Theor Appl Genet (2004) 109: 1406–1416
DOI 10.1007/s00122-004-1752-3

ORIGINAL PAPER

K. Yoshida · N. T. Miyashita · T. Ishii

Nucleotide polymorphism in the *Adh1* locus region of the wild rice *Oryza rufipogon*

Received: 24 November 2003 / Accepted: 2 June 2004 / Published online: 6 October 2004

Abstract Nucleotide variation in the alcohol dehydrogenase (*Adh1*) locus region of the wild rice *Oryza rufipogon* and its related species was analysed to clarify the maintenance mechanism of DNA variation in *Oryza* species. The estimated nucleotide diversity in the *Adh1* locus region of *O. rufipogon* was 0.002, which was one of the lowest values detected in nuclear loci of plant species investigated so far. Tests of neutrality detected significantly negative deviation from the neutral mutation model for the coding region, especially for replacement sites. When each of the ADH1 domains was considered, significance was detected only for the catalytic domain 1. These results suggest purifying selection in the *Adh1* coding region. In the phylogenetic tree of *Oryza* species based on *Adh1* variation, cultivated rice *O. sativa* subspp. *japonica* and *indica* were included in the cluster of *O. rufipogon*. The genetic distance of the *Adh1* region between *O. rufipogon* and *O. sativa* was as low as the nucleotide diversity of *O. rufipogon*. These results imply that *O. rufipogon* and *O. sativa* cannot be classified based on the nucleotide variation of *Adh1*. No replacement divergence between *O. rufipogon* and the other three A-genome species (*O. glumaepatula*, *O. barthii* and *O. meridionalis*) were detected, indicating that ADH1 is conserved in the A-genome species. On the other hand, between *O. rufipogon* and the E-genome species *O. australiensis*, replacement changes were detected only in the catalytic domain 1. The difference in replacement substitutions between the A- and E-genome species may be related to adaptive changes in the ADH1 domains, reflecting environmental differences where the species encounter anaerobic stress.

Communicated by O. Savolainen

K. Yoshida · N. T. Miyashita (✉)
Laboratory of Plant Genetics,
Graduate School of Agriculture,
Kyoto University, Sakyo-ku,
Kyoto, 606-8502, Japan
E-mail: arabis@kais.kyoto-u.ac.jp
Tel.: +81-75-7536138
Fax: +81-75-7536486

T. Ishii
Laboratory of Plant Breeding, Faculty of Agriculture,
Kobe University, Nada-ku, Kobe, 657-8501, Japan

Introduction

The A-genome species of rice are composed of two cultivated and five wild species. The cultivated species are *Oryza sativa* and *O. glaberrima*, and wild species include *O. rufipogon*, *O. barthii*, *O. longistaminata*, *O. glumaepatula* and *O. meridionalis* (Vaughan 1989). Among the wild species,*O. rufipogon* ($2n = 24$) is considered to be the ancestor of the cultivated rice *O. sativa* (Oka and Chang 1959). *O. rufipogon* is distributed over tropics and subtropics in monsoon Asia from South China to West India and Indonesia (Morishima et al. 1984). It generally grows in swampy lowlands, such as ponds, canals and roadside ditches, and is classified into annual and perennial types according to its habitat. The annual type can be found in the dry and open spaces during the dry season and is known to produce many seeds after the season. On the other hand, the perennial type inhabits a watery environment all year long.

Many studies have been conducted to clarify the evolutionary history of *O. sativa*, for example, for questions like whether the ancestor of *O. sativa* is an annual or a perennial type of *O. rufipogon*, and from where *O. sativa* originates. A comparison of morphological traits showed that the intermediate life form between annual and perennial types could be the ancestor of *O. sativa* (Morishima et al. 1961). Studies of allozyme and chloroplast DNA variations suggested that *O. sativa* subspp. *japonica* and *indica* could have originated independently from *O. rufipogon* (Second 1982; Ishii et al. 1988). Similarly, a study of the SINE insertion polymorphism indicated that *O. sativa* subsp. *japonica* was closely related to the *O. rufipogon* perennial strains,

while *O. sativa* subsp. *indica* was more similar to the *O. rufipogon* annual strains (Cheng et al. 2003). Recently RAPD, RFLP and simple sequence length polymorphism analysis of chloroplast DNA showed that *O. sativa* subsp. *japonica* was the most genetically similar to Chinese strains of *O. rufipogon* among investigated *O. rufipogon*, including South Asian strains (Bautista et al. 2001). This result suggests that *O. sativa* subsp. *japonica* originated in China. Despite extensive studies on the origin of *O. sativa*, decisive conclusions for the above questions have not been obtained.

The wild species *O. rufipogon* provides important genetic resources for rice breeding because it has many agriculturally advantageous characters, such as high grain yield and resistance to rice blast disease (Xiao et al. 1998; Utami et al. 2001). It is important to estimate the level of genetic diversity in natural populations of *O. rufipogon*, since assessment of genetic diversity could be necessary to construct experimental and management plans for searching beneficial genes for rice breeding and conserving genetic resources in natural populations of *O. rufipogon*. Allozyme variation has been investigated to estimate the level of genetic diversity in natural populations of *O. rufipogon* (Second 1985; Barbier 1989; Morishima and Barbier 1990; Akimoto et al. 1999; Gao and Hong 2000; Gao et al. 2001). Average heterozygosities for investigated 40 allozyme loci of the South Asian and Chinese *O. rufipogon* were 0.25 and 0.19, respectively (Second 1985). These values were higher than the average heterozygosity (0.15) of allozyme variation (average number of loci = 16.5) for 473 plant species (Hamrick and Godt 1990), indicating that *O. rufipogon*, with a relatively large allozyme variation, is expected to be potent for rice breeding. However, nucleotide sequence variation in specific genes of *O. rufipogon* has not been analysed so intensively, even though it is important to reveal the molecular nature of allozyme variation. So far, nucleotide sequence polymorphism of *O. rufipogon* has been investigated only for prolamin and phytochrome genes (Barbier and Ishihama 1990; Barbier et al. 1991). The estimated frequency of segregating sites was 0.009 for the prolamin and frequencies for exons and introns of the phytochrome gene were 0.002 and 0.012, respectively. Because these values are different from each other of the genes, it is not possible to conclude if the level of nucleotide polymorphism is relevant to that of allozyme variation in *O. rufipogon*.

