RNA Folding in Drosophila Shows a Distance Effect for Compensatory Fitness Interactions

Wolfgang Stephan and David A. Kirby

Department of Zoology, University of Maryland, College Park, Maryland 20742

Manuscript received February 26, 1993
Accepted for publication May 15, 1993

ABSTRACT

Phylogenetic-comparative analysis was used to construct a secondary structure of *Adh* precursor messenger RNA (pre-mRNA) in Drosophila. The analysis revealed that the rate of coevolution of base-pairing residues decreases with their physical distance. This result is in qualitative agreement with a model of compensatory fitness interactions which assumes that mutations are individually deleterious but become harmless (neutral) in appropriate combinations. This model predicts that coupled mutations can become fixed in a population under mutation pressure and random genetic drift, when the mutations are closely linked. However, the rate of joint fixation drops as distance between sites increases and recombination breaks up favorable combinations. RNA secondary structure was also used to interpret patterns of linkage disequilibrium at *Adh*.

THE question of whether evolution can proceed on a surface of selected values from one adapted peak to another one, passing through a valley of intermediate deleterious states, was first raised by HALDANE (1931) and WRIGHT (1932). KIMURA (1985) has formulated this problem at the molecular level. Consider a model with two loci (nucleotide sites). Mutations at these sites are assumed to be individually deleterious but harmless (neutral) in appropriate combinations. This model of compensatory fitness interactions predicts that double mutants can easily become fixed in a population under continued mutation pressure and random genetic drift, when the two sites are closely linked. However, the rate of joint fixation may drop dramatically, as distance between sites increases and recombination breaks up favorable combinations. Compensatory fitness interactions may therefore play an important role in determining the size of a coevolvable genetic unit ("gene").

RNA secondary structure provides a unique opportunity to test this model. Mutations occurring in a pairing region of a secondary structural element are individually deleterious if they destabilize a functionally important structure. But fitness can be restored, when a compensatory mutation occurs that reestablishes the pairing potential. To examine the model of compensatory neutral mutations, we constructed a secondary structure of a large RNA, *Adh* pre-mRNA in Drosophila.

At this time, the most powerful *a priori* approach to elucidating the secondary structure of a large RNA has proven to be the use of phylogenetic DNA sequence comparisons (FOX and WOESE 1975; JAMES *et al.* 1988). In this method, possible pairing regions are identified by seeking the equivalent pairing in a homologous RNA from a different species (in which the sequences vary from species to species because of base substitutions during evolutionary time). Covariation of paired residues in a putative pairing region, such that the pairing potential is preserved, offers support for the structure. So far, the phylogenetic method has only been used to infer secondary structures of transfer RNAs, ribosomal RNAs and ribozymes.

MATERIALS AND METHODS

For phylogenetic comparison we used five species of the subgenus *Sophophora* (*Drosophila melanogaster* (KREITMAN 1983), *Drosophila teissieri* (JEFFS and ASHBURNER 1991) and *Drosophila erecta* (ASHBURNER 1992) of the *melanogaster* species subgroup; *Drosophila pseudoobscura* (SCHAEFFER and AQUADRO 1987) and *Drosophila ambigua* (MARFANY and GONZALEZ-DUARTE 1991) of the *obscura* group), four species of the subgenus *Drosophila* (*Drosophila mulleri* (*Adh-2*) (FISCHER and MANIATIS 1985) and *Drosophila hydei* (*Adh-2*) (MENOTTI-RAYMOND, STARMER and SULLIVAN 1991) of the *repleta* group, and *Drosophila affinidisjuncta* (ROWAN and DICKINSON 1988) and *Drosophila silvestris* (ROWAN and HUNT 1991) of the Hawaiian picture-wing species group), and one species of the subgenus *Scaptodrosophila* (*Drosophila lebanonensis*; JUAN, PAPACEIT and QUINTANA 1990). These species are from the subfamily Drosophilinae. Sequences from drosophilids outside this subfamily or from families related to Drosophilidae are not known. The species were chosen to be sufficiently different to provide numerous instances of sequence variation such that pairing possibilities could be tested. The subgenus *Scaptodrosophila* is considered to be close to the base of the Drosophilinae radiation (THROCKMORTON 1975). Sequence alignments within species groups were constructed with the programs GAP and BESTFIT (DEVEREUX, HAEBERLI and SMITHIES 1984). Then, alignments were done successively between species groups and subgenera. The alignment of coding sequences presented no ambiguities. But in the introns and the 3′- and

5′-flanking regions, the alignment had to be done manually based on the pairwise and multiple alignments within groups.

