Mol. Cells *27*, 459-465, April 30, 2009
DOI/10.1007/s10059-009-0061-8

Molecules and Cells

Sequence Diversity of a Domesticated Transposase Gene, *MUG1*, in *Oryza* Species

Soon-Jae Kwon[1,5], Kyong-Cheul Park[1,5], Jae-Han Son[1], Thomas Bureau[2], Cheul-Ho Park[3], and Nam-Soo Kim[1,4,*]

***MUG1* is a MULE transposon-related domesticated gene in plants. We assessed the sequence diversity, neutrality, expression, and phylogenetics of the *MUG1* gene among *Oryza ssp*. We found *MUG1* expression in all tissues analyzed, with different levels in *O. sativa*. There were 408 variation sites in the 3886 bp of *MUG1* locus. The nucleotide diversity of the *MUG1* was higher than functionally known genes in rice. The nucleotide diversity (π) in the domains was lower than the average nucleotide diversity in whole coding region. The π values in nonsynonymous sites were lower than those of synonymous sites. Tajima D and Fu and Li D* values were mostly negative values, suggesting purifying selection in *MUG1* sequences of *Oryza ssp*. Genome-specific variation and phylogenetic analyses show a general grouping of *MUG1* sequences congruent with *Oryza ssp.* biogeography; however, our *MUG1* phylogenetic results, in combination with separate B and D genome studies, might suggest an early divergence of the *Oryza ssp.* by continental drift of Gondwanaland. *O. longistaminata MUG1* divergence from other AA diploids suggests that it might not be a direct ancestor of the African rice species.**

INTRODUCTION

The genus *Oryza* consists of 22 wild species and two cultivated species represented by 9 genome groups (A-, B-, BC-, CD-, E-, F-, G-, HJ-, and HK-) (Guo and Ge, 2005). On the basis of morphological traits and habitats, *Oryza ssp.* are divided into four main groups: (1) *O. sativa* and its relatives, which are AA genome diploid species, (2) the *O. officinalis* complex, which contains species with BB, CC, EE, FF, BBCC, and CCDD genomic constitutions, (3) the *O. ridleyi* complex, which contains HHJJ tetraploids, and (4) the *O. granulate* complex, which is mainly GG diploid species (Vaughan et al., 2005; 2008). Because the genus *Oryza* contains agronomically important species and genetic model species for cereal grasses, genetics and phylogenetics of the *Oryza* species have been well established (Ge et al., 1999; Guo and Ge, 2005; Khush, 1997; Vaughan et al., 2008). However, our understanding of the genome differentiation and nucleotide diversity in the wild *Oryza* species is still incomplete.

Genome projects show that transposable elements (TEs) are the major components of eukaryotic genomes and are significantly involved in genetic evolution (Kidwell and Lisch, 2001). TEs are additionally involved in cellular processes of their hosts (Agrawal et al., 1998; Cowan et al., 2005; Hudson et al., 2003). Transposon-related domesticated genes are genes that acquire host function after loss of transposon features. The best known domesticated transposase-related genes are the *FHY3* and *FAR1*genes found in *Arabidopsis* (Hudson et al., 2003). The *FHY3* and *FAR1*, which are related to *Mutator*-like element (MULE) transposases, encode transcription factors regulating phytochrome A-signaling pathway (Lin et al., 2007). MULEs are autonomous class II transposons which are mobilized by the transposase, MURA, encoded in the *mudrA* gene (Lisch et al., 1999). MURA binds to the terminal invert repeats (TIR) of ~200 bp of the MULEs to mobilize the elements (Benito and Walbot, 1997). Upon insertion into the genome, the MULEs create 9-11 bp target site duplication (TSDs). Some MULEs lack the terminal symmetry to be termed non-TIR MULEs (Hoen et al., 2006; Yu et al., 2000). Depending on the retention of TIRs, TSD sequences and other transposon features, *mudrA* genes have been grouped into three categories: (1) those with TSD sequences (TAMs, transposon-associated *mudrA* genes), (2) those lacking TSDs but retaining other transposon features (pTAM, potentially transposon-associated *mudrA* genes), and (3) those lacking all transposon features (TDMs, transposon-dissociated *mudrA* genes) (Cowan et al., 2005).

