

# Tracking Selection Signatures Based on Variation in *OsLEA27* within Myanmar Landraces of Upland and Dryland Rice

Wunna<sup>1</sup>, Syed Abdullah Gilani<sup>2</sup>, Makoto Kawase<sup>3</sup>, Ryo Ohsawa<sup>1</sup>, Kazuo N. Watanabe<sup>1\*</sup>

<sup>1</sup>Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

<sup>2</sup>Department of Biological Sciences and Chemistry, University of Nizwa, Nizwa, Sultanate of Oman

<sup>3</sup>Gene Bank, National Institute of Agrobiological Sciences, Tsukuba, Japan

Email: \*[nabechan@gene.tsukuba.ac.jp](mailto:nabechan@gene.tsukuba.ac.jp)

Received 17 May 2015; accepted 15 August 2015; published 18 August 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

To track the selection evident along the genome segment of *OsLEA27* gene, a member of dehydrin gene family, 2.9-kbp nucleotide sequence containing the promoter 5' upstream and transcribed region of *OsLEA27* was determined for 35 upland and dryland Myanmar landraces from drought-prone areas. Nucleotide diversity, neutrality tests, haplotype network analysis, and linkage disequilibrium (LD) analysis were performed to infer the impact of selection and to investigate non-random associations of SNPs within all or part of the entire *OsLEA27* region. The evidence for LD, the presence of two distinct haplotype groups across four different geographical regions, and the significant values obtained in a sliding-window analysis of mutation-drift tests all suggest the effects of selection on *OsLEA27* in a set of 30 landraces. The neutrality test values for 5' upstream region of *OsLEA27* were significantly negative ( $p < 0.05$ ) within the set of 30 accessions. The entire *OsLEA27* region was significantly negative in accessions of the northern group, indicating a recent increase in population size or selection pressure. This evidence for selection signatures at *OsLEA27* in this study sample provides insight into the roles of selection, crop adaptation, and genetic diversity in establishing present-day variation at the *OsLEA27* locus.

## Keywords

Dehydrine, Linkage Disequilibrium (LD), Nucleotide Diversity, Haplotype, Selection Signature, Upland Rice, Dryland Rice

---

\*Corresponding author.

## 1. Introduction

Rice (*Oryza sativa* L.) is well adapted to semiaquatic conditions but highly sensitive to water shortage during its growth and pre-reproductive period. Adaptation to water-shortage conditions is a desirable trait, especially for farmers from drought-prone areas. Studies of rice from drought-prone regions would open avenues to understanding of the influence of selection on the genetic variation found in such materials.

In Myanmar, rice landraces are generally grown in two different ecosystems: upland and dryland. The term “dryland rice” refers to landraces or local varieties that have high tolerance to water deficiency whereas “upland rice” refers to those landraces or local varieties having fair-to-good levels of drought tolerance. Since each of these ecotypes has distinct characteristics, landraces or local varieties from both upland and dryland regions are valuable resources for future utilization as germplasm and for study of drought tolerance. Ecosystems, morphological variations and complex physiological responses, local conditions, and poor cultural practices and management have played a vital role in shaping the diversity patterns of drought tolerance genes through natural selection [1]-[3].

During dehydration, cells undergo great physical and chemical changes, which may become irrecoverable and ultimately lead to cell death. Desiccation tolerance is a trait associated with LEA proteins [4]. Desiccation-tolerant species synthesize group II LEA proteins known as dehydrin proteins [5] [6], RAB (responsive to ABA), or D-11. Dehydrins are induced by abiotic stress. Under conditions of dehydration, the presence of dehydrin proteins causes deformation of the cell and prevents it from rupturing during drastic fluctuations in plant water content [7].

Wang *et al.* identified eight dehydrin genes, *OsLEA22* to *OsLEA29*, with bioinformatics approaches and performed expression analysis in rice. Expression of *OsLEA27* and *OsLEA28* was up-regulated by drought treatment as revealed by microarray analysis. However, the expression of *OsLEA28* under both the normal conditions and the treatments could not be detected by sqRT-PCR [8]. Therefore, we used the *OsLEA27* gene to study the role of genetic variation in Myanmar upland and dryland rice landraces.

The regulation of dehydrin gene expression in rice has been extensively studied, and cis elements in the dehydrin promoters have been identified because dehydrin genes are important for tolerance to abiotic stress. An understanding of the nucleotide diversity and selection signatures (*i.e.*, evidence of selection) in the *OsLEA27* gene within upland and dryland rice landraces would provide information to know the molecular nature of adaptation from the view point of population and evolutionary genetic. The contribution of selection in shaping the genetic variation among individuals in landraces would provide valuable information because some kind of selection events associated with adaptation and the evolution of the new profile of species [9]-[11]. For these reasons, focusing on the role of selection in shaping the patterns of diversity in *OsLEA27* may be also one approach to know an evolution of drought tolerance traits (adapting to domestication or strong selection pressure) within Myanmar upland and dry-land rice landrace. The objectives of this study were to investigate nucleotide diversity and linkage disequilibrium (LD) and to examine the possible factors that influence selection on the *OsLEA27* gene within Myanmar upland and dryland rice landraces. Additionally, tracking along the genome segment of *OsLEA27* gene to detect signs that potentially indicate the action of natural or artificial selection within specific drought faced population is the main theme of this study.

## 2. Materials and Methods

### 2.1. Materials and Phenotyping

A total of 35 Myanmar rice landraces were used in the present study. These included 27 locally wide adapted landraces from upland and dryland areas (seven accessions from the central dry zone regions and the remaining 20 from northern and western upland regions) and accepted from Department of Agriculture, Myanmar. Phenotyping of drought adaptable traits was conducted in Plant Biotechnology Center, Myanmar at summer season. These 27 accessions were averagely assumed as an intermediate level of tolerance to drought accordingly to their panicle harvest index; 60 and above 60 percentage and yield per plant; about 8.4 g, by comparing with the lowland cultivar during the treatment of drought condition. Moreover, eight rainfed and irrigated local cultivars; five high-eating quality rice cultivars (panicle harvest index; 40) and three traditional local cultivars (panicle harvest index; less than 40) from the Ayeyarwadee delta were added (**Supplementary Table 1**).

## 2.2. Genomic DNA Extraction, PCR Amplification and Sequencing

The seeds were acclimatized at room temperature for 2 days, and then seed dormancy was broken by placing the seeds in an oven at 50°C for 5 days. Plants of each accession were grown in a controlled environment of 28°C and 14-h light conditions. At three week old seedling stage, DNA was extracted by using a modified CTAB method [12] and used for PCR amplification of the genomic region of *OsLEA27*. To amplify a 2.9-kbp region containing the promoter and transcribed region of *OsLEA27*, the six primer sets (Supplementary Table 2) were designed based on Os11g26760.1 locus (genbank NC008404.2, 17154224 to 17157109) using Primer 3 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>). Amplifications were carried out as one initial denaturation step at 95°C for 5 min, followed by 40 cycles of denaturation step at 95°C for 30 s, primer annealing at 55°C - 60°C for 30 s, and primer extension at 72°C for 30 s.

Sub-cloning of PCR products and transformation of competent *E. coli* were done according to the pGEM-T Easy Vector System Technical Manual (TM042) by using a T-Vector cloning kit (A1360; Promega Corporation, Madison, WI, USA). DNA sequencing was performed using an ABI PRISM Genetic Analyzer 3130 (Applied Biosystems Japan, Tokyo, Japan) according to the manufacturer's instructions.