Two types of complementary sequences are likely to occur in large RNAs: (i) short-range pairings that create hairpins and (ii) long-range pairings that order several short-range pairings into discrete structural units (JAMES *et al.* 1988). The long-range pairings include also pseudoknots, which arise by pairing of sequences within a secondary structural element (*e.g.*, a loop of a stemloop structure) with sequences from outside this structure (PLEIJ, RIETVELD and BOSCH 1985). Based on this classification, we constructed the *Adh* secondary structure in a hierarchical way. First, we searched within the *melanogaster* subgroup for inverted repeats that were ≤50 bp apart from each other by the program STEMLOOP. These potential pairing regions were then subjected to phylogenetic analysis. In the second round, we searched the unpaired regions of the locus for long-range pairings, starting with the nearest pair of still unpaired regions and moving to more distant pairs.

The phylogenetic analysis was done manually based on the following criterion: a putative helix was considered "proven" if two or more covariations, caused by independently occurring base substitutions in the complementary sequences of a putative helix, were found (FOX and WOESE 1975; JAMES *et al.* 1988). Under simplifying conditions, this criterion may be justified as follows. Consider two species that have been separated for a very long time. Assume that a Watson-Crick pair (A-U or G-C) in a putative helix is changed by substitutions into one of 15 other possible states. Under the assumption that the two species are separated for a very long time, the substitution process along each lineage can be considered stationary. Therefore, if the mutation process followed a one-parameter model (JUKES and CANTOR 1969), such that transitions and transversions were equally frequent, all 15 combinations of nucleotide pairs (into which a Watson-Crick pair is changed) should be equally likely. Under stationarity, the same was true for a two-parameter model (KIMURA 1980a), in which transitions and transversion are allowed to occur at different rates. [Under stationary conditions, the frequency of each of the four mutational states is 1/4, if transition and transversion rates are positive. First, it can be immediately seen that 1/4 is a stationary state because of symmetry in the mutation scheme. Second, the stationary solution is unique, because the graph associated with the two-parameter mutation model is connected, such that any two pairs of vertices are connected by at least one sequence of edges (HAKEN 1977, Chap. 4).] Of these 15 possible states, three maintain Watson-Crick pairing. Thus, the probability of a change from one Watson-Crick pair to another one is $3/15 = 1/5$. Furthermore, assuming that substitutions in a stem are independent of each other, the probability of observing two such Watson-Crick covariations in a region will be the product of this probability ($1/5 \times 1/5 = 0.04$). This value is below the significance level of 0.05. To our knowledge, a more general criterion that would allow for arbitrary mutation schemes and for comparison of a range of species with different genetic distances has not been worked out. In the present form, the criterion for a proven helix does not take detailed phylogenetic information into account. In particular, potentially important information on intermediate states (non-Watson-Crick pairs) is disregarded.

The phylogenetic method is usually applied to a sufficiently wide range of species such that enough substitutions could have occurred since their common ancestor. A wide range of species is needed to account for the different rates at which different parts of a molecule may evolve. If only one covariation in a putative helix is detected because of limitations in the species range and/or sequence conservation, information on intermediate states becomes important. We considered the presence of G·U pairs at at least one other covarying site as supporting evidence, because intermediate G·U pairs are almost always present in one or more species, when a transition from one Watson-Crick pair to another one occurs (ROUSSET, PÉLANDAKIS and SOLIGNAC 1991). In addition, covariations caused by insertion/deletion in the complementary sequences that did not disrupt a pairing region were also considered as support for the structure.

