction (Lynch and Crease 1990), D_{xy} nucleotide divergence with Jukes and Cantor correction (Nei 1987) between rice cultivars and wild rice, θ Watterson's estimator of θ/basepar (Watterson 1975) calculated on the total number of polymorphic sites; θ_{sil} Watterson's estimator of θ/basepar (Watterson 1975) calculated on the silent sites, θ_a Watterson's estimator of θ/basepar (Watterson 1975) calculated on the nonsynonymous substitution sites

additional times, once at 5°C lower then the annealing temperature (45–50°C). If it still could not obtain PCR products, the PCR reaction was repeated using a newly designed primer pair. For most of the selected genes, PCR products were directly sequenced in cultivated rice. For multicopy genes (*Pib* locus), the PCR products were cloned into *pGEM*-T Easy Vector (Promega A1360, Madison, WI, USA), and ≥6 colonies for each cultivar were sequenced separately until no new homologue sequences could be identified. For wild relatives of rice, in which either homozygous or heterozygous individuals exist, the PCR fragments were cloned into *pGEM*-T Easy Vector (Promega A1360), and ≥4 colonies were then sequenced individually. All PCR products were sequenced using an ABI 3100A automated sequencer. DNA sequences were visually aligned, and all polymorphisms were rechecked from chromatograms or by resequencing, with special attention paid to low-frequency polymorphisms. Sequence data were deposited into GenBank under the accession numbers FJ709611 to FJ710042 and EF641964 to EF642487.

Sequence Analysis

Multiple sequence alignments were performed using ClustalW1.83 (Thompson et al. 1994). The nucleotide alignments were analyzed using DnaSP version 4.0 (Rozas et al. 2003). Insertion/deletions (indels) were excluded from all estimates. Nucleotide diversity was estimated by π with the Jukes and Cantor correction (Lynch and Crease 1990) and by θ from the number of polymorphic segregating (S) sites (Watterson 1975). The divergences between species were obtained by D_{xy} with the Jukes and Cantor correction (Nei 1987). Phylogenetic trees were constructed based on the bootstrap neighbor-joining (NJ) method with a Kimura two-parameter model by MEGA version 4.0 (Tamura et al. 2007). The stability of internal nodes was assessed by bootstrap analysis with 1,000 replicates.

Analysis of Genetic Structure

Two distinct classes of tests—the haplotype-based statistical test (χ^2 statistic) and the sequence-based statistical test (F_{st} and S_{nn})—were applied to detect genetic differentiation between wild and cultivated rice or *indica* and *japonica* rice subspecies. The F_{st} statistic (Weir 1996), which measures the genetic variance between population divided by the total genetic variance of the entire population, was used to quantify the degree of the genetic differentiation between cultivated and wild rice or *indica* and *japonica* subspecies from 22 individual *R*-genes and 10 non–*R*-genes genes using the AELEQUIN version 3.11 software (http://www.lgb.unige.ch/arequin). The statistical significance (p value) of pairwise F_{st} was determined by permuting the data 1,000 times. The nearest-neighbors

statistic (S_{nn}), which measures the "nearest-neighbors" (in sequence space) of sequences—which appears to be the most powerful statistic (or nearly as powerful as the best statistic under all conditions examined) (Hudson 2000)—was used to test for genetic differentiation as described in Hudson (2000). The statistical significance of pairwise S_{nn} values was determined by permuting the data 1,000 times in DnaSP v 4.0. The genetic differentiation was also estimated using χ^2 test based on haplotype frequencies (Nei 1987), which can be directly adapted to use with nucleotide variation by treating each distinct haplotypes as an allele.

Results

Nucleotide Polymorphism of *R*-Genes in Cultivated and Wild Rice

On the sampled 22 NBS-LRR loci, average nucleotide diversity (π) was 0.0104 in wild rice relatives, 0.0076 in rice cultivars, and 0.0101 between them (Table 1). Approximate 73.1% of polymorphism in wild rice, a significantly smaller variation ($p < 0.05$), was maintained in rice cultivars; however, this number is larger than those (56.0% of wild rice polymorphism on average) found on 10 non–*R*-loci (Table 1).