In this report, nucleotide sequence variation in the alcohol dehydrogenase (*Adh1*) locus region of *O. rufipogon* was analysed. The alcohol dehydrogenase gene is one of the most investigated nuclear regions in plant species from the viewpoint of molecular population genetics (Gaut and Clegg 1993; Innan et al. 1996; Miyashita et al 1996; Charlesworth et al. 1998; Cumming and Clegg 1998; Lin et al. 2001). In *O. sativa*, there are two alcohol dehydrogenase loci (*Adh1* and *Adh2*), both of which are located in the short arm of chromosome 11 (Tarchini et al. 2000). The ADH enzymes encoded by these two genes, both of which are composed of 379 amino acids, are induced under anaerobic conditions and catalyse the NAD^+-dependent oxidation of alcohol produced in the process of anaerobic metabolism (Xie and Wu 1989; Nakazono et al. 2000). The level of expression of *Adh1* is high in leaf tissue, while *Adh2* is expressed mainly in root tissue (Xie and Wu 1989). Experimental evidence that ADH is adaptively important has been obtained. A mutant of *O. sativa* that lost function of ADH encoded by *Adh1* was more vulnerable to anaerobic stress than the wild type, since the mutant could not recover the activity of ADH after long-term anaerobic stress (Matsumura et al. 1998). In addition, allozyme study has shown that the alcohol dehydrogenase gene itself had a lower level of variation ($H_e = 0$ in *O. sativa*, $H_e = 0.02$ in South Asian *O. rufipogon*) than the average heterozygosity in *O. sativa* and South Asian *O. rufipogon* (Second 1982, 1985). The low heterozygosity of ADH might be related to the adaptive importance in anaerobic environment and stress in the tropics and subtropics in monsoon Asia, where *O. rufipogon* is found. It is of interest from the viewpoint of population genetics to investigate how the adaptive importance is reflected on nucleotide sequence variation in the *Adh1* region of *O. rufipogon* and its related species.

Materials and methods

Plant materials

Seventeen plants of *O. rufipogon* (one individual each of its related species) and plants from two cultivars of *O. sativa* were used (Table 1). Their seeds and DNA samples are maintained in the Laboratory of Plant Breeding, Faculty of Agriculture, Kobe University. They were originally provided by the National Institute of Genetics (Japan), Shizuoka University (Japan) and the International Rice Research Institute (The Philippines). Seeds were sterilized by using benomyl (Dupont, Wilmington, Del.) at 28°C overnight and germinated in the dark at 28°C. Plants of each accession were grown in a pot under 28°C and 14-h light conditions.

DNA extraction, PCR amplification and sequencing

Total DNA was extracted from leaves and stems with modification of the CTAB method (Weising et al. 1991) and used for PCR amplification of the *Adh1* region. The primers for the PCR amplification for *O. rufipogon*, *O. sativa* Indica and *O. barthii* were 5′-CCATCCTCCATCCTCTCCTT-3′ and 5′-CCAGGATACACAGAAGAACC-3′, which are located in 5′- and 3′-flanking regions, respectively. By using these primers, a 3.2-kb segment of the *Adh1* region was amplified. The primers for PCR amplification for *O. glumaepatula*, *O. meridionalis* and *O. australiensis* were 5′-ATGGCGACCGCAGGGAAGGT-3′ and 5′-AGTTCTCCATGCGGAT-

Table 1 Plant materials used in this study

Species	Genome	Strain	Origin	Source[a]
Oryza rufipogon	A	W108	India	1
	A	W120	India	1
	A	W593	Malaysia	1
	A	W1956	China	1
	A	W1965	China	1
	A	W1972	Indonesia	1
	A	W1976	Indonesia	1
	A	CB22	Cambodia	2
	A	CB27	Cambodia	2
	A	CT56	Vietnam	2
	A	LV61	Laos	2
	A	VT51	Laos	2
	A	SN	Thailand	2
	A	KA	Thailand	2
	A	W630	Myanmar	1
	A	YG2A	Myanmar	2
	A	PT1A	Myanmar	2
O. sativa subsp. *japonica*	A	YT1A	Japan	3
O. sativa subsp. *indica*	A	435	Sri Lanka	1
O. glumaepatula	A	W1167	Cuba	1
O. barthii	A	W607	Guinea	1
O. meridionalis	A	W1627	Australia	1
O. australiensis	E	101397	Australia	4

[a] *1* National Institute of Genetics, Japan; *2* Shizuoka University, Japan; *3* GenBank accession AAF34414; *4* International Rice Research Institute, Philippines

GATG-3′, which are located in the first and tenth exons, respectively. By using these primers, a 3.1-kb segment of the *Adh1* region, except for 5′- and 3′-flanking regions, was amplified. These four primers were designed based on a published sequence of the *Adh1* region of *O. sativa* subsp. *japonica* YT1A (GenBank accession AAF34414). The PCR products were cloned into the plasmid pUC118*Hin*cII/BAP (TaKaRa), which was used as a template for the sequencing reaction. Sequence reaction was conducted by using the Thermo Sequenase fluorescent-labelled cycle sequencing kit with 7-deaza-dGTP (Amersham/Pharmacia Biotech, Piscataway, N.J., USA). We mixed the three plasmid clones at almost the same molarity to eliminate PCR artefacts. Sequencing primers were designed at about 500-bp intervals. Newly determined sequences were deposited in the DDBJ databank under accession numbers AB118850–AB118852 and AB118948–AB118966.

Data analysis

DnaSP program, version 3.50 (Rozas and Rozas 1999), was used to analyse intra- and inter-specific variations. Nucleotide diversity (π) and the proportion of segregating sites θ ($=4N_e\mu$, Watterson 1975) were estimated after removing indels. Tests of Tajima (Tajima 1989), Fu and Li (Fu and Li 1993) and McDonald and Kreitman (McDonald and Kreitman 1991) were conducted to investigate the departure from neutrality. To examine the difference in levels of DNA variation between polymorphism and divergence, an HKA test (Hudson et al 1987) was performed between different functional domains by using *O. australiensis* as an outgroup species. A neighbour-joining (NJ) tree was constructed by using the MEGA program, version 2.1 (Kumar et al. 2001).

Results

Pattern and level of DNA polymorphism in the *Adh1* region of *O. rufipogon*

In the 3.2-kb *Adh1* region of *O. rufipogon*, there were 38 variations (33 segregating sites and five indels), of which 24 variations (22 segregating sites and two indels) were singletons (Fig. 1). The coding region had 12 polymorphic sites (eight synonymous and four replacement sites) and no indels. All four replacement variations were singletons. Amino acid changes caused by these replacement variations were conservative with respect to physicochemical difference, according to Miyata et al. (1979).

π for the entire region was 0.0020 (Table 2), which is one of the lowest values estimated for nuclear genes in plants. Although π for the 5′-flanking region (0.0041) and synonymous sites (0.0058) was relatively high, the other functionally different regions had similar values around 0.0020, except for replacement sites (0.0005). Because of the high proportion of singletons, the estimated π was smaller than θ for all the regions considered. To examine change in nucleotide variation along the entire region of *Adh1*, sliding-window analysis in terms of silent π was conducted (Fig. 2). No outstanding change in silent π was detected along the entire region of *Adh1*. The small number of segregating sites, most of which were singletons, caused this result.

Difference in the level of nucleotide polymorphism among the *Adh1* domains

To identify the domain structure of ADH1 of *O. rufipogon*, we compared amino acid sequences of ADH1 of *Zea mays* (Dennis et al. 1984) and *O. rufipogon* (Fig. 3). The ADH1 of *Z. mays* has a catalytic domain 1, a co-enzyme-binding domain and a catalytic domain 2 (Dennis et al. 1984). Since the amino acid sequence of ADH1 of *O. rufipogon* was 97% identical with that of *Z. mays*, and they have the same number of amino acids, the three domains of ADH1 in *O. rufipogon* were clearly defined (Fig. 3). In addition, there are two groups of amino acids, three amino acid residues (47th, cysteine; 69th, histidine and 177th, cysteine) and four cysteine residues (99th, 102nd, 105th and 113th), which are essential for catalytic activity of ADH1 by binding to two different zinc ions. These amino acids were also conserved between the two species. Amino acids located at the boundary between domains were also conserved between *Z. mays* and *O. rufipogon*. The level of nucleotide variation was estimated for each of the three domains of *Adh1* of *O. rufipogon* (Table 3). It is clear that the level of nucleotide variation differs among the three

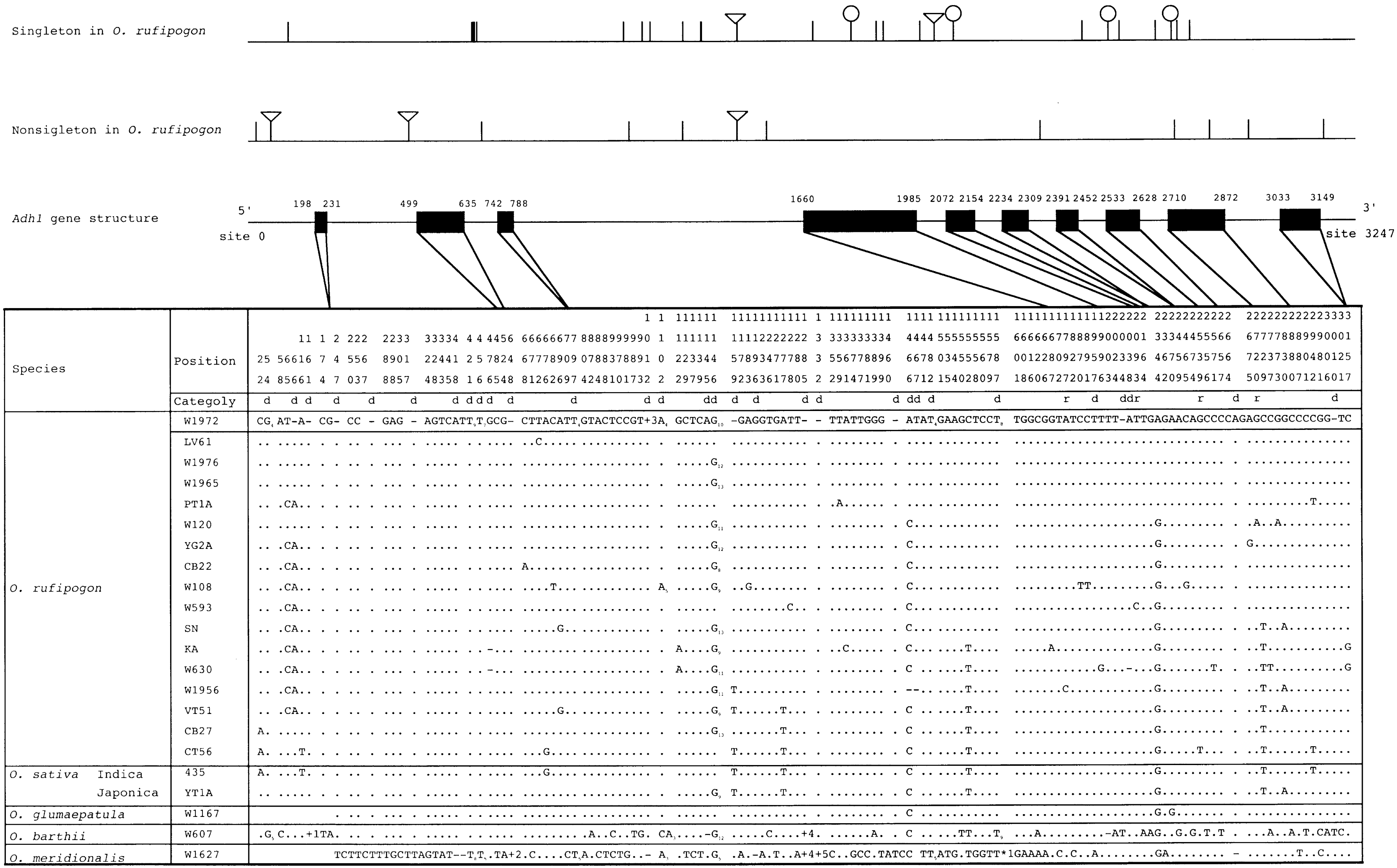

Fig. 1 Summary of DNA variations in the 3.2-kb alcohol dehydrogenase (*Adh1*) region of *Oryza rufipogon*, *O. sativa*, *O. glumaepatula*, *O. barthii* and *O. meridionalis*. The structure of the *Adh1* region from the sites 0–3,247 is shown at the *centre* of the figure. The *black boxes* indicate exons. Singletons and non-singletons 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 the figure*, where *dots* indicate similarities to the variation of the W1972 sequence. *D* and *r* are abbreviations for indels and replacement changes, respectively. A_n, T_n and G_n indicate that adenine, thymine and guanine repeat *n* times, respectively. +*1* TTGGT, +*2* CATCACACC, +*3* GTGTGTTG, +*4* CTAGTCAACAGTACAATTGCATG, +*5* CTTGGTAC, **1* TTGTT

Table 2 Summary of DNA variation of the *Adh1* region of *O. rufipogon* and its related species. *S* Substitutions, π nucleotide diversity, θ proportion of segregating sites, *D* Tajima's *D* statistic, *D** Fu and Li's *D** statistic, *K* nucleotide divergence between two species shown in subscript: $_{R}O.$ *rufipogon*, $_{SJ}O.$ *sativa* subsp. *japonica*, $_{SI}O.$ *sativa* subsp. *indica*, $_{G}O.$ *glumaepatula*, $_{B}O.$ *barthii*, $_{M}O.$ *meridionalis*, $_{A}O.$ *australiensis*