RESULTS

***Adh* pre-mRNA secondary structure:** Two different transcripts have been identified in adult and larval tissues of *D. melanogaster* that differ in their 5′-noncoding leader sequences (BENYAJATI *et al.* 1983). We had to limit our phylogenetic analysis to sequences homologous to *D. melanogaster* larval pre-mRNA, because the physical organization of the 5′-region of the *Adh* locus is very different in the subgenera *Drosophila* and *Sophophora* (MENOTTI-RAYMOND, STARMER and SULLIVAN 1991). Based on the analysis of the long-range structure of total larval pre-mRNA, we were able to divide the molecule into two parts, the 5′-leader sequence and the rest comprising the translated portion of Exon 2, Intron 2, Exon 3, Intron 3, Exon 4 and the 3′-untranslated region (coordinates 778–1858; KREITMAN 1983). The 5′-leader sequence (in *D. melanogaster* at 708–777) could be treated separately because no RNA-RNA interaction between 5′-leader sequences and the rest of the *Adh* pre-mRNA was detected.

Based on sequence comparison of 10 species, we identified one putative pairing region in the 5′-leader sequence and 22 putative pairing regions in the rest of the larval pre-mRNA of *Adh* comprising approximately 1080 nucleotides. A summary of the phylogenetic support for these pairings is given in Table 1. Most of them are short-range pairings (hairpins). The physical distances between the residues forming the bottom pair of these stems are usually less than 30 bp. In contrast, the long-range pairings involve sequences that may be separated by much larger distances. We were unable to detect any pseudoknots. As indicated in Table 1, pairings were found in both coding and noncoding portions of the *Adh* pre-mRNA molecule.

The secondary structure models of *Adh* pre-mRNA vary between the 10 species compared. There is some structural variation in the long-range pairings that order the molecule into discrete subdomains. Table 1 indicates that not all long-range pairings are conserved across the whole species range examined. Most structural variation is in the sequence and stem length of the hairpins. The secondary structure of the hairpin in Intron 2 is given as an example. Figure 1 shows that the hairpin structures detected in Intron 2 exist

TABLE 1

Phylogenetic support for pairings in the *Adh* pre-mRNA secondary structure model

Paired regions	Species range	Covariations	Comments
(i) Hairpins			
717–719/743–745	S/leb	2/0/2	5′ leader
821–828/836–843	all	1/0/2	
846–849/855–858	all	2/0/1	
899–904/911–916	all	5/1/2	Intron 2
946–951/955–961	all	2/0/2	
1017–1021/1038–1042	all	1/0/1	
1099–1106/1113–1120	all	1/0/1	
1207–1217/1231–1241	all	0/1/2	Structural change due to two substitutions in one stem sequence
1253–1263/1268–1278	all	1/0/1	
1434–1441/1446–1453	all	1/0/3	
1467–1470/1482–1485	S/H/leb	1/0/3	
1530–1535/1539–1543	M/D	2/0/0	
1562–1568/1608–1615	all	1/0/1	
1639–1648/1666–1676	all	1/0/2	
1712–1714/1719–1721	S	1/0/1	Present only in *Sophophora*
1772–1774/1802–1804	S/leb	2/0/3	3′ untranslated region
1826–1831/1836–1841	M/D	2/0/2	3′ untranslated region
(ii) Long-range pairings			
783–793/1847–1855	M	0/1/2	Present only in *Sophophora*
809–819/1756–1766	all	1/0/2	
991–1000/1158–1167	all	1/1/2	
1050–1056/1132–1139	all	1/0/3	
1188–1199/1313–1324	all	1/0/1	
1459–1463/1630–1634	all	1/0/1	G·U ↔ U·G "covariation"

The first column contains the coordinates of the pairing regions in *D. melanogaster* according to KREITMAN (1983). The second one shows the species range over which the helix is conserved. The abbreviations are: D, *Drosophila*; H, Hawaiian species group; leb, *D. lebanonensis*; M, *melanogaster* species subgroup; S, *Sophophora*. The numbers in the third column give (i) the number of covariations due to individual base substitutions, (ii) the number of structural covariations due to insertion/deletion events and/or changes of longer sequences in one or both sequences of a stem and (iii) the number of Watson-Crick pairs showing G·U intermediates. The fourth column gives the reasons why certain putative helices are not included in Figure 2. Furthermore, the pairings found in the 5′ and 3′ untranslated regions and in the introns are indicated.

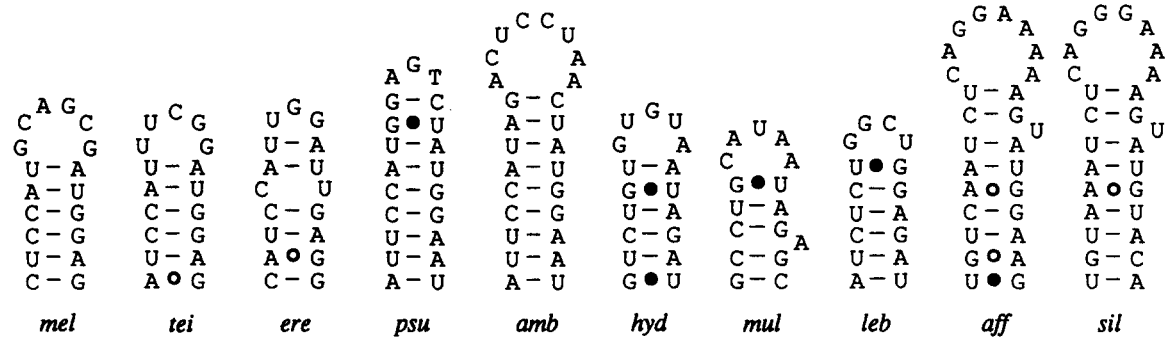

FIGURE 1.—Variations of hairpin structures in Intron 2. Alignment was done separately for the *Sophophora/Scaptodrosophila* subgenera (including the *repleta* group) and for the Hawaiian picture wing species group. In the first group of species, the pairing is supported by five covariations, in the latter one by a covariation and a U·G ↔ U-A change. The structures were drawn using the program LOOPDLOOP (GILBERT 1992).

for all 10 species. They occur in two basic forms, one of which is present in the two species of the Hawaiian picture wing group, the other one in the eight species of the *Sophophora/Scaptodrosophila* subgenera and the *repleta* group. This structural conservation is remarkable given that the primary sequence is rather variable.

We have evidence that the phylogenetic method was not able to pick up all possible helices in the *Adh* primary transcript. Because of the limited species range considered, sequences in some putative helices are too conserved to exhibit enough covariations to satisfy our criteria. On the other hand, there is evidence that structural elements were not detected because the pairing is too weak such that the structure is present only in one species, but absent in other species of the same species group. As discussed below, such weakly conserved structures can often be picked

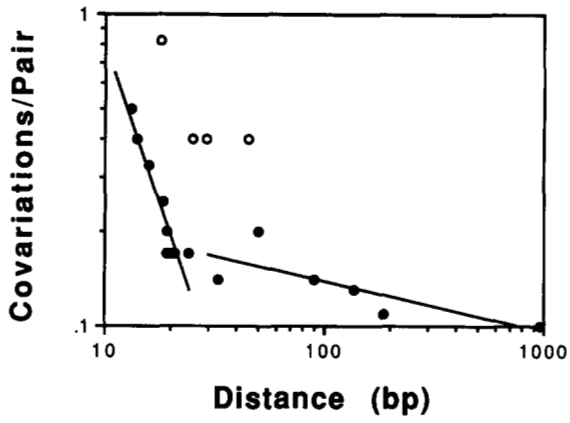

FIGURE 2.—Covariations per base-pairing residues *vs.* their physical distance (in base pairs). The log-log plot shows the *Adh* data from Drosophila (Table 1), where the filled circles represent pairing regions involving coding pre-mRNA and the open circles noncoding regions. The lines represent linear regression fits of the log-log transformed data for distances <30 nucleotides and also for greater distances. The regression lines are fitted only to the data from coding regions. Distances were calculated between the residues forming the bottom pair of a stem by averaging over the subgenera (with equal weