Theor Appl Genet (2005) 111: 1215–1228
DOI 10.1007/s00122-005-0054-8

ORIGINAL PAPER

Kentaro Yoshida · Naohiko T. Miyashita

Nucleotide polymorphism in the *Adh2* region of the wild rice *Oryza rufipogon*

Received: 28 March 2005 / Accepted: 12 July 2005 / Published online: 17 August 2005
© Springer-Verlag 2005

Abstract DNA variation in the alcohol dehydrogenase (*Adh2*) region of the wild rice *Oryza rufipogon* and its related species was analyzed to clarify maintenance mechanisms of the DNA variation in these species. A dimorphic pattern was detected in the *Adh2* region of *O. rufipogon*. The silent nucleotide diversity (π) in the *Adh2* region in *O. rufipogon* was 0.011, which was higher than that of the *Adh1* region in *O. rufipogon*. Especially, a high nucleotide diversity was detected at synonymous sites of the catalytic domain 1. Average nucleotide diversity at silent sites within each of the dimorphic sequence types of the *Adh2* region was similar to that in the *Adh1* region, indicating that the high level of silent polymorphism in the *Adh2* region was caused by the difference between the dimorphic sequence types. On the other hand, the level of replacement polymorphism in the *Adh2* region was as low as that in the *Adh1* region. The neutrality test of Fu and Li indicated significantly negative deviation from the neutral mutation model for the replacement sites of the *Adh2* region. This result suggests purifying selection on the replacement sites of the *Adh2* region, as detected for the *Adh1* region. Significant linkage disequilibria (16.4% of the tests) were detected between the *Adh1* and *Adh2* regions. Even when nonrandom association was tested for the strains belonging to one of the divergent sequence types of the *Adh2* region, significant interlocus linkage disequilibria were detected. The close physical distance and/or epistasis between the two *Adh* regions could be invoked to explain these nonrandom associations.

Communicated by T. Sasaki

K. Yoshida (✉) · N. T. Miyashita
Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University, Sakyo-ku, Kyoto, 606-8502 Japan
E-mail: k720@kais.kyoto-u.ac.jp
Tel.: +81-75-7536145
Fax: +81-75-7536486

Introduction

Rice plants including the cultivated *Oryza sativa* and wild *Oryza rufipogon* are well adapted to semiaquatic conditions by acquisition of high tolerance to anaerobiosis (Sauter 2000). The adaptation is accomplished by the avoidance of acidification in the cytoplasm, which leads to cell death (Menegus et al. 1991). Plant species with low tolerance to anaerobiosis, such as maize, initially use the pathway of lactic fermentation as glucose metabolism under anaerobiosis. The accumulation of lactic acids produced in this pathway results in the acidification in the cytoplasm (Menegus et al. 1989). Instead of the lactic fermentation, rice mainly uses alcoholic fermentation as glucose metabolism under anaerobiosis (Menegus et al. 1991).

Alcohol dehydrogenase is one of the key enzymes in the alcoholic pathway, and catalyzes the reversible reaction from ethanol to acetaldehyde (Sauter 2000). *O. sativa* has two alcohol dehydrogenase loci (*Adh1* and *Adh2*), which are located on the chromosome 11. The physical distance between the two *Adh* regions was about 29 kbp (Tarchini et al. 2000). The ADH1 and ADH2 of *O. sativa* are mainly expressed in the leaf and root tissues, respectively (Xie and Wu 1989). The function of the ADH1 has been studied by using a mutant of *O. sativa* lacking the activity of the ADH1 (Matsumura et al. 1998). The mutant could not recover plant growth after long-term anaerobic stress, implying that the ADH1 is important for the adaptation to anaerobiosis.

Recently, DNA polymorphism in the *Adh1* region of *O. rufipogon* has been studied to elucidate the maintenance mechanism of DNA polymorphism in natural populations of *O. rufipogon* (Yoshida et al. 2004). Nucleotide diversity (π, Nei and Tajima 1981) in the entire region of the *Adh1* in *O. rufipogon* was 0.0020 (π at silent sites: 0.0025), which is one of the lowest values in the *Adh* regions of the plant species studied so far (Gaut and Clegg 1993a, b; Innan et al. 1996; Miyashita et al. 1996; Cummings and Clegg 1998; Lin et al. 2002;

Chiang et al. 2003). The coding region of the *Adh1* of *O. rufipogon* can be classified into three domains: catalytic domain 1 (CD1), co-enzyme-binding domain (CBD) and catalytic domain 2 (CD2). The CD1 had a lower level of polymorphism ($\pi = 0.0009$) than the other domains. Tests of neutrality for the CD1 indicated significantly negative deviation from the neutral mutation model. These results suggest that purifying selection operates on the CD1, reducing the level of DNA polymorphism in the *Adh1* region of *O. rufipogon*.

However, the low level of DNA polymorphism in the *Adh1* region could be related to small effective population size. So far, DNA polymorphism in the *Adh* regions has also been analyzed for five grass species that are intolerant of submergence in water. Among the five grass species, nucleotide diversity for the *Adh* regions in three selfing species *Hordeum vulgare*, *Pennisetum glaucum*, and *Miscanthus sinensis* (π) is 0.0018, 0.0020 and 0.0062, respectively (Gaut and Clegg 1993a; Cummings and Clegg 1998; Chiang et al. 2003). On the other hand, outcrossing grass species *Zea mays* and *M. condensatus* have nucleotide diversity of 0.0174 and 0.0197, respectively (Gaut and Clegg 1993b; Chiang et al. 2003). Since selfing species have a smaller effective population size than outcrossing species (Wright 1938), the lower levels of DNA polymorphism in the selfing species might be due to smaller effective population size. *O. rufipogon* has breeding system of partial selfing and also propagates clonally (Morishima et al. 1961; Xie et al. 2001). We could not reject the possibility that the low level of DNA polymorphism in the *Adh1* region of *O. rufipogon* was caused by small effective population size. Since small effective population size would influence the level of polymorphism over the entire genome, other nuclear regions of *O. rufipogon* should be studied to examine this possibility.

The interspecific comparisons in the *Adh1* region between *O. rufipogon* and its related species showed that the amino acid sequences were conserved among the A genome species of Oryza (Yoshida et al. 2004). On the other hand, the level of replacement divergence between *O. rufipogon* and *O. australiensis* (E genome) was high. One of the replacement substitutions in the CD1 caused physicochemical change of amino acid, according to Miyata et al. (1979), which might influence the ADH1 activity. *O. rufipogon* is found in watery environments, while *O. australiensis* is found on the edge of ditches, and does not seem to be deeply submerged in water even in rainy season (Morishima 2002). The replacement substitutions between *O. rufipogon* and *O. australiensis* might be related to adaptive change in the ADH1, reflecting environmental differences where the Oryza species encounter anaerobiosis.

In this report, we analyzed nucleotide variations in the *Adh2* regions of *O. rufipogon* and its related species (including the cultivated rice). One of the purposes is to clarify maintenance mechanisms of DNA polymorphism in the *Adh2* region. We compared levels of DNA polymorphism between the *Adh1* and *Adh2* regions to examine the possible influence of small effective population size on the low level of nucleotide variation detected in the *Adh1* region. The second purpose is to examine the involvement of replacement substitutions in the *Adh2* region with the adaptation in anaerobic environments, as suggested for the *Adh1* region. We compared level of replacement divergence of the *Adh2* region between *O. rufipogon* and the other A genome species with that between *O. rufipogon* and *O. australiensis*. The third purpose is to elucidate epistatic interaction between the *Adh1* and *2* regions of *O. rufipogon*. It is possible that two adjacent *Adh* genes could be coregulated for anaerobic adaptation. Epistatic interaction could cause linkage disequilibrium (LD) between loci in natural populations (Kimura 1956). We test LD between the two *Adh* regions in *O. rufipogon*. In addition, *O. rufipogon* is considered to be the wild ancestor of *O. sativa* (Oka and Chang 1959). *O. sativa* Japonica and Indica are assumed to be originated from different strains of *O. rufipogon* (Second 1982; Ishii et al. 1988). The study of DNA polymorphism in the *Adh2* region of *O. rufipogon*, *O. sativa* Japonica and Indica would contribute to clarifying the phylogenetic relationship between these species.