	Number of sites	S	π	θ	*D*	*D**	$K_{R\text{-}SJ}$[a]	$K_{R\text{-}SI}$	$K_{SJ\text{-}SI}$	$K_{R\text{-}G}$	$K_{R\text{-}B}$	$K_{R\text{-}M}$	$K_{R\text{-}A}$
Entire region	3,185	33	0.0020	0.0031	−1.467	−1.843	0.0017 (5.29)	0.0025 (7.94)	0.0016 (5.00)	0.0014 (4.06)	0. 0098 (30.94)	0.0189 (53.35)	0.0707 (181.06)
Coding region	1,137	12	0.0018	0.0031	−1.638	−2.411*	0.0018 (2.00)	0.0012 (1.35)	0.0009 (1.00)	0. 0010 (1.18)	0.0019 (2.18)	0.0046 (5.18)	0.0352 (39.06)
Synonymous sites	265.4	8	0.0058	0.0089	−1.251	−1.843	0.0067 (1.76)	0.0042 (1.12)	0.0038 (1.00)	0.0036 (0.94)	0.0074 (1.94)	0. 0189 (4.94)	0.1397 (33.82)
Replacement sites	871.6	4	0.0005	0.0014	−1.843*	−2.469*	0.0003 (0.24)	0.0003 (0.24)	0.0000 (0.00)	0.0003 (0.24)	0.0003 (0.24)	0. 0003 (0.24)	0.0060 (5.24)
Non-coding region	2,049	21	0.0021	0.0031	−1.245	−1.290	0.0016 (3.29)	0.0032 (6.59)	0.0020 (4.00)	0.0017 (2.88)	0. 0143 (28.77)	0.0285 (48.18)	0.0980 (142.00)
5′-flanking region	192	3	0.0041	0.0047	−0.345	−0.063	0.0025 (0.47)	0.0112 (2.12)	0.0105 (2.00)	NA[b]	0.0239 (4.47)	NA	NA
Intron	1,756	17	0.0019	0.0029	−1.368	−1.627	0.0016 (2.71)	0.0025 (4.35)	0.0011 (2.00)	0.0017 (2.88)	0. 0140 (24.18)	0.0285 (48.18)	0.0980 (142.00)
3-′flanking region	101	1	0.0022	0.0029	−0.491	0.677	0.0012 (0.12)	0.0012 (0.12)	0.0000 (0.00)	NA	0.0012 (0.12)	NA	NA

Significance level: $*P < 0.05$

[a] All these genetic distances were calculated by Jukes and Cantor method. Average numbers of nucleotide differences between species were shown in *parentheses*

[b] *NA* Not applied

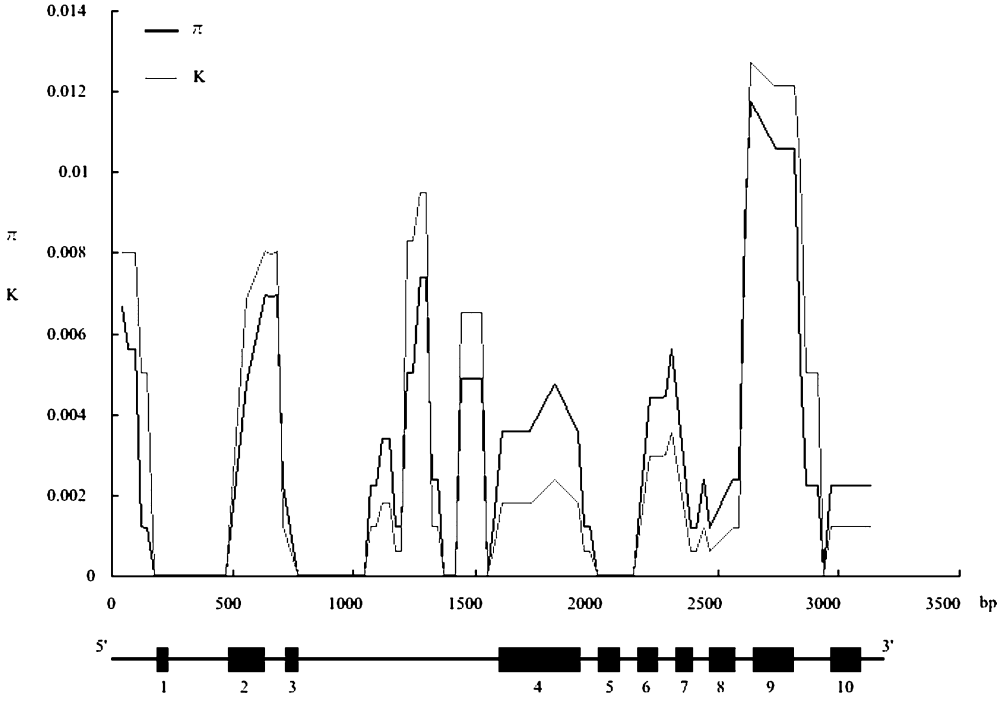

Fig. 2 Sliding-window plot for the 3.2-kb *Adh1* region. The *black line* represents polymorphism in *O. rufipogon*. The *grey line* represents divergence between *O. rufipogon* and *O. sativa*. The window size is 100 bp, step size is 25 bp, and the unit is silent site. The gene structure is shown at the *bottom of the figure*. The *black boxes* indicate exons

domains. Catalytic domain 1 has a lower level of nucleotide variation than the other domains.

Test of neutrality for nucleotide polymorphism in the *Adh1* region of *O. rufipogon*

To examine the neutrality of nucleotide polymorphism, tests of Tajima (1989) and Fu and Li (1993) were conducted for functionally different regions (Table 2). Because of the large number of singletons, test statistics were negative for most of the comparisons. Significant deviation from test assumptions of neutrality and population equilibrium was detected for the coding region by the Fu and Li test, and for replacement sites by both tests (Table 2). These results are indicative of purifying selection in the coding region, especially for replacement sites, consistent with the low ratio of synonymous and replacement polymorphism.

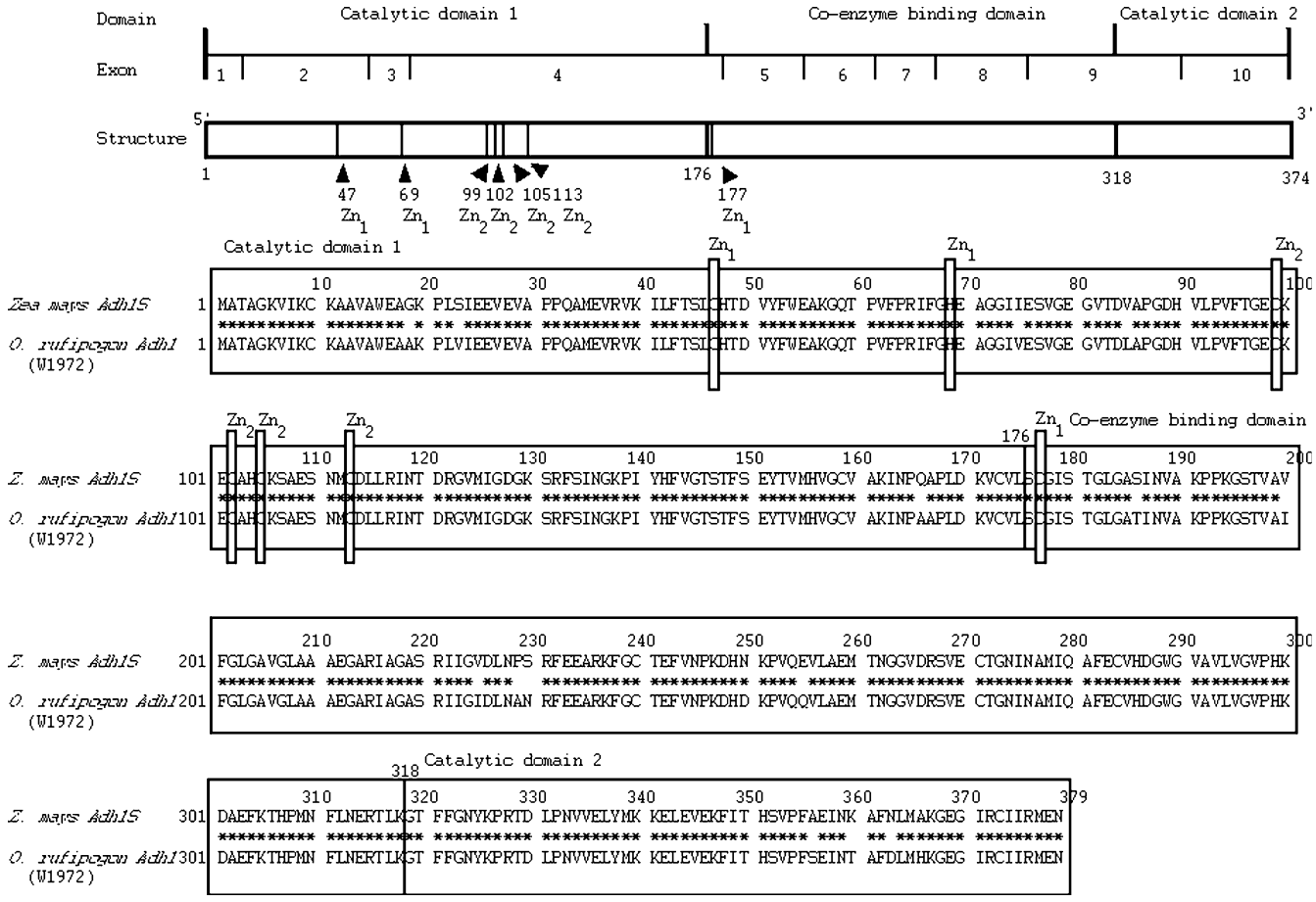

Fig. 3 Amino acid-sequence alignment of *Adh1* between *O. rufipogon* and *Z. mays*. The gene structure of coding regions of *Adh1* is shown at the *top of the figure*. Zn_1 indicates the location of amino acid residues—46th, cysteine; 69th, histidine and 177th, cysteine—for binding one zinc ion. Zn_2 indicates the location of four cysteines for binding the other zinc ion. *Asterisks* indicate that an amino acid of *O. rufipogon* is identical to that of *Z. mays*

When these tests were applied to each of the three domains (Table 3), the tests of Tajima (1989) and Fu and Li (1993) gave significantly negative values for catalytic domain 1 (Table 3). Although significance was not detected for the co-enzyme-binding domain, test statistics were fairly large and nearly significant. Since the number of segregating sites was one in catalytic domain 2, the tests are not meaningful. When segregating sites were classified into synonymous and replacement sites, significance was not detected in any domain. This is probably due to the lack of power of the tests.

Divergence between *O. rufipogon* and the cultivated rice *O. sativa*

Between the *Adh1* regions of *O. rufipogon* and *O. sativa*, there was no species-specific difference (different between species, but monomorphic in species, Fig.1). Six different nucleotide sites detected between *O. sativa* subspp. *japonica* and *indica* were also segregating in *O. rufipogon*. This observation implies that these different sites between the two types of cultivated rice had been already polymorphic in *O. rufipogon* before establishment of cultivated *O. sativa* and agrees with the hypothesis that *O. sativa* subspp. *indica* and *japonica* originated independently from different *O. rufipogon* lines (Bautista et al. 2001).

We estimated levels of divergence (K) for functionally different regions of *Adh1* among *O. rufipogon*, *O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica* (Table 2). The levels of divergence between pairs of these species were similar to those of polymorphism (π) in *O. rufipogon*. It is noted that the distance between *O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica* was generally smaller than that of *O. rufipogon* and *O. sativa*. This result agrees with the hypothesis that *O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica* originated from *O. rufipogon*. From the results shown here, it is not possible to say whether *O. sativa* subsp. *japonica* or subsp. *indica* originated earlier, since K values vary over different regions compared. When the estimation was applied to each of the domains (Table 3), no difference in distance was detected between *O. rufipogon* and *O. sativa* for catalytic domain 1 and the co-enzyme-binding domain. Between *O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica*, nucleotide sequences of the catalytic domain 1 and co-enzyme-binding domain were identical. Although there were some differences for catalytic domain

Table 3 Summary of DNA variation in each domain of the *Adh1* region of *O. rufipogon* and its related species

	Number of sites	S	π	θ	D	D^*	$K_{R\text{-}SJ}$	$K_{R\text{-}SI}$	$K_{SJ\text{-}SI}$	$K_{R\text{-}G}$	$K_{R\text{-}B}$	$K_{R\text{-}M}$	$K_{R\text{-}A}$
Catalytic domain 1													
All sites	525.0	4	0.0009	0.0023	−1.840*	−2.469*	0.0005 (0.24)	0.0005 (0.24)	0.0000 (0.00)	0. 0005 (0.24)	0.0005 (0.24)	0.0081 (4.24)	0.0456 (23.24)
Synonymous sites	124.0	3	0.0029	0.0071	−1.705	−2.255	0.0014 (0.18)	0.0014 (0.18)	0. 0000 (0.00)	0.0014 (0.18)	0.0014 (0.18)	0.0345 (4.18)	0.1627 (18.18)
Replacement sites	400.6	1	0.0003	0.0007	−1.163	−1.477	0.0002 (0.06)	0.0002 (0.06)	0. 0000 (0.00)	0.0002 (0.06)	0.0002 (0.06)	0.0002 (0.06)	0.0127 (5.06)
Co-enzyme-binding domain													
All sites	429.0	7	0.0029	0.0048	−1.420	−2.267	0.0022 (0.94)	0.0022 (0.94)	0.0000 (0.00)	0. 0018 (0.76)	0.0041 (1.76)	0.0018 (0.76)	0.0204 (8.65)
Synonymous sites	101.6	4	0.0086	0.0117	−0.816	−1.569	0.0075 (0.76)	0.0075 (0.76)	0. 0000 (0.00)	0.0058 (0.59)	0.0152 (1.59)	0.0058 (0.59)	0.0884 (8.47)
Replacement sites	327.4	3	0.0011	0.0027	−1.706	−2.255	0.0005 (0.18)	0.0005 (0.18)	0. 0000 (0.00)	0.0005 (0.18)	0.0005 (0.18)	0.0005 (0.18)	0.0005 (0.18)
Catalytic domain 2													
All sites	183.0	1	0.0017	0.0016	0.090	0.677	0.0045 (0.82)	0.0010 (0.18)	0.0055 (1.00)	0.0010 (0.18)	0. 0010 (0.18)	0.0010 (0.18)	0.0403 (7.18)
Synonymous sites	39.5	1	0.0078	0.0075	0.085	0.677	0.0211 (0.82)	0.0045 (0.18)	0.0258 (1.00)	0.0045 (0.18)	0. 0045 (0.18)	0.0045 (0.18)	0.2083 (7.18)
Replacement sites	143.4	0	0.0000	0.0000	NA	NA	0.0000 (0.00)	0.0000 (0.00)	0.0000 (0.00)	0. 0000 (0.00)	0.0000 (0.00)	0.0000 (0.00)	0.0000 (0.00)

2, the number of sites compared for the domain was one for the comparisons among *O. rufipogon*, *O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica*, so the difference is not significant.

To investigate changes in the level of divergence between *O. rufipogon* and *O. sativa* along the entire region of *Adh1*, a sliding-window analysis was conducted (Fig. 2). Positions of peaks were identical for π in *O. rufipogon* and divergence between *O. rufipogon* and *O. sativa*. This is an expected result because there was no species-specific variation, and polymorphic sites in *O. rufipogon* were shared with *O. sativa*. Taken together, these results indicate that *O. sativa* and *O. rufipogon* could not be classified based on the nucleotide sequence variations in the *Adh1* region.

Divergence between *O. rufipogon* and its related species

Between *O. rufipogon* and the other three A-genome species, there were 1, 33 and 68 species-specific differences for *O. glumaepatula, O. barthii* and *O. meridionalis*, respectively (Fig. 1). All of these differences were either synonymous changes in coding regions or DNA variations (nucleotide substitutions and indels) in non-coding regions. None of these species-specific differences caused changes in the amino acid sequence of ADH1, indicating that ADH1 is highly conserved among these A-genome species.

On the other hand, between *O. rufipogon* and the E-genome species *O. australiensis*, there were 39 species-specific differences (six replacement and 33 synonymous differences) in the coding regions of *Adh1*. In addition, there were 140 site and 75 indel differences in non-coding regions (introns). All the fixed replacement changes between the species were detected only in catalytic domain 1, while no replacement changes were detected in the other domains. Notably, one of the six replacement changes in catalytic domain 1 caused physico-chemical amino acid changes from glutamic acid (hydrophilic) in *O. rufipogon* to alanine (neutral) in *O. australiensis*, according to Miyata et al. (1979). This result contrasts with the low level of polymorphism and significantly negative results of neutrality tests for catalytic domain 1 of *O. rufipogon*.

The estimated divergence between *O. rufipogon* and related species for functionally different regions of *Adh1* (Table 2) indicated that the E-genome species *O. australiensis* was highly divergent from *O. rufipogon*. This result is especially clear for replacement sites, while the level of divergence for replacement sites between the A-genome species is fairly constant. The level of divergence was also estimated for each of the three domains of *Adh1* (Table 3). Consistent with the above observation, the two species are highly divergent for replacement sites in catalytic domain 1. On the other hand, divergence at replacement sites was not detected for the co-enzyme-binding domain. These results indicate that the E-genome species *O. australiensis* is divergent from *O. rufipogon* with respect to not only silent (including synonymous) sites but also replacement sites, especially in catalytic domain 1.

Tests of contingency between polymorphism and divergence

If the neutral mutation theory holds, it is expected that levels of polymorphism in a species and divergence between related species be correlated. To examine the ratio

Table 4 McDonald and Kreitman tests in the *Adh1* region

	Synonymous	Silent[a]	Replacement	χ^2_{syn} [b]	χ^2_{silent} [c]
Polymorphism in *O. rufipogon*	8	25[d]	4		
Fixed between *O. rufipogon* and *O. glumaepatula*	0	1	0	NA	0.159
Fixed between *O. rufipogon* and *O. barthii*	1	23	0	0.481	3.437
Fixed between *O. rufipogon* and *O. meridionalis*	4	50	0	1.778	7.264**
Fixed between *O. rufipogon* and *O. australiensis*	33	173	6	1.875	6.674**

Significance levels: $*P < 0.05$, $**P < 0.01$

[a]Silent sites consisting of synonymous site(s) and non-coding region(s)

[b]χ^2 value when synonymous sites were considered

[c]χ^2 value when silent sites were considered

[d]When the test between *O. rufipogon* and *O. australiensis* was conducted, the number of silent polymorphisms in *O. rufipogon* was 23. The tests were conducted in the region that could be aligned between *O. rufipogon* and the four related species

of the numbers of polymorphic replacement and synonymous sites in *O. rufipogon* relative to divergence between *O. rufipogon* and its related species, McDonald and Kreitman (MK) tests (McDonald and Kreitman 1991) were performed (Table 4). The tests did not give significant values against any interspecific comparisons. These results indicate that the rate of replacement to synonymous substitutions in the coding region of *Adh1* is correlated between polymorphism and divergence.

When all the silent substitutions both in non-coding and coding regions were considered, MK tests against *O. meridionalis* and *O. australiensis* were significant (Table 4). These results were caused by high, silent, fixed differences between *O. rufipogon* and these species and imply that even silent polymorphism in *O. rufipogon* is suppressed. When an MK test was conducted for each of the three domains separately, significance was not detected for catalytic domains 1 and 2 in any interspecific comparisons (data not shown). For the co-enzyme-binding domain, the test between *O. rufipogon* and *O. australiensis* gave a marginally significant value ($\chi^2 = 3.82$, $P = 0.05$). This result is due to the high number of synonymous fixed differences in the co-enzyme-binding domain.

Since a difference in the level of nucleotide variation was detected among the three domains in *O. rufipogon* (Table 3), it is of interest to examine if the difference is also observed in interspecific comparisons. HKA tests were used to check correlations of levels of polymorphism and divergence between domains of *Adh1* (Table 5). None of the HKA tests between *O. rufipogon* and A-genome species gave significant results. Only in the comparison between *O. rufipogon* and *O. australiensis* did the HKA test for the replacement sites between catalytic domain 1 and the co-enzyme-binding domain give significant results ($P < 0.05$). This significance was probably caused by the high level of replacement divergence in catalytic domain 1 ($K_{RA} = 0.0127$) and the low level of replacement divergence in the co-enzyme domain ($K_{RA} = 0.0005$) (Table 3). However, if a multiple-test correction were applied, this test would not be significant. These results of the HKA tests imply that difference detected between domains in *O. rufipogon* is generally correlated with divergence between species, especially against the A-genome species.

Phylogenetic relationship of *O. rufipogon* and its related species based on *Adh1* variation

A phylogenetic tree was constructed by using the NJ method based on the nucleotide variation in the entire region of *Adh1* (Fig. 4). For *O. rufipogon*, the topology of the tree was star-like, and internal branches connecting strains of *O. rufipogon* were short. Among *O. rufipogon*, no clear association between geographic origin and clustering in the tree was detected. For example, Laotian strains LV61 and VT51, Chinese strains W1965 and W1956 and Kampuchean strains CB22 and CB27 were separated into different clusters. These observations suggest that *O. rufipogon* has recently spread over the tropics and subtropics of monsoon Asia.

The cultivated species *O. sativa* subspp. *japonica* and *indica* were included in the cluster of *O. rufipogon* (Fig. 4). This result supports the hypothesis that *O. sativa* originated from *O. rufipogon*. Although *O. sativa* subspp. *japonica* and *indica* were not clearly separated into different clusters, they were relatively close to each other, considering the entire range of variation. Despite the difference in geographic distribution area, the South American species *O. glumaepatula*, which is morphologically similar to *O. rufipogon*, was also found in the cluster of *O. rufipogon* of Southeast Asia. This result supports the hypothesis that the American wild rice

Table 5 HKA test among domains of *Adh1* between *O. rufipogon* and its related species. χ^2 values are shown. *OR*, *O. rufipogon*; *OG*, *O. glumaepatula*; *OM*, *O. meridionalis*; *OAO. australiensis*

		OR vs OG	OR vs OB	OR vs OM	OR vs OA
CD1 vs CBD[a]	Coding region	0.570	0.170	2.788	3.549
	Synonymous sites	0.159	0.345	2.194	1.268
	Replacement sites	0.998	0.000	1.002	5.052*
CD1 vs CD2	Coding region	0.164	0.051	0.832	0.010
	Synonymous sites	0.066	0.047	0.904	0.007
	Replacement sites	NA	NA	NA	NA
CBD vs CD2	Coding region	0.016	0.013	0.016	1.878
	Synonymous sites	0.002	0.071	0.002	0.845
	Replacement sites	NA	NA	NA	NA

[a]*CD1* Catalytic domain 1, *CBD* co-enzyme-binding domain, *CD2* catalytic domain 2

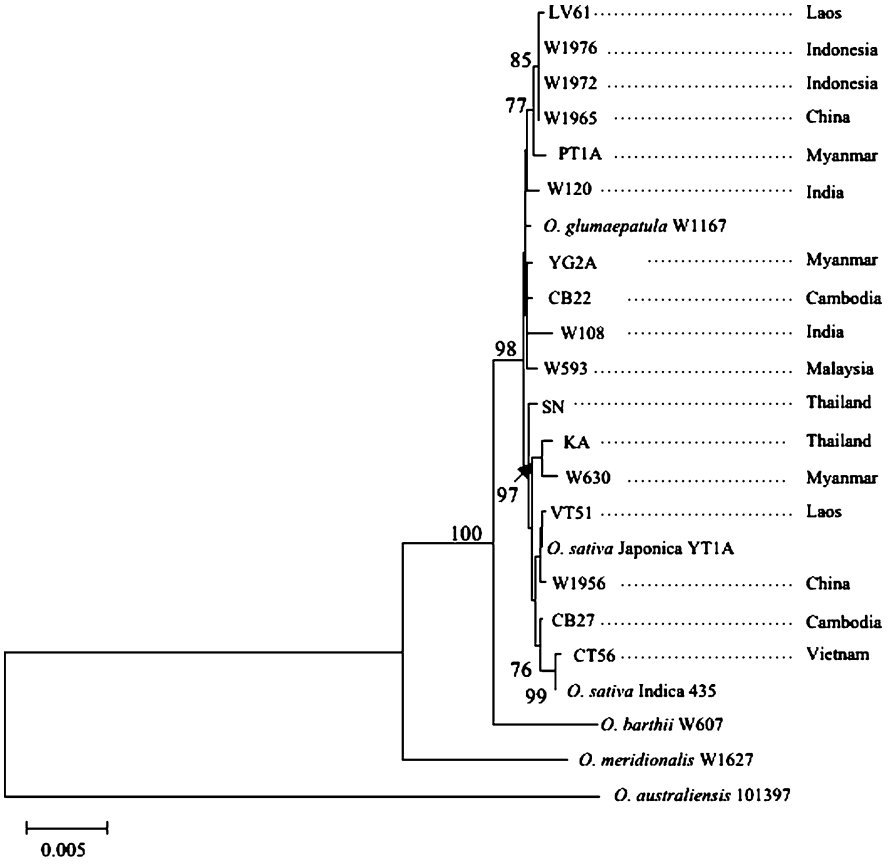

Fig. 4 Neighbour-joining tree for 17 strains of *O. rufipogon* and its related species, using genetic distance calculated by the Jukes and Cantor method on the basis of DNA variation in the entire region of *Adh1*. Bootstrap probabilities >70% from 500 replications are shown. The scale bar of genetic distance is shown at the *bottom of the tree*. Strains of *O. rufipogon* also show their geographic origins

O. glumaepatula could have been recently introduced to the American continents from Asia by human beings (Ishii et al. 1996a). Between the other two A-genome species, *O. barthii* was closer to *O. rufipogon* than *O. meridionalis*. As expected from the large *K* values (Table 2), the E-genome species *O. australiensis* was far distant from all the A-genome species. Phylogenetic relationships among species investigated in this study are generally consistent with those obtained by analyses of RFLP of nuclear DNA, RAPD (Ishii et al. 1996a, b; Bautista et al. 2001), AFLP (Aggarwal et al. 1999), SINE insertion polymorphism (Cheng et al. 2002, 2003) and *Adh1* sequence variation (Ge et al. 1999).

Discussion

Polymorphism in the *Adh1* region of *O. rufipogon*

This study demonstrated that the level of DNA polymorphism in the entire *Adh1* region of *O. rufipogon* was one of the lowest among nuclear loci in plant species. The low level of DNA polymorphism was in accordance with the low heterozygosity of ADH1 allozyme variation in *O. rufipogon*. One of the possible explanations for the low level of nucleotide variation in the *Adh1* region is purifying selection. The tests of neutrality gave significantly negative values for the coding region, especially for replacement sites. When each of the three domains was tested, significance was detected for catalytic domain 1, where the lowest level of nucleotide variation was detected among the three domains. Significantly negative values are indicative of purifying selection, which is also consistent with the observation that *Adh1* plays an important role for recovering plant growth after long-term submergence stress for *O. sativa* (Matsumura et al. 1998). Since *O. rufipogon* is sometimes submerged in water in nature and consequently subjected to anaerobic stress (Morishima 2002), DNA variations in the *Adh1* region could be under natural selection to adapt to the submerged condition.

The other possible hypothesis is the small, effective population size of *O. rufipogon*. Under a population equilibrium and neutral mutation model, if the mutation rate is constant, the level of polymorphism is proportional to the effective population size (Kimura 1983). To examine this hypothesis, other nuclear regions of *O. rufipogon* should be studied, since a small effective size would influence the level of polymorphism over the entire genome.