Average nucleotide diversity (θ), which was estimated by the number of polymorphic segregating (S) sites (Watterson 1975), better measured the richness of genetic variation among populations because this type of variation was less affected by the frequency of nucleotide substitutions. Particularly, the silent diversity (θ_{sil}), which is believed to evolve neutrally, can clearly reflect the proportion of maintained variation. In *R*-genes, the average θ_{sil} in wild relatives of rice was 0.0160, which was significantly higher than the average θ_{sil} of 0.0091 in rice cultivars ($p < 0.001$). These data suggest that domesticated rice has approximately 56.8% (θ_{sil}) of the variation found in its progenitor. The nonsynonymous substitution sites (θ_a) essentially reflect the richness of functional variation between wild and cultivated rice. The average θ_a was 0.0050 in rice cultivars, which was 58.8% of that (0.0085) in wild rice and significantly less than that in its wild progenitors ($p < 0.001$).

Our analyses demonstrate that there is significantly more genetic variation of *R*-genes in wild rice. These results agree with the previous inference that *O. sativa* originated from common wild rice (Zhu and Ge 2005). Further analysis, however, shows that a much higher proportion (83.1, 71.2, or 65.0%, respectively), of π, θ_{sil}, and θ_a is maintained in nonconserved cultivar *R*-genes, which have a higher level (e.g., $\pi > 0.002$) of variation (Table 1). The other conserved *R*-genes ($\pi < 0.002$) only show 5.7, 5.4 and 14.9% of π, θ_{sil} and θ_a, respectively. Chi-square test shows that there are significant differences between conserved and nonconserved *R*-genes ($p < 0.001$), indicating that their mechanism of variation maintenance is different.

Estimation of the Founder Effect from Polymorphic Levels in Cultivated and Wild Rice

Genetic diversity in domesticated crops is largely determined by founder effects (Ross-Ibarra et al. 2007). The founder effect can be directly estimated by the proportion of maintained polymorphism in rice cultivars. For example, the proportions mentioned previously (5.4 to 14.9% for conserved and 65 to 83.1% for nonconserved *R*-genes) may largely result from the amount of polymorphisms in founder plants during domestication. Because the difference between these estimates was large, we analyzed the linear regression of π, θ_{sil}, and θ_a in wild rice against these parameters in rice cultivars to provide a more accurate estimate.

Our calculations listed in Table 1 demonstrate that there is a strong positive correlation in the values of π, θ_{sil}, and θ_a between wild and cultivated rice ($r > 0.90$ and $p < 0.001$). This significant correlation occurred in all of the genes studied, not just in the *R*-genes (Fig. 1). For π, the slope = 1.06, $r = 0.96$, and $p < 0.0001$. For θ_{sil}, the slope = 1.08, $r = 0.93$, and $p < 0.0001$. For θ_a, the slope = 1.15, $r = 0.97$, and $p < 0.0001$. From these data, the maintained variations in rice cultivars are approximately 63% for θ_a and 60% for θ_{sil}, respectively, which are similar to the data on *R*-genes (approximately 59% for θ_a and 56% for θ_{sil}), suggesting that there is a similar proportion of variation maintained by both *R*- and non–*R*-genes.

Meanwhile, the linear regression analyses showed that there were slopes >1 and intercepts >0 for the parameters of π, θ_{sil} and θ_a between the rice cultivars and its wild relatives. Theoretically, when lacking artificial selection on the rice cultivars, the slope is expected to be equal to one, and the intercept should be zero. Therefore, both the slope and intercept can reflect artificial evolutionary forces during rice domestication. The reciprocal of the slope can provide an indication of the founder effect, and the intercept can result from strong directed selection on genes with a low variation.

Notably, the highest correlation coefficient was obtained from θ_a ($r = 0.976$ and $p < 0.0001$), a parameter directly associated with the phenotypes of individual plants. This suggests that artificial selection may start from the stage when the founder is selected and may continue during the rice domestication process. Because of strong artificial selection, θ_a was the best parameter by which to detect evolutionary forces acting on the maintenance of *R*-genes.