The *MUSTANG* (*MUG*) gene family has TDMs present in a diverse group of angiosperms (Cowan et al., 2005). One member of the *MUG* gene family, *MUG1,* shows high amino acid sequence similarity among plant species. *MUG1* is expressed in a variety of tissues (rice and *Arabidopsis spp.* calli, corn endosperm and ears, developing seeds, inflorescent meristem, and seedling leaves) by matching sequences with either cDNAs, ESTs, or both (Cowan et al., 2005). The *MUG1 d*N/*d*S

[1]Department of Molecular Biosciences, School of Biosciences and Biotechnology, Kangwon National University, Chunchon 200-701, Korea, [2]Department of Biology, McGill University, Montreal, Quebec, H3A 1B1, Canada, [3]Department of Plant Biotechnology, Kangwon National University, Chunchon 200-701, Korea, [4]Institute of Biosciences and Biotechnology, Kangwon National University, Chunchon 200-701, Korea, [5]These authors contributed equally to this work.
*Correspodence: kimnamsu@kangwon.ac.kr

Received January 6, 2009; revised February 24, 2009; accepted March 3, 2009; published online April 13, 2009

Keywords: domesticated gene, *MUG1*, nucleotide diversity, *Oryza,* phylogenetics, purifying selection

Table. 1. Name of the species, genomes, accessions and origin of the *Oryza* species analysed

Species	Genome	Accession number	Origin
Oryza sativa ssp. *Japonica*	AA	Jinmibyeo	Korea
Oryza sativa ssp. *Indica*	AA	IR57313-106-2-3	IRRI
Oryza rufipogon	AA	IRGC 100657	Taiwan
Oryza nivara	AA	IRGC 100593	Taiwan
Oryza glaberrima	AA	IRGC 100984	Nigeria
Oryza barthii	AA	IRGC 86447	Guinea
Oryza glumaepatala	AA	IRGC 100184	Cuba
Oryza longistaminata	AA	IRGC 101202	Nigeria
Oryza meridionalis	AA	IRGC 101145	Australia
Oryza officinalis	CC	IRGC 105328	China
Oryza minuta	BBCC	IRGC 101081	Philippines
		IRGC 103865	Philippines
Oryza alta	CCDD	IRGC 100161	Brazil
		IRGC 105222	Surinam
Oryza grandiglumis	CCDD	IRGC 101405	Brazil
		IRGC 105560	Brazil
Oryza latifolia	CCDD	IRGC 100172	Guatemala
		IRGC 105145	Colombia

value was significantly < 1 among the orthologs, implying that *MUG1* maintains a type of host function when it is expressed. To test this, we analyzed sequence diversity and *MUG1* loci relationships among different species in *Oryza* genus.

MATERIALS AND METHODS

Plant materials and DNA extraction

We obtained *Oryza ssp*. seeds from the International Rice Genebank Collection (IRGC), IRRI, Los Baňos, Philippines, of cultivated and wild rice accessions (Table 1). We extracted genomic DNA from the young leaves of newly planted seeds using methods of Dellaporta et al. (1983).

PCR amplification and sequencing

We designed two pairs of primers to amplify the *MUG1* locus from the 5′-flanking region to the 3′-flanking region. One pair of primers (P1F 5′-TTG ATC TGC CTT CAT ACT CCA A-3′, P1R 5′-TTG ATC TGC CTT CAT ACT CCA A-3′) amplified part of the 5′-flanking sequence and 100 bp of the 5′-coding region of the *MUG1* locus. Another pair (P2F 5′-GAC GCT AGG TCA GTC ACT GGT T-3′, P2R 5′- TAG AGA GTC AGA AAC CGC ATG TAG-3′) amplified the entire coding region. We overlapped the amplified fragments from the P1 and P2 pairs by 100 bp in order to connect the sequences of P1 and P2 amplicons. The PCR reaction was: 94°C for 2 min, 35 cycles of 94°C for 30 s, 55°C for 30 s, 68°C for 4 min, and 68°C for 10 min. The PCR cocktail contained 20 μl of template DNA (15 ng), primers (0.05 μM), dNTPs (0.2 mM), $MgSO_4$ (2 mM), and one unit of Taq polymerase (Invitrogen Plantinum High Fidelity, USA). We ligated the amplified fragments with a pGEM T-Easy Vector (Promega Inc., USA) and cloned the ligated fragments using standard protocols.

We selected 10 clones from each accession of diploid species and 20 clones from each accession of tetraploid species. We used an ABI377 automatic DNA sequencer (Applied Biosystems, USA) to sequence each species, and then we aligned the sequences using CLUSTALX 1.81 (Thompson et al., 1997), which requires manual editing. An NCBI domain searching program (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) identified the following domains: MuDR, transposase, and Zinc finger.

Expression analysis

Using a Trizol extraction kit (Invitrogen, USA), we isolated total RNAs from the following developmental tissue stages of *O. sativa* ssp. *japonica*: inflorescences, buds from germinating seeds, shoots from early seedlings, germinating embryos removed of roots and shoots, roots from early seedlings, leaves from young plants, and stems from young plants. We used a SuperScript™ III First-Strand synthesis kit (Invitrogen, USA) to convert mRNAs to cDNAs (You et al., 2007). We designed complementary primers (RTPF 5′-CAC GGC TAG GTT TTG GAG TG-3′ and RTPR 5′-GCC CAA AGT CAT GGT GAT TC-3′) for the 5′-UTR and 5′-ORF regions to amplify the 5′-region. PCR conditions were the same as previously described.