## 2.3. Sequence Alignment and Statistical Analysis

The *OsLEA27* gene sequences from 35 rice accessions, Nipponbare and Kasalath were aligned and compared using BioEdit software v.7.0.9.0 [13]. Based on the pairwise differences, a neighbor-joining bootstrap consensus tree was constructed using MEGA v.4.1 [14]. The proportion of segregating sites ( $\theta = 4Ne\mu$ , where  $Ne$  is the effective population size and  $\mu$  is the per-generation mutation rate [15], the average number of pairwise nucleotide differences among sequences ( $k$ ) [16], the average number of pairwise nucleotide substitutions per site between two sequences (*i.e.*, the nucleotide diversity  $\pi$ ) [16], the number of polymorphic sites ( $S$ ), the number of different haplotypes ( $H$ ), and the haplotype diversity ( $Hd$ ) [17] were computed by DnaSP v5 [18]. Haplotype networks, representing unique alleles separated by mutational steps, were constructed using the Network 4.5.1.6 program (<http://www.fluxus-technology.com/sharenet.htm>) with the median-joining method [19]. LD within and upstream of the *OsLEA27* gene was measured as  $r^2$  (squared allele frequency correlation) [20] using DnaSP v5 and TASSEL v2.1 [21]. Fisher's exact test and  $\chi^2$  test implemented in the DnaSP v5 program were used to calculate the  $P$  values for all pairwise comparisons of polymorphic sites within and upstream of the *OsLEA27* gene. Tajima's  $D$  [22] and Fu & Li's  $D^*$  and  $F^*$  [23] were also calculated with DnaSP for testing deviation from neutrality. To understand whether different regions were subjected to different evolutionary forces, the neutrality tests were applied by a sliding-window method. Test statistics were analyzed on windows of 250 bp with a step size of 25 bp by using the function implemented in DnaSP v5.

## 3. Results

### 3.1. Patterns of DNA Polymorphism in Various Regions of *OsLEA27*

We began by attempting to sequence the entire *OsLEA27* region (*i.e.*, the gene locus region plus the 5' promoter region) in 35 rice accessions. Five of the 35 accessions failed to show amplification of the 5' promoter region. Therefore, 35 sequences were used for gene locus region (890 bp) analysis but only the 30 sequences representing the entire *OsLEA27* region (2885 bps) were used to search for selection signatures and for LD analysis.

Sequence analysis in the *OsLEA27* gene locus (NC008404.2 (17,154,224 to 17,155,114 bp)) of the 35 accessions showed 49 single-nucleotide polymorphisms (SNPs; 7 singletons and 42 parsimony informative sites) variations. The frequency distribution of SNPs was 1 SNP per 17 bp and  $\pi$  was 0.0121 (Table 1). Highly variable sites and three indels were found in the intron region. Haplotype diversity ( $Hd$ ) was 0.818, with 12 haplotypes (*i.e.*,  $H = 12$ ). In the coding region, there were seven segregating sites (four nonsynonymous and three synonymous) and no indels, with lower nucleotide diversity ( $\pi = 0.0033$ ) than in the noncoding transcribed regions ( $\pi = 0.0285$ ) (Table 2). The ratio of  $\pi_{\text{nonsyn}}/\pi_{\text{syn}}$  was less than 1 in coding region ( $n = 35$ ).

In the 5' upstream region of *OsLEA27* (1.9 kbp; NC008404.2 (17,155,115 to 17,157,109)), which could be analyzed in 30 of the 35 accessions, there were 30 variable sites and  $\pi$  of 0.0029.  $Hd$  was 0.825 and  $H$  was 12. The average SNP frequency distribution was 1 SNP per 66 bp. Within the entire *OsLEA27* region, 78 SNPs were found, with  $\pi = 0.00592$ . A higher level of nucleotide variation was found in the *OsLEA27* gene locus ( $\pi = 0.0136$ ,  $\theta = 0.01481$ ) than in the 5' upstream region ( $\pi = 0.00292$ ,  $\theta = 0.00387$ ) (Table 1). In a sliding-window

**Table 1.** Comparison of population and genetic parameters between the 5' upstream region, *OsLEA27* locus, and entire *OsLEA27* region.

Data	35 landrace accessions ( <i>OsLEA27</i> gene locus)		30 landrace accessions	
		5' upstream region	<i>OsLEA27</i> gene locus	Entire <i>OsLEA27</i> region
Length (bp)	890	1995	890	2885
Nucleotide diversity ( $\pi$ )	0.0121	0.0029	0.0136	0.0059
No. of polymorphic sites ( <i>S</i> )	49	30	48	78
No. of haplotypes ( <i>H</i> )	12	12	11	16
Haplotype diversity ( <i>Hd</i> )	0.818	0.825	0.841	0.899
Watterson's nucleotide diversity ( $\theta$ )	0.01454	0.00387	0.01481	0.00683
Tajima's <i>D</i>	-0.5958 (ns)	-0.8864 (*)	-0.3047 (ns)	-0.5462 (ns)

\*Significant  $p < 0.05$ ; (ns) non-significant.

**Table 2.** Summary of DNA variation in the entire region of *OsLEA27* gene locus of 30 landraces.

No	Specific region	L (bp)	<i>S</i>	$\pi$	$\theta$	<i>D</i>	<i>D</i> *	<i>H</i>	<i>Hd</i>
1	Entire region	2885	78	0.00592	0.00683	-0.5462	0.42361	16	0.899
2	Coding region	495	7	0.00335	0.00357	-0.17920	-0.81656	7	0.752
3	5' UTR, intron and 3' UTR	395	41	0.0285	0.03044	-0.31535	1.14215	8	0.789
4	Intron region	95	37	0.1079	0.1128	0.1579*	1.6**	4	0.651
5	5' upstream region	1995	30	0.00292	0.00387	-0.8864*	-0.2576*	12	0.825

L, length; *S*, number of polymorphic sites;  $\pi$ , nucleotide diversity;  $\theta$ , Watterson's  $\theta$ ; *H*, no of haplotypes; *Hd*, haplotype diversity; *D*, Tajima's *D*; *D*\*, Fu & Li's *D*\*. \*Significant at  $p < 0.05$  and \*\*significant at  $p < 0.01$ .

analysis along the entire *OsLEA27* region, one prominent peak of nucleotide variation was found in the intron region. No notable peaks were detected in the 5' upstream region (**Supplementary Figure 1**).

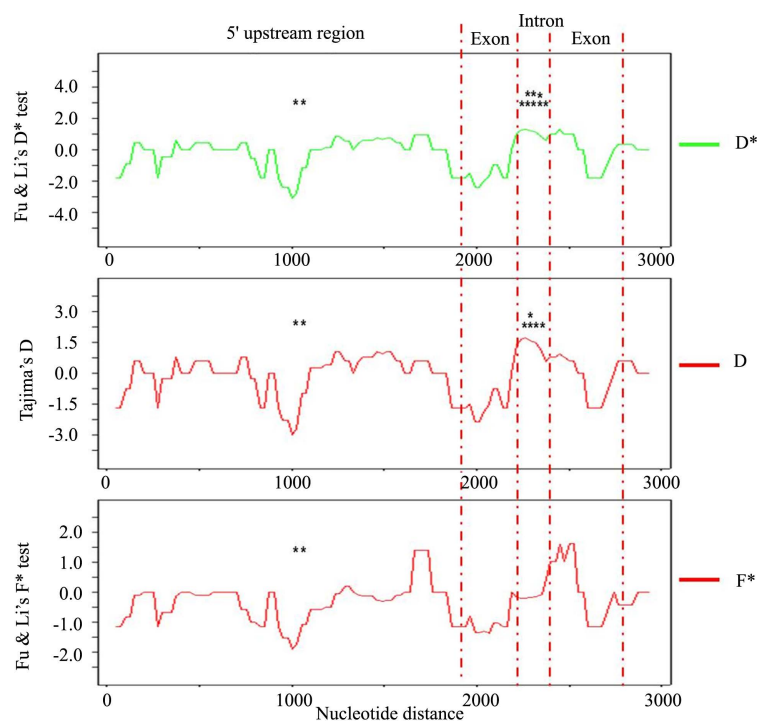
When the landraces were grouped by geographical region into four subpopulations, the western upland (hereafter, "western") group showed the highest value of nucleotide diversity ( $\pi = 0.0127$ ), followed by the central dryland ("central"), northern upland ("northern"), and delta (rainfed and irrigated groups) (**Table 3**). Similarly, the average number of pairwise nucleotide differences (*k*) within each subpopulation followed the same trend. The peak position for  $\pi$  was identical in the 30 accessions for which the entire *OsLEA27* region could be analyzed, and divergence between accessions from the western group (highest value of nucleotide diversity ( $\pi$ ); **Table 3**) and a combined group consisting of the remaining accessions in sliding-window analysis (**Supplementary Figure 1**). This finding implied that there were no significant nucleotide variations among landraces. Thus, the 30 landraces could not be classified by geographic region based on nucleotide sequence variation in the *OsLEA27* region.

### 3.2. Tests of Neutrality in the Entire *OsLEA27* Region

Tajima's *D*, Fu & Li's *D*\*, and *F*\* were used to detect drift from the standard neutral model of evolution. The significant negative value and positive value of *D* and *D*\* was found in 5' upstream region ( $D = -0.8864$  ( $p < 0.05$ ),  $D^* = -0.2576$  ( $p < 0.05$ )) and intron region ( $D = 0.1579$  ( $p < 0.05$ ),  $D^* = 1.6$  ( $p < 0.01$ )) within the set of 30 accessions (**Table 1** and **Table 2**). Although the frequency of two observed alleles was greater than the expected frequency within 15 segregating sites (**Supplementary Figure 2**), the Tajima's *D*, Fu & Li's *D*\* test and Fu & Li's *F*\* test did not detect any deviation from neutrality in the entire *OsLEA27* region.

In contrast, sliding-window analysis of the mutation-drift equilibrium tests also showed that value of three neutrality tests were significant and were detected in the 5' upstream region. Moreover, positive values of *D*\* that were significant at the 0.05 and 0.01 levels were detected in the intron region (**Figure 1**). The trends of *D*, *D*\*, and *F*\* were not uniform throughout the entire sequence because they are each calculated in slightly different

ways. When analyzed by geographic location, accessions from the northern groups significantly deviated from selective neutrality ( $p < 0.05$ ) for all three test statistics (Table 3).



**Figure 1.** Sliding-window analysis of neutrality tests along the entire *OsLEA27* gene: (a) Fu & Li's  $D^*$ , (b) Tajima's  $D$ , and (c) Fu & Li's  $F^*$ . The window size is 250 bp, and step size is 25 bp. Confidence intervals at 95% significance levels for each of the three tests are shown for a sample size of 30. Stars indicate the center of windows showing a significant statistic. Significance levels are shown as \* (0.05 level) and \*\* (0.01 level).

**Table 3.** Population variability parameters and genetic parameters for the entire *OsLEA27* region.

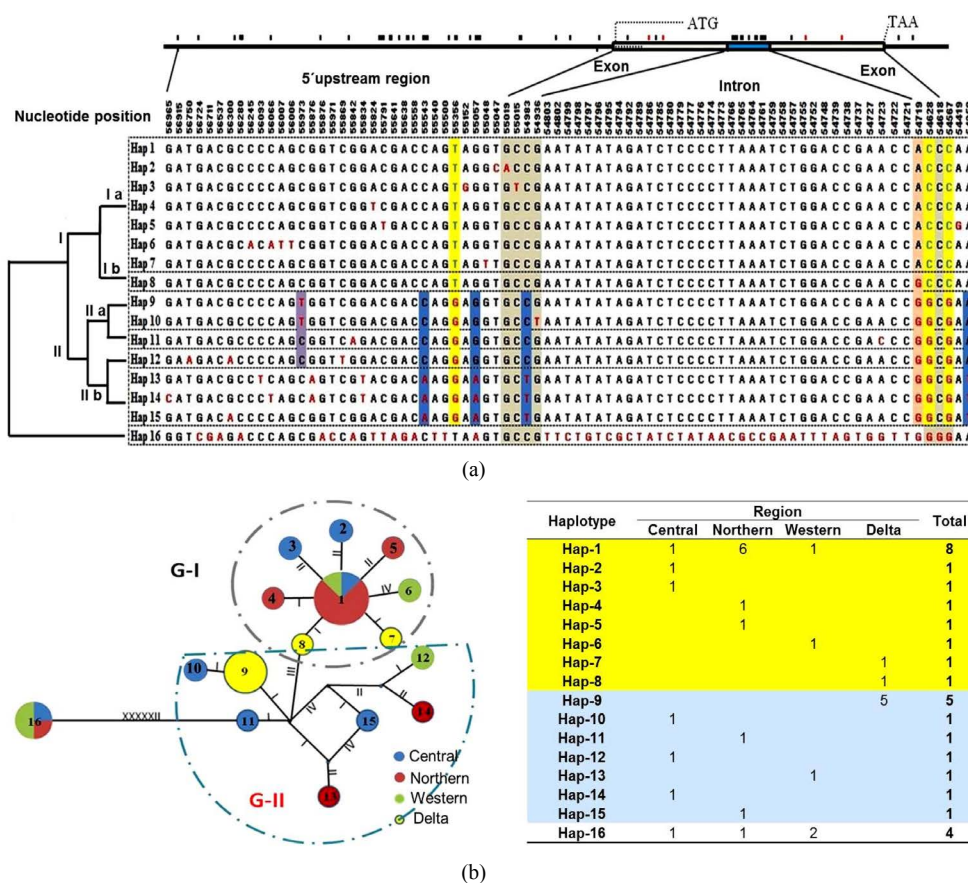
No.	Genetic parameter	Central	Northern	Western	Delta	Total
1	No. of samples ( $n$ )	7	11	5	7	30
2	Segregating sites ( $S$ )	71	61	66	6	78
3	Singletons ( $s$ )	66	53	11	2	15
4	Parsimony informative sites	5	8	55	4	65
5	Indel sites ( $I$ )	10	10	11	2	12
<b>Population statistic</b>						
6	No. of haplotype ( $H$ )	7	6	4	3	16
7	Nucleotide diversity ( $\pi$ )	0.00750	0.00442	0.01278	0.00085	0.00592
8	Haplotype diversity ( $Hd$ )	1	0.727	0.9	0.524	0.899
9	Watterson's $\theta$	0.01005	0.00722	0.01083	0.00084	0.00683
10	Tajima's $D$ ( $D$ )	-1.47723	-1.88962*	1.36718	0.05645	-0.54620
11	Fu & Li's $D^*$ test ( $D^*$ )	-1.60550	-2.16668*	-1.36718	0.43876	0.42361
12	Fu & Li's $F^*$ test ( $F^*$ )	-1.74653	-2.37912*	-1.48346	0.38497	0.11993

\*Significant at  $p < 0.05$ .

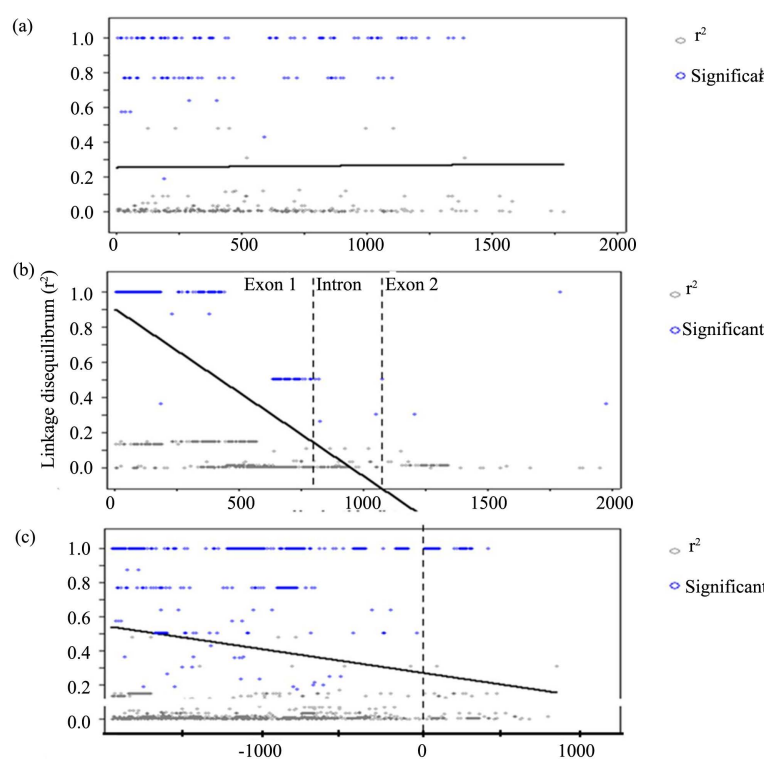
### 3.3. Haplotype Diversity and Haplotype Network Analyses of the Entire *OsLEA27* Region

Analysis of the entire *OsLEA27* region revealed 16 haplotypes (Figure 2). Thirteen (81%) of the haplotypes were represented by a single landrace; these were designated as “private haplotypes”. The highest *Hd* was found in landraces from the central region (*Hd* = 1, indicating that each one had a unique haplotype) and the lowest diversity was observed in accessions from the delta region (*Hd* = 0.524) (Table 3). Overall, the 16 haplotypes contained a total of 78 variable sites across all the accessions, which we further subjected to network analysis.

The 16 haplotypes were categorized into two groups (I and II) by three parsimony informative sites (T55356G, C54628G, and C54567G). Group I was again divided into subgroups Ia (seven haplotypes) and Ib (Hap 8 only). Group II also comprised two subgroups (IIa and IIb) on the basis of four parsimony informative sites (C55543A, A55057G, T54983C, and T54258A) (Figure 3(a)). Group II contained six private haplotypes originating from three different regions along with one unique haplotype (Hap 9) represented only by accessions from the delta region. Hap 16 formed a separate group, distinguished from the two major haplotype groups by a high number of polymorphic sites. Approximately 50% and 37% of all the accessions had haplotypes from groups I and II, respectively, with the remaining 13% belonging to Hap 16 (Figure 2(b)). Accessions from group I have high



**Figure 2.** Nucleotide polymorphisms among haplotypes and haplotype network. (a) Position of Nucleotide polymorphisms among haplotypes at the entire region of *OsLEA27*. The exact position of substitutions on NC008404.2 (17154224 -17157109 bp) are indicated on the top base. Upper horizontal line represents gene structure of *OsLEA27*; short black and red vertical lines represent SNPs and replacement sites, respectively. Colored shading indicates parsimony informative sites, which differentiated the haplotype groups. Haplotypes were based on 78 SNPs. (b) Haplotype network at the entire region of *OsLEA27*. Each haplotype is represented by a circle; the number in each circle is the haplotype number, and its size reflects the number of accessions with that haplotype. Dashed shapes encircle the major haplotype groupings. Roman numerals on the haplotype network figure indicate the number of SNP sites between the connected haplotypes. Number of landraces from each geographic region is shown in table at right. Yellow color in the table represent Group I haplotypes and blue is Group II.



**Figure 3.** Scatterplot diagram of pairwise distance and LD in three regions of *OsLEA27*: (a) 5' upstream region, (b) *OsLEA27* gene locus, and (c) entire region of *OsLEA27* gene. LD based on  $r^2$  between all SNPs in the 5' upstream region, *OsLEA27* gene locus, and entire region of *OsLEA27* in 30 landraces. The logarithmic trend lines represent the average LD decay.

performance of drought adaptable trait (panicle harvesting index; PHI% = about 70% or over 70%) except two accession AC-33 and AC-31 from delta region (**Supplementary Table 1**). Members of both major haplotype groups were found in all four geographic regions.

### 3.4. Linkage Disequilibrium in the Entire *OsLEA27* Region

Linkage disequilibrium (LD) was studied by calculating the p value of Fisher's exact test for all pairwise comparisons of polymorphic sites. **Figure 3** shows the decay of LD with increasing nucleotide pairwise distance for 30 rice accessions in the 5' upstream region (**Figure 3(a)**), the *OsLEA27* gene locus (**Figure 3(b)**), and the entire *OsLEA27* region (**Figure 3(c)**). For the 78 polymorphic sites in the entire *OsLEA27* region, there were 2926 pairwise comparisons: 1318 (45%) showed significant LD between pairs of polymorphic sites. This percentage was higher than for the upstream region (27%) and lower than for the *OsLEA27* gene locus (59%) (**Table 4**). The higher percentage of significant pairwise comparisons in the *OsLEA27* gene locus relative to other regions is noteworthy and might represent a real difference from normal LD decay. Even more striking was the low frequency of significant comparisons in the 5' upstream region, normal decay of LD was detected along the entire *OsLEA27* region.

The value  $r^2$  is the squared allele frequency correlation to fit the logarithm curve across distances for all SNPs loci examined. In the 5' upstream region, the  $r^2$  value for LD appeared to increase slightly with increasing distance (**Figure 3(a)**). In the *OsLEA27* gene locus (890 bp), rapid decay of LD was found between 400 bp and 450 bp. Almost all of the significant pairwise comparisons (640 of the 1081 tests) were scattered within exon 1 and the intron region (**Figure 3(b)**).