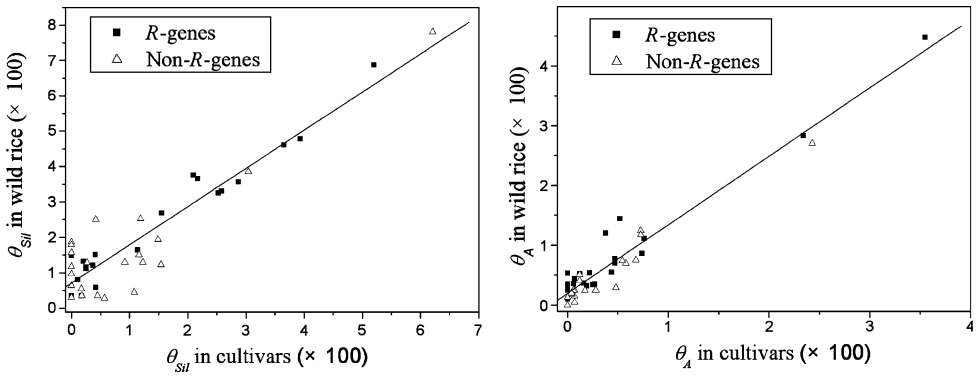

Fig. 1 The correlation of nucleotide silent substitution sites (θ_{sil}) or nucleotide nonsynonymous substitution sites (θ_a) between cultivated and wild rice. For θ_{sil}, the slope = 1.08, $r = 0.93$, and $p < 0.0001$. For θ_a, the slope = 1.15, $r = 0.97$, and $p < 0.0001$

Genetic Relation of the *R*-Loci Between Cultivated and Wild Rice

To clarify the phylogenetic relation of *R*-genes between cultivated and wild rice, NJ trees were constructed based on nucleotide variations. In most *R*-gene trees, there was no significant differentiation between cultivated and wild rice or between *indica* and *japonica* subspecies. However, in most non–*R*-loci, significant groupings were detected between the subspecies and between the wild rice and rice cultivars (Figs. 2, S1, and S2).

The χ^2 statistic, F_{st}, and S_{nn} were used to investigate genetic differentiation, and, as listed in Table S2, similar results were obtained using the 3 different methods. On the whole, significant genetic differentiation ($p < 0.05$) between cultivated and wild rice was detected in approximately 22.7% to 36.3% of the *R*-loci. Among these, 4 *R*-loci showed significant genetic differentiation between rice cultivars and its wild relatives by 3 statistical methods, and 2 *R*-loci showed significant variation by 2 statistical methods. Similarly, only 2 of 22 *R*-genes (9.1%) showed significant genetic differentiation between *O. sativa* L. *indica* and *japonica* subspecies ($p < 0.05$; Table S2). The low proportion of differentiated *R*-loci indicates that there was intensive gene flow or gene introgression at these *R*-loci between cultivated and wild rice or between the two subspecies. In addition, ongoing recurrent mutations may have occurred in the rice cultivars after domestication.

In contrast, 65% to 70% of non–*R*-gene loci showed significant genetic differentiation between rice cultivars and wild relatives (Table S2). The F_{st}, S_{nn}, and χ^2 statistical tests indicated that there was significant divergence between the 2 subspecies in 17 of 20 non–*R*-genes ($p < 0.05$). These results suggest that *R*-genes and non–*R*-genes evolved in distinct manners.

In the phylogenetic analysis, the *indica* and *japonica* subspecies were clearly separated into different clades in most of the non–*R*-gene trees (17 of 20; Fig. S2). Significant genetic differentiation between the two subspecies was also detected in these genes (Table S2). However, detailed examination showed that the differentiation between the subspecies is caused by two-directional selection. As shown in Fig. S2, the two divergent types of alleles, the *indica* and *japonica* subspecies, exist in wild rice, and few new mutations are found in these two types of genes. Therefore, the two distinct types probably arose from two-directional selection, not from real differentiation after domestication.

Test for Natural Selection in *R*-Genes

Higher rates of nonsynonymous over synonymous substitutions were found in *R*-genes compared with those in non–*R*-genes (Table S3). In *R*-genes, $\pi_a/\pi_s = 0.539$, 0.508, and 0.590 on average in rice cultivars, wild relatives, and all sequences, respectively. These values were significant higher than those at non–*R*-genes, where $\pi_a/\pi_s = 0.196$, 0.236, and 0.227 respectively ($p < 0.005$, paired Student *t test*; Table S3). A few *R*-genes exhibited exceptionally high levels of nonsynonymous substitution, such as *Os02g25900* ($\pi_a/\pi_s = 1.62$ on average), *Os08g09430* ($\pi_a/\pi_s = 4.88$ on average), and *Os04g02110* ($\pi_a/\pi_s = 5.34$ on average; Table S3), indicating that there was significant positive selection on these genes. No genes with $\pi_a > \pi_s$ were detected at non–*R*-genes loci, suggesting that there was purifying selection on these genes. In addition, a larger proportion of shared nonsynonymous substitutions was maintained at *R*-genes (56.0%) than at non–*R*-genes (27.5%; Table S3). The excess of nonsynonymous substitutions in *R*-loci suggests that positive selection acted on them, which likely resulted from the diversifying selection on the alleles and/or the balancing selection that maintained the polymorphisms.

Tajima's *D* was used to measure allele frequency changes by comparing *R*-genes with non–*R*-loci in cultivated and wild rice. The results showed that the *D* values in both cultivar *R*-genes (0.59 on average) and non–*R*-loci (0.76 on average) were significantly higher than the *D* values measured in wild rice (–0.39 at *R*-genes and –1.50 at non–*R*-loci; $P < 0.01$; Table S3). These results suggest that

there is an excess of low-frequency nucleotide polymorphisms in wild rice and more intermediate-frequency polymorphisms in rice cultivars. This explanation is consistent with the theoretic expectation that some low-frequency variants have been preferentially lost in rice cultivars because of the recent bottleneck during domestication.

Detection of Shared Mutations Between Cultivated and Wild Rice

The polymorphism patterns of rice cultivars and their wild relatives may allow us to infer their evolutionary histories. Assuming that cultivated rice was derived from common wild rice, more variations will be found in wild rice, and shared polymorphisms will be detected. Therefore, the polymorphic sites of wild-rice unique, shared, and cultivar unique could reflect the process of domestication. A clear difference among these sites was detected between *R*- and non–*R*-genes (Table S3). In non–*R*-genes, the overwhelming majority of polymorphisms (66.7%) were contributed by common wild rice (Table S3), which is consistent with the hypothesis that cultivated rice developed from common wild rice. There were only 15.4% shared mutations between cultivated and wild rice, and 17.9% of the mutations were present in rice cultivars alone. No fixed mutations were detected in the non–*R*-loci (Table S3), suggesting that the rice cultivars had recently originated from common wild rice.

In contrast, at *R*-loci, there were more shared mutations (43.1%), and fewer mutations were present only in wild rice (40.0%). The percent of polymorphisms present only in cultivated rice (16.9%) was similar to that observed for non–*R*-genes (Table S3). Notably, we detected two fixed single-nucleotide polymorphisms in *R*-genes, located at the *Os06g06380* and *Os06g06390* loci, respectively, which have been demonstrated to be under direct artificial selection (Yang et al. 2008).

Discussion

Effect of Domestication on the Maintenance of Rice *R*-Genes

Asian cultivated rice was domesticated from common wild rice, and it has long been believed that bottleneck or founder effect was the primary force shaping evolutionary patterns during the process of rice domestication (Ross-Ibarra et al. 2007; Zhu et al. 2007). Our data show that there is a strong positive correlation in π, θ_{sil}, and θ_a values between rice cultivars and its wild relatives and that θ_a is the best parameter for detecting artificial evolutionary forces acting on the maintenance of *R*- and non–*R*-gene polymorphisms.

Founder effect was usually estimated by the proportion of maintained polymorphisms in rice cultivars (Zhu et al. 2007). In the absence of founder effects and artificial selection, the regression analysis between θ_a of rice cultivars and wild rice lines is expected to result in a slope of one and an intercept of 0. However, both the slope and intercept were greater than the expected values. Obviously, the intercept with a value >0 is caused by quite a few zero values in rice cultivars (Table 1 and Fig. 1). Four of these zero values, *Sh4* and three *R*-genes, were confirmed to be under strong direct artificial selection (Li et al. 2006; Yang et al. 2008). These examples suggest that the genes with low θ_a values may also be under direct artificial selection. The existence of genes under direct artificial selection will also result in a larger intercept and a smaller slope. Analysis of only these genes can therefore result in an underestimate of maintained variation.

Thus, the reciprocal of slopes with values >1 can provide a better estimate of maintained variation and an indication of founder effects. From our analysis, the maintained variation in rice cultivars, especially the θ_a, is approximately 60%. To complete our analysis, we included other 10 non–*R*-gene sequences with a higher degree of nucleotide polymorphism diversity from Tang et al. (2006). Highly positive correlations between the wild rice and rice cultivars were also observed in our analysis of 22 *R*-genes and 20 non–*R*-genes (Fig. 1). From these data, the maintained variations in rice cultivars are approximately 63% for θ_a and 60% for θ_{sil}, which are similar to the data on *R*-genes (approximately 59% for θ_a and 56% for θ_{sil}), suggesting a similar proportion of variation maintained by both *R*- and non–*R*-genes.

Two-Directional Selection in Cultivated Rice

Several estimates have been made for the time of divergence between *indica* and *japonica* subspecies, all of which place the time of divergence from the most recent common ancestor at >100,000 years ago. However, this divergence time is an order of magnitude larger than the oldest estimates of when rice domestication occurred (Vitte et al. 2004; Ma and Bennetzen 2004). We detected significant differentiation between *indica* and *japonica* subspecies, with three different statistical tests in both *R*- and non–*R*-genes (Table S2 and Fig. S3). However, the detail analysis provides an alternative explanation for the differentiation.

Theoretically, differentiation must meet the following criteria. One criterion is the occurrence of new mutations, meaning that there are new, differentiated alleles that cannot be found in wild rice, which is the ancestor of the domesticated rice cultivars. A second criterion is that the

differentiation must be statistically significant, meaning that extensive nucleotide changes must be present between the differentiated alleles. In our analysis, few cultivar-unique mutations were detected, and almost no new mutations were discovered between *indica* and *japonica* subspecies (Table S3). Therefore, based on these criteria, almost all of the genes that were significantly different in our sequence data have no new mutations or have a <0.001 mutational change (Table S3). These results demonstrate that the significantly differentiated loci actually reflect a two-directional selection in these two types of alleles.

The origin of *indica* and *japonica* subspecies is one of the long-lasting mysteries of rice domestication. The debate centers over whether Asian cultivated rice originated monophyletically or polyphyletically. The fact that genes are under two-directional selection can shed light on this question. The prototypes of differentiated genes in both *indica* and *japonica* subspecies have had a longstanding existence in wild rice. A wild rice line could carry some ancestral genes, and some of these genes could have become the *indica* subspecies, whereas others became the *japonica* subspecies. Both *indica* and *japonica* subspecies indeed originated from wild rice; however, no wild rice line could be identified as the ancestor for either of them. Two-directional selection in *indica* and *japonica* subspecies supports a model of "mosaic origin," which could be the key to understand rice domestication.

Introgression May Play an Important Role in the Evolution of Rice *R*-Genes

The investigation of polymorphic patterns showed that a significantly higher proportion of shared mutations between cultivated and wild rice was detected in *R*-loci that in non–*R*-genes (Table S3). This suggests that the shared genetic variants in *R*-genes represent polymorphisms present in their common ancestor, which have been maintained either by chance or by some form of balancing selection. Another possibility is that these genetic variants resulted from introgression between species or from recurrent mutations.

In the *R*-loci comparison, recurrent mutations were insufficient to account for the large number of putative ancestral mutations, indicating that most of them resulted from individual mutation events. Although shared polymorphisms can arise by recurrent mutation (homoplasy), this phenomenon can only account for a small fraction of the shared polymorphisms we observed. In addition, most of the *R*-gene trees showed that alleles from the same species or subspecies were not clustering together (Figs. S1 and S2), and no significant genetic differentiations were detected, which suggests gene flow. Thus, the most parsimonious explanation of the observations is that ancestral mutations in rice cultivars were reacquired by introgression (Table S3). Estimations of population recombination parameters also showed that frequent recombination events between rice cultivars and its wild relatives were detected at *R*-gene loci.

Previous studies have shown that introgression plays an important role in rice domestication. The introgression of valuable genes, including *R*-genes, from wild to cultivar species is a common breeding practice for crop improvement (Brar and Khush 1997; Meyers et al. 2005; Amante-Bordeos et al. 1992). Some functional *R*-genes, such as *Xa21*, *Xa23*, *Xa27*, *Pi-9*, and *Pib*, have successfully been introgressed into varieties of cultivated rice from its wild relatives (Meyers et al. 2005; Amante-Bordeos et al. 1992). Therefore, the recent and future introgression of beneficial genes from the wild rice gene pool to cultivated varieties by way of conventional and molecular breeding programs can be viewed as the continuation of domestication (Brar and Khush 1997).

Maintenance of Variation and Pressure of Selection at Rice *R*-Loci

Selection can also play an important role in the maintenance of the variation added by introgression at *R*-loci. Previous studies demonstrated that *R*-genes exhibit evidence of balancing selection by maintaining high numbers of alternative alleles that exhibit high levels of recombination (Bakker et al. 2006, 2008). In our study, *R*-genes clearly maintained high level of variations and an increased recombination frequency. In addition, most of the *R*-genes had highly diverged alleles at intermediate frequencies as well as high levels of silent polymorphism. These observations strongly suggest that these *R*-genes are under the influence of long-term balancing selection (Table S3). In contrast, the non–*R*-genes displayed lower levels of nonsynonymous nucleotide diversity and had high levels of low-frequency nucleotide polymorphism, suggesting that non–*R*-genes experience purifying selection or functional constraint (Table S3).

Clearly, most non–*R*-genes showed low levels of amino acid polymorphism, and the test of neutrality showed that purifying selection was the prevailing selective pressure operating at these loci (Table S3). The θ_{sil} in the *R*-genes was only 1.5-fold greater than that in 10 non–*R*-genes in wild rice, suggesting that a slightly higher variation rate could occur in *R*-genes (Table 1). However, the average θ_a at *R*-genes was approximately 4-fold higher than that in non–*R*-genes in wild rice (Table 1), suggesting that amino acid substitutions might drive the evolution of *R*-genes.

One possible explanation for the increased number of amino acid polymorphisms and detection of pseudogenes at *R*-gene loci is relaxed constraint. However, relaxed

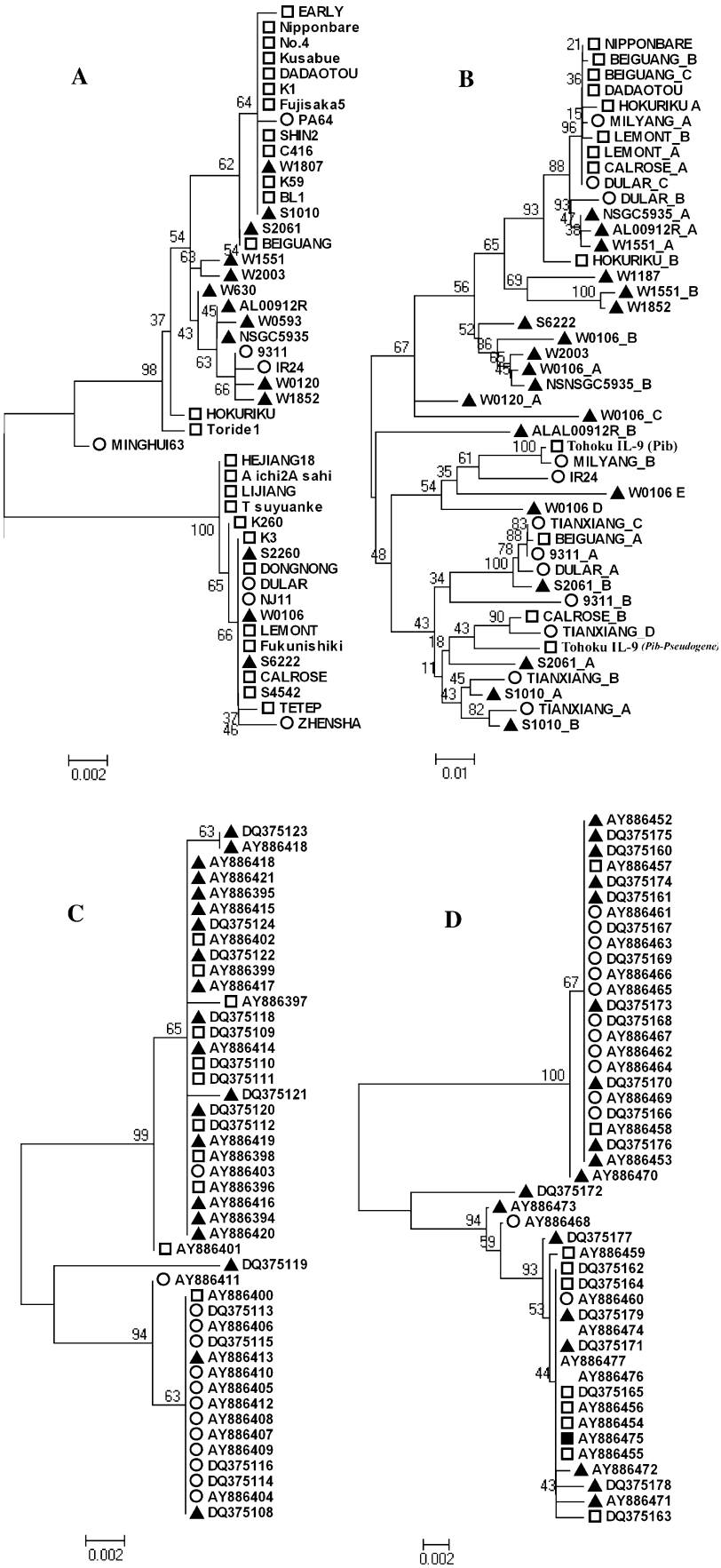

Fig. 2 The phylogenetic trees of two selected *R*-loci (**a** *Os10g22484* homologs, and **b** *Pib* homologs) and two selected non–*R*-loci (**c** *AK100849* homologs and **d** *AK102890* homologs). The phylogenetic trees of all *R*-loci and all non–*R*-loci are exhibited in Figs. S1 and S2, respectively. The full name of wild rice and subspecies name of rice cultivars in the phlogenetic trees are listed in Table S4. The *empty squares* denote *japonica*; the *empty circles* represent *indica*, and the *filled triangles* denote wild rice in the figures

constraint can not explain the increased shared polymorphisms, the increased recombination frequencies, the highly divergent alleles at intermediate frequencies, and the high level of silent polymorphism in cultivated and wild rice. Indeed, the pattern of variation detected at *R*-loci could be better explained by balancing selection because those characters were consistent with the evolutionary scenario of balancing selection (Bakker et al. 2008).

For a small number of *R*-genes (Table S3), significant $\pi_a/\pi_s > 1$ were detected, suggesting that these genes were under positive selection, which might be involved in the specific recognition of pathogen isolates. In contrast, significantly higher rates of nonsynonymous than synonymous substitutions were detected in *R*-genes than in non–*R*-genes, suggesting that diversifying selection was an evolutionary response to selective pressure to resistant pathogens (Table 1).

During the process of domestication, genes important for domestication were subjected to conscious or unconscious directional selection, resulting in decreased variation. In our selected 22 *R*-loci, no cultivar polymorphism was detected at 5 of them (Table 1), suggesting that there is a possible selection sweep or artificial selection on those genes. A similar result was observed at the nonshattering *sh4* locus (Table 1), which is under a strong selection sweep or artificial selection at the major shattering locus (Li et al. 2006). Interestingly, the differences of these gene loci in rice cultivars seem to have originated only once.

In conclusion, our data showed that a similar proportion of nucleotide variation in wild rice was retained in cultivated rice for both *R*-genes and non–*R*-genes. We demonstrated that there is a larger founder effect than previously reported and a strong direct-selection effect in rice genes. In addition, two-directional selection was commonly found in differentiated genes between *indica* and *japonica* rice subspecies, which may explain the mosaic origins of these varieties. Furthermore, in most *R*-genes, no significant differentiation between rice cultivars and its wild relatives was detected. We found evidence for genetic introgression from wild rice, which may have played an important role during the domestication of rice *R*-genes.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grants No. 30570987 and 30870176) to D. T. or J.-Q. C.

References

Allen RL, Bittner-Eddy PD, Grenville-Briggs LJ, Meitz JC, Rehmany AP et al (2004) Host-parasite coevolutionary conflict between Arabidopsis and downy mildew. Science 306:1957–1960

Amante-Bordeos A, Stich LA, Nelson R, Dalmacio RD, Oliva NP et al (1992) Transfer of bacterial blight and blast resistance from the tetraploid wild rice *Oryza minuta* to cultivated rice. Theor Appl Genet 84:345–354

Bakker EG, Toomajian C, Kreitman M, Bergelson J (2006) A genome-wide survey of R gene polymorphisms in Arabidopsis. Plan Cell 18:1803–1818

Bakker EG, Traw MB, Toomajian C, Kreitman M, Bergelson J (2008) Low levels of polymorphism in genes that control the activation of defense response in Arabidopsis thaliana. Genetics 178:2031–2043

Brar DS, Khush GS (1997) Alien introgression in rice. Plant Mol Biol 35:35–47

Ding J, Araki H, Wang Q, Zhang P, Yang S, Chen JQ, Tian D (2007a) Highly asymmetric rice genomes. BMC Genomics 8:154

Ding J, Zhang W, Jing Z, Chen JQ, Tian D (2007b) Unique pattern of *R*-gene variation within populations in *Arabidopsis*. Mol Genet Genomics 277:619–629

Elli J, Dodds P, Pryor T (2000) Structure function and evolution of plant disease resistance genes. Curr Opin Plant Biol 3:278–284

Hudson RR (2000) A new statistic for detecting genetic differentiation. Genetics 155:2011–2014

Jiang H, Wang C, Ping L, Tian D, Yang S (2007) Pattern of LRR nucleotide variation in plant resistance genes. Plant Sci 173:253–261

Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. Science 311:1936–1939

Liu G, Lu G, Zeng L, Wang GL (2002) The two broad-spectrum blast resistance genes, Pi9(t) and Pi2(t), are physically linked on rice chromosome 6. Mol Genet Genomics 267:472–480

Londo JP, Chiang YC, Huang 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

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Mol Biol Evol 7:377–394

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101:12404–12410

Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Curr Opin Plant Biol 8:129–134

Moffat AS (2001) Finding new ways to fight plant diseases. Science 292:2270–2273

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

Nimchuk Z, Eulgem T, Holt BF 3rd, Dangl JL (2003) Recognition and response in the plant immune system. Annu Rev Genet 37:579–609

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc Natl Acad Sci USA 104:8641–8648

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

Sang T, Ge S (2007) The puzzle of rice domestication. J Intergr Plant Biol 49:760–768

Second G (1982) Origin of the genetic diversity of cultivated rice (*Oryza spp.*): Study of the polymorphism scored at 40 isozyme loci. Jpn J Genet 57:25–57

Song Z, Li B, Chen J, Lu B-R (2005) Genetic diversity and conservation of common wild rice (*Oryza rufipogon*) in china. Plant Species Biol 20:83–92

Sun X, Zhang Y, Yang S, Chen JQ, Hohn B, Tian D (2008) Insertion DNA promotes ectopic recombination during meiosis in Arabidopsis. Mol Biol Evol 25:2079–2083

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software, version 4.0. Mol Biol Evol 24:1596–1599

Tang T, Lu J, Huang J, He J, McCouch SR, Shen Y et al (2006) Genomic variation in rice: genesis of highly polymorphic linkage blocks during domestication. PLoS Genet 2:el99

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-speciWc gap penalties, and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tian D, Wang Q, Zhang P, Araki H, Yang S, Kreitman M, Nagylaki T, Hudson T, Bergelson J, Chen JQ (2008) Single-nucleotide mutation rate increases close to indels in eukaryotes. Nature 455:105–108

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol Genet Genomics 272:504–511

Wang X, Sun C (1996) The origin and differentiation of Chinese cultivated rice. China Agricultural University Press, Beijing

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

Weir BS (1996) Genetic data analysis II, 2nd edn. Sinauer, Sunderland, MA

Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y, Chen JQ, Tian D (2006) Genome-wide investigation on the genetic variations of rice disease resistance genes. Plant Mol Biol 62:181–193

Yang S, Gu T, Pan C, Feng Z, Ding J, Hang Y, Chen JQ, Tian D (2008) Genetic variation of NBS-LRR class resistance genes in rice lines. Theor Appl Genet 116:165–177

Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in rice reveals significant expansion of divergent non-TIR NBS Genes. Mol Genet Genomics 271:402–415

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

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