The Evolution of the Alcohol Dehydrogenase Gene Family by Loss of Introns in Plants of the Genus *Leavenworthia* (Brassicaceae)

**D. Charlesworth, F.-L. Liu, and L. Zhang**

Department of Ecology and Evolution, University of Chicago

We present results showing that several species in the plant genus *Leavenworthia*, in the Brassica family, have three alcohol dehydrogenase loci, unlike *Arabidopsis thaliana*, which has only a single classical (class P) alcohol dehydrogenase locus. Based on a portion of the sequence, the alcohol dehydrogenase loci of *Leavenworthia* show about 92%–93% amino acid sequence identity to that of the *A. thaliana* alcohol dehydrogenase. The great majority of the sequence differences from the *A. thaliana Adh*-coding sequence, and also between three different *Leavenworthia* species, are synonymous, suggesting that all are currently functional (or have been in the recent evolutionary past). The loci differ in the numbers of introns present, with one locus (*Adh-3*) having no introns present. RT-PCR tests detect expression of all three loci. Linkage data using variant alleles identified by single-strand conformation polymorphism analysis show that the three *Leavenworthia* loci are not closely linked. The results therefore suggest that the *Adh-3* locus may have arisen via an mRNA intermediate but, despite loss of the introns, is expressed.

**Introduction**

Processed pseudogenes are thought to be quite rare in plants (Drouin and Dover 1987), despite the lack of a separate lineage of germ line cells (Poethig 1989) and despite the well-documented lability of introns (e.g., Logsdon and Palmer 1994). As part of our work on sequence polymorphism in the plant genus *Leavenworthia* (in the plant family Brassicaceae), we have cloned and characterized the alcohol dehydrogenase loci, and we here report evidence for differences in the numbers of introns between the loci, with one expressed locus having no introns.

It is well known that gene duplications have occurred in plants, although in many cases, these originated by polyploidy (e.g., Soltis and Soltis 1990; Gottlieb 1982), but there are at present few studies of the detailed nature of the changes involved in the evolution of such small gene families (Thomas, Laudancia-Chingcuanclo, and Gottlieb 1992; Thomas et al. 1993; Ford, Thomas, and Gottlieb 1995). Such occurrences in members of gene families can provide information on the ways in which new loci can arise in evolution. Alcohol dehydrogenase loci have been studied in a wide variety of taxonomic groups, including plants (e.g., Trick et al. 1988; Yokoyama and Harry 1993). Two types of alcohol dehydrogenase have been identified: classical alcohol dehydrogenases (termed class P in plants), and the very different formaldehyde-active class III alcohol dehydrogenases (Shafquat et al. 1996). *Arabidopsis thaliana* has only a single class P alcohol dehydrogenase locus, but most plants have two to three classical *Adh* loci (reviewed by Clegg, Cummings, and Durbin 1997). Intron positions in alcohol dehydrogenase loci are generally conserved, even between angiosperms and gymnosperms. The nine introns in four of the seven expressed alcohol dehydrogenase loci that were characterized in *Pinus banksiana* are in the same positions as those in maize (Perry and Furnier 1996), and six of them are found in *A. thaliana*. Our findings add some new information to this picture.

**Materials and Methods**

To study the *Leavenworthia* loci, sequences of alcohol dehydrogenases (ADH; EC 1.1.1.1) were obtained from GenBank from the following dicotyledon species: *A. thaliana*, strawberry (*Fragaria ananassa*), and potato (*Solanum tuberosum*) (accession numbers X77943, X15588, and M25154, respectively). The aligned nucleotide sequences were examined to find conserved regions of at least 20 bases, and two internal primers were designed (S1: 5′-GATGT(T,C,A)TACTTCTGGGAAGC-3′ and S2: 5′-ATC(G,A)TGACACATTCAAAATGC-3′).

Plants of the species *Leavenworthia stylosa* (Gray), *L. crassa* (Rollins), and *L. uniflora* (Michx.) Britton were grown from seeds supplied by L. L. Lyons and T. E. Hemmerley, and additional collections of these species, plus *L. alabamica* (Rollins) and *L. exigua var. lutea* (Rollins) were collected in the field in spring 1995 and grown in the University of Chicago greenhouse until seeds could be harvested. For seed germination and plant culture methods, see Charlesworth, Lyons, and Litchfield (1994).