Nucleotide diversity and neutrality test

Prior to any of the following analyses, we removed the indel mutations from the samples. We analyzed nucleotide diversity with the DnaSP version 4.5 (http://www.ub.es/dnasp/) (Rozas and Rozas, 1999) and determined neutrality statistics, *d*N, and *d*S, with the MEGA program, Version 4.0 (http://www.megaso-ftware.net) (Kumar et al., 2001). We analyzed nucleotide diversity (π), ratio of segregating site (θ) (Watterson, 1975), D (Tajima, 1989), and D* (Fu and Li, 1993) for the entire region (including 5′- and 3′-UTR sequences of the *MUG1* locus), for the coding region, and for the domain sequences, all with pair-wise comparisons between intra- and inter-genomic sequences.

Phylogenetic analysis

We conducted a multiple alignment of the nucleotide sequences from 17 *MUSTANG* genes with the MAFFT program

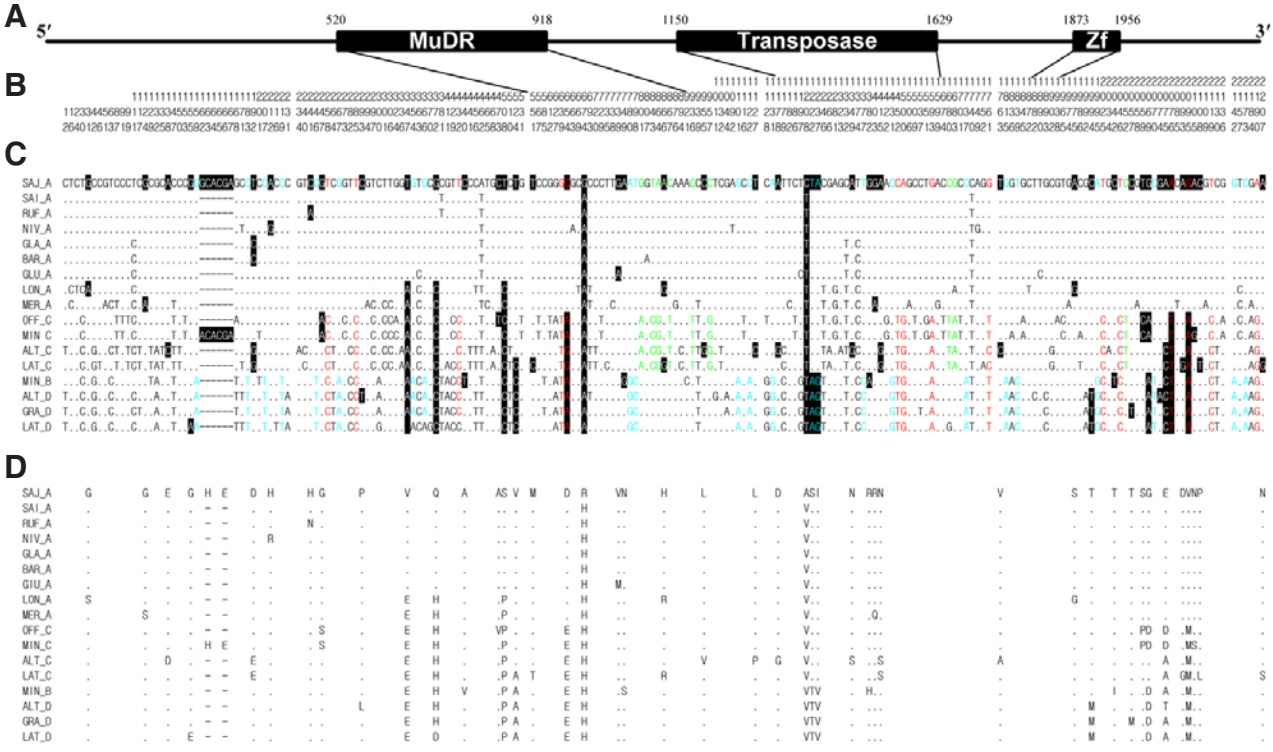

Fig. 1. Summary of DNA variation sites of *MUG1* in *Oryza ssp.* (A) *MUG1* structure at sites 0-2,250 bp. The black boxes indicate domains. (B) Polymorphic sites of *MUG1* in *Oryza ssp.* (C) Polymorphic nucleotide sequences and genome-specific variation. The black boxes indicate non-synonymous sites. The red, green, and blue letters indicate (A), (C), and (B) (or D) genome-specific variation sites, respectively. (D) *MUG1* amino acid sequence alignment in *Oryza ssp.*

(http://align.bmr.kyushu-u.ac.jp/mafft/online/server/). We constructed a phylogenetic tree from 1,000 bootstrap repetitions of the "neighbor joining-JTT method", and retrieved the tree from MAFFT raw data using the TreeView program, Version 1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

RESULTS

MUG1 sequence amplification

We isolated *MUG1* sequences from ten different species of *Oryza*. The isolated *MUG1* locus spanned 3,885 bp. The 5′-flanking sequence was 482 bp. The 5′-UTR (un-translated region) sequence was 713 bp, with 198 bp of intronic sequence, and the 3′-UTR sequence was 440 bp. Coding sequence was 2,250 bp and it had three functional domains: *MuDR* (339 bp), transposase (480 bp), and Zinc finger (84 bp) (Fig. 1).

We analyzed *Oryza ssp.* that included diploids of AA (*O. sativa, O. rufipogon, O. nivara, O. glaberrima, O. barthii, O. glumaepatula, O. longistaminata*, and *O. merionalis*) and CC (*O. officinalis*), and allotetraploids of BBCC (*O. minuta*) and CCDD (*O. alta, O. grandiglumis,* and *O. latifolia*). We confirmed all the genome-specific haplotypes in the allopolyploids, except for the C haplotype in *O. grandiglumis* (CCDD), by the comparison of the haplotypes of diploids with those of corresponding allopolyploids.

Patterns and levels of DNA polymorphism

We found 408 variation sites (407 segregating sites and one 6 bp indel) in the *MUG1* regions of the *Oryza ssp.* (Fig. 1, Table 2). The segregating sites included 134 singleton variable sites with two variants, 244 parsimony informative sites with two variants, 5 singleton variable sites with three variants, and 24 parsimony informative sites with three variants. The 5′ flanking region [of the *MUG1* gene] had the highest nucleotide diversity (π) (Fig. 2). In coding sequences, both 5′- and 3′-end regions revealed higher nucleotide diversity than the rest of the regions. The domains of *MuDr* and the transposase had similar nucleotide diversity that was low compared to other regions in the *MUG1* locus. The π values in synonymous sites were slightly higher than those in non-synonymous sites over the entire re-

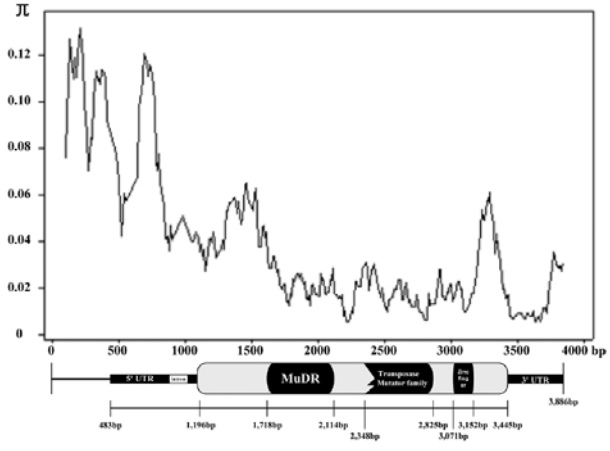

Fig. 2. A sliding-window plot of the nucleotide variations in the *MUG1* locus. Window size is 100 bp and step size is 25 bp. Gene structure is shown at the bottom of the figure.

Table. 2. Summary of DNA variation of *MUG1* locus in *Oryza* species

	Number of sites	S	π	θ	D	D*
Entire region	3886	408	0.03611	0.03812	-0.22796	-0.27037*
5′ flanking	482	82	0.09985	0.09207	0.35905	0.61798*
5′ UTR	713	95	0.06230	0.07036	-0.48740	-0.73101
Intron	198	15	0.06262	0.07156	-0.48134	-1.05299
3′ UTR	440	31	0.01644	0.02206	-1.04268*	-1.78409*
Coding region	2250	200	0.02624	0.02768	-0.22328	-0.16011
Synonymous sites	520.73	152	0.31849	0.31440	0.05580	0.30115
Replacement sites	1720.27	43	0.22917	0.31418	-1.13114	-1.57839
MuDR region	399	29	0.01947	0.02230	-0.51586	-0.09888
Synonymous sites	88.46	22	0.28772	0.30865	-0.27222	0.32692
Replacement sites	307.53	7	0.25000	0.33805	-0.92118	-0.79891
Transposase region	480	29	0.01716	0.01849	-0.29310	-0.27188
Synonymous sites	111.95	21	0.31723	0.30988	0.09456	0.20945
Replacement sites	368.05	8	0.19669	0.29579	-1.18493	-1.33662
Zinc finger region	84	5	0.01050	0.01761	-1.30165	-1.09665
Synonymous sites	21.41	5	0.17647	0.29579	-1.30165	-1.09665
Replacement sites	62.59	0				

*Significance level: $P < 0.05$

S, Number of polymorphic sites; π, Nucleotide diversity; θ, Proportion of segregating sites; D, Tajima's D statistics; D*, Fu and Li's D* statistics.

gion and within each domain (Table 2), suggesting purifying selection likely occurs in the entire *MUG1* locus of *Oryza ssp.* D and D* statistical tests showed nucleotide polymorphism neutrality values to be mostly negative in the entire region, the coding region, and the domain regions, with the exception of D in the synonymous sites of the coding region and D* in the transposase region, MuDR, and domains.

The *d*N and *d*S values among *MUG1* sequences within genomes were lower than those between genomes (Table 3). Average *d*S values were 0.025, 0.081, and 0.014 in A, C, and D genomes, respectively, and 0.102 for all *Oryza ssp.* Average *d*N values were 0.002, 0.006, and 0.002 in A, C, and D genomes, respectively, and 0.006 for all *Oryza ssp.*

Genome specific variation

We detected 200 variable sites in the coding region, of which 43 were non-synonymous mutations and 152 were synonymous mutations. Five variable sites (1,127; 1,128; 2,074; 2,075; and 2,076) caused complex codon changes (i.e., more than two variable sites in a codon caused multiple amino acid changes) and accounted for 152 synonymous mutations instead of 157 synonymous mutations. We detected 48 genome-specific variable sites in the coding region, of which 43 sites were in the synonymous sites and five were in the non-synonymous sites (Fig. 1). These variable sites included 15 in the A genome, 24 in the C genome, and 9 in the B and D genomes specific sites.

Expression pattern of *MUG1*

The RT-PCR analysis revealed *MUG1* expression in every tissue analyzed. The levels of expression varied among tissue type. Inflorescences and SAM (shoot apical meristem) showed high level expression. Germinating buds, shoots from germinating seeds, and stems from young plants showed intermediate level expression. The roots from germinating seeds, germinating embryos, and leaves of young plants showed low level expression (Fig. 4).

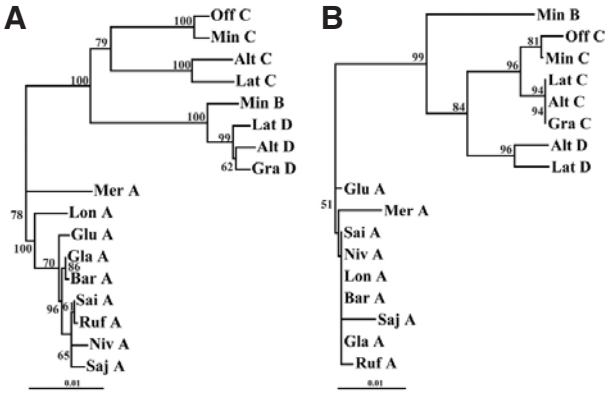

Fig. 3. Phylogenetic trees of *MUG1* (A) and *Adh1* (B), with loci based on nucleotide sequence variation in *Oryza ssp.* The numbers at the nodes are bootstrap values. The bars at the bottom are substitution rates.

Phylogenetic analysis of the *MUG1* locus in the *Oryza* species

We constructed phylogenetic trees of *Oryza ssp. MUG1* and *Adh1* loci (Fig. 3). We used nucleotide sequence variations of entire coding regions in the *MUG1* analysis and cDNA sequence variations of exon 2 to exon 7 in the *Adh1* analysis. *Adh1* is one of the most investigated plant nuclear genes in population genetic studies of *Oryza ssp* (Ge et al., 1999; Yoshida et al., 2004). The base substitution rate (bars in Fig. 3) shows that the base changes in *MUG1* are much higher than those in *Adh1*. Although the basic topology of the two phylogenetic trees was as similar as the genome-specific clustering, its details were incongruent with clustering. The rate of base changes, denoted by the branch length, was higher in polyploids, implying that purifying selection in the polyploids might have been low compared to the diploids. In the *Adh1* tree, the

Table. 3. Estimates of the number of synonymous mutation at synonymous sites (*d*S) and the number of nonsynonymous mutation at nonsynonymous sites (*d*N) in *MUG1* ORF region in *Oryza* species

	Synonymous mutation (*d*S)																
	Saj_A	Sai_A	Ruf_A	Niv_A	Gla_A	Bar_A	Glu_A	Lon_A	Mer_A	Min_B	Off_C	Min_C	Alt_C	Lat_C	Gra_D	Alt_D	Lat_D
Saj_A		0.006	0.006	0.012	0.012	0.014	0.016	0.033	0.062	0.131	0.128	0.128	0.136	0.138	0.149	0.151	0.146
Sai_A	0.001		0	0.01	0.01	0.012	0.014	0.031	0.06	0.125	0.126	0.126	0.133	0.135	0.142	0.144	0.14
Ruf_A	0.002	0.001		0.01	0.01	0.012	0.014	0.031	0.06	0.125	0.126	0.126	0.133	0.135	0.142	0.144	0.14
Niv_A	0.002	0.001	0.001		0.016	0.018	0.02	0.037	0.066	0.136	0.133	0.133	0.14	0.138	0.154	0.151	0.146
Gla_A	0.001	0	0.001	0.001		0.002	0.008	0.025	0.054	0.127	0.119	0.119	0.124	0.126	0.14	0.142	0.137
Bar_A	0.001	0	0.001	0.001	0		0.01	0.027	0.056	0.129	0.121	0.121	0.126	0.129	0.142	0.144	0.14
Glu_A	0.002	0.001	0.001	0.001	0.001	0.001		0.025	0.054	0.125	0.119	0.119	0.127	0.129	0.138	0.139	0.135
Lon_A	0.005	0.004	0.004	0.004	0.004	0.004	0.004		0.052	0.138	0.117	0.117	0.124	0.126	0.147	0.144	0.144
Mer_A	0.004	0.003	0.004	0.004	0.003	0.003	0.004	0.003		0.157	0.126	0.126	0.134	0.14	0.17	0.168	0.168
Min_B	0.01	0.008	0.009	0.009	0.008	0.008	0.009	0.008	0.008		0.159	0.152	0.149	0.151	0.028	0.034	0.032
Off_C	0.007	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.005	0.006		0.014	0.11	0.108	0.172	0.174	0.167
Min_C	0.007	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.005	0.006	0.001		0.114	0.11	0.17	0.172	0.165
Alt_C	0.01	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.008	0.009	0.008	0.008		0.025	0.143	0.145	0.143
Lat_C	0.01	0.009	0.009	0.009	0.009	0.009	0.009	0.007	0.008	0.008	0.008	0.008	0.007		0.15	0.147	0.145
Gra_D	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.007	0.003	0.006	0.006	0.008	0.007		0.014	0.016
Alt_D	0.009	0.008	0.009	0.009	0.008	0.008	0.009	0.008	0.008	0.004	0.006	0.006	0.009	0.008	0.002		0.014
Lat_D	0.009	0.008	0.009	0.009	0.008	0.008	0.009	0.008	0.008	0.004	0.006	0.006	0.009	0.008	0.002	0.002	
	Non-ynonymous mutation (*d*N)																

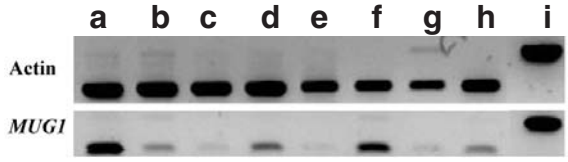

Fig. 4. *MUG1* expression in different tissues of *O. sativa*. The mRNAs were from inflorescence (a), germinating buds (b), roots from germinating seeds (c), shoots from germinating seeds (d), germinating embryo (e), SAM (f), leaves from young plants (g), and stems from young plants (h). Lane i = genomic DNA amplified with Actin primers and *MUG1* (control).

O. minuta B genome was out-grouped from C and D genomes; however, in the *MUG1* tree, the *O. minuta* B genome was closely clustered to D genome species. Since *Adh1* genetic variation among the A genome species was low, *Adh1* variation could not be used to define species relationships. In contrast, *MUG1* variation was sufficient to cluster species by geographical origin: Asian species (*O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica, O. rufipogon*, and *O. nivara*) and African species (*O. glaberrima* and *O. barthii*). *O. meridionalis* was out-grouped in the *MUG1* tree.

DISCUSSION

The *MUG1* is a MULE transposon-related gene (Cowan et al., 2005). MULEs are mobilized by the MuDR transposase that is encoded in the *mudrA* gene (Lisch et al., 1999). Since the *MUG1* locus in the current study lost its transposon activity from a lack of transposon features, the *mudrA* gene in the *MUG1* may have undergone decay by accumulation of mutations. Alternatively, it may have retained some function through purifying selection and unknown host function. Cowan et al. (2005) show that *MUG1* is present and has unknown host function in diverse angiosperm species. We analyzed *MUG1* sequences in the *Oryza* genus, which has well-characterized phylogenetic relationships developed by DNA markers and sequence variations (Ge et al., 1999; Kim et al., 2003; Park et al., 2003).

The nucleotide diversities (π) of the *MUG1* loci among the *Oryza ssp.* are higher than those of 10 functional genes in Asian rice species (Zhu et al., 2007) and higher than those of *Adh1* in *O. rufipogon* (Yoshida et al., 2004). In the domains, however, lower diversity and higher π values in the synonymous sites as compared to the non-synonymous sites imply that purifying selection is maintained on the domains of the *MUG1* locus. Negative D and D* values in those regions also suggests purifying selection in the *MUG1* locus. In all *Oryza ssp. MUG1* loci, the MuDR and Zinc finger domains contained whole domain sequences, and the transposase domain carried a truncated domain sequence; however, all three domains had similar π, D, and D* values, suggesting they may be under the same level of evolutionary constraint. Further study on the selection constraint of the truncated transposase domain could verify this claim. Although the sequence changes of the *MUG1* are higher than those of functionally known genes, the *d*N/*d*S ratio is lower than 1 (Table 3), suggesting *MUG1* is under the influence of purifying selection. The function of the *MUG1* in plants is not clear at the moment. *MUG1* expresses itself constitutively in many different tissues and at many different levels. Expression patterns in rice are similar to *Arabidopsis thaliana,* as shown in the MPSS database (data not shown). *MUG1* function in plants, therefore, may have been preserved even before the divergence of the monocotyledonous and dicotyledonous plant species.

The *MUG1* genome-specific variation and phylogenetic relationships that we found in this study do not correlate with the results of previous studies. In this study, the B genome of *O. minuta,* and the D genomes of *O. latifolia, O. alta,* and *O. grandiglumis,* share enough genome-specific variation to show their close phylogenetic relationships. However, studies on *Adh1* variation (Ge et al., 1999), molecular markers (Kim et al., 2003), and transposon displays (Panaud et al., 2002; Park et al., 2003) show *O. minuta* (BBCC) to be phylogenetically closer to *O. officinalis* (CC or BBCC) than to *O. latifolia* (CCDD), *O. alta* (CCDD), or *O. grandiglumis* (CCDD). *O. latifolia, O. alta,* and *O. grandiglumis* are found in South America, and *O. minuta* and *O. officinalis* are found in South Asia (Vaughan, 1994). Currently, two hypotheses on the origin and evolution of the genus *Oryza* exist. One hypothesis proposes that the *Oryza ssp.* originated at least 130 million years ago (MYA) in Gondwanaland as wild grasses and then drifted into different continents when the super continent broke up into Asia, Africa, America, and Australia (Chang, 1976; Khush, 1997). Another hypothesis proposes that the origin and dispersal of *Oryza ssp.* occurred 70 MYA and that *Oryza ssp.* were dispersed by animals, including humans and birds, rather than by continental drift (Vaughan et al., 2005; 2008). Our result of the B genome *MUG1* sequence variation best supports the first hypothesis with close relationship of *O. minuta, O. latifolia, O. alta,* and *O. grandiglumis.*

Evolutionary divergence of Asian and African AA genome rice is disputable. *O. glaberrima*, a cultivated African rice, is believed to be a direct ancestor of *O. longistaminata*, an African perennial (Khush, 1997). Ma and Bennetzen (2004), however, showed Asian species (*O. sativa, O. nivara*, and *O. rufipogon*) and African species (*O. glaberrima* and *O. barthii*) likely diverged before *O. longistaminata*. Furthermore, our data shows that *O. longistaminata* diverged earlier than *O. sativa* and *O. glaberrima*, which is congruent with the results of molecular marker analyses by Wang et al. (1992) and Agrawal et al. (1998).

Although the general topology of the phylogenetic trees was similar between *MUG1* and *Adh1* among the *Oryza* species, the detailed differences were present. While the *Adh* gene expressed during the stress conditions (Chung and Ferl, 1999; Dolferus et al., 1994), the molecular function of the *MUG 1* has not been clearly established at this moment. The sequence variation of the *Adh1* locus has been used widely in plant population studies (Ge et al., 1999; Yoshida et al., 2004). The general consistency of phylogenetic trees between two genes may support the purifying selection from the data of *d*N/*d*S (Table 3). The unclear molecular function of the *MUG1* is being characterized by us.

In conclusion, the domesticated *MUG1* sequence diversity revealed in this study suggests *MUG1* was likely under purifying selection during the evolution of *Oryza ssp.* Although *MUG1* diversity levels were higher than the functionally known genes of *Oryza ssp*., constitutive expression at different levels in different tissues suggests *MUG1* has a host function. Our *MUG1* phylogenetic pattern is congruent with some of the species relationships documented in other studies; however, since it is not congruent with all relationships, we recommend further phylogenetic analyses.

ACKNOWLEDGMENT

Part of this work was carried out in TB's laboratory with a sabbatical grant to NSK from Korean Research Foundation. KSJ, PKC, and SJH are recipients of the post Brain Korea 21 program from Ministry of Education of Korea. The authors thank to Drs. K. Jena and S. Choi for critical reading of the manuscript.

REFERENCES

Agrawal, A., Eastman, Q.M., and Schatz, D.G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature *394*, 744-751.

Benito, M.I., and Walbot, V. (1997). Characterization of the maize Mutator transposable element MURA transposase as a DNA-binding protein. Mol. Cell. Biol. *17*, 5165-5175.

Chang, T.T. (1976). The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. Euphytica *25*, 425-441.

Chung, H.J., and Ferl, R.J. (1999). Arabidopsis alcohol dehydrogenase expression in both shoots and roots is conditioned by root growth environment. Plant Physiol. *121*, 429-436.

Cowan, R.K., Hoen, D.R., Schoen, D.J., and Bureau, T.E. (2005). *MUSTANG* is a novel family of domesticated transposase genes found in diverse angiosperms. Mol. Biol. Evol. *22*, 2084-2089.

Dellaporta, S., Wood, J., and Hicks, J. (1983). A simple and rapid method for plant DNA preparation. Version II. Plant Mol. Biol. Rep. *1*, 19-21.

Dolferus, R., Jacobs, M., Peacock, W.J., and Dennis, E.S. (1994). Differential interactions of promoter elements in stress responses of the Arabidopsis Adh gene. Plant Physiol. *105*, 1075-1087.

Fu, Y.X., and Li, W.H. (1993). Statistical tests of neutrality of mutations. Genetics *133*, 693-709.

Ge, S., Sang, T., Lu, B.R., and Hong, D.Y. (1999). Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. USA *96*, 14400-14405.

Guo, Y., and Ge, S. (2005). Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes[1]. Am. J. Bot. *92*, 1548-1558.

Hoen, D.R., Park, K.C., Elrouby, N., Yu, Z., Mohabir, N., Cowan, R.K., and Bureau, T.E. (2006). Transposon-mediated expansion and diversification of a family of ULP-like genes. Mol. Biol. Evol. *23*, 1254-1268.

Hudson, M.E., Lisch, D.R., and Quail, P.H. (2003). The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant J. *34*, 453-471.

Khush, G.S. (1997). Origin, dispersal, cultivation and variation of rice. Plant. Mol. Biol. *35*, 25-34.

Kidwell, M.G., and Lisch, D.R. (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution *55*, 1-24.

Kim, N.H., Lee, J.K., Shin, Y.B., Park, C.H., Cho, E.G., and Kim, N.S. (2003). Phylogenetic relationship and genetic variation among *Oryza* species revealed by AFLP analysis. Kor. J. Breed. *35*, 78-85.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics *17*, 1244-1245.

Lin, R., Ding, L., Casola, C., Ripoll, D.R., Feschotte, C., and Wang, H. (2007). Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. Science *318*, 1302-1305.

Lisch, D., Girard, L., Donlin, M., and Freeling, M. (1999). Functional analysis of deletion derivatives of the maize transposon MuDR delineates roles for the MURA and MURB proteins. Genetics *151*, 331-341.

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

Panaud, O., Vitte, C., Hivert, J., Muzlak, S., Talag, J., Brar, D., and Sarr, A. (2002). Characterization of transposable elements in the genome of rice (*Oryza sativa* L.) using representational difference analysis (RDA). Mol. Genet. Genomics *268*, 113-121.

Park, K.C., Lee, J.K., Kim, N.H., Shin, Y.B., Lee, J.H., and Kim, N.S. (2003). Genetic variation in *Oryza* species detected by MITE-AFLP. Genes Genet. Syst. *78*, 235-243.

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

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

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *25*, 4876-4882.

Vaughan, D. (1994) *The wild relatives of rice*, IRRI Manila, Philippines.

Vaughan, D., Kadowaki, K., Kaga, A., and Tomooka, N. (2005). On the phylogeny and biogeography of the genus *Oryza*. Breed. Sci. *55*, 113-122.

Vaughan, D., Lu, B., and Tomooka, N. (2008). The evolving story of rice evolution. Plant Sci. *174*, 394-408.

Wang, Z., Second, G., and Tanksley, S. (1992). Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. *83*, 565-581.

Watterson, G.A. (1975). On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. *7*, 256-276.

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

You, M.K., Oh, S.I., Ok, S.H., Cho, S.K., Shin, H.Y., Jeung, J.U., and Shin, J.S. (2007). Identification of putative MAPK kinases in *Oryza minuta* and *O. sativa* responsive to biotic stresses. Mol. Cells *23*, 108-114.

Yu, Z., Wright, S.I., and Bureau, T.E. (2000). Mutator-like elements in *Arabidopsis thaliana*: Structure, diversity and evolution. Genetics *156*, 2019-2031.

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