DNA Variation in the 5’ Upstream Region of the Adh Locus of the Wild Plants Arabidopsis thaliana and Arabis gemmifera

Naohiko T. Miyashita
Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

To investigate the level and pattern of DNA polymorphism in the noncoding regulatory region in the plant nuclear genome, 2.4 kb of nucleotide sequence of the 5’ upstream region of Adh was determined for 14 ecotypes of Arabidopsis thaliana and five accessions of Arabis gemmifera. Using this data set and previously determined Adh sequence data, DNA variation was analyzed in a 4.4-kb region of the locus. Two divergent sequence types detected in the transcriptional unit of Adh were not present in the 5’ region of the Adh gene in A. thaliana. Nucleotide diversity of the entire 5’ region was estimated to be 0.0040, which is lower than that in the transcriptional unit. The level of variation was not uniform. There were peaks of variations in a ~400-bp region where cis-regulatory elements for Adh expression were clustered and in exon 4. In interspecific comparison with A. gemmifera, lower divergence was observed in the 5’ flanking region than in the exons. High peaks of divergence in the 400-bp regulatory region and exon 4 were also detected, although there were many other peaks. These results indicate that regions of functional importance have a high level of polymorphism and divergence in the Adh locus of these genera. The possibility of balancing selection in the Adh gene of these plants is discussed.

Introduction

Alcohol dehydrogenase (ADH) plays a key enzymatic function in the response to anaerobic conditions in plants (Sachs, Subbaiah, and Saab 1996). Other stimuli, including low temperature, dehydrogenation, and the phytohormone abscisic acid, also induce its expression (Dolferus, Peacock, and Dennis 1994; De Bruxelles et al. 1996). ADH is induced in response to low levels of oxygen, and the induced ADH enzyme increases the level of NAD+ by reducing acetaldehyde to ethanol; this allows glycolysis to proceed and a normal level of ATP to be maintained in plant cells exposed to low-oxygen conditions.

The alcohol dehydrogenase gene (Adh) of Arabidopsis thaliana is a single-copy gene (Dolferus and Jacobs 1984) and was one of the first nuclear genes in this plant species to be cloned (Chang and Meyerowitz 1986). The regulation of Adh expression in A. thaliana has been extensively studied (Hoeren et al. 1998), and the cis-regulatory sequence elements in the Adh promoter have been identified (Dolferus, Peacock, and Dennis 1994; Hoeren et al. 1998). Two anaerobic response elements (AREs) similar to those in maize are present in the 400-bp region upstream of Adh. A Myb transcription factor (AtMYB2) binds to one of the AREs and is involved induction of Adh under low-oxygen conditions (Hoeren et al. 1998).

Because induction of ADH in plants is essential for defense against environmental stress, ADH is thought to be adaptively important. DNA variation in the Adh locus of A. thaliana and related species has been analyzed from the viewpoint of population and evolutionary genomics (Hanfstingl et al. 1994; Innan et al. 1996; Miyashita, Innan, and Terauchi 1996; Miyashita et al. 1998). In a worldwide sample of A. thaliana, a high level of nucleotide variation was detected, and there were six distinct sequence types in the Adh locus. This pattern of DNA polymorphism can be explained by four intragenic recombination events between the two most divergent Adh sequence types that occurred during the evolutionary history of this plant species (Innan et al. 1996). Two divergent sequence types (dimorphism) were also observed at several other loci (Kawabe et al. 1997; Stahl et al. 1999; Kawabe and Miyashita 1999), suggesting that dimorphism could be one of characteristics of DNA polymorphism in the A. thaliana nuclear genome. However, results of neutrality tests and levels and patterns of polymorphism varied at different loci, so the origin of the dimorphism and the mechanism by which it is maintained remain to be clarified.

In this study, nucleotide sequence variation was analyzed in a 2.4-kb upstream region of the Adh locus in A. thaliana and Arabis gemmifera. This study had two purposes. One was to investigate the level and pattern of DNA polymorphism in the 5’ region where regulatory sequence elements exist. It is expected that levels of polymorphism and divergence between species will be negatively correlated with functional importance (Kimura 1983). The well-characterized 5’ upstream region of A. thaliana Adh provides a good opportunity to examine this expectation, especially for a noncoding region in a plant species. The second purpose was to reveal the extent of the dimorphism detected in the transcriptional unit of Adh. If dimorphism in the Adh region is simply a consequence of neutral evolution, such as accumulation of neutral mutations in an isolated population structure, it is expected that dimorphism would extend farther into the 5’ upstream region to some extent, although intensity of natural selection and recombination would influence the extent. Based on the results presented here, possible mechanisms for maintaining dimorphism at Adh are discussed.

Materials and Methods

Plant Materials

Fourteen ecotypes of A. thaliana and five accessions of A. gemmifera were used in this study (table 1).
Twelve of the *A. thaliana* ecotypes were described (Innan et al. 1996). Two ecotypes (Cvi-0 and Kas-1) were newly chosen. Seeds were provided by N. Goto, Sendai Arabidopsis Seed Stock Center (SASSC), Miyagi University of Education, Sendai, Japan. The growth conditions for *A. thaliana* were described previously (Innan et al. 1996). The plants of *A. gemmifera* were collected in 1995. The *Adh* sequences of the five *A. gemmifera* individuals were described (Miyashita, Innan, and Terauchi 1996; Miyashita et al. 1998).

### Sequencing Method

Total DNA was extracted from mature plants by a modified CTAB method and used for PCR amplification. To amplify a 2.4-kb 5′ upstream region of *Adh* of the two species, PCR primers were designed based on an *A. thaliana* (ecotype Col-0) BAC (F22K20) sequence (GenBank accession number AC002291). The primer sequences were 5′-AAC ACT TAT TCC AGC CTC TT-3′, which is in a hypothetical protein gene located about 2.4 kb upstream of *Adh*, and 5′-CGG TGG TAG ACA TTA TCA AC-3′, which is in exon 1 of *Adh*. The PCR product was cloned into the plasmid pUC18. Sequencing reactions were conducted by using the Auto Read Sequencing Kit, the Cy5 Thermo Sequenase fluorescent labeled cycle sequencing kit with 7-deaza-dGTP (Amersham/Pharmacia Biotech, Piscataway, N.J.). To minimize PCR artifacts, the sequencing template was prepared by mixing three independently isolated plasmid DNA samples for each plant. Sequencing primers were designed at ~400-bp intervals. The 2.2-kb *Adh* sequence of the two *A. thaliana* ecotypes, Cvi-0 and Kas-1, was determined by using primers described in Innan et al. (1996). Newly determined sequences were deposited in the DDBJ database under accession numbers AB048383–AB048400.

### Data Analyses

Fourteen *A. thaliana* and five *A. gemmifera* sequences were used, including *A. thaliana* ecotype Col-0. The *Adh* sequences of the two species described in Innan et al. (1996), Miyashita, Innan, and Terauchi (1996), and Miyashita et al. (1998) were included in the data set for analysis so that approximately 4.4 kb from the *Adh* region was analyzed. This 4.4-kb region can be divided into three parts: the 5′ flanking region (2,366 bp), the transcriptional unit (1,973 bp from the transcription initiation site to the transcription termination site), and the 3′ flanking region (103 bp). The first nucleotide of the transcriptional unit was assigned as coordinate position 1. Program package DnaSP, version 2.0 (Rozas and Rozas 1997), was used to analyze intra- and interspecific variation via the estimation of nucleotide diversity (π; Nei and Li 1979), 4Nμ (θ; Watterson 1975), the number of recombination events (Rm; Hudson and Kaplan 1985), and 4Nc (C; Hudson 1987) and the tests of Hudson, Kreitman, and Aguadé (1987; HKA test), Tajima (1989a), and Fu and Li (1993). The run statistic (K_b), the Kolmogorov-Smirnov statistic (DKS), and the maximum sliding G statistic (G_{max}) were used to examine heterogeneity between polymorphism and divergence using the program DNA slider (McDonald 1996, 1998).

### Results

Polymorphism in the 5′ Upstream Region of the *Adh* Locus of *A. thaliana*

In the 5′ upstream region of the *Adh* locus in *A. thaliana*, a total of 80 (42 site and 38 indel) variations were detected, of which 20 (15 site and 5 indel variations) were found more than once in the sample (fig. 1). Nonsingleton variations were detected throughout the
Fig. 1.—Summary of DNA variation in the 5' upstream region of Adh of Arabidopsis thaliana. Regulatory regions and elements (Hoeren et al. 1998) are indicated by open boxes. Locations of site and indel variations are shown below the map. Circles and numbers indicate variations detected more than once in the sample, which are summarized in the box at the bottom. Variations 13–20 had already been reported in Innan et al. (1996). An asterisk indicates that a sequence is identical to that of Landsberg, and a dash indicates absence. The sequence of deletion 1 (variation 2, d2) is GTGTTTAAGATTTTGGGA.

region, although the occurrence was highest in the sequence proximal to the coding region of Adh. The presence of divergent sequence types (dimorphism) was truncated at the 3' region of the second ARE, and did not extend further to the 5' upstream region. No variations were detected within any of the four regulatory elements, except for a singleton deletion in ARE2 and a singleton nucleotide substitution in GBOX 1. This result indicates that complete sequence conservation in these regulatory elements may not be required for Adh expression in nature, assuming that these ecotypes are wild-type with respect to Adh expression.

The level of nucleotide variation was estimated for functionally different regions of Adh (table 2). Because a different set of ecotypes was used in this study, the estimated values were not exactly the same as those in Innan et al. (1996) and Miyashita et al. (1998). However, the pattern observed previously, namely, a higher level of variation in the transcriptional unit (or exons) than in other regions, was confirmed in this study. The overall level of variation was lower in the nontranscribed 5' flanking region than in the transcriptional unit. For the further-upstream region (FUR; coordinate positions between −2366 and −369) excluding the region of regulatory elements, a lower level of nucleotide variation was found ($\pi = 0.0026$ and $\theta = 0.0048$) than for the entire Adh upstream region.

Sliding-window analysis was conducted to examine changes in the level of variation along the entire 4.4-kb region (fig. 2). There were three peaks of nucleotide variation: the 5' region, where regulatory sequence elements exist; exon 4; and the nontranslated 3' flanking region. No notable peaks were detected in the 5' upstream region. This result indicates that regions of functional importance in Adh have high levels of polymorphism in A. thaliana, although the functional significance of the nontranslated 3' flanking region is not clear.

When Adh sequence types 1 and 6 (table 1) were analyzed separately, these three peaks disappeared as described in Innan et al. (1996). This result indicates that the peaks detected in the whole sample are caused mainly by nucleotide variation between the different sequence types.

The neutrality tests performed for functionally different regions did not indicate significant deviation from test assumptions, namely, neutrality and population equilibrium, and only the transcriptional unit gave positive values for the test statistics (table 2). However, it should be mentioned that for the FUR, Tajima's test gave a significant result ($D = -1.94, P < 0.05$), indi-
Table 2

Level of DNA Polymorphism in the Adh Region of Arabidopsis thaliana and Arabis gemmifera in this Study

<table>
<thead>
<tr>
<th></th>
<th>5' Flanking Region</th>
<th>Transcriptional Unit</th>
<th>3' Flanking Region</th>
<th>Entire Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. thaliana (n = 14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size of region* (bp)</td>
<td>2,366</td>
<td>1,973</td>
<td>103</td>
<td>4,442</td>
</tr>
<tr>
<td>No. of polymorphic sites</td>
<td>42</td>
<td>43</td>
<td>1</td>
<td>86</td>
</tr>
<tr>
<td>θ (silent sites)</td>
<td>0.0040</td>
<td>0.0126 (0.0024)b</td>
<td>0.0014</td>
<td>0.0065</td>
</tr>
<tr>
<td></td>
<td>(0.0008)c</td>
<td>(0.0070)c</td>
<td>(0.0057)c</td>
<td>(0.0063)c</td>
</tr>
<tr>
<td>Tajima's test (D)</td>
<td>−1.38 NS</td>
<td>0.61 NS</td>
<td>−1.16 NS</td>
<td>−0.41 NS</td>
</tr>
<tr>
<td>Fu and Li's test (D*)</td>
<td>−1.59 NS</td>
<td>0.06 NS</td>
<td>−1.40 NS</td>
<td>−0.80 NS</td>
</tr>
<tr>
<td>A. gemmifera (n = 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size of region* (bp)</td>
<td>2,201</td>
<td>2,000</td>
<td>23</td>
<td>4,224</td>
</tr>
<tr>
<td>No. of polymorphic sites</td>
<td>9</td>
<td>12</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>θ (silent sites)</td>
<td>0.0018</td>
<td>0.0035 (0.0025)b</td>
<td>0</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>(0.0027)c</td>
<td>(0.0027)c</td>
<td>(0.0022)c</td>
<td>(0.0025)c</td>
</tr>
<tr>
<td>Tajima's test (D)</td>
<td>−0.53 NS</td>
<td>−1.10 NS</td>
<td>NA</td>
<td>−0.88 NS</td>
</tr>
<tr>
<td>Fu and Li's test (D*)</td>
<td>−0.53 NS</td>
<td>−0.98 NS</td>
<td>NA</td>
<td>−0.81 NS</td>
</tr>
</tbody>
</table>

NOTE.—NS = nonsignificant; NA = not applicable.

* Includes gaps.

b For nonsynonymous sites.

c For all sites.

...indicating that less frequent sites are in excess. Because the level of variation is not uniform (fig. 2), a sliding-window analysis was conducted to examine the change in Tajima’s D value (fig. 3). There was a drastic difference in D values between the 5’ upstream region and the transcriptional unit. In the 5’ upstream region, the D values were negative, and 38% of window points (29 of 77) in the FUR were beyond the critical value of at least 10% significance. On the other hand, in the transcriptional unit D values were mainly positive. Although only four points were beyond the critical value of 10% significance, the points corresponded with peaks of nucleotide variation in the 400-bp region of regulatory elements and exon 4. It should be added here that if Bonferroni multiple correction was applied, none of test points were significant.

Recombination parameters (Rm and C) were estimated for functionally different regions of Adh (table 3). Compared with those obtained for other genes (table 2 in Kawabe and Miyashita 1999), it was noted that the 5’ region of Adh had a high level of recombination, while nucleotide mutation was relatively constant over different regions. In the FUR, the estimate of C per site was much larger (0.1028). The high level of recombination in this region was consistent with the absence of dimorphism in the 5’ upstream region.

Polymorphism in the 5’ Upstream Region of the Adh Locus from A. gemmifera

In the 5’ upstream region of the Adh locus in A. gemmifera, a total of 23 (9 site and 14 indel) variations...
were detected, of which four (2 site and 2 indel variations) were found more than once in the sample. In addition, three indels were detected more than once in the 69-bp region between the transcription initiation site and exon 1. No variations were found within any of the four regulatory elements. A 130-bp insertion was detected 5' upstream of the TATA box in accession Ashibi56 (fig. 4). The FASTA homology search (Pearson and Lipman 1988) using the sequence of the insertion identified several noncoding sequences in the A. thaliana nuclear genome, which had about 70% identity. For example, one of them had sequence identity to the (TAAA) n repeat unit. No clear peak was detected in sliding-window analysis of the entire region (data not shown). To examine peaks of divergence between the two species, sliding-window analysis was applied to the interspecific comparison (fig. 5). The highest peak was in the first intron, which was difficult to align between the two species due to many indel changes. There were many other peaks of divergence, including a peak in the 5' upstream region of the Adh regulatory elements and a peak in exon 4, which correspond with regions of high polymorphism in A. thaliana. When Adh sequence types 1 and 6 of A. thaliana were analyzed separately, the peaks still existed, indicating that the peaks of divergence were not due to dimorphic variations in A. thaliana.

There were 47 fixed differences (37 silent and 10 replacement) in the 7 exons of Adh, of which 23 (18 silent and 5 replacement differences) were in exon 4. Among the 10 replacement changes, 5 caused drastic physicochemical differences (Miyata, Miyazawa, and Yasunaga 1979). Although there was no statistically significant difference in the ratio of conservative to drastic amino acid changes for exon 4 and the remaining exons, four of the five drastic amino acid changes were in exon 4.

### Table 3

<table>
<thead>
<tr>
<th>Region</th>
<th>Silent Divergence</th>
<th>Polymorphism</th>
<th>Entire Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' Flanking</td>
<td>0.0013</td>
<td>0.0044</td>
<td>0.0044</td>
</tr>
<tr>
<td>Transcriptional Unit</td>
<td>0.0057</td>
<td>0.0080</td>
<td>0.0080</td>
</tr>
<tr>
<td>Entire Region (2,277 bp)</td>
<td>0.0031</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
<tr>
<td>3' Flanking</td>
<td>0.0080</td>
<td>0.0057</td>
<td>0.0057</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Region</th>
<th>Polymorphism</th>
<th>Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire Region</td>
<td>0.0040</td>
<td>0.0272</td>
</tr>
<tr>
<td>Transcriptional Unit</td>
<td>0.0031</td>
<td>0.0080</td>
</tr>
<tr>
<td>5' Flanking</td>
<td>0.0013</td>
<td>0.0044</td>
</tr>
<tr>
<td>3' Flanking</td>
<td>0.0057</td>
<td>0.0080</td>
</tr>
</tbody>
</table>

**Note:** NA = not applicable.

---

**FIG. 4.**—Sequence alignment of a 500-bp upstream region of Adh between Arabidopsis thaliana (Col-0) and Arabis gemmifera (Okunikawa1). Regulatory sequences are boxed. An arrow indicates the position of a 130-bp insertion in A. gemmifera (Ashibi56).
Discussion
The Level and Pattern of DNA Polymorphism in the 5′ Upstream Region of the *A. thaliana* Adh Gene

This study demonstrates that the 5′ upstream region of *Adh* has a lower level of DNA polymorphism than the transcriptional unit in both *A. thaliana* and *A. gemmifera*. Especially in the FUR of *A. thaliana*, the level of nucleotide variation is much lower, and Tajima’s test suggests purifying selection. This could possibly mean that there is some constraint on the 5′ upstream region. One possibility is that there are undiscovered regulatory elements for *Adh* and/or a hypothetical protein-coding region, because an open reading frame was not detected in the upstream region. Another possibility is a structural constraint, although this kind of constraint has not often been considered. It has been shown that positioning of nucleosomes influences gene expression (see Wolffe 1994 for review). Vega-Palas and Ferl (1995) reported that when *Adh* of *A. thaliana* is expressed, a nucleosome binds loosely to the 400-bp regulatory region, and the surrounding regions maintain a tight chromatin structure. Thus, the low level of nucleotide variation in the FUR may be required to maintain nucleosome positioning and chromatin structure elements that are required for normal expression of *Adh*.

This study also shows that dimorphism in the *Adh* locus does not extend into the FUR. This result seems to be inconsistent with the hypothesis that dimorphism is due to accumulation of neutral mutations in an isolated population structure, and suggests that the population structure could not be the cause of dimorphism. However, if the constraint mentioned above exists, variations would not have accumulated even in an isolated population structure. Recombination also influences the extent of dimorphism. Recombination parameters were high in the 5′ region of *Adh*. A high level of recombination would have randomized divergent sequence types, even if two divergent sequence types had existed before recombination occurred. Because of significant deviation from neutrality and a high level of recombination detected in the 5′ upstream region, the origin of dimorphism at the *Adh* transcriptional unit remains unclear.

Peaks of Nucleotide Variation in the *A. thaliana* Adh Gene

There are two regions in which peaks of variation were detected: the 400-bp upstream region, including promoter regulatory elements, and exon 4. Although these two peaks of variation were previously reported in Innan et al. (1996), by extending the scope of the earlier data set, this study shows that these are the only notable peaks of variation in the *Adh* region of *A. thaliana*. No clear peaks were detected in the FUR. Thus, regions of functional importance have high levels of DNA polymorphism in the *Adh* region of *A. thaliana*. This is contrary to the negative correlation between level of variation and functional importance expected from neutral mutation theory. As discussed below, it seems reasonable to consider that some kind of balancing selection had acted on *Adh*. Here, one could argue against the functional importance of the two regions. If the two regions are not functionally important, the two peaks of variation would be consistent with the neutral mutation theory, reflecting higher mutation rates in the regions, for example. However, it is certain that the sequence elements in the 400-bp upstream region are responsible for normal and induced expression of *Adh* (Hoeren et al. 1998) and that exon 4 codes for approximately 42% of the 379 amino acids of ADH. Replacement polymorphic sites in exon 4 are associated with ADH allozyme variations, which show clear ADH activity differences (Dolferus and Jacobs 1984). As far as ADH activity is necessary for *A. thaliana* in nature, it is unlikely that the two regions of high variation are free from functional constraints.

---

**Table 4**

<table>
<thead>
<tr>
<th>Region</th>
<th>5′ Flanking Region</th>
<th>Transcriptional Unit</th>
<th>3′ Flanking Region</th>
<th>Entire Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of nucleotides</td>
<td>2,049</td>
<td>1,907</td>
<td>23</td>
<td>3,979</td>
</tr>
<tr>
<td>No. of fixed sites</td>
<td>195</td>
<td>151</td>
<td>2</td>
<td>348</td>
</tr>
<tr>
<td>K (silent sites)</td>
<td>0.0991</td>
<td>0.1492 (0.0822)*</td>
<td>0.0900</td>
<td>0.1157 (0.0934)</td>
</tr>
</tbody>
</table>

* For pooled data.
In the 400-bp region of regulatory elements, the peak is caused by polymorphic site variations located outside of the regulatory elements themselves. Two indel changes in this region could cause spatial differences between these elements and the coding region of Adh, which may have some effect on Adh expression. However, both indels are single-nucleotide changes, so the spatial change is quite small. It seems unlikely that the variations in the regulatory region have a strong influence on Adh expression. Therefore, the regulatory significance of detected polymorphic variations, both indel and nucleotide changes, and, consequently, the reason for the peak of variation are not clear.

Balancing Selection in the Adh Region from A. thaliana

There are four observations that are relevant to a consideration of the genetic mechanisms acting on the Adh region of A. thaliana. First, the dimorphic variations in exon 4, especially replacement polymorphic sites, are associated with allozyme variation in the Adh gene (Hanfstingl et al. 1994; Miyashita et al. 1998). Second, as shown in Innan et al. (1996) and this study, the two peaks of variation in the Adh locus are caused by dimorphic variation in A. thaliana. This result is suggestive of balancing selection between the two sequence types (Hudson and Kaplan 1988). Third, by investigating DNA variation at the entire genome level with AFLP analysis, Miyashita, Kawabe, and Innan (1999) concluded that polymorphisms occurring at intermediate frequencies were rare in the A. thaliana nuclear genome. Therefore, dimorphism at an intermediate frequency in the Adh locus, especially in exon 4, could be unusual in the A. thaliana nuclear genome. Finally, this study showed drastic change in Tajima's D value along the entire region investigated. Positive peaks of Tajima's D value were detected in the regulatory region and exon 4. Because the D value in the transcriptional unit was positive (table 2), statistical significance of the peaks of D values can be examined by a one-tailed Tajima's test with an alternative hypothesis of \( D > 0 \) (E. Tajima, personal communication). If so, the four peak points in the transcriptional unit can be regarded as significantly larger than zero (fig. 3). Also, considering that the present A. thaliana population has expanded its habitat recently (Price, Palmer, and Al-Shehbaz 1994; Innan, Terauchi, and Miyashita 1997; Miyashita, Kawabe, and Innan 1999), Tajima's D value would be expected to be even lower than that under neutrality (Tajima 1989b). In this case, more points in the transcriptional unit would become significantly positive, although there is the problem of multiple correction.

These observations suggest that balancing selection could be invoked to explain the dimorphism at an intermediate frequency in exon 4 of Adh. A possible selection model is diversifying selection, which assumes variable direction and intensity of natural selection in subpopulations such that balancing selection is detected at the species level. However, this seems to be unlikely, because there was no association between sequence types in Adh and geographic origin (Innan et al. 1996). If selection varies in different subpopulations, some correlation between ecotype and origin would be expected. This is not the case. Another explanation is overdominance. It may be difficult to imagine balancing selection in a selfing species like A. thaliana. In natural populations of this plant, almost complete homozygosity is detected (Abbot and Gomes 1989; Todokoro, Terauchi, and Kawano 1996; Bergelson et al. 1998). However, Innan et al. (1996) showed that there were at least four intragenic recombinations in Adh in the evolutionary history of this plant species. To have intragenic recombination, it is obvious that heterozygotes must have been formed between sequence types. Although this could be extremely rare, heterozygotes formed between sequence types might have some advantage over the parental homozygotes. An experiment to measure ADH activity and Adh expression is under way to examine the possibility of overdominance in this selfing plant.

Divergence in the Adh Region Between A. thaliana and A. gemmifera

Between the two species, the 5' flanking region has lower divergence than the transcriptional unit. Also, the regions of high polymorphism in A. thaliana, that is, the 400 bp regulatory region and exon 4, have high divergence as well, although the peak of divergence in the regulatory region is not particularly high. Unlike polymorphism in A. thaliana, the peaks of divergence were not caused by dimorphic variation in A. thaliana. However, this could be expected, since the level of divergence between species is much greater than the difference between the two most divergent sequence types in A. thaliana. Therefore, the peaks of divergence reflect solely the mechanism operating on the speciation process of the two species.

An explanation for the correlation between polymorphism and divergence in the two regions is that adaptive nucleotide changes occurred in Adh, especially in exon 4, in the divergence process to adapt to the ecological niche of each species. Drastic amino acid substitutions in exon 4 may be related to this adaptation process. There were many gaps in the sequence alignment of the regulatory region between the two species, despite a high sequence identity (>90%). This result may suggest that differences in the spatial distance between regulatory elements and the Adh coding region may be related to changes in expression of Adh in the two species to compensate the possible adaptive changes in exon 4.

Acknowledgments

I express my thanks to N. Goto for A. thaliana seeds, and to M. Aguadé, F. Tajima, R. Terauchi, S. Nasuda, H. Innan, and A. Kawabe for comments. This study is contribution 561 from the Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University.
LITERATURE CITED


FUMIO TAJIMA, reviewing editor

Accepted October 4, 2000