Genomic DNA was prepared from *Leavenworthia* leaves of individual plants by a plant miniprep method using the CTAB method, or from single seeds by the PureGene DNA isolation method (Gentra Systems Inc.) modified by adding two chloroform extractions of the lysates after protein precipitation. This modification helped to remove enzyme-inhibiting contaminants in the seeds (Murray and Thompson 1980). To obtain alcohol dehydrogenase sequences from *Leavenworthia*, PCR reactions were performed using primers S1 and S2, under the following conditions: 2ng/μl primers, 3.0 mM MgCl₂, 2 U/100 μl Taq DNA polymerase, and 10–50 ng of *Leavenworthia* DNA. After 2 min at 94°C, the
following cycle was repeated 30 times: 30 s at 94°C, 30 s at 52°C, 2 min at 70°C. The products of the reactions were run on 1% agarose gels with 1 × TAE buffer and stained with ethidium bromide.

Three bands were seen with all Leavenworthia species studied (see below). These were cloned into TA vector using the prokaryotic TA cloning kit (Invitrogen). Clones were amplified using universal primers of the vector (M13 reverse primer and M13-20 primer), and the product was purified using QiAquick-spin columns. Sequencing templates were prepared from amplified DNA fragments or amplified cloned inserts. Sequencing reactions were prepared using template-specific primers for analysis on an ABI 373 automated sequencer using the Taq DyeDeoxy Terminator Cycle Sequencing Kit (ABI Applied Biosystems, Inc.) to obtain sequences for both strands. For sequencing from cloned PCR products, at least two clones were generated for sequencing for each locus from each species studied.

To obtain further sequences 3' and 5' to those obtained with the primer pair S1 and S2, the method of bubble PCR (Riley et al. 1990) was used, with a slight modification by J. A. McDonald of the linker sequences containing the “bubble,” as follows: Vec-35K: 5'-CGTGTGTGTGATACGTACCAATGTCACCGC-3' (kinased) and Vec-33: 5'-GCCGTGTGATTGGTACGGACATTTCGAGCAACAAC-3'.

The loci in Leavenworthia were used to design locus-specific primers. The Adh loci from the three Leavenworthia species were aligned using the CLUSTAL W multiple-alignment program, and suitable conserved regions were sought. The following primers were designed: Adh-1-specific—Adh1.P1: 5'-TGTCACTCTGAGGAATCCA-3', Adh1.P3R: 5'-CACAACACTCAGTCACACCA-3', and Adh1.Cp4R: 5'-ACAACCCCAACCTGAAAC-3'; Adh-2-specific—Adh2.P3: 5'-AACCAACACTGAGACGACG-3' and Adh2.P4R: 5'-TGGTTTACAAACTCAGTA-3'; Adh-3-specific—Adh3.P1: 5'-GCTGGAGGGATTGTGGAG-3' and Adh3.P3R: 5'-TATCTACGACCTTCCGGAC-3'.

These locus-specific primers were used to test DNA samples from different individuals, using the method of “cold SSCP,” a single-strand conformation polymorphism detection method which is expected to be capable of detecting single differences in PCR products up to about 350 bases (Hongyo et al. 1993). Electrophoresis was done using 1 × TBE buffer in a cold room (room temperature is around 4°C to 6°C) at a constant voltage of 250 V for 2–4 h, depending on the gene fragment under study. Before running the PCR products on the gels, restriction enzyme digestion was sometimes used to obtain smaller fragments, making it possible to classify alleles that were otherwise hard to distinguish. For Adh-3, the enzyme Rsa I was used, as the sequence of the locus showed a site for this enzyme such that digestion should produce two fragments of sizes 244 and 194 bp. For linkage analysis, variants at the loci of interest that were detectable by SSCP were used.

RNA Extraction and RT-PCR

RNA was extracted from fresh or frozen leaves for RT-PCR as follows. Material was reduced to a fine powder by vortexing at maximum speed for 30 s to 1 min with one chilled RNase-free metal bead plus several small glass beads per RNase-free microfuge tube. RNA was prepared from the powder using Qiagen Rneasy Plant Mini kits, digested with DNase I to remove genomic DNA, and treated with “micropure enzyme remover” (Amicon) to remove the DNase. Finally, first-strand cDNA synthesis was performed using an Invitrogen cDNA Cycle kit for RT-PCR.

Sequence Comparison and Estimation of Phylogenetic Relationships of the Adh Sequences

Sequences to be compared were aligned using the SeqMan program of the DNASTAR software, and numbers of synonymous and nonsynonymous differences between pairs of coding sequences were calculated using the MEGA software (Kumar, Tamura, and Nei 1994), version 1.01. Numbers of substitutions were estimated without Jukes-Cantor correction, since some comparisons were between distantly related species or loci, while others were between the same locus within the genus Leavenworthia. The latter differences were thus small and needed no correction, while the former were too large for correction to provide reliable estimates. Nevertheless, our conclusions are unaffected by Jukes-Cantor correction. Differences were calculated on a per-site basis in order to be able to include sequences of various lengths and thus to include the largest regions possible of the loci sequenced, but the analysis was also done using a small number of longer sequences to make sure that they were not affected by differences in substitution patterns in different regions of the locus; the results obtained were very similar to those shown.

To attempt to establish the origin of the Adh-3 gene, we estimated a phylogenetic tree based on the sequence information from the region of the gene for which sequences were available from all three Adh loci of the three Leavenworthia species studied. This reduced the region to 351 nt of coding sequence from exons 4 and 5. Arabidopsis thaliana and two Arabis taxa were used as outgroups. The initial set of sequences included a set of allelic sequences obtained in a study of within- and between-population sequence diversity in natural populations of Leavenworthia species (Liu, Zhang, and Charlesworth 1998). To reduce the number of sequences, only one of each different sequence found within each species was used for each locus, resulting in 89 sequences including the three outgroup sequences X (GenBank accession numbers AF037472–AF037564).

For neighbor-joining (NJ) analysis, pairwise distances were calculated using maximum likelihood (ML) by Dnadist of the PHYLIP package, which computes distances between species from nucleotide sequences, using the default settings of a 2.0 transitions/transversions ratio, empirical base frequencies as the expected frequencies of the four nucleotides, and a single category of substitution rates (Kishino and Hasegawa 1989). A consensus neighbor-joining tree (Saito and Nei 1987) was then constructed based on the distance matrix using PHYLIP (Felsenstein 1989), with A. thaliana Adh as an outgroup. Version 3.5 of this software was used, with
the input order of the sequences (or taxa) randomized. One hundred bootstrap iterations (Felsenstein 1985) were run. A consensus treefile was created using the Consense program in the PHYLIP package and was then plotted using PAUP (Swofford 1993).

For parsimony analysis, the number of sequences was reduced, to speed up bootstrap calculations, by the following procedure. After alignment of all sequences of a given locus from a given species using the SeqMan program of the DNASTAR software, a consensus sequence was constructed for each locus within each species, again using DNASTAR software. This was done using the IUB codes (IUPAC-IUB Commission on Biochemical Nomenclature 1970) for polymorphic bases, except for sites at which there was a majority nucleotide and the alternative base was a singleton; since this is phylogenetically noninformative, in such cases the majority nucleotide among the alleles was used. Maximum-parsimony analysis of aligned Leavenworthia Adh consensus sequences, together with three outgroup Adh sequences (A. thaliana, Arabis lyrata, and Arabis petraea), was conducted using the PAUP heuristic search procedure with 1,000 bootstrap iterations to generate a 50% majority-rule consensus tree (Swofford 1993). Parsimony analysis of a smaller set of longer aligned Leavenworthia Adh sequences, together with two outgroup Adh sequences (A. thaliana and Arabis gemmifera, for which sequence was available for the same regions as for Leavenworthia), gave similar results.

Isozyme Electrophoresis

Cellulose acetate electrophoresis was performed by the method of Hebert and Beaton (1989) using Tris-glycine buffer. Electrophoresis was performed for 20 min at 185 V. Plates were stained for alcohol dehydrogenase by standard methods (Soltis et al. 1983) using agar overlays.

Results

Characterization of Leavenworthia Alcohol Dehydrogenase Loci

Using the primer pair S1 and S2, three alcohol dehydrogenase bands were seen for all five Leavenworthia species studied. They will be referred to as Adh-1, Adh-2, and Adh-3, with sizes of approximately 970, 880, and 700 bp, respectively (fig. 1). After cloning, sequences of all three loci were obtained for the middle portion of the locus spanned by the primers for all three species. The complete sequences of the L. stylosa Adh-2 and Adh-3 genes were also obtained, as was the complete sequence of L. uniflora Adh-1 (except for the first 12 bases and the last exon). After alignment, the Leavenworthia sequences were compared with that of A. thaliana alcohol dehydrogenase (Chang and Meyerowitz 1986). Arabidopsis thaliana Adh has six introns, and Adh-1 from L. uniflora has the same number, in identical positions in the amino acid sequence, although most of the intron sizes are slightly different. The same is true for the portions of the Adh-1 sequence which are available from L. stylosa, i.e., from intron 2 to the 3' end of the sequence. Intron 4 is missing from the Adh-2 locus, and no introns are present in the L. stylosa Adh-3 (fig. 1); we presume that they are missing from the loci in the other species, since the band sizes of the PCR products are very similar in all species tested. The L. stylosa Adh-2 coding sequence codes for 379 amino acids, the same size as in A. thaliana, while Adh-3 has one more amino acid near the start of the protein sequence. There are no stop codons or frameshifts in the coding sequence of the Adh-3 locus.

Table 1 compares portions of the coding sequences of Leavenworthia alcohol dehydrogenase loci with the A. thaliana sequence (Landsberg erecta). Entire sequences were not compared because of differences in intron lengths. The Adh-3 sequence differs slightly more from the A. thaliana Adh sequence than does the Adh-1 sequence. Most of the sequence differences are synonymous, although all three loci also differ by a number of amino acid substitutions from the A. thaliana sequence.

Table 2 compares portions of coding sequences of all three loci from the three Leavenworthia species. The three Leavenworthia loci differ considerably, including multiple amino acid differences. The number of silent differences is lower than that between the Leavenworthia...
Alcohol Dehydrogenase Genes of *Leavenworthia* 555

Table 1
Comparison of 669 Nucleotides of Coding Sequences of *Leavenworthia* Alcohol Dehydrogenase Loci with the *Arabidopsis thaliana* Locus

<table>
<thead>
<tr>
<th>Locus</th>
<th>Species</th>
<th>Synonymous Sites</th>
<th>Nonsynonymous Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Adh-1</em></td>
<td><em>L. crassa</em></td>
<td>0.392</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td><em>L. stylosa</em></td>
<td>0.370</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td><em>L. uniflora</em></td>
<td>0.392</td>
<td>0.033</td>
</tr>
<tr>
<td><em>Adh-2</em></td>
<td><em>L. stylosa</em></td>
<td>0.495</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td><em>L. crassa</em></td>
<td>0.521</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td><em>L. stylosa</em></td>
<td>0.503</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td><em>L. uniflora</em></td>
<td>0.496</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Note.—Differences are expressed as probabilities of difference between the sequences per site without Jukes-Cantor correction.

*thia and *A. thaliana* sequences. *Adh-3* differs slightly less between species than the other two alcohol dehydrogenase loci in terms of either synonymous or amino acid differences. It differs more from *Adh-1* than does *Adh-2*, while its difference from *Adh-2* is similar to the *Adh-112* difference.