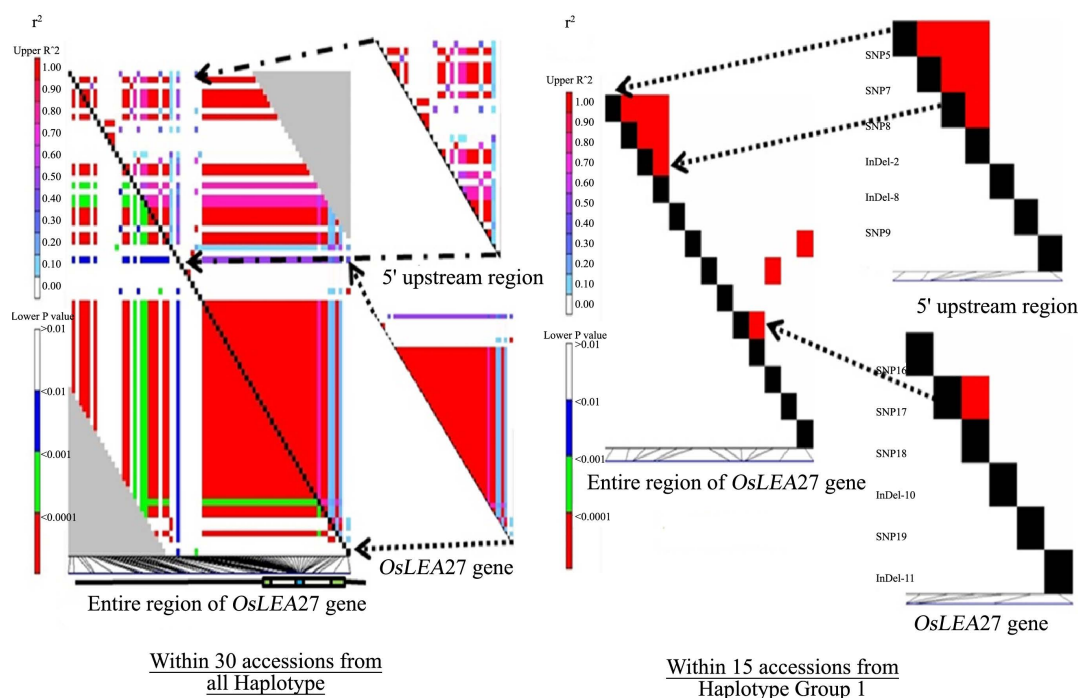
The occurrence of rapid decay LD in *OsLEA27* locus depends on the sample size and its population history. To investigate LD decay along the complete *OsLEA27* region, 32 rice accessions (the 30 with available 5' region sequence, *japonica* control; Nipponbare and *indica* control; Kasalath) were used for phylogenetic analysis.

Four accessions were grouped with Nipponbare as a separate group. Fifteen accessions from group I haplotypes were clustered with Kasalath (data not shown). LD along the 5' upstream and *OsLEA27* locus was determined by  $\chi^2$  tests for the 30 accessions. In the group I haplotypes composed with fifteen accessions, the numbers of polymorphic sites in the 5' upstream and *OsLEA27* locus region were seven and six, respectively. Six of the 21 tests within the 5' upstream region were significant, whereas one out of 15 in the *OsLEA27* gene locus region was significant (Figure 4). These results indicated that the large number of significant pairwise comparisons within the *OsLEA27* locus region was caused by the dimorphic pattern (i.e., the clustering into two distinct groups).

## 4. Discussion

### 4.1. Nucleotide Diversity, Phylogenetic Analysis, and Haplotype Networks in a Group of Myanmar Landraces

In general, selection, demography, effective population size, and mating are the factors that affect SNP frequency and nucleotide diversity, two important measures of genetic variation [17]. In Myanmar upland and dryland rice landraces, a lower level of nucleotide variation was found in the 5' upstream region of the *OsLEA27* gene ( $\pi = 0.00292$ ,  $\theta = 0.00387$ ) than in the *OsLEA27* gene locus region. Human selection may act on regulatory elements in the 5' upstream region that affect the transcription of *OsLEA27*, whereas purifying selection may affect



**Figure 4.** Tests of linkage disequilibrium between upstream region and *OsLEA27* gene locus of 30 landraces and 15 accessions from haplotype group I with chi-square ( $\chi^2$ ) test. Number and percent of significant comparisons tests at three regions of *OsLEA27* gene between the accessions from Group I haplotypes and whole accessions are shown table.

**Table 4.** Assessment of linkage disequilibrium with Fisher’s test.

Region	Length (bp)	No. of polymorphic sites	Fisher’s test		
			No. of pairwise comparison tests	No. of significant comparisons	Significant percentage
5' upstream region	1995	30	435	117	27%
<i>OsLEA27</i> gene locus	890	48	1081	640	59%
Entire <i>OsLEA27</i> region	2885	78	2926	1318	45%



the coding region (which had a low level of variation and  $\pi_{\text{nonsyn}}/\pi_{\text{syn}} < 1$ ). High nucleotide variation in the intron region was found in the *OsLEA27* gene locus and consequently affected the phylogenetic analysis. These segregating sites in the intron played a role in grouping the studied samples and shaped the genetic variation towards selection. The observed phylogenetic relationships and haplotype network analysis indicated mixing of the *japonica* subspecies into the predominantly *indica* gene pool of upland and dryland landraces.

Considering the entire *OsLEA27* region, two distinct haplotype groups are found in rice from several distinct geographic areas, providing further evidence for directional selection. Hap 1 is the predominant haplotype due to its high frequency in accessions from northern upland, central dryland, and western upland regions. Accessions from Hap 1 showed high performance in testing the drought adaptable trait (Panicle harvesting index; PHI% = above 70%) (**Supplementary Table 1**). The predominance of this haplotype seems to be the consequence of strong selection.

Even if clustering were totally independent of geographic origin, we would need to consider an uneven level in adaptation to stress conditions within geographic groups. Generally, rice has adapted to humans extending its range of cultivation and changes in several environmental features, and the effects of selection by humans would be one possible and important factor [24].

#### 4.2. Tracking Linkage Disequilibrium along the Entire *OsLEA27* Region

The LD curve representing the 5' upstream region suggests that there may have been a decrease in effective population size, leading to an increase in inbreeding. These two possible events can be correlated with domestication and a population bottleneck. It is interesting to note that the increase in LD in this region pointed to consider impact of selection pressure. The lower level of nucleotide diversity in the 5' upstream region and the higher LD showed the possible effects of selection, associated with environmental factors and/or human interaction. These results depict the correlation between nucleotide diversity and LD, which may provide a means to search for evidence of selection in particular parts of the *OsLEA27*; 5' upstream regions [25]. Because the actual force of selection may depend on the expression level of *OsLEA27*, the promoter of which is active under abiotic stress [26].

When we analyzed the complete set of 35 rice accessions, dimorphism from wide geographical regions also showed that LD decay occurred rapidly. Rapid decay of LD within the *OsLEA27* gene locus depends on its population history. A wide geographic sample of germplasm has more chances to decay rapidly in LD [27].

#### 4.3. Factors Affecting the Evidence of Selection Detected with Neutrality Tests

In a standard neutrality test, Tajima *D*, Fu & Li's  $D^*$  and Fu & Li's  $F^*$  tests in the entire *OsLEA27* region failed but 5' upstream region was significant with negative value which indicate a recent expand in population size or selection pressure. That significant value area was 5' upstream region where cis-regulatory element be there; such regions facilitate transcription of a particular gene for functional adaptation.

Significant positive values of Fu & Li's  $D^*$  test were found at an intron. The  $D^*$  test statistic is based on the distribution of mutations in the genealogy and compares the number of old and new mutations [28]. These results showed an excess of intermediate allelic frequency in intron-splicing sites. We support here that recent admixture between two highly divergent subpopulations may shape the diversifying selection model. In sliding-window analysis, sites responsible for significant negative and positive test values were identified in the promoter region and intron, respectively.

Accessions from the northern group deviated from mutation-drift equilibrium because of recent population size expansion or a recent selective sweep, as evidenced by the significant negative value of Tajima's *D*. We speculate here that individuals from several subpopulations intermated and that new haplotypes appeared in their progeny. The adaptation of rice, typically a semiaquatic species, to mountainous areas with limited water may be due in part the effect of the *OsLEA27* gene, its promoter region, and/or the effects of these regions on one another.

