Letter to the Editor

Intra- and Interspecific Variation of the Alcohol Dehydrogenase Locus Region in Wild Plants *Arabis gemmifera* and *Arabidopsis thaliana*

Naohiko T. Miyashita, Hideki Innan,* and Ryohei Terauchi†*
*Laboratory of Genetics, Faculty of Agriculture and †Laboratory of Plant Systematics, Faculty of Science, Kyoto University

With the advent of molecular techniques in biology, especially PCR amplification, it is possible to compare the patterns of intra- and interspecific variation in homologous genes between related species without difficulty. This kind of study on sibling species of *Drosophila* has opened new areas in population and evolutionary genetics (McDonald and Kreitman 1991; Ohta 1993). The evolutionary history of *Drosophila* species and genetic mechanisms in the speciation process have been discussed in great detail. In order to infer the genetic and evolutionary mechanisms of organic evolution in general, studies on organisms other than *Drosophila,* especially plants, are necessary. Population studies of plant species have been hampered because of experimental difficulties associated with plants. However, the situation has been changing as the molecular biology on plants grows rapidly.

Here we report intra- and interspecific variations in the alcohol dehydrogenase (*Adh*) locus of wild plant species *Arabis gemmifera* and *Arabidopsis thaliana.* Molecular biology of *A. thaliana* is one of the fastest growing fields of biological science. It is possible to take advantage of accumulating information on *A. thaliana* for population studies of *A. thaliana* itself and related plant species. Alcohol dehydrogenase (ADH; alcohol: NAD^+ oxidoreductase, EC 1.1.1.1) is one of the most investigated enzymes in plant species including *Arabidopsis,* wheat, rice, maize, and pearl millet (Gottlieb 1982; Dennis et al. 1984; Chang and Meyerowitz 1986; Gaut and Clegg 1993; Hanfstingl et al. 1994). It is known that *Adh* in plants is expressed in response to anaerobiosis (Sachs, Freeing, and Okimoto 1980). In nature, this gene could be adaptively important for responding to environmental changes such as flooding. Because the genus *Arabis* is phylogenetically classified as close to *Arabidopsis thaliana* ($2n = 2x = 10$), it was expected that PCR amplification based on an *A. thaliana* sequence could be used to clone and sequence genes of interest from the *Arabis* species. By using PCR primers based on the *A. thaliana Adh* sequence (Chang and Meyerowitz 1986), a 2.5-kbp PCR product containing the entire coding sequence was obtained from five *Arabis* species sampled from around Japan. Among the five *Arabis* species, *A. gemmifera* ($2n = 2x = 16$) was chosen for a population study based on its wide distribution and diploidy. *Arabis gemmifera* is a wild crucifer commonly found in the mountain areas of the temperate zone and sea shores of the frigid zone in the Japanese Islands and Korean Peninsula.

Eight samples of *A. gemmifera* were collected in May 1994 in Hiroshima (accession code Uga), Okayama (Hukiya), Hyogo (Ohtani), Osaka (Minou), Kyoto (Umenoki), Yamagata (Hukasawa and Mazawa), and Akita (Yatate) Prefectures in Japan. Sampling locations are at least 40 km apart, and farthest apart Hiroshima and Akita samples are about 1,000 km away in a straight line. Seeds of five wild ecotypes of *A. thaliana* sampled in Germany (accession code Aa-0), Libya (Mt-0), Canada (Pog-0), and Japan (Shokei and Hiroshima) were obtained from the Miyagi Arabidopsis Stock Center, Miyagi Kyoiku University, Sendai, Japan. Total DNA was extracted by a modified CTAB method (Terauchi and Konuma 1994). The number of *Adh* genes per genome in *A. gemmifera* was examined by digesting total DNAs with several 6-cutter restriction enzymes. Hybridization with the *A. thaliana Adh* clone resulted in one or two bands in autoradiograms. For the restriction enzymes producing two bands, the presence of a single restriction site for those enzymes in the amplified region was confirmed in obtained nucleotide sequences. These results indicate that the copy number of *Adh* in *A. gemmifera* genome is one, as in *A. thaliana* (Chang and Meyerowitz 1986). Thus, the PCR amplification and nucleotide sequencing of the products are not complicated with the multiplicity of this gene in the genome. The primer sequences for PCR amplification for the entire 2.5-kb *Adh* region are 5′ AAT GTC TAC TAT CCC TTA AT 3′ (sense) and 5′ TGG CGA CTC GAA ACG GCT 3′ (antisense). Templates for sequencing were PCR products amplified from the total DNAs and/or cloned PCR products in plasmid pUC18. Primers for sequencing reactions were designed in 250–300-bp interval based on *A. thaliana* and *A. gemmifera* sequences. Sequences were determined in both directions with a DNA sequencer (ABI 373). The ODEN package (DDBJ, National Institute of

Key words: *Arabis gemmifera, Arabidopsis thaliana,* alcohol dehydrogenase, polymorphism, divergence.

Address for correspondence and reprints: Naohiko T. Miyashita, Laboratory of Genetics, Faculty of Agriculture, Kyoto University, Sakyo-ku, Kyoto 606-01, Japan.

Mol. Biol. Evol. 13(2):433–436. 1996

Genetics, Mishima, Japan) created by Ina (1992) was used to align and analyze obtained sequences. Sequences of *A. thaliana* ecotypes, Landsberg (Chang and Meyerowitz 1986) and Columbia (Hanfstingl et al. 1994), were included in the analysis. In total, eight *A. gemmifera* and seven *A. thaliana* sequences were analyzed. Newly determined 13-nucleotide sequences are deposited in the DDBJ/GENBANK/EMBL databases with the following accession numbers: #D63452–#D63459 for *Arabis gemmifera* and #D63460–#D63464 for *Arabidopsis thaliana.*

It was shown that the *Adh* gene of *Arabis gemmifera* has the exon–intron structure identical to that of *A. thaliana.* There are seven exons and six introns. Total length of the coding sequence is 1,137 bp, which is the same as that of *A. thaliana.* Although the first intron of *A. gemmifera* is about 40 bp longer than that of *A. thaliana,* the other introns differ only by a few base pairs. All the introns start with dinucleotide GT and end with AG, as typical in eukaryote genes. Among eight *A. gemmifera* sampled in Japan, 51 polymorphic nucleotide sites and 12 insertion/deletion polymorphisms were detected in the aligned 2,290-bp region. Forty-seven of the 51 polymorphic sites were found only once in the sample. The number of unique polymorphic sites ranged from 0 (Yamagata–Mazawa sample) to 14 (Kyoto) with an average of 5.9. Estimated nucleotide diversity (Tajima and Nei 1984) for the entire region was 0.0075, which is a little larger than that of *Adh*1 in pearl millet but smaller than that of maize *Adh*1 (Gaut and Clegg 1993). In the 1,137-bp coding region, 23 polymorphic nucleotide sites and no insertion/deletion polymorphisms were detected. Twenty-one of 23 polymorphic sites were unique. Nucleotide diversity for the coding region was 0.0055. Tajima's test (1989*a*) was conducted for nucleotide polymorphisms in the entire region ($D = -1.69$, $P < 0.05$) and the coding region ($D = -1.57$, $P < 0.10$). Although tests are marginally significant, results suggest the deviation from neutral mutation process in the *Adh* gene of *A. gemmifera.* The distribution of nucleotide polymorphisms is obviously skewed toward more rare alleles. Note that we have not detected any possible structure of subpopulations as long as we look at a phylogenetic tree of these gene sequences (data not shown).

In the seven *A. thaliana* ecotypes, 56 polymorphic nucleotide sites and 11 insertion/deletion polymorphisms were detected in the 2,378-bp region aligned. Thirty-six polymorphic sites were found to be unique, 32 of which were detected in a Japanese ecotype sampled in Hiroshima. Interestingly, the Libyan ecotype has the completely identical sequence to the ecotype Landsberg. The estimated nucleotide diversity for the entire region (0.0101) of *A. thaliana* was larger than that of *A. gemmifera.* This larger value is due to the presence of two dominant haplotypes (Columbia and Landsberg types) among seven ecotypes of *A. thaliana,* as noted by Hanfstingl et al. (1994). For Columbia ($n = 4$) and Landsberg ($n = 3$) types, the nucleotide diversity estimates for the entire region were 0.0115 and 0.0003, respectively. The inflated nucleotide diversity for Columbia types was due to the Hiroshima ecotype with its many unique polymorphic sites. This is because Nei's nucleotide diversity is defined by $\Sigma\, d_{ij}$ where d_{ij} is the nucleotide differences per site for a given pairwise comparison. In the coding region, 14 polymorphic sites were detected, of which four were unique in Hiroshima ecotype. Nucleotide diversity for the coding region of *A. thaliana* was 0.0060 ($n = 7$). For Columbia types the estimate was 0.0017, and 0.0000 for Landsberg types. Variation in Columbia types is due to the Hiroshima ecotype, and no polymorphic sites were found in Landsberg types. Tajima's test was not significant ($D = -0.83$ for the entire region and $D = 0.66$ for the coding region).

Table 1
Intra- and Interspecific Variation in the *Adh* Coding Region of *Arabis gemmifera* and *Arabidopsis thaliana*

	No.	
	Nonsynonymous	Synonymous
Polymorphic within *A. gemmifera*	17	6
Polymorphic within *A. thaliana*	4	10
Fixed between *A. gemmifera* and *A. thaliana*	9	34

NOTE.–The numbers of sites are shown.

The interspecific distance for *Adh* coding sequences between a randomly chosen *A. thaliana* (Landsberg ecotype; Chang and Meyerowitz 1986) and a randomly chosen *A. gemmifera* (Osaka sample) was examined by calculating the numbers of synonymous (Ks = 0.182) and nonsynonymous (Kn = 0.015) substitutions (Nei and Gojobori 1986). Corresponding values for randomly chosen coding sequences of the *Adh* in *Drosophila melanogaster* (GENBANK: DMRI32ADH) and *Drosophila yakuba* (GENBANK: DYADH) are Ks = 0.150 and Kn = 0.012. Thus, the distance for the *Adh* genes in these two plant species belonging to two different genera can be regarded similar to that of two sibling species of *Drosophila,* studied by McDonald and Kreitman (1991).

Table 1 summarizes the intra- and interspecific variation of the *Adh* coding region of *A. gemmifera* and *A. thaliana.* In *A. gemmifera,* nonsynonymous polymorphic sites are found more frequently than synonymous ones, whereas this pattern was reversed in *A.*

thaliana. In the interspecific comparison, fixed synonymous substitutions were detected more frequently than nonsynonymous ones. Fisher's exact test was used to test the correlation between patterns of polymorphism and divergence. Tests involving within-*A. gemmifera* polymorphism gave significant results (between vs. within *A. gemmifera,* $P < 10^{-4}$ and within *A. gemmifera* vs. within *A. thaliana,* $P < 0.05$). Significance was not detected in the comparison of between species vs. within *A. thaliana.* Obviously, a higher level of nonsynonymous polymorphic sites in *A. gemmifera* is responsible for the discordance and is unexpected from the previous population studies at the DNA level (Kimura 1983). However, a relatively high level of nonsynonymous polymorphic sites has been detected in some plant species. In *Clarkia lewisii,* Thomas et al. (1993) reported that five nucleotide changes detected in the entire coding region between two allozymic forms of the phosphoglucose isomerase (*Pgi*C) were all nonsynonymous. In a short segment (262 bp) of the fourth exon of *Adh* in *A. thaliana,* 6 out of 14 polymorphic nucleotides were found nonsynonymous (Hanfstingl et al. 1994). Further investigation is needed to obtain a general picture of molecular polymorphism in plants.

The patterns of polymorphism in *A. thaliana* and interspecific divergence between *A. thaliana* and *A. gemmifera* are typical in the proportions of synonymous and nonsynonymous changes (Kimura 1983). It can be assumed that the *Adh* gene had been under purifying selection in the speciation process of these two plants and in the micro-evolutionary lineage of *A. thaliana.* Synonymous substitutions could have been allowed to accumulate more than nonsynonymous substitutions because of selection. In order to account for the unexpected high proportion of nonsynonymous site polymorphisms in *A. gemmifera,* two explanations would be possible. One is the change of selection regime. After *A. gemmifera* was established as a species in its present ecological niche, the selection pressure might have been relaxed or reversed to start retaining nonsynonymous mutations in populations. Actually, these two species occupy different ecological niches. *Arabidopsis thaliana* is found in dry and open space, whereas *A. gemmifera* prefers shady and moist habitats. Their different ecologies could be related to the different selection scheme. The other explanation is the reduction of effective size and limited migration after the establishment of *A. gemmifera.* Nonsynonymous (maybe slightly deleterious) mutations could have been fixed in populations of reduced effective size (Ohta 1973). Skewed distribution of polymorphic nucleotide sites toward rare alleles is consistent with this explanation (Tajima 1989*b*). At present, a survey on a larger scale with more samples of *A. gemmifera* and the other *Arabis* species on the *Adh* region is under way to discuss the evolutionary history and population structure of *Arabis* species in more detail.

Acknowledgments

The authors are indebted to Professor S. Kawano, Faculty of Science, Kyoto University, for introducing the *Arabis* species as materials for plant molecular population genetics. The authors also express their thanks to Professor N. Goto, Miyagi Kyoiku University, for *A. thaliana* seeds and to Drs. R. T. Endo, T. Gojobori, F. Tajima, and Y. Yasui for comments. Contribution #537 from the Laboratory of Genetics, Kyoto University.

LITERATURE CITED

CHANG, C., and E. M. MEYEROWITZ. 1986. Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase. Proc. Natl Acad. Sci. USA **83**:1408–1412.

DENNIS, E. S., W. L. GERLACH, A. J. PRYOR, J. L. BENNETZEN, A. INGLIS, D. LLEWELLYN, M. M. SACHS, and W. J. PEACOCK. 1984. Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. Nucleic Acids Res. **12**: 3983–4000.

GAUT, B. S., and M. T. CLEGG. 1993. Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135**:1091–1097.

GOTTLIEB, L. D. 1982. Conservation and duplication of isozymes in plants. Science **216**:373–380.

HANFSTINGL, U., A. BERRY, E. A. KELLOGG, J. T. COSTA III, W. RUTIGER, and F. M. AUSUBEL. 1994. Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for both balancing and directional selection? Genetics **138**:811–828.

INA, Y. 1992. ODEN. National Institute of Genetics, Mishima, Japan.

KIMURA, M. P. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

MCDONALD, J. H., and M. KREITMAN. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila.* Nature **351**:652–654.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

OHTA, T. 1973. Slightly deleterious mutant substitutions in evolution. Nature **246**:96–98.

———. 1993. Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size. Proc. Natl. Acad. Sci. USA **90**:4548–4551.

SACHS, M. M., M. FREEING, and R. OKIMOTO. 1980. Anaerobic proteins of maize. Cell **20**:761–767.

TAJIMA, F. 1989*a*. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585–595.

———. 1989*b*. The effect of change in population size on DNA polymorphism. Genetics **123**:597–601.

TAJIMA, F., and M. NEI. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1**: 269–285.

TERAUCHI, R., and A. KONUMA. 1994. Microsatellite polymorphism in *Dioscorea tokoro,* a wild yam species. Genome **37**:794–801.

THOMAS, B. R., V. S. FORD, E. PICHERSKY, and L. D. GOTTLIEB. 1993. Molecular characterization of duplicate cytosolic phosphoglocose isomerase genes in Clarkia and comparison to the single gene in Arabidopsis. Genetics **135**: 895–905.

TAKASHI GOJOBORI, reviewing editor

Accepted October 26, 1995