Phylogenetic relationship between *O. rufipogon* and *O. sativa*

The phylogenetic relationship among *Oryza* species based on nucleotide variations in the *Adh1* region supported the idea that *O. sativa* originated from *O. ruf-*

ipogon (Oka and Chang 1959). Cultivated *O. sativa* subspp. *japonica* and *indica* were included in the cluster of *O. rufipogon*, and species-specific differences between *O. sativa* and *O. rufipogon* were not detected. Nucleotide differences between *O. sativa* subspp. *japonica* and *indica* in the *Adh1* region were found to be polymorphic in *O. rufipogon*. This observation suggests that *O. sativa* subspp. *japonica* and *indica* originated independently from different lines of *O. rufipogon*, supporting the hypothesis of Second (1982) and Ishii et al. (1988) concerning the origin of *O. sativa* subspp. *japonica* and *indica*.

O. sativa subsp. *indica* and a Vietnamese *O. rufipogon* formed a pair supported with a high bootstrap probability, suggesting that Vietnam (Southeast Asia) could be the birthplace of *O. sativa* subsp. *indica*. On the other hand, *O. sativa* subsp. *japonica* was close to Chinese and Laotian *O. rufipogon*, but bootstrap probability did not support this clustering. However, for the entire cluster of *O. rufipogon*, internal branches are short, reflecting the low level of nucleotide polymorphism in the *Adh1* region. Any clear association between geographic origin and clustering in the tree was not detected, making inference of the origin of *O. sativa* somehow dubious. From the tree obtained, it is difficult to conclude where *O. sativa* subspp. *japonica* and *indica* originated.

Star-like topology and short internal branches of the entire cluster of *O. rufipogon* suggest that *O. rufipogon* has recently spread over the present distribution area. However, this observation can also be explained by purifying selection in *Adh1*. If purifying selection were strong enough to keep the level of nucleotide polymorphism low in a region, it would be difficult to reveal a phylogenetic relationship from polymorphism data of the region. To clarify the evolutionary history of *O. sativa* subspp. *japonica* and *indica*, further studies on more genes and accessions are necessary in the future.

Levels of divergence between *O. rufipogon* and its related species

There were no replacement substitutions in the coding region of *Adh1* between *O. rufipogon* and the other A-genome species, indicating that the ADH1 amino acid sequence is conserved among these A-genome species. On the other hand, between *O. rufipogon* and the E-genome species *O. australiensis*, there were six replacement changes in the coding region (only in catalytic domain 1). One of the replacement changes causes physico-chemical differences in ADH1 between the two species. This high level of divergence in catalytic domain 1, especially replacement sites, contrasts with the conserved nature of ADH1 among the A-genome species. Additionally, within *O. rufipogon*, catalytic domain 1 has the lowest level of polymorphism among the three domains of *Adh1*.

The high divergence between *O. rufipogon* and *O. australiensis* in catalytic domain 1, especially at replacement sites, might reflect adaptive evolution in different environmental conditions between *O. rufipogon* and *O. australiensis*. *O. australiensis*, which is found on the edge of ditches, does not seem to be deeply submerged in water, even in the rainy season (Morishima 2002). This habitat contrasts with that of the A-genome species, which are generally found in watery environments (Morishima 2002). The highly conservative nature of ADH1 among the A-genome species may be caused by the common environmental condition that they share. Similarly, it might be possible to hypothesize that the difference in replacement substitutions in catalytic domain 1 between the A-genome species and *O. australiensis* is related to differences in their habitats.

Acknowledgements The authors express their thanks to the National Institute of Genetics, Japan; Dr. Y.I. Sato, Shizuoka University, Japan; the International Rice Research Institute, The Philippines, for providing plant materials for this study; and R. Terauchi, Iwate Biotechnology Research Center, Japan, for comments on an early version of this manuscript. This is contribution number 578 from the Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University.

References

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

Akimoto M, Shimamoto Y, Morishima H (1999) The extinction of genetic resources of Asian wild rice, *Oryza rufipogon* Griff.: a case study in Thailand. Genet Resour Crop Evol 46:419–425

Barbier P (1989) Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. I. Population differentiation in life-history traits and isozymic loci. Jpn J Genet 64:259–271

Barbier P, Ishihama A (1990) Variation in the nucleotide sequence of a prolamin gene family in wild rice. Plant Mol Biol 15:191–195

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

Bautista NS, Solis R, Kamijima O, Ishii T (2001) RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice. Genes Genet Syst 76:71–79

Charlesworth D, Liu FL, Zang L (1998) The evolution of the alcohol dydrogenase gene family by loss of introns in plants of the genus *Leaveworthia* (Brassicaceae). Mol Biol Evol 15:552–559

Cheng C, Tsuchimoto S, Ohtsubo H, Ohtsubo E (2002) Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. Genes Genet Syst 77:323–334

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

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

Dennis ES, Gerlach WL, Pryor AJ, Bennetzen JL, Inglis A, Llewellyn D, Sachs MM, Ferl RJ, Peacock WJ (1984) Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. Nucleic Acids Res 12:3983–4000

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

Gao Lz, Hong DY (2000) Allozyme variation and population genetic structure of common wild rice *Oryza rufipogon* Griff. in China. Theor Appl Genet 101:494–502

Gao LZ, Wei C, Yang Q, Hong DY, Ge H (2001) Intra-population genetic structure of *Oryza rufipogon* Griff. in Yunnan, China. J Plant Res 114:107–113

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

Ge S, Sang T, Lu BR, Hong DY (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc Natl Acad Sci USA 96:14400–14405

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: Brown HD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer, Sunderland, pp 43–63

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

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

Ishii T, Nakano T, Maeda H, Kamijima O (1996a) Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis. Genes Genet Syst 71:195–201

Ishii T, Nakano T, Maeda H, Kamijima O, Khush GS (1996b) Phylogenetic relationships betweeen cultivated and wild species of rice as revealed by DNA polymorphisms. In: Rice genetics III. Proceedings of the Third International Rice Genetics Symposium, International Rice Research Institute, Manila, 16–20 October 1995

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

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) *MEGA2*: molecular evolutionary genetics analysis. Bioinformatics 17:1244–1245

Lin JZ, Brown AHD, 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

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

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652–654

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

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

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, Barbier P (1990) Mating system and genetic structure of natural populations in wild rice *Oryza rufipogon*. Plant Species Biol 5:31–39

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

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

Nakazono M, Tsuji H, Li Y, Saisho D, Arimura S, Tsutsumi N, Hirai A (2000) Expression of a gene encoding mitochondrial aldehyde dehydrogenase in rice increases under submerged conditions. Plant Physiol 124:587–598

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

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

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

Second G (1985) Evolutionary relationships in the *Sativa* group of *Oryza* based on isozyme data. Genet Sel Evol 17:89–114

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 *Adh1–Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381–391

Utami DW, Moeljopawiro S, Septiningsih EM, Aswidinnoor H, Sujiprihati SD (2001) Introgression of the resistance trait of blast from *Oryza rufipogon* to IR64. J Bioteknol Pertanian 6:51–58

Vaughan DA (1989) The genus *Oryza* L. Current status of taxonomy. IRRI Res Pap Ser 138:1–21

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

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

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