#### 4.4. Detailed Analysis of Selection Signatures within Myanmar Upland and Dryland Landraces from Different Geographic Groups

One target of this study was to analyze the selection signatures within landrace accessions from diverse geo-

graphical regions. Even though no significant deviations from neutrality were initially observed with three neutrality tests in the entire *OsLEA27* region, purifying selection was observed in the coding region. In fact, most protein-coding genes have evolved under purifying selection of varying strength [29]. Therefore, purifying selection in the transcribed region was not sufficient evidence of artificial selection. We were interested to learn whether analysis of the promoter region would give results opposite to those obtained for the transcribed region. Sequence variation in promoter regions results in differential regulation of transcription [30] [31]. Therefore, we assume that the increased LD and significant of neutrality tests in the promoter region of *OsLEA27* has resulted from the effects of artificial selection on that area, resulting in decreased recombination in that region.

Furthermore, the presence of highly variable sites in the intron with positive significant values of Fu & Li's  $D^*$  indicates mixing of two highly divergent subpopulations. A dimorphism effect, *i.e.*, mixing of the *japonica* and *indica* gene pools, was detected in the intron-splicing region. In my unpublished works of NCBI BLAST search showed that 11% of intron-splicing region sequences were nearly identical to *japonicaOsLEA27* and 88% were similar to *indicaOsLEA27*. If Paw San group accessions from the delta region were the same accession identified as *japonica* type (**Supplementary Table 1**), which have similar fragment as *indicaOsLEA27* according to a BLAST search, these results might suggest that there is diversifying selection between the *japonica* and *indica* sequence types. That haplotype dimorphism effect related with the selection events on candidate genes during the domestication of one or both of two subspecies [32].

The selection signatures in landraces from the northern group revealed significant positive selection. The mountainous ecosystem is natural selector for upland landraces, which remain under continuous selection pressure on slope areas that are unable to retain water. We speculate that natural selection was followed by artificial selection, which fixed the alleles of *OsLEA27* that were common to the landraces from the northern group. Therefore, the characteristics of this study sample may result from intermating of individuals from different subpopulations and/or artificial selection imposed on this landrace group. For landraces from the western and central regions, more detailed studies are still needed to make firmer conclusion.

## 5. Conclusion

In this study, we have analyzed nucleotide diversity and selection signatures by using trait-oriented sequence analysis. The pattern of variation within *OsLEA27* revealed weak selection pressure along the entire region. However, the significant excess of low allelic frequency in the northern group suggested that the haplotype structure and population differentiation had been caused by the action of positive natural selection. The high variability observed in the intron region may represent a signature of diversifying selection within these upland and dryland landraces that are very unique genotypes. Additionally, although this study was based on the 2.9 kbp genome segment of one specific gene, outcome information provided the powerful proof of concept for evidence of selection within specific population. The results of selection signature analysis of the *OsLEA27* locus enhance our understanding of the roles of selection, crop adaptation and genetic diversity on the variability of *OsLEA27* in Myanmar landraces and will motivate further study of this locus and its association with drought tolerance.

## Acknowledgements

This research was supported in part by Grants-in-Aid #21405017 & #25257416 of the Japan Society of Promotion of Sciences. The germplasm was provided by the Ministry of Agriculture and Irrigation of the Republic of the Union of Myanmar under SMTA of FAO IT PGR FA. The authors thank Drs. Khing Maung Thet and Pa Pa Aung for their assistance in obtaining germplasm information.

## References

- [1] Arrauudeau, M.A. (1995) Upland Rice: Challenges and Opportunities in the Less Favorable Ecosystem. *Geojournal*, **35**, 325-328. <http://dx.doi.org/10.1007/BF00989139>
- [2] Khush, G.S. (1997) Origin, Dispersal, Cultivation and Variation of Rice. *Plant Molecular and Biology*, **35**, 25-34. [http://dx.doi.org/10.1007/978-94-011-5794-0\\_3](http://dx.doi.org/10.1007/978-94-011-5794-0_3)
- [3] Tun, Y.T. (2006) Characterization of Myanmar Rice Landraces Based on Analyses of Genetic Diversity of Various Traits. Ph.D. Thesis, Tokyo University of Agriculture, Tokyo, 87-94.

- [4] Hoekstra, F.A., Golovina, E.A. and Buitink, J. (2001) Mechanisms of Plant Desiccation Tolerance. *Trends in Plant Science*, **6**, 431-438. [http://dx.doi.org/10.1016/s1360-1385\(01\)02052-0](http://dx.doi.org/10.1016/s1360-1385(01)02052-0)
- [5] Baker, J., Steel, C. and Dure III, L. (1988) Sequence and Characterization of 6 LEA Proteins and Their Genes from Cotton. *Plant Molecular Biology*, **11**, 277-291. <http://dx.doi.org/10.1007/bf00027385>
- [6] Collett, H., Shen, A., Gardner, M., Farrant, J.M., Denby, K.J. and Illing, N. (2004) Towards Transcript Profiling of Desiccation Tolerance in *Xerophyta humilis*: Construction of a Normalized 11 k X. Humilis cDNA Set and Microarray Expression Analysis of 424 cDNAs in Response to Dehydration. *Physiologia Plantarum*, **122**, 39-53. <http://dx.doi.org/10.1111/j.1399-3054.2004.00381.x>
- [7] Layton, B.E., Boyd, M.B., Tripepi, M.S., Bitonti, B.M., Dollahon, M.R. and Balsamo, R.A. (2010) Dehydration-Induced Expression of a 31-kDa Dehydrin in *Polypodium polypodioides* (Polypodiaceae) May Enable Large, Reversible Deformation of Cell Walls. *American Journal of Botany*, **97**, 535-544. <http://dx.doi.org/10.3732/ajb.0900285>
- [8] Wang, X.S., Zhu, H.B., Jin, G.L., Liu, H.L., Wu, W.R. and Zhu, J. (2007) Genome-Scale Identification and Analysis of LEA Genes in Rice (*Oryza sativa* L.). *Plant Science*, **172**, 414-420. <http://dx.doi.org/10.1016/j.plantsci.2006.10.004>
- [9] Li, C., Zhou, A. and Sang, T. (2006) Rice Domestication by Reducing Shattering. *Science*, **311**, 1936-1939. <http://dx.doi.org/10.1126/science.1123604>
- [10] Sweeney, M.T., Thomson, M.J., Cho, Y.G., Park, Y.J., Williamson, S.H., Bustamante, C.D. and McCouch, S.R. (2007) Global Dissemination of a Single Mutation Conferring White Pericarp in Rice. *PLoS Genetics*, **3**, 1418-1424. <http://dx.doi.org/10.1371/journal.pgen.0030133>
- [11] Olsen, K.M., Caicedo, A.L., Polato, N., McClung, A., McCouch, S. and Purugganan, M.D. (2006) Selection under Domestication: Evidence for a Sweep in the Rice Waxy Genomic Region. *Genetics*, **173**, 975-983. <http://dx.doi.org/10.1534/genetics.106.056473>
- [12] Weising, K., Kaemmer, D., Weigand, F., Epplen, J.T. and Kahl, G. (1991) Oligonucleotide Fingerprinting Reveals Various Prove-Dependent Levels of Informativeness in Chickpea (*Cicer arietinum*). *Genome*, **35**, 436-442. <http://dx.doi.org/10.1139/g92-065>
- [13] Hall, T.A. (1999) BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.
- [14] Tamura, K., Dudley, J., Nei, N. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599. <http://dx.doi.org/10.1093/molbev/msm092>
- [15] Watterson, G.A. (1975) On the Number of Segregating Sites in Genetical Models without Recombination. *Theoretical Population Biology*, **7**, 256-276. [http://dx.doi.org/10.1016/0040-5809\(75\)90020-9](http://dx.doi.org/10.1016/0040-5809(75)90020-9)
- [16] Tajima, F. (1983) Evolutionary Relationship of DNA Sequences in Finite Populations. *Genetics*, **105**, 437-460.
- [17] Nei, M. (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.
- [18] Librado, P. and Rozas, J. (2009) DnaSP V5: A Software for Comprehensive Analysis of DNA Polymorphism Data. *Bioinformatics*, **25**, 1451-1452. <http://dx.doi.org/10.1093/bioinformatics/btp187>
- [19] Bandelt, H., Forster, P. and Rohl, A. (1999) Median-Joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology and Evolution*, **16**, 37-48. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026036>
- [20] Hill, W.G. and Robertson, A. (1968) Linkage Disequilibrium in Finite Populations. *Theoretical and Applied Genetics*, **38**, 226-231. <http://dx.doi.org/10.1007/BF01245622>
- [21] Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M. and Ramdoss, Y. (2007) TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics*, **23**, 2633-2635. <http://dx.doi.org/10.1093/bioinformatics/btm308>
- [22] Tajima, F. (1989) Statistical Methods for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, **123**, 585-595.
- [23] Fu, Y.X. and Li, W.H. (1993) Statistical Tests of Neutrality of Mutations. *Genetics*, **133**, 693-709
- [24] Ding, Z., Wang, C., Chan, S. and Yu, S. (2011) Diversity and Selective Sweep in *OsAMT1;1* Genomic Region of Rice. *BMC Evolutionary Biology*, **11**, 61. <http://dx.doi.org/10.1186/1471-2148-11-61>
- [25] Nielsen, R. (2005) Molecular Signatures of Natural Selection. *Annual Review of Genetics*, **39**, 197-218. <http://dx.doi.org/10.1146/annurev.genet.39.073003.112420>
- [26] Yang, Y., He, M., Zhu, Z., Li, S., Xu, Y., Zhang, C., Singer, S.D. and Wang, Y. (2012) Identification of the Dehydrin gene Family from Grapevine Species and Analysis of Their Responsiveness to Various Forms of Abiotic and Biotic Stress. *BMC Plant Biology*, **12**, 140. <http://dx.doi.org/10.1186/1471-2229-12-140>
- [27] Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F. and Gaut, B.S. (2001) Patterns of DNA Sequence Polymorphism along Chromosome 1 of Maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy*