Finally, in all comparisons, the frequencies of synonymous differences per synonymous site (*P*<sub>S</sub>) are about 10-fold higher than those of nonsynonymous differences per replacement site (*P*<sub>R</sub>), much lower than would be expected if *Adh-3* were evolving neutrally (in which case *P*<sub>R</sub> should be similar in value to *P*<sub>S</sub>; see Nei 1987). Together with the absence of stop codons or frameshifts in the coding sequence of the *Adh-3* locus, this suggests that *Adh-3* either is still expressed and under selective constraints or is a recent duplication of another alcohol dehydrogenase locus and, even if it is not now expressed, has not yet accumulated mutations that have destroyed its function. Recent origin seems unlikely given the large amount of allele sequence diversity within each of the species studied. All three loci, including *Adh-3*, exhibit considerable silent-site diversity (table 2; see also Liu, Zhang, and Charlesworth 1998), implying that there has been time for differences to accumulate under genetic drift.

**Phylogenetic Relationships of the Adh Sequences**

In an attempt to establish the origin of the *Adh-3* gene, we estimated a phylogenetic tree for the sequences obtained, using coding sequence from exons 4 and 5 that was available for several alleles at each of the *Adh* loci. A tree constructed by the NJ algorithm is shown (fig. 2). Consistent with the observation that all *Leavenworthia* species examined have three alcohol dehydrogenase loci, the tree based on the amounts of sequence divergence between their coding sequences shows that the origin of *Adh-3* predates the diversification of these species. As this conclusion is based on only a short sequence, a parsimony tree was also constructed based on 669 nt of coding sequences from exon 3 to exon 5 that were available for a small number of *Leavenworthia* *Adh* alleles. *Arabidopsis thaliana Adh* was used as an outgroup. The topologies produced with the two sets of sequences analyzed were identical in that *Adh-3* appears to be more closely related to *Adh-2* than to *Adh-1*.

The phylogeny of these sequences, based on multiple nucleotide and amino acid differences between *Adh-3* alleles and those of other *Leavenworthia* alcohol dehydrogenases, supports the conclusion that this locus does not have a very recent origin. These findings also rule out the possibility of exchange of sequence information between the loci (for instance, by gene conversion), which might have preserved an apparently functional sequence in this locus, even if it were no longer expressed. Thus, the clear evidence for selective constraints during the evolution of *Adh-3* sequences allows us to conclude that this locus is still expressed despite its lack of introns.

Table 2
Mean Pairwise Numbers of Synonymous Differences per Synonymous Site *P*<sub>S</sub> (above the diagonal) and of Nonsynonymous Differences per Nonsynonymous Site *P*<sub>R</sub> (below the diagonal) Between the Exons of *Adh-1*, *Adh-2*, and *Adh-3* Alleles of *Leavenworthia crassa* (CR), *Leavenworthia stylosa* (ST), and *Leavenworthia uniflora* (UN)

<table>
<thead>
<tr>
<th></th>
<th>CR</th>
<th>ST</th>
<th>UN</th>
<th>CR</th>
<th>ST</th>
<th>UN</th>
<th>CR</th>
<th>ST</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR-<em>Adh-1</em></td>
<td>0.02282</td>
<td>0.03638</td>
<td>0.04541</td>
<td>0.24729</td>
<td>0.26263</td>
<td>0.28767</td>
<td>0.33246</td>
<td>0.31530</td>
<td>0.31941</td>
</tr>
<tr>
<td>ST-<em>Adh-1</em></td>
<td>0.00177</td>
<td>0.03097</td>
<td>0.05058</td>
<td>0.25387</td>
<td>0.27041</td>
<td>0.29426</td>
<td>0.33813</td>
<td>0.32152</td>
<td>0.32532</td>
</tr>
<tr>
<td>UN-<em>Adh-1</em></td>
<td>0.00479</td>
<td>0.00612</td>
<td>0.00360</td>
<td>0.02697</td>
<td>0.02833</td>
<td>0.03043</td>
<td>0.34747</td>
<td>0.33047</td>
<td>0.33728</td>
</tr>
<tr>
<td>CR-<em>Adh-2</em></td>
<td>0.01841</td>
<td>0.01947</td>
<td>0.01606</td>
<td>0.04335</td>
<td>0.05116</td>
<td>0.07925</td>
<td>0.24438</td>
<td>0.23586</td>
<td>0.23467</td>
</tr>
<tr>
<td>ST-<em>Adh-2</em></td>
<td>0.02088</td>
<td>0.02189</td>
<td>0.01830</td>
<td>0.00580</td>
<td>0.02371</td>
<td>0.10390</td>
<td>0.26286</td>
<td>0.25390</td>
<td>0.25439</td>
</tr>
<tr>
<td>UN-<em>Adh-2</em></td>
<td>0.02774</td>
<td>0.02880</td>
<td>0.02540</td>
<td>0.01071</td>
<td>0.01140</td>
<td>0.00000</td>
<td>0.28195</td>
<td>0.27755</td>
<td>0.27383</td>
</tr>
<tr>
<td>CR-<em>Adh-3</em></td>
<td>0.02496</td>
<td>0.02601</td>
<td>0.02276</td>
<td>0.01991</td>
<td>0.02221</td>
<td>0.02617</td>
<td>0.01731</td>
<td>0.03213</td>
<td>0.02308</td>
</tr>
<tr>
<td>ST-<em>Adh-3</em></td>
<td>0.02902</td>
<td>0.03007</td>
<td>0.02654</td>
<td>0.01958</td>
<td>0.02027</td>
<td>0.02408</td>
<td>0.00484</td>
<td>0.00253</td>
<td>0.03743</td>
</tr>
<tr>
<td>UN-<em>Adh-3</em></td>
<td>0.02647</td>
<td>0.02748</td>
<td>0.02362</td>
<td>0.01889</td>
<td>0.02046</td>
<td>0.02410</td>
<td>0.00291</td>
<td>0.00255</td>
<td>0.02340</td>
</tr>
</tbody>
</table>
FIG. 2.—Consensus neighbor-joining tree of *Leavenworthia*, *Arabidopsis thaliana*, and *Arabis* alcohol dehydrogenase sequences. Sequences are labeled with the species name, followed by the locus number and an identification number for each allele. The lengths of terminal branches are not proportional to the distance values, because branch lengths of the tree are proportional to the number of inferred nucleotide site changes according to the reconstruction, rather than to the distance values. For this reason, and for clarity, sequences in sets of very similar sequences are not individually labeled, but groups of alleles are indicated in the figure. Percentages of bootstrap replicates supporting the branches are shown where the values exceeded 70%.

Linkage Relationships of the *Leavenworthia* Alcohol Dehydrogenase Loci

The absence of introns suggests that *Adh-3* is a product of a reverse transcription event involving an mRNA intermediary. Three hundred base pairs of flanking sequence 3’ to the stop codon contained no evidence of a poly-A sequence. If such a tract is present, it must thus be more distant from the stop codon. Attempts to find a nearby poly-A sequence by PCR amplification from *Leavenworthia* DNA using one primer internal to the *Adh-3* locus and a poly-T primer with one other base at the end also failed to amplify a product. There is thus no evidence of any extant poly-A sequence close to the end of the *Adh-3* coding sequence.

An origin for *Adh-3* by reverse transcription interpretation would most likely mean that it would be unlinked to the other alcohol dehydrogenase loci. To test for linkage between these two loci, we therefore identified *L. stylosa* plants that were heterozygous at these loci using SSCP analysis. Primers specific for *Adh-1*, *Adh-2*, and *Adh-3* that amplify small portions of the relevant loci (see fig. 1) were used. The *Adh-3* products were also digested with the restriction enzyme *Rsa* I, and those from *Adh-2* were digested with *Dra* I. These procedures yielded bands that identified each allele of the doubly or triply heterozygous parent of each of our crosses and enabled us to score segregation of the alleles in their progeny. These parents were crossed with plants carrying alleles not present in the heterozygous parent. Individual progeny seeds were obtained, DNA was extracted from them, and their genotypes were scored for the alleles derived from the doubly or triply heterozygous parent.
Expression of the Adh-3 Locus

The evidence given above suggests that Adh-3 is still expressed. To obtain further evidence on the expression of Adh-3, RT-PCR was also done on cDNA prepared from leaves of plants of four *Leavenworthia* species. Expression in leaves of mature plants may appear surprising, as alcohol dehydrogenases are expressed at low levels in leaves of many plants, although some are induced by anoxic stress. Differential regulation of loci has, however, been demonstrated in tomato (Tanksley and Jones 1981), so it is quite possible that *Leavenworthia* species, which grow in conditions where flooding is common (Quartermaster 1950), may show unusual expression. Even in *A. thaliana*, expression occurs in cells in the vascular bundles of leaves (Dolferus and Jacobs 1984). Furthermore, expression was detected in all three *Leavenworthia* species tested as faint staining on cellulose acetate gels, a system which has high sensitivity to detect enzymes, even at low activity.

The RT-PCR results are shown in figure 3. Primers from coding sequence that are specific for Adh-1 yielded a single band of the expected size, 420 bp, from cDNA of several *Leavenworthia* species (lanes 1–4 of fig. 3), and a single larger band (about 507 bp) from genomic DNA (lane 5) of the same species, as expected, since the region amplified includes intron 4. There is thus no detectable contamination of the cDNA with genomic DNA. The primer pair Adh1.P1 and Adh1.Cp4R (which includes some sequence from intron 5) yielded no visible bands from cDNA (lanes 14–17 of fig. 3), but only from genomic DNA (lane 18 of fig. 3), and the size corresponded with that expected. The results for Adh-2 are similar to those for Adh-1 (only *L. stylosa* and *L. uniflora* were tested): primers P3 and P4, for the region of Adh-2 where the missing intron in Adh-1 is located, yielded the expected 388-bp bands from both types of DNA source (lanes 6–8 of fig. 3). Clear bands were obtained from either genomic or cDNA with primers P1 and P3R designed for the Adh-3 locus, and the bands were the size expected (438 bp) given the absence of introns (lanes 9–13 of fig. 3).

Conclusions and Discussion

In most plant species studied, there are two or three class P Adh loci (e.g., Trick et al. 1988; Yokoyama and Harry 1993). Differences in the number of loci encoding a given type of enzyme are not uncommon (e.g., D’Hondt et al. 1997), and differences in the numbers of introns are known in several cases of type P alcohol dehydrogenase loci. For instance, the maize Adh-1 locus has nine introns, but barley Adh-2 and Adh-3 have fewer (Trick et al. 1988), and *A. thaliana* has only six introns. Changes in intron length are common (e.g., Denda et al. 1995). It is, however, surprising for all introns to be absent, and we are not aware of any previous such case in plants.

On the evidence available so far, it seems unlikely that the Adh-3 of *Leavenworthia* is a pseudogene. Apart from a single extra codon near the beginning of the sequence, there are no insertions, deletions, or stop codons. Adh-3 is present in all five species studied, suggesting that it has been present in their genomes since before the species split from a common ancestor. Despite considerable sequence differences from the other alcohol dehydrogenase loci, consistent with a long time period for its evolution, most of the changes in the coding portions of the sequences are synonymous, rather than causing amino acid substitutions. This is true for differences from the *A. thaliana* Adh sequence, and also in comparisons of Adh-3 between different *Leavenworthia* species or between the different alcohol dehydrogenase loci in these species, suggesting functional constraints on this locus. Expression is confirmed by the RT-PCR results. We have so far been unable to find any evidence that the Adh-3 locus has a poly-A sequence in the 3′ flanking region, as would be expected if it were
a processed pseudogene. It is possible that further sequence from this region will reveal a poly-A sequence farther downstream than is usual for plant mRNAs (Wu, Ueda, and Messing 1995). In the messenger RNA for A. thaliana Adh, the poly-A sequence starts after 310 bp, and an even longer interval is found in the barley Adh-1 mRNA (Good, Pelcher, and Crosby 1988). However, accumulation of mutations downstream of the coding sequence is likely, given the observed synonymous-site divergence of Adh-3 from Adh-1 or Adh-2, so it may no longer be possible to detect the presence of this sequence.

At present, the origin of the Adh-3 locus is unknown. Two questions are of interest. One is whether this locus is a tandem duplication of one of the other alcohol dehydrogenase loci. If it arose from an mRNA precursor, as in the case of processed pseudogenes, its location would most likely not be close to the locus from which it originated, because the insertion event could occur anywhere in the genome. We have established that Adh-3 is not closely linked to either the Adh-1 or the Adh-2 locus.

A second question is when the Adh-3 of Leavenworthia species lost the introns. This could be a relatively recent loss of introns that were present in an ancestral locus (presumably either Adh-1 or Adh-2), or the locus might represent a gene that has persisted without introns from a common ancestor of both Leavenworthia and Arabidopsis (with the lineage leading to the latter species having lost this locus). The first of these possibilities would be supported if the Leavenworthia sequences cluster together and differ from other alcohol dehydrogenase sequences. If, on the other hand, the Leavenworthia sequence for one or another of the loci with introns present is strongly similar to the Arabidopsis Adh, and Adh-3 differs from this sequence, the Adh-3 gene might represent a distinct locus that could have been present in a common ancestor. Since the three Leavenworthia Adh loci differ roughly equally from Arabidopsis Adh, and our phylogenetic analysis suggests that Adh-2 and Adh-3 form a clade, this is evidently not the case. It therefore appears most likely that the loss of introns occurred after the split between Leavenworthia and Arabidopsis. At present, no suitable outgroup exists other than A. thaliana and its close relative Arabis. It will also be helpful to study the taxonomic distribution of the intronless state. On the current evidence, we should expect it to be confined to Leavenworthia and to be absent from other species of the Brassica family and other plant families.

In L. stylosa, Adh-3 has one more amino acid than Adh-2, the extra residue being either codon 3 or 4, depending on the alignment chosen. The Adh loci of other plants also differ by the presence or absence of an amino acid residue in this position. In maize, barley, and A. thaliana, for example, exon 1 is 34 bp long, as in L. stylosa Adh-2, while in Vitis vinifera Adh1, a strawberry Adh, and L. esculentum Adh2, it is 37 bp long, as in L. stylosa Adh-3. It is not clear whether these differences represent several changes during the evolution of the angiosperms or whether there are two ancient lineages, one with the 34- and the other with the 37-bp condition, but it is clear that the 37-bp state does not characterize a locus resembling the Leavenworthia Adh-3, because no other locus among this set is intronless.

Whatever the origin of Adh-3, it is interesting that a locus that belongs to a gene family can lack the introns present in other members of the family. A somewhat similar case was recently discovered in Drosophila yakuba and Drosophila tessieri. The locus, now known as jingwei, which has sequence homology to Adh, is on chromosome 3R, rather than the usual Adh location on 2L (Jeffs and Ashburner 1991). The jingwei locus was initially thought to be a processed pseudogene because of the absence of introns and the apparent lack of evident promoter sequences and of the initiating codon, but several findings were not entirely consistent with this interpretation. These include the absence of stop codons or length differences in the coding sequence and the fact that most of the sequence differences from Adh are silent substitutions. Recently, Long and Langley (1993) reported evidence for expression in the mRNA population. They also showed that there are coding sequences 5' to the Adh-derived sequences and that an initiating codon is indeed present in this upstream region. Furthermore, while polymorphism levels of this gene are similar to those for the normal Adh, sequence evolution has been more rapid than for the normal locus, suggesting that the jingwei locus has evolved a new function. In the case of Leavenworthia Adh-3, we also find no stop codons or frameshifts in the coding sequence, and the majority of substitutions between and within species are silent, so the gene appears to be expressed. However, differences between Leavenworthia species for Adh-3 appear to be fewer than those for Adh-1 (table 2). Thus, although the difference is not statistically significant, there is certainly no evidence that this locus has evolved faster than the other alcohol dehydrogenase loci.

Acknowledgments

We thank the staff of the University of Chicago greenhouses for excellent plant care; Drs. E. E. Lyons, G. Hilton, and T. E. Hemmerly for plant material; and P. Awadalla for running MEGA. This work was supported by NIH Program Project “Genetic Mechanisms of Evolution, Constraints on Molecular Variation and Evolution” (with M. Kreitman and B. Charlesworth, PO1 GM-50355). We thank Marty Kreitman for training in molecular methods, discussions, and comments on an earlier version of this paper.

LITERATURE CITED


Jeffrey Powell, reviewing editor

Accepted January 22, 1998