Materials and methods

Plant materials

Twenty Asian wild rice species *O. rufipogon*, four of its related species and five cultivated rice *O. sativa* were used for analyzing DNA polymorphism in the *Adh2* region (Table 1). Eight strains of *O. rufipogon* were examined for allozyme variation of ADH. Seeds and DNAs are maintained in the Laboratory of Plant Breeding, Faculty of Agriculture, Kobe University. They were originally provided from the National Institute of Genetics (Japan), Sizuoka University (Japan), and the International Rice Research Institute (Philippines). The seeds were sterilized by using the benomyl (Du pont, Wilmington, DE, USA) at 28°C over night, and germinated in the dark at 28°C. The plants of each accession used for DNA extraction were grown in a pot under 28°C and 14-h light conditions. The plants used for the allozyme experiment were grown in a green house during August 2002 in Kyoto, Japan.

DNA extraction, PCR amplification and sequencing

Total DNA was extracted from leaves and stems with the modification of the CTAB method (Weising et al. 1991), and used for PCR amplification of the *Adh2* region. The primers for the PCR amplification for *O. rufipogon*, *O. sativa* Japonica, *O. sativa* Indica, *O. barthii* and *O. meridionalis* are ORADH2-1: 5′-TCCTCCTTGTCTTCACTCTG-3′ and ORADH2-

Table 1 Plant materials

Species	Genome	Strain	Origin	Life form[a]	Allozyme[b]	Source[c]
Oryza rufipogon	A	W108	India	Perennial	○	1
	A	W120	India	Perennial	○	1
	A	W593	Malaysia	Perennial	○	1
	A	W1956	China	Perennial	○	1
	A	W1965	China	Perennial	○	1
	A	W1972	Indonesia	Perennial	○	1
	A	W1976	Indonesia	Perennial	○	1
	A	CB22	Cambodia	Perennial		2
	A	CB27	Cambodia	Perennial		2
	A	CT51	Vietnam	Perennial		2
	A	CT56	Vietnam	Perennial		2
	A	CT57A	Vietnam	Perennial		2
	A	LV27	Laos	Perennial		2
	A	LV61	Laos	Annual		2
	A	VT51	Laos			2
	A	SN	Thailand	Perennial		2
	A	KA	Thailand	Annual		2
	A	W630	Myanmar	Annual	○	1
	A	YG2A	Myanmar	Perennial		2
	A	PT1A	Myanmar	Perennial		2
O. sativa Japonica	A	Nourin22	Japan	Annual		1
	A	YT1A	Japan	Annual		3
	A	Nipponbare	Japan	Annual		4
O. sativa India	A	435	Sri Lanka	Annual		1
	A	IR36		Annual		1
O. barthii	A	W607	Guinea	Annual		1
O. glumaepatula	A	W1167	Cuba			1
O. meridionalis	A	W1627	Australia	Annual		1
O. australiensis	E	101397	Australia	Perennial		5

[a]The annual and perennial strains of *O. rufipogon* in this table have been referred to as *O. nivara* and *O. rufipogon*, respectively (Sharma and Shastry 1965)

[b]The strains of *O. rufipogon* examined allozyme variation indicate the circle

[c]1: National Institute of Genetics, Japan, 2: Sizuoka University, Japan, 3: Genbank accession AF172282, 4: Genbank accession AC123515, 5: International Rice Research Institute, Philippines

101: 5′-GCCACAATGCTGACAATAAA-3′, which are located in the 5′ and 3′ flanking regions, respectively. By using these primers, a 3.2-kbp region of the *Adh2* was amplified. The primers for the PCR amplification for *O. glumaepatula* are ORADH2-3.1: 5′- ATGGCGACAGCCGGGAAGGT-3′, which is located in the exon 1, and ORADH2-101. By using these primers, a 2.8-kbp region of the *Adh2* was amplified. The primers for the PCR amplification for *O. australiensis* are ORADH2-3.1 and ORADH2-103: 5′-CGTCCCCTTGAGCGTCTTCT-3′. A 2.3-kbp region of the *Adh2* from exons 1–9 was obtained, which lacks a part of exon 9 and 10. These four primers were designed using published sequence information of the *Adh2* region of *O. sativa* (Genbank accession AF172282). Taq polymerase (Roche Applied Science) was used for the PCR reaction. The PCR products were cloned into the plasmid pUC118 (TaKaRa), which was used as template for sequencing reaction. Sequence reaction was conducted by using the Thermo Sequenase fluorescent-labeled cycle-sequencing kit with 7-deaza-dGTP (Amersham/Pharmacia Biotech, Piscataway, NJ, USA). Sequence was determined by a Pharmacia ALFred sequencer. We mixed three plasmid clones at almost the same molarity to eliminate PCR artifacts. Sequencing primers were designed at about 500-bp intervals. Newly determined sequences were deposited in the DDBJ databank under accession numbers AB208516-AB208542.

Allozyme of ADH

A 12.5% starch gel was prepared, and a slit was made near the cathodal end of gel. Leaves and roots at 1 month after germination were ground with extraction buffer grade I^+, separately. The buffer grade I^+ is composed of 0.1 M Tris–HCl pH7.5, 5% sucrose (w/v), 10 mM diethyldithiocarbamate, 21 mM mercaptoethanol (0.15% v/v), 0.2% bovine serum albumin (w/v) and 5% PVP-40 (w/v). A piece of filter paper was used to absorb the extract, which was inserted in the slit. Electrophoresis was conducted for 5 h at 160 volts under 4°C. After electrophoresis, the starch gel was stained for ADH as described by Glaszmann et al. (1988). The ADH bands were determined on the basis of the report by Xie and Wu (1989). The fast band is ADH2 homodimer, the middle band ADH1-ADH2 heterodimer and the slow band ADH1 homodimer.

Data analysis

DnaSP program version 3.50 (Rozas and Rozas 1999) was used to analyze intra- and interspecific DNA variations. Nucleotide diversity (π) and θ (4Neμ: Watterson 1975) were estimated after removing indels. Tests of Tajima (Tajima 1989) and Fu and Li (Fu and Li 1993) were conducted to investigate departure from the neutrality. Genetic distance (K) between *O. rufipogon* and its related species was calculated by Jukes and Cantor method (Jukes and Cantor, 1969). Genetic distance for the 5′ flanking region of the *Adh2* between *O. rufipogon* and *O. glumaepatula* could not be estimated due to missing sequence information on the 5′ flanking region of *O. glumaepatula*. Also we could not estimate genetic distance for both flanking and CD2 regions of *Adh2* between *O. rufipogon* and *O. australiensis*, because we could not determine these sequences of *O. australiensis*. MEGA program version 2.1 (Kumar et al. 2002) was used to construct neighbor-joining (NJ) tree. Maximum parsimony (MP) tree was constructed with a heuristic search using PAUP 3.1.1 (Swofford 1993). Informative indels were included as a fifth base. Heuristic search was also performed to estimate the number of replacement substitutions on each of tree branches of *O. rufipogon* and *O. australiensis*. Homology plot analysis between the *Adh1* and *Adh2* regions of *O. rufipogon* was conducted by using EMBOSS GUI v 1.12 dottup (Rice et al. 2000). Intra- and interlocus LDs for polymorphic variations detected in the *Adh1* and *Adh2* regions of 17 *O. rufipogon* analyzed for both regions were examined by χ^2 test implemented in the DnaSP program. For the χ^2 tests, we included indels as DNA variations irrespective of their length to test LD, but SSR (simple sequence repeat) polymorphisms were excluded from the LD analysis.

Results

Polymorphic sites in the *Adh2* region of *O. rufipogon*

In the 3.3-kbp *Adh2* region of *O. rufipogon*, 152 nucleotide variations were detected (95 sites and 57 indels) (Fig. 1). In the 5′ flanking region, there were 31 nucleotide variations (13 sites and 18 indels), none of which were located in the TATA box or putative regulatory elements, involved in the adjustment of *Adh2* expression in anaerobic conditions (Xie and Wu 1989). This result suggests that the nucleotide variations in the 5′ flanking region do not change the transcriptional regulation of the *Adh2* gene of *O. rufipogon*. In the coding region, there were 24 polymorphic sites (16 synonymous and 8 replacement) and no indel variations, of which 9 polymorphic sites (8 synonymous and 1 replacement) were found more than once in the samples. The nonsingleton replacement (CTG (Val) and CTC (Leu)) at the position 1566 (CD1) did not cause drastic physicochemical change of amino acid, according to Miyata et al. (1979). The singleton replacements at the positions 1540 (CD1), 1654 (CD1), 2322 (CBD) and 2499 (CBD) had amino acid distances ranging between 2.37 and 3.06 (Miyata et al. 1979), which indicates distinct physicochemical changes without causing the structural disruption of the protein. The other three replacements did not cause any physicochemical change or structural disruption. Electrophoresis of ADH protein detected no changes in the mobility among strains, which have these singleton replacement sites (data not shown).

Pattern of nucleotide polymorphism in the *Adh2* region of *O. rufipogon*

A dimorphic pattern was detected in the *Adh2* region of *O. rufipogon* (Fig. 1). From the distribution of polymorphic variations, seven distinct sequence types can be defined. The *Adh2* region can be divided into seven blocks on the basis of the partition pattern of the strains, although the boundary between the blocks 6 and 7 was not clear because of the unclear dimorphic pattern. There was no correlation between these sequence types and their geographic origins, implying that the dimorphic pattern was not caused by geographic isolation among local populations of *O. rufipogon*. No association between the dimorphic pattern and life form of *O. rufipogon* was detected either (Fig. 1). For example, both LV61 and KA strains of *O. rufipogon* were annual types. The LV61 and KA strains belonged to the sequence types 1 and 7 separately, which were the most divergent among the seven sequence types (Fig. 1).

Since the sequence types 1 and 7 were the most divergent, these two sequence types could be ancestral. The other sequence types could be recombinants between the sequence types 1 and 7. Assuming two double recombinations in the history of the *Adh2* region in *O. rufipogon*, the other five sequence types could be produced by five intragenic recombinations (Fig. 2a). The times of recombination events were estimated by using the method of Innan et al. (1996) on the basis of the number of nucleotide differences at silent sites (including synonymous sites) between recombinants and ancestral sequence types. Given synonymous substitution rate of 6.5×10^{-9} per site per year for grass *Adh* genes (Gaut et al. 1996), it was shown that these recombination events occurred 0.14 ~ 0.81 MYA (Fig. 2b). Since the two ancestral sequence types were estimated to have diverged 1.73 MYA, the intragenic recombinations in the *Adh2* region of *O. rufipogon* would have occurred only recently.

Level of nucleotide polymorphism in the *Adh2* region of *O. rufipogon*

Level of nucleotide polymorphism in the *Adh2* region of *O. rufipogon* was estimated (Table 2). The levels of polymorphism (π and θ) in the entire region of the *Adh2*

Fig. 1 Summary of DNA variations in the 3.3-kb *Adh2* region of *O. rufipogon* and *O. sativa* Japonica and Indica. The structure of the *Adh2* region from the sites 0 to 3,325 is shown at the *center* of figure. The *black*, *white* and *gray boxes* indicate the exons coding the catalytic domain 1, the co-enzyme-binding domain and the catalytic domain 2, respectively. Singleton and nonsingleton nucleotide sites in *O. rufipogon* are shown as *vertical bars*. *Open circles* show replacement polymorphism. *Triangles* on the top of the *vertical bar* show indels. DNA variations are summarized at the *bottom* of figure, where *dots* indicate the variation identical to W1976 sequence. The *d*, *m*, *s* and *r* are abbreviations for indels, microsatellite, synonymous and replacement polymorphism, respectively. + and − indicate indels, respectively. A_n, T_n and G_n indicate that adenine, thymine and guanine repeated *n* times, respectively. m_n indicates microsatellite with *n* repeat. *A* indicates the annual type of *O. rufipogon*

were 0.008 and 0.009, respectively, which are higher than those of the *Adh1* of *O. rufipogon* ($\pi = 0.002$ and $\theta = 0.003$: Yoshida et al. 2004). Polymorphism was high in the 5′ flanking region and introns of the *Adh2*. In the coding region, synonymous sites had a high level of variation, whereas the level of variation at replacement sites was as low as that of the *Adh1* region of *O. rufipogon* (*ibid.*). These results indicate that the difference in the level of nucleotide polymorphism between the *Adh1* and *Adh2* regions was caused by the difference at silent sites.

The level of polymorphism for the sequence type 1 was estimated (Table 3). We could not estimate level of polymorphism for the sequence type 7, which was detected only once in the sample. The level of polymorphism was low over all the functionally different regions. Especially, a low nucleotide diversity was detected at replacement sites.

Each block (Fig. 1) could be classified into partition A identical to the sequence type 1 and partition B identical to the sequence type 7. Average silent nucleotide diversity (π) for the partitions A and B was 0.004 and 0.003, respectively (Fig. 3). These values were smaller than that ($\pi = 0.011$) at silent sites of the entire *Adh2* region (Table 2). Taken together, these results indicate that the high level of polymorphism at the silent sites of the entire *Adh2* region was caused by nucleotide differences between the two divergent sequence types.

Difference in the level of nucleotide polymorphism among the ADH2 domains

We estimated the level of nucleotide polymorphism for subcoding region corresponding to each of the three domains (CD1, CBD and CD2) of the ADH2 in *O. rufipogon* (Table 4). The highest nucleotide diversity was detected in the CD1, while nucleotide diversity in the CBD and CD2 was low. This result is different from that of the ADH1, where the CD1 had a lower diversity than the other domains (Yoshida et al. 2004). When only the sequence type 1 was analyzed, all the domains had a low level of