- of Sciences of the United States of America*, **98**, 9161-9166. <http://dx.doi.org/10.1073/pnas.151244298>
- [28] Audigeos, D., Buonamici, A., Belkadi, L., Rymer P., Boshier, D., Scotti-Saintagne, C., Giovanni, G., Vendramin, G.G. and Scotti, I. (2010) Aquaporins in the Wild: Natural Genetic Diversity and Selective Pressure in the PIP Gene Family in Five Neotropical Tree Species. *BMC Evolutionary Biology*, **10**, 202. <http://dx.doi.org/10.1186/1471-2148-10-202>
- [29] Koonin, E.V. and Wolf, Y.I. (2010) Constraints and Plasticity in Genome and Molecular Phenome Evolution. *Nature Reviews Genetics*, **11**, 487-498. <http://dx.doi.org/10.1038/nrg2810>
- [30] Crawford, D.L., Segal, J.A. and Barnett, J.L. (1999) Evolutionary Analysis of TATA-Less Proximal Promoter Function. *Molecular Biology and Evolution*, **16**, 194-207. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026102>
- [31] Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T. and Yano, M. (2006) An SNP Caused Loss of Seed Shattering during Rice Domestication. *Science*, **312**, 1392-1395. <http://dx.doi.org/10.1126/science.1126410>
- [32] Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L., Li, J., He, W., Zhang, G., Zheng, X., Zhang, F., Li, Y., Yu, C., Kristiansen, K., Zhang, X., Wang, J., Wright, M., McCouch, S., Nielsen, R., Wang, J. and Wang, W. (2012) Resequencing 50 Accessions of Cultivated and Wild Rice Yields Markers for Identifying Agronomically Important Genes. *Nature Biotechnology*, **30**, 105-111. <http://dx.doi.org/10.1038/nbt.2050>

**Supplementary Table 1.** List of upland and dryland rice landraces in this study and their drought adaptable characters.

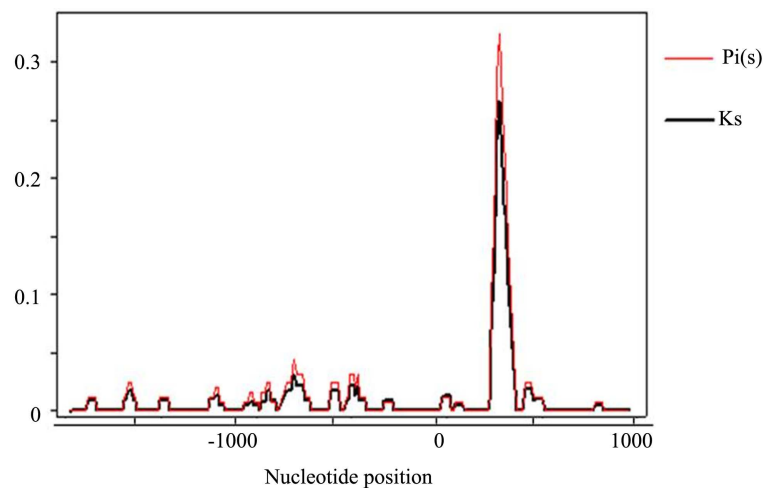
Code No.	Accessions name	Source	Ecotype	Type	Haplotype group	PH	RL	TN	SW	RW	PHI (%)	YPP	Remarks
AC1	Ma La Byat	Central; Sagaing	Dryland	Indica	I	91	18	13	21.8	17	<b>80</b>	26	-
AC2	TaungYar SABA	Central; Magway	Dryland	Indica	II	110.2	21	10	21	9.1	<b>60</b>	13	-
AC3	YaThay	Central; Magway	Dryland	Japonica	I	167.3	26.6	10	69.3	24.6	<b>65</b>	7.3	-
AC4	Chin Lay	Central; Magway	Dryland	Japonica	I	160.6	14.3	8	73.3	20.1	<b>80</b>	11.9	-
AC5	YaTha Pu	Central; Magway	Dryland	-	II	163.8	25.2	9	70.5	31.9	<b>60</b>	17.8	-
AC6	Sheik PhaPhu	Central; Mandalay	Dryland	Indica	II	132	23.5	19	61.2	18.3	<b>63</b>	54.2	-
AC7	Unknown	Central; Mandalay	Dryland	Japonica		104.2	15.8	13	49	18.5	<b>70</b>	20.4	-
AC8	Jin Ma Yin	Northern; Kachin	Upland	Japonica	II	127.6	19	5	24	11	<b>60</b>	9.4	-
AC9	La Chit Mum	Northern; Kachin	Upland	Indica	I	182.6	26.4	16	57.6	24	<b>70</b>	23.5	-
AC10	BwinAkyan	Northern; Kachin	Upland	Indica		107.6	14.2	10	25.1	16	<b>84</b>	18.4	*
AC11	Taung YarNgaCheik	Northern; Kachin	Upland	Indica	I	140.2	25	14	33.6	20.2	<b>80</b>	22.4	-
AC12	Do SABA	Northern; Kachin	Upland	Indica		98.2	16.8	14	32.8	14	<b>74</b>	17.4	*
AC13	Shaka Mam	Northern; Kachin	Upland	Indica	I	106.8	21	22	98.2	30	<b>80</b>	36.3	-
AC14	Padee Pu	Northern; Kachin	Upland	Japonica	I	92.2	16.4	10	27	15	<b>80</b>	9.7	-
AC15	Nyan Sa	Northern; Kachin	Upland	Indica	I	104.2	14.6	18	78.2	21.6	<b>80</b>	29.5	-
AC16	Kyaw San	Northern; Kachin	Upland	Indica	I	176.3	24.3	8	72.5	35.5	<b>73</b>	37.8	-
AC17	Sa Pa Ni	Northern; Kachin	Upland	Japonica		136	16.7	8	30	9.6	<b>60</b>	9	-
AC18	Maw Yaw KaukHnyin	Northern; Kachin	Upland	Indica	I	155.6	15.6	9	39	22	<b>80</b>	10.6	-
AC19	NaungHtu KaukHnyin	Northern; Kachin	Upland	Japonica	II	143.8	26.4	12	58.8	18	<b>60</b>	9.7	-
AC20	Kaung Kaung Akhin	Northern; Kachin	Upland	Indica	I	153.6	25.3	13	52	16.8	<b>80</b>	18.3	-
AC21	KyeitPhat	Western; Chin	Upland	Indica	II	98.2	16.8	14	32.8	14	<b>70</b>	17.4	-
AC22	Sar Pa Ni	Western; Chin	Upland	-	I	123.2	17	9	31	20	<b>72</b>	16.4	-
AC23	Eai Si	Western; Chin	Upland	Japonica		184	19	10	53.1	21.7	<b>60</b>	10.9	-
AC24	Sin E Kayi-2	Western; Chin	Upland	Indica		144.5	26	9	43	19	<b>60</b>	9.2	*
AC25	Traung	Western; Chin	Upland	Indica	I	99	17.2	13	48	30	<b>70</b>	16.4	-
AC 26	Yar	Western; Chin	Upland	-		110	19.5	11	52	29	<b>64</b>	12.7	*
AC27	San Bial	Western; Chin	Upland	Indica		127	16	7	26.3	12.6	<b>60</b>	8.4	-
AC28	Sit Pwa	Delta; Ayeyarwadee	Local cultivar	Indica	II	45	11	4	28	11	<b>30</b>	3	-
AC29	Ta Daung Po	Delta; Ayeyarwadee	Local cultivar	Indica	II	54	9.5	3	34	16	<b>20</b>	4.5	-
AC30	TaungHti	Delta; Ayeyarwadee	Local cultivar	Indica	II	38	11	3	26	15	<b>30</b>	2	-
AC31	Paw San Yin	Delta; Ayeyarwadee	Local cultivar	Japonica	I	99.2	19.5	5	31	21	<b>50</b>	57.2	-
AC32	Paw San TaungPyung	Delta; Ayeyarwadee	Local cultivar	Japonica		107	21	7	58	20.3	<b>45</b>	10	*
AC33	Paw San Gyi	Delta; Ayeyarwadee	Local cultivar	Japonica	I	110	17.6	5	49	19.8	<b>54</b>	13	-
AC34	Paw San Baykyar	Delta; Ayeyarwadee	Local cultivar	Japonica	II	97	19	8	57	20	<b>46</b>	7	-
AC35	Paw San Hmwe	Delta; Ayeyarwadee	Local cultivar	Japonica	II	121	21	6	60	23	<b>41</b>	4	-

PH, plant height (cm); RL, root length (cm); TN, tiller number; SW, straw weight (gm); RW, root weight (gm); PHI is panicle harvest index and YPP is yield per plant (gm). \*Accessions could not be detected with two set of primers for the 5' upstream region of *OsLEA27*.

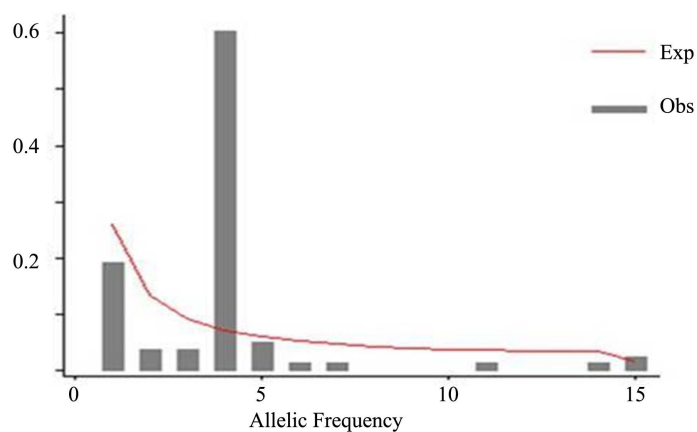
**Supplementary Table 2.** Primer pairs for amplification of entire *OsLEA27* region.

No.	Primer pair name	Sequence accession number	Sequence (5' - 3')		Molecular weight (bp)	$T_m$ (°C)
			Forward	Reverse		
1	<i>OsLEA27</i> (E1)	LOCOs11g26760.1	F-GTACCCTTCCTGACCTACACT	R-CCTTGATTCCCTTCTTCCTC	540	56
2	<i>OsLEA27</i> (E2)	LOCOs11g26760.1	F-ATGGGAGGGAGGAGGAAGA	R-CCGAATCTAACTGCGAGGAA	619	60
3	U1-Lea27	LOCOs11g26760.1	F-TGTTGGGCTTTATTTGTGG	R-AGGCCAGCAAGAAGGTATGT	730	57
4	U2-Lea27	LOCOs11g26760.1	F-CCAATCCGTACATGAGAGAAAA	R-AGTGCCGTACCAAGCCTCT	637	55
5	U3-Lea27	LOCOs11g26760.1	F-GCAGATGAGAGCGACGATT	R-GCTCCACAGTTCACCTTTT	615	57
6	U4-Lea27	LOCOs11g26760.1	F-TTTTCCGGCATAATGTGGT	R-ACTGATCTTGGAGGTTGGAGA	670	56

$T_m$  (°C) = Annealing temperature; U1-Lea27 primer pairs did not amplify in five accessions.



**Supplementary Figure 1.** Sliding-window plot for the entire *OsLEA27* region (2.9 kbp). The thin red line represents polymorphism in 30 accessions. The thick black line represents divergence between the western group accessions and the remaining accessions at each nucleotide position. The window size is 100 bp, the step size is 25 bp, and the unit is silent. Nucleotide position “+1” indicates the transcription start site.



**Supplementary Figure 2.** Frequency spectrum of derived alleles. Obs, observed number of SNPs; Exp, expected number of SNPs. The expected frequency spectrum is calculated according to Watterson (1975). That is, at the neutral equilibrium, the expected number of SNP sites is given by  $4Ne\mu/i$ , where  $Ne$  and  $\mu$  are the effective population size and mutation rate, respectively and the derived allele is assumed to be present  $i$  times in the sample. The formula to estimate  $4N\mu$  is the number of observed sites divided by  $(1 + 1/2 + 1/3 + \dots + 1/n - 1)$ , where  $n$  is the number of chromosomes in each population.