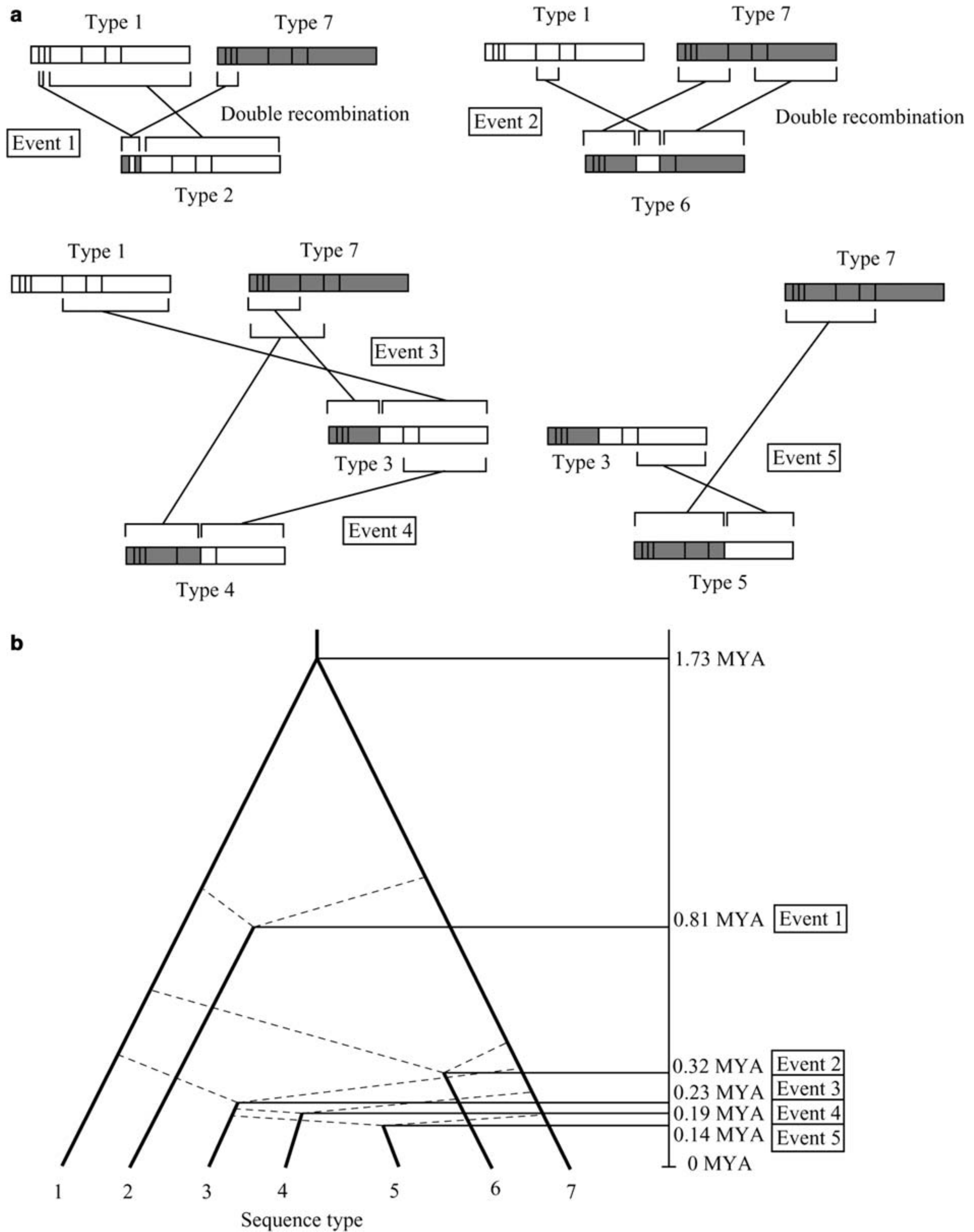


Fig. 2 **a** Diagram of recombination events for the five recombinant sequence types detected in this study. **b** Estimated recombination time. *Dashed lines* connect the recombinant to the parental sequence types

polymorphism (data not shown). This result indicates that the high nucleotide diversity in the CD1 was mainly due to the difference between the two divergent sequence types.

Table 2 Summary of DNA variation in the *Adh2* of *O. rufipogon* and its related species

(# of samples = 20)	# of sites	S	π	θ	D[a]	D*[b]	$K^{c}_{R\text{-}SJ}$	$K^{c}_{R\text{-}SI}$	$K^{c}_{SJ\text{-}SI}$	$K^{c}_{R\text{-}G}$	$K^{c}_{R\text{-}B}$	$K^{c}_{R\text{-}M}$	$K^{c}_{R\text{-}A}$
Entire region	3325	95	0.008	0.009	−0.29	−1.06	0.007	0.008	0.001	0.007	0.016	0.020	0.071
Silent sites	2158.2	89	0.011	0.012	−0.14	−0.84	0.009	0.010	0.001	0.010	0.022	0.030	0.106
Coding region	1137	24	0.004	0.006	−1.08	−1.79	0.003	0.004	0.002	0.004	0.009	0.012	0.040
Synonymous sites	262.2	16	0.014	0.017	−0.70	−1.05	0.011	0.015	0.006	0.010	0.034	0.047	0.138
Replacement sites	874.8	8	0.001	0.003	−1.57	−2.62*	0.001	0.001	0.001	0.002	0.002	0.002	0.012
Noncoding region	2188	71	0.011	0.011	0.00	−0.76	0.009	0.009	0.001	0.009	0.021	0.025	0.100
5′ Flanking region	409	13	0.013	0.011	0.55	−0.67	0.014	0.012	0.003	–[d]	0.013	0.018	–
Introns	1559	56	0.012	0.012	−0.06	−0.62	0.009	0.010	0.000	0.011	0.025	0.030	0.100
3′ Flanking region	220	2	0.001	0.003	−1.51	−2.05	0.000	0.000	0.000	0.000	0.005	0.005	–

*$P < 0.05$
[a]Tajima's D
[b]Fu and Li's D*
[c]K is divergence between two species shown in subscript; R indicates *O. rufipogon*; SJ indicates *O. sativa* Japonica; SI indicates *O. sativa* Indica; G indicates *O. glumaepatula*; M indicates *O. meridionalis*; A indicates *O. australiensis*
[d]Divergence could not be estimated

Tests of neutrality for nucleotide polymorphism in the *Adh2* region of *O. rufipogon*

Tests of Tajima and Fu and Li were conducted to examine the neutrality of nucleotide polymorphism for the functionally different regions of the *Adh2* (Table 2). Fu and Li's D* test gave a significantly negative value for replacement sites. This result was consistent with the low level of polymorphism at the replacement sites of the *Adh2* region (Table 2), indicating that purifying selection operates on the replacement sites of the *Adh2* region.

When the neutrality tests were conducted only for the sequence type 1, significantly negative values were detected in the coding region (Table 3). The Tajima's D and Fu and Li's D* were also largely negative for replacement sites, although the tests did not give significant values. Considering the lowest nucleotide diversity at the replacement sites, these negative values for the coding region and replacement sites of the sequence type 1 could be explained by purifying selection.

When these tests were applied to each of the three domains, Tajima's D value was significantly negative for the CBD (Table 4). Both tests gave significantly negative values for replacement sites in the CBD. Considering the low level of polymorphism in the CBD (Table 4), these results indicate that purifying selection operates on the CBD. When the tests of neutrality were conducted for each of the three domains of the sequence type 1, no significant value was obtained (data not shown). This is probably due to lack of the power in the tests, because the number of segregating sites was small.

Table 3 Summary of DNA variation for the sequence type 1 of *O. rufipogon* and its related species

	# of sites	S	π	θ	D[a]	D*[b]	$K^{c}_{R\text{-}SJ}$	$K^{c}_{R\text{-}SI}$	$K^{c}_{R\text{-}G}$	$K^{c}_{R\text{-}B}$	$K^{c}_{R\text{-}M}$	$K^{c}_{R\text{-}A}$
Entire region	3325	22	0.002	0.003	−1.42	−1.71	0.001 (0.016)	0.002 (0.017)	0.002 (0.016)	0.016 (0.019)	0.018 (0.023)	0.069 (0.072)
Silent sites	2315.2	18	0.002	0.003	−1.28	−1.58	0.001 (0.021)	0.002 (0.023)	0.003 (0.022)	0.021 (0.024)	0.024 (0.032)	0.100 (0.105)
Coding region	1137	8	0.002	0.003	−1.80*	−2.03*	0.001 (0.007)	0.002 (0.008)	0.002 (0.008)	0.010 (0.008)	0.011 (0.011)	0.036 (0.036)
Synonymous sites	262.2	4	0.003	0.006	−1.61	−1.80	0.003 (0.027)	0.007 (0.031)	0.002 (0.027)	0.037 (0.023)	0.045 (0.043)	0.127 (0.151)
Replacement sites	874.8	4	0.001	0.002	−1.61	−1.80	0.001 (0.002)	0.001 (0.002)	0.002 (0.002)	0.003 (0.003)	0.002 (0.002)	0.012 (0.012)
Noncoding region	2053	14	0.002	0.003	−1.07	−1.37	0.001 (0.021)	0.002 (0.021)	0.002 (0.021)	0.019 (0.025)	0.021 (0.030)	0.095 (0.098)
5′ Flanking region	409	3	0.003	0.003	−0.55	−0.73	0.002 (0.023)	0.002 (0.026)	–[d]	0.004 (0.023)	0.010 (0.026)	–
Introns	1559	10	0.002	0.002	−1.07	−1.37	0.001 (0.023)	0.001 (0.024)	0.003 (0.025)	0.024 (0.029)	0.026 (0.036)	0.100 (0.105)
3′ Flanking region	220	1	0.001	0.002	−1.09	−1.19	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.005 (0.009)	0.005 (0.005)	–

Nine strains of *O. rufipogon* belonging to the sequence type 1 were used for the analysis of polymorphism. Divergence between sequence type 7 and its related species is shown in parentheses
*$P < 0.05$
[a]Tajima's D
[b]Fu and Li's D*
[c]K is divergence between two species shown in subscript; R indicates *O. rufipogon*; SJ indicates *O. sativa* Japonica; SI indicates *O. sativa* Indica; G indicates *O. glumaepatula*; M indicates *O. meridionalis*; A indicates *O. australiensis*
[d]Divergence could not be estimated

Sequence type	Partition	Block 1	2	3	4	5	6	7	Average
1 (n=9)	A	0	0.002	0.006	0.007	0.004	0.005	0.003	0.004
2 (n=1)									
3 (n=2)									
4 (n=2)									
5 (n=2)	B	0	0.006	0.005	0.001	0.001			
6 (n=1)							0.000	0.004	0.003
7 (n=1)									
All (n=20)		0	0.021	0.019	0.030	0.010	0.017	0.005	0.011

Fig. 3 Summary of nucleotide diversity in each block. The sequences in the each block are classified into the partition (**A**) identical to the sequence types 1 and the partition (**B**) identical to the sequence type 7

Phylogenetic relationship between *O. rufipogon* and its related species

To clarify the phylogenetic relationship between *O. rufipogon* and its related species, a NJ tree was constructed based on the nucleotide variations in the entire region of the *Adh2* (Fig. 4). *O rufipogon* was separated from *O. barthii*, *O. meridionalis* and the E-genome species *O. australiensis*. Especially, *O. australiensis* was highly diverged from *O. rufipogon*. The sequence type 1 of *O. rufipogon* was grouped with *O. glumaepatula*, *O. sativa* Japonica and Indica. A MP tree was also constructed. The topology of the MP tree was almost the same as that of the NJ tree (data not shown).

As observed for the *Adh1* region (Yoshida et al. 2004), *O. sativa* Japonica and Indica were not clearly separated into the different clusters in the tree (Fig. 4). One of *O. sativa* Japonica (Nourin 22) was closer to *O. sativa* Indica (435 and IR36) than the other Japonica strains (YT1A and Nipponbare), which were close to a Myanmar *O. rufipogon* (YG2A). Two of *O. sativa* Indica (435 and IR 36) and an Indian *O. rufipogon* (W120) formed a single cluster, but the bootstrap probability did not support this clustering. From this analysis, it is difficult to conclude the birth place of *O. sativa*.

Species-specific difference between *O. sativa* Japonica and Indica was detected in the region of the block 1 of

Table 4 Summary of DNA variation in each domain of the *Adh2* of *O. rufipogon* and its related species

(# of samples = 20)	# of sites	S	π	θ	D[a]	D*[b]	$K^c_{R\text{-}SJ}$	$K^c_{R\text{-}SI}$	$K^c_{SJ\text{-}SI}$	$K^c_{R\text{-}G}$	$K^c_{R\text{-}B}$	$K^c_{R\text{-}M}$	$K^c_{R\text{-}A}$
Catalytic domain1													
All sites	525	14	0.007	0.008	−0.326	−1.390	0.006	0.006	0.001	0.005	0.008	0.014	0.031
Synonymous sites	122.8	10	0.023	0.023	−0.036	−0.966	0.011	0.020	0.004	0.018	0.030	0.058	0.097
Replacement sites	402.2	4	0.002	0.003	−0.827	−1.693	0.001	0.001	0.000	0.001	0.002	0.001	0.011
Co-enzyme-binding domain													
All sites	429	9	0.003	0.006	−1.855*	−1.719	0.001	0.005	0.004	0.004	0.005	0.008	0.052
Synonymous sites	99.9	5	0.008	0.014	−1.392	−0.413	0.004	0.013	0.010	0.004	0.034	0.031	0.195
Replacement sites	329.1	4	0.001	0.003	−1.868*	−2.627*	0.001	0.002	0.002	0.004	0.001	0.001	0.013
Catalytic domain2													
All sites	183	1	0.001	0.002	−1.164	−1.540	0.002	0.000	0.002	0.000	0.022	0.017	–[d]
Synonymous sites	39.5	1	0.003	0.007	−1.164	−1.540	0.001	0.001	0.000	0.001	0.081	0.054	–
Replacement sites	143.5	0	0.000	0.000	NA	NA	0.002	0.000	0.002	0.000	0.007	0.007	–

*$P < 0.05$ [a]Tajima's D

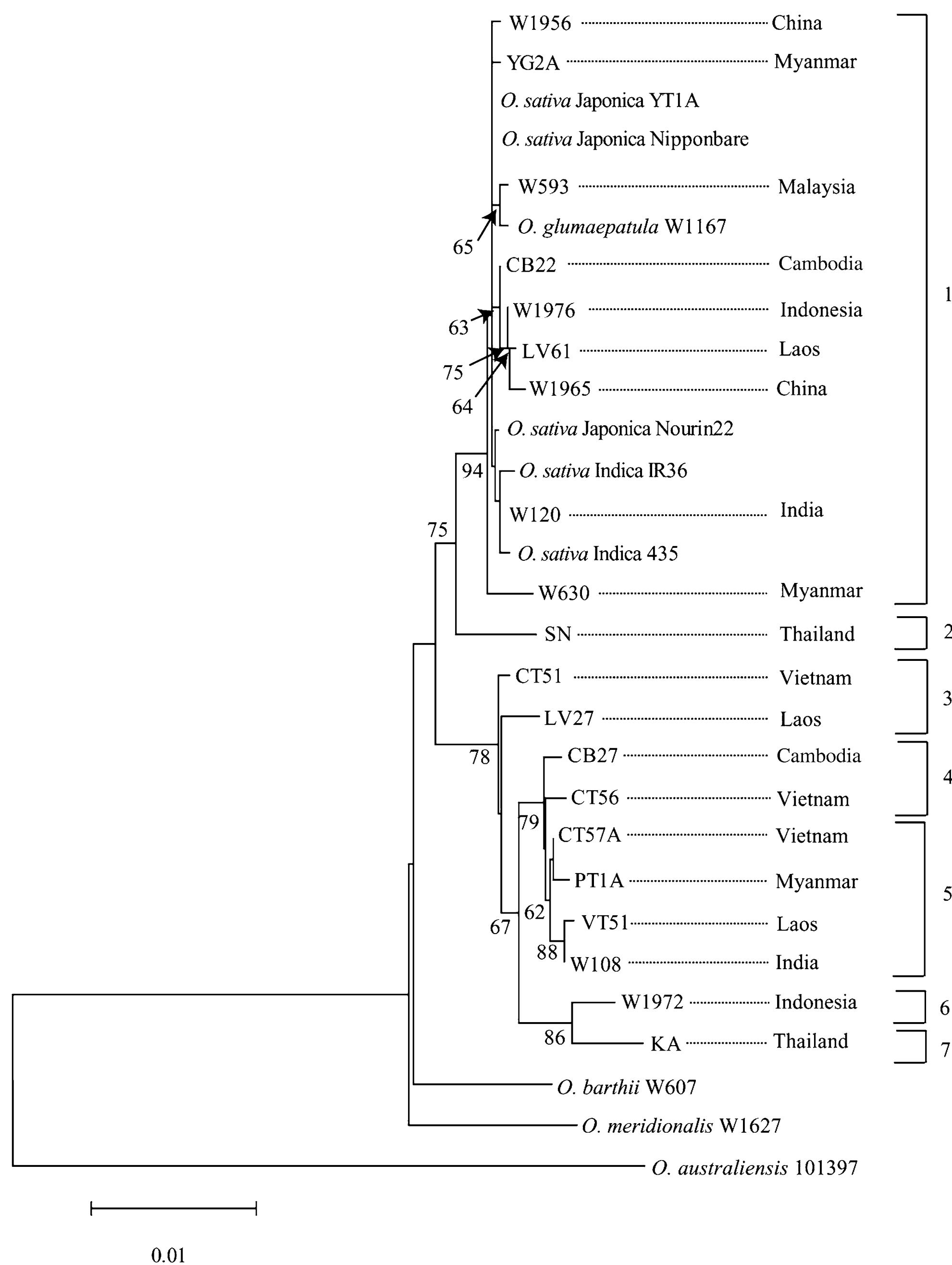
[b]Fu and Li's D*

[c]K is divergence between two species shown in subscript; R indicates *O. rufipogon*; SJ indicates *O. sativa* Japonica; SI indicates *O. sativa* Indica; G indicates *O. glumaepatula*; M indicates *O. meridionalis*; A indicates *O. australiensis*

[d]Divergence could not be estimated

NA not applied

Fig. 4 Neighbor-joining tree for 20 strains of *O. rufipogon* and its related species using genetic distance calculated by Jukes and Cantor method on the basis of DNA variation in the entire region of *Adh2*. Bootstrap probabilities >60% from 1,000 replications are shown. The *scale bar* of genetic distance is shown at the *bottom* of the tree. Strains of *O. rufipogon* also show their geographic origins. The number on the right side of the figure indicates the sequence type of *O. rufipogon*



O. rufipogon (Fig. 1). Nucleotide sequence in all the blocks of *O. sativa* Indica was similar to that of the sequence type 1 of *O. rufipogon*. On the other hand, the sequence in the block 1 of *O. sativa* Japonica was identical with that of the sequence type 7, and the sequence in the other blocks was largely identical with that of the sequence type 1. This observation implies that the *Adh2* region of *O. sativa* Japonica is a recombinant between the sequence types 1 and 7. The time of the recombination event was estimated to be 0.10 MYA, given synonymous substitution rate of 6.5×10^{-9} per site per year for grass *Adh* genes (Gaut et al. 1996). This recombination event occurred before the domestication of *O. sativa*, which is estimated 0.01 MYA (Wasano 1995).

Nucleotide divergence in the *Adh2* region between *O. rufipogon* and its related species

Level of divergence (K) between *O. rufipogon* and its related species was estimated for the functionally different regions of the *Adh2* (Table 2). In all the regions, *O. australiensis* is highly diverged from the other species. In the coding region, the level of divergence at replacement sites was lower than that at synonymous

sites. Particularly, the level of replacement divergence was constant over the A genome species. When level of divergence between each of the sequence types 1 and 7 of *O. rufipogon* and its related species was estimated, the replacement divergence between *O. rufipogon* and A genome species was also at a low level (Table 3). These results indicate that the amino acid sequences of the ADH2 were conserved among the A genome species.

The level of divergence was estimated for the three domains of the ADH2 (Table 4). When *O. rufipogon* and *O. australiensis* were compared, genetic distance at synonymous sites of the CBD was larger than that of the CD1. The distance at replacement sites of the CBD was a little larger than that of the CD1 between *O. rufipogon* and *O. australiensis*. The similar result was obtained when each of the divergent sequence types of *O. rufipogon* and *O. australiensis* was compared (data not shown). These results contrast with the observations for the *Adh1* region that the levels of synonymous and replacement divergence in the CD1 were higher than those in the CBD between *O. rufipogon* and *O. australiensis* (Yoshida et al. 2004). In other words, different domains are more diverged in the two *Adh* genes.

Heuristic search for replacement substitutions between *O. rufipogon* and *O. australiensis*

To determine the branches on which the observed replacement substitutions in the CD1 and CBD of the ADH2 occurred, a heuristic search was conducted for *O. rufipogon* and *O. australiensis* (Table 5). When *H. vulgare* was used as outgroup species, the number of replacement substitutions on the branch of *O. rufipogon* and *O. australiensis* was six and one, respectively. When *Z. mays* was used, five and two replacement substitutions were detected on the branch of *O. rufipogon* and *O. australiensis*, respectively. These results indicate that the number of replacement substitutions on the branch of *O. rufipogon* was larger than that of *O. australiensis*. However, the χ^2 tests did not support the difference statistically (*Z. mays* is used as outgroup: $\chi^2_{df=1} = 1.29$ NS, *H. vulgare*: $\chi^2_{df=1} = 3.57$ NS). This may be because the number of substitutions was small.

Two of the replacement substitutions on the branch of *O. rufipogon* were detected in the CD1 (Table 5). One of the replacement substitutions was located at the 103rd amino acid position of the ADH2. This amino acid position is involved in the formation of a lobe that binds the second zinc atom of the subunit of ADH (Eklund et al. 1976). The second zinc atom is essential for the catalytic activity. The other was located at the 130th amino acid position, which is not involved in the secondary structure of the CD1. The amino acid distance between *O. rufipogon* and *O. australiensis* at these positions was more than one, according to Miyata et al. (1979). Therefore, both replacement substitutions caused distinct physicochemical changes of amino acid. Considering the amino acid position of these substitutions, the replacement substitution at the 103rd position might be related to the activity of the ADH2. Three replacement substitutions (Table 5) in the α helix of the CBD of the ADH2 did not cause any distinct physicochemical changes of amino acid.

On the other hand, for the *Adh1* region, the number of replacement substitutions on the branch of *O. rufipogon* was not high, and depended on the outgroup species in the heuristic search (Table 5). Two replacement substitutions in the CD1 caused physicochemical changes of amino acid. One of these substitutions occurred on the branch of *O. rufipogon*, irrespective of the choice between *Z. mays* and *H. vulgare* as outgroup species. We discussed possible reasons why the replacement substitutions with physicochemical changes of amino acid in the ADH1 and ADH2 occurred more on the branch of *O. rufipogon* in Discussion.

Table 5 Summary of amino acid substitution between *O. rufipogon* and *O. australiensis*

	Amino acid position	Domain	Amino acid (codon)		Branch of substitution[a]		Distance[b]	Structure of protein[c]
			R[d]	*A*[e]	(*H. vulgare*)	(*Z. mays*)		
Adh2	4	CD1	Ala (GCC)	Thr (ACG)	A	A	0.90	
	103	CD1	Asp (GAT)	Ala (GCT)	R	R	2.37	Second zinc-binding lobe
	130	CD1	Lys (AAG)	Gln (CAG)	R	R	1.06	
	136	CD1	Lys (AAG)	Asn (AAC)	R	A	1.84	
	179	CBD	Phe (TTC)	Ile (ATC)	R	R	0.61	α-helix
	183	CBD	Phe (TTT)	Leu (CTT)	R	R	0.63	α-helix
	187	CBD	Val (GTG)	Leu (CTG)	R	R	0.91	α-helix
Adh1	6	CD1	Lys (AAG)	Glu (GAG)	A	A	1.14	
	19	CD1	Ala (GCG)	Gly (GAG)	R	R	0.91	
	62	CD1	Val (GTG)	Met (ATG)	R	A	0.62	
	103	CD1	Ala (GCC)	Pro (CCC)	R	A	0.06	Second zinc-binding lobe
	108	CD1	Ala (GCA)	Pro (CCA)	A	A	0.06	Second zinc-binding lobe
	166	CD1	Ala (GCA)	Glu (GAG)	R	R	2.46	

[a] The branch of *O. rufipogon* or *O. australiensis* that amino acid substitutions occurred on
[b] Amino acid distance (Miyata et al. 1979)
[c] The structure of protein that amino acid is involved in
[d] R indicates *O. rufipogon*
[e] A indicates *O. australiensis*

Divergence between the *Adh1* and *Adh2* regions of *O. rufipogon*

Homology plot analysis between the investigated regions of the *Adh1* and *Adh2* of *O. rufipogon* was conducted to examine gene conversion between these regions. Homology was detected only in the exons (data not shown). When the *Adh1* region and each of the two divergent sequence types of the *Adh2* region were compared, homology was detected again only in the exons. These results indicate no gene conversion between the two *Adh* regions in *O. rufipogon*.

Since only the sequences for the exons of the *Adh1* and *Adh2* regions can be aligned, NJ tree based on the nucleotide variations in the exons was constructed to clarify the phylogenetic relationship of these loci between *O. rufipogon* and grass family (data not shown). Clearly, the two *Adh* genes of *O. rufipogon* were divided into different clades, where those of other grass species clustered separately. These observations agree with the idea that the *Adh1* and *Adh2* duplicated before the divergence of the grass family (Gaut et al. 1999).

Linkage disequilibrium in the *Adh1* and *Adh2* regions of *O. rufipogon*

Intra- and interlocus LD was examined by χ^2 tests to investigate the possibility of epistasis within and between the *Adh1* and *Adh2* regions of *O. rufipogon* (Fig. 5). The number of informative polymorphic variations including indels was 14 and 68 in the *Adh1* and *Adh2* regions, respectively. The number of tests within the *Adh1* region was 91, of which 26 (28.6%) gave significant results. In the *Adh2* region, 1,785 of the 2,278 tests (78.4%) were significant. When only the strains of *O. rufipogon* belonging to the sequence type 1 of the *Adh2* were used in the analysis, the number of informative polymorphic variations was five and three in the *Adh1* and *Adh2* regions, respectively. Five of the ten tests were significant within the *Adh1* region, whereas significance was not detected within the *Adh2* region. These results indicate that the large number of significant pairwise comparisons within the *Adh2* region was caused by the dimorphic pattern.

The number of tests between loci was 952, of which 156 (16.4%) were significant. Among these significant comparisons, 56 (35.9%) were detected between polymorphic variations in the introns. This percent of significant pairs in the introns was higher than that within each region (*Adh1*: 26.9%, *Adh2*: 34.2%). Even when the strains of the sequence type 1 were used, four comparisons in the noncoding region were significant. This result contrasted with the observation that no significance was detected within the *Adh2* region. These interlocus LDs might be caused by epistatic interaction between the two *Adh* genes.

Discussion

The level and pattern of polymorphism in the *Adh2* region of *O. rufipogon*

The level of polymorphism (π) in the entire region of the *Adh2* of *O. rufipogon* was 0.008, which is higher than that of the *Adh1* of *O. rufipogon*. A high nucleotide diversity was detected at silent sites, while replacement diversity was low. When polymorphism for the sequence type 1 was analyzed, a low nucleotide diversity was detected even at silent sites. This result indicates that the high level of silent polymorphism was caused by the dimorphic pattern. On the other hand, the low level of variation at the replacement sites could be explained by purifying selection. The neutrality test of Fu and Li for the replacement sites indicated significantly negative deviation from the neutral mutation model. When the tests of Tajima, and Fu and Li were conducted for each domains of ADH2, significantly negative values were detected at replacement sites of the CBD. These results were indicative of purifying selection on the replacement sites, especially of the CBD.

Comparison of the level and pattern of polymorphism between the *Adh1* and *Adh2* regions of *O. rufipogon*

A large difference in the nucleotide diversity (π) between the *Adh1* and *2* regions was detected at silent sites of the entire region (*Adh1*: 0.003, *Adh2*: 0.011). Considering that there was no dimorphic pattern in the *Adh1* region (Yoshida et al. 2004), it is clear that the difference in the nucleotide diversity between the two *Adh* genes was related to the dimorphic pattern of the *Adh2* region.

Taking into account the low level of polymorphism and significantly negative value in the neutrality test of Fu and Li at replacement sites of the *Adh1* and *2* regions, it could be concluded that purifying selection operates on both enzymes. Since the ADH1 and ADH2 were induced in leaf and root under anaerobic conditions respectively (Xie and Wu 1989), each of the enzymes would contribute to tolerance to anaerobiosis in the respective tissues.

The low level of polymorphism in the *Adh1* region of *O. rufipogon* could be also explained by small effective population size (Yoshida et al. 2004). Under this hypothesis, the *Adh2* region is expected to have a low level of variation. Contrary to this expectation, the nucleotide diversity in the *Adh2* region was higher than that of the *Adh1* region. However, when level of polymorphism was estimated for the sequence type 1, a low nucleotide diversity was detected over all the functionally different regions of the *Adh2*. Therefore, from this study, it is not possible to conclude whether the low level of polymorphism in the *Adh1* region of *O. rufipogon* is due to small effective population size. To examine this

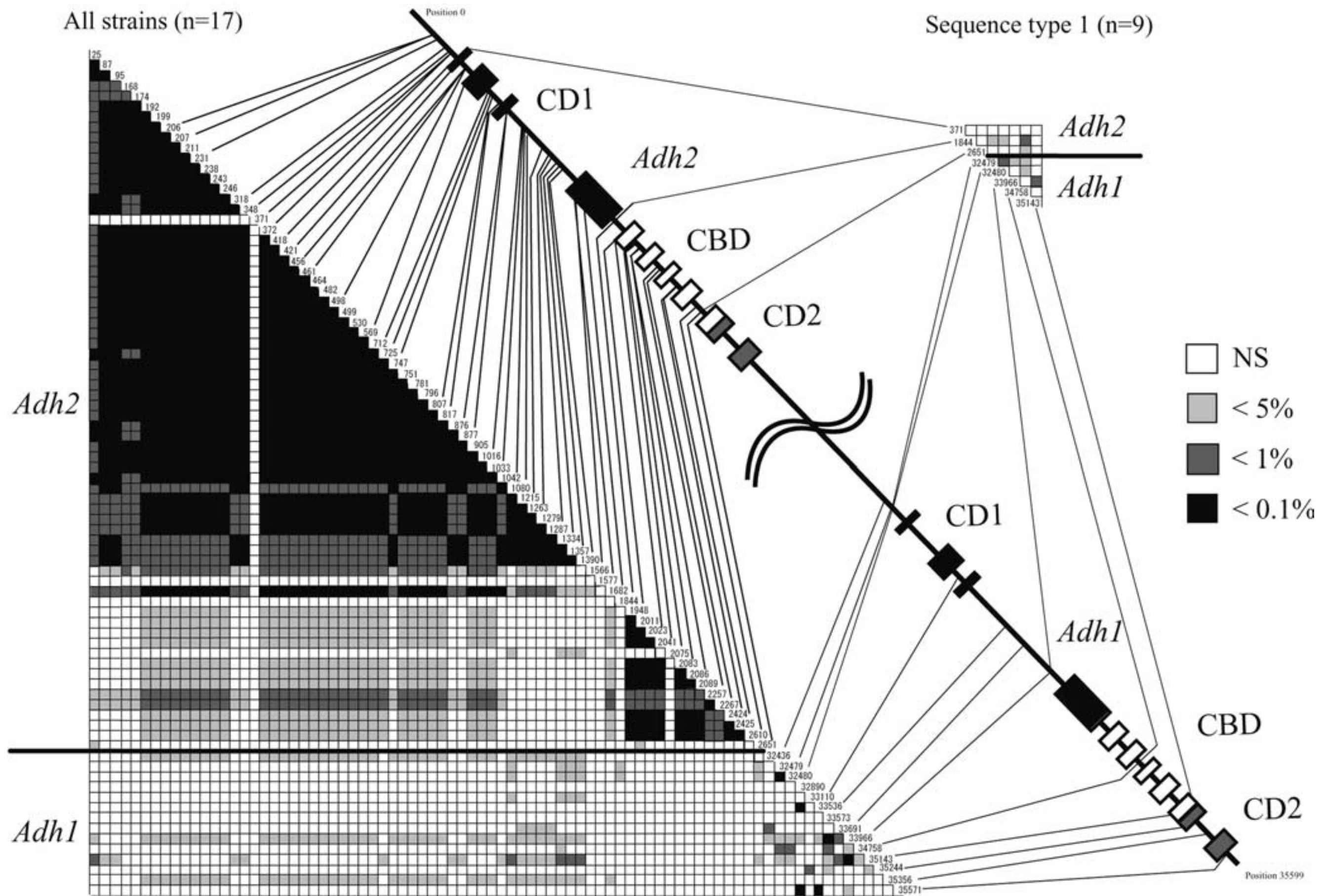


Fig. 5 Summary of linkage disequilibrium within and between *Adh1* and *Adh2* regions of *O. rufipogon*. Significance detected by Chi-square test is shown. The test resulted for the nine strains of *O. rufipogon* belonging to the sequence type 1 of the *Adh2* is shown at the *upper right*. *The number* indicates a position in the *Adh1* and *Adh2* regions. *Black line* connects to the position of the segregating site in the gene structure

possibility, nucleotide polymorphism for other loci of *O. rufipogon* needs to be analyzed.

Divergence between *O. rufipogon* and its related species

As shown for the *Adh1* region (Yoshida et al. 2004), interspecific comparisons of the *Adh2* region revealed that amino acid sequences were conserved among the A genome species, whereas the E genome species *O. australiensis* was highly diverged from *O. rufipogon*. The divergence between *O. rufipogon* and *O. australiensis* was high in the CD1 and CBD of the ADH2. The heuristic search showed that the replacement substitutions on the branch of *O. rufipogon* largely contributed to the high level of replacement divergence in the two domains between *O. rufipogon* and *O. australiensis*. Especially, replacement substitutions in the CD1 caused distinct physicochemical changes of amino acid. Similarly, replacement substitutions on the branch of *O. rufipogon* were detected in the CD1 of the ADH1. One of these substitutions caused physicochemical changes of amino acid. The habitat of *O. rufipogon* is more deeply submerged in water than that of *O. australiensis* (Morishima 2002), while *Z. mays* and *H. vulgare* are adapted to dry conditions (Sauter 2000). Considering the different habitats of the four grass species, the replacement substitutions with the physicochemical changes of amino acid in the ADH1 and ADH2 on the branch of *O. rufipogon* may be related to the adaptation to the anaerobic environments.

Linkage disequilibrium between the *Adh1* and *Adh2* regions of *O. rufipogon*

This study showed that there were significant LDs between the two *Adh* regions of *O. rufipogon*. Significant LDs were detected even when the strains of *O. rufipogon* belonging to the sequence type 1 were used in the analysis. The two *Adh* regions are located at the same position in the genetic map of *O. sativa*, and the physical distance between them is only about 29 kbp (Tarchini et al. 2000). These observations suggest that the physical linkage between these regions could be responsible for the observed LDs. However, it was noted that the percent of significant pairs in the intron between the two *Adh* genes was higher than that within each gene. It is known that intron sequences in actin of *O. sativa* and

Adh1 of *Z. mays* regulate gene expression (Koziel et al. 1996). The introns 1, 2 and 6 in the *Adh1* of *Z. mays* increase gene expression in transient assays (Callis et al. 1987; Mascarenhas et al. 1990), although its molecular mechanism has not been fully identified (Clancy and Hannah 2002). If the introns of the two *Adh* genes of *O. rufipogon* synergistically regulate the gene expression of both genes, epistasis would be an attractive candidate to explain the significant interlocus LDs in the introns. In future, it is necessary to experimentally demonstrate that intron sequences of the *Adh1* and *Adh2* are involved in synergistic regulation of the two *Adh* gene expressions.

The origin of *O. sativa* Japonica and Indica

The study of the nucleotide variation in the *Adh2* region of *O. rufipogon*, *O. sativa* Japonica and Indica showed that the sequence type of *O. sativa* Japonica was a recombinant between the sequence types 1 and 7 of *O. rufipogon*. The sequence of *O. sativa* Indica was similar to that of the sequence type 1. These results support the hypothesis that *O. sativa* Japonica and Indica would be originated from different *O. rufipogon* ancestors (Second 1982; Ishii et al. 1988). The estimated recombination time indicates that the ancestors of *O. sativa* Japonica and Indica would be separated at least 0.1 MYA. Considering that the perennial type of *O. rufipogon* has a higher outcrossing rate than the annual type of *O. rufipogon* (Morishima et al. 1984), the observed trace of recombination in the *Adh2* region of *O. sativa* Japonica also supports the idea that the ancestor of *O. sativa* Japonica is the perennial type of *O. rufipogon* (Morishima et al. 1984; Cheng et al. 2003).

Acknowledgements We are grateful to T. Ishii, Kobe University, for seeds and DNAs of *O. rufipogon*, and we thank K. Yamane for the technical advice for allozyme examination. This article is contribution number 586 from the Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University.

References

Callis J, Fromm M, Walbot V (1987) Introns increase gene expression in culture maize cells. Genes Dev 1:1183–1200

Chang, TT (1976) The origin, evolution, cultivation, dissemination, and diversification of Asian and Africa rices. Euphytica 25:425–441

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

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

Clancy M, Hannah LC (2002) Splicing of the maize Sh1 first intron is essential for enhancement of gene expression, and a T-rich motif increases expression without affecting splicing. Plant Physiol 130:918–929

Cummings MP, Clegg MT (1998) Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. Proc Natl Acad Sci USA 95:5637–5642

Eklund H, Nordstrom B, Zeppezauer E, Soderlund G, Ohlsson I et al (1976) Three-dimensional structure of horse liver alcohol dehydrogenase at 2–4 A resolution. J Mol Biol 102:27–59

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

Gaut BS, Clegg MT (1993b) Molecular evolution of the *Adh1* locus in the genus *Zea*. Proc Natl Acad Sci USA 90:5095–5099

Gaut BS, Clegg MT (1993a) Nucleotide polymorphism in the Adh1 locus of pearl millet (*Pennisetum glaucum*) (*Poaceae*). Genetics 135:1091–1097

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbc*L. Proc Natl Acad Sci USA 93:10274–10279

Gaut BS, Peek AS, Morton BR, Clegg MT (1999) Patterns of genetic diversification within the Adh gene family in the grasses (Poaceae). Mol Biol Evol 16:1086–1097

Glaszmann JC, Delosreys BG, Kusch GS (1988) Electrophoretic variation of isozymes in plumules of rice (Oryza sativa L.): a key to the identification to 76 alleles at 24 loci. IRRI Res Paper Ser 134:1–14

Innan H, Tajima F, Terauchi R, Miyashita NT (1996) Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana*. Genetics 143:1761–1770

Ishii T, Terachi T, Tsunewaki K (1988) Restriction endonuclease analysis of chloroplast DNA from A-genome diploid species of rice. Jpn J Genet 63:523–536

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (eds) Mammalian protein metabolism. Academic, New York, pp 21–132

Kimura M (1956) A model of a genetic system which leads to closer linkage under natural selection. Evolution 10:278–287

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Koziel MG, Carozzi NB, Desai N (1996) Optimizing expression of transgenes with an emphasis on post-transcriptional events. Plant Mol Biol 32:393–405

Kumar S, Tamura K, Jakobsen I, Nei M (2002) *MEGA2*: Molecular evolutionary genetics analysis, version 2.0. Pennsylvania and Arizona State Universities, University Park, Pennsylvania and Tempe. Arizona

Lin JZ, Brown AH, Clegg MT (2001) Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). Proc Natl Acad Sci USA 98:531–536

Lin JZ, Morrell PL, Clegg MT (2002) The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol dehydrogenase loci in wild barley (*Hordeum vulgare* ssp. *spontaneum*). Genetics 162:2007–2015

Mascarenhas D, Mettler IJ, Pierce DA, Lowe HW (1990) Intron-mediated enhancement of heterologous gene expression in maize. Plant Mol Biol 15:913–920

Matsumura H, Takano T, Takeda G, Uchimiya H (1998) *Adh1* is transcriptionally active but its translational product is reduced in a *rad* mutant of rice (*Oryza sativa* L.), which is vulnerable to submergence stress. Theor Appl Genet 97:1197–1203

Menegus F, Cattaruzza L, Chersi A, Fronza G (1989) Differences in the anaerobic lactate-succinatc production and in the changes of cell sap pH for plants with high and low resistance to anoxia. Plant Physiol 90:29–32

Menegus F, Cattaruzza L, Mattana M, Beffagna N, Ragg E (1991) Response to anoxia in rice and wheat seedlings. Changes in the pH of intracellular compartments, glucose-6-phosphate level, and metabolic rate. Plant Physiol 95:760–767

Miyashita NT, Innan H, Terauchi R (1996) Intra- and interspecific DNA variation of the alcohol dehydrogenase locus region of wild plants *Arabis gemmifera* and *Arabidopsis thaliana* species. Mol Biol Evol 13:433–436

Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitutions in protein evolution. J Mol Evol 12:219–236

Morishima H (2002) Reports of the study-tours for investigation of wild and cultivated rice species. Conducted by the staff of National Institute of Genetics, Japan and coworkers during 1957–1997 Part I

Morishima H, Oka HI, Chang WT (1961) Directions of differentiation in populations of wild rice, *Oryza perennis* and *O. sativa* f. *spontanea*. Evolution 15:326–339

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

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. Genetics 97:145–163

Oka HI, Chang WT (1959) The impact of cultivation on populations of wild rice, *Oryza* f. *spontanea*. Phyton 13:105–117

Oka HI, Morishima H (1967) Variations in the breeding systems of a wild rice, *Oryza perennis*. Evolution 21:249–258

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16:276–277

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Sauter M (2000) Rice in deep water: ''How to take heed against a sea of troubles''. Naturwissenschaften 87:289–303

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

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

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign

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

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *Adh1adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381–391

Wasano K (1995) Genetical and archaeological investigation on the origin of cultivated rice and ancient rice culture in east Asia. Saga University, Saga Japan (in Japanese)

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

Weising K, Kaemmer D, Weigand F, Epplen JT, Kahl G (1991) Oligonucleotide fingerprinting reveals various prove-dependent levels of informativeness in chickpea (*Cicer arietinum*). Genome 35:436–442

Wright S (1938) Size of population and breeding structure in relation to evolution. Science 87:430–431

Xie Y, Wu R (1989) Rice alcohol dehydrogenase genes: anaerobic induction, organ-specific expression and characterization of cDNA clones. Plant Mol Biol 13:53–68

Xie ZW, Lu YQ, Ge S, Hong DY, Li FZ (2001) Clonality in wild rice (*Oryza rufipogon*, Poaceae) and its implications for conservation management. Am J Bot 88:1058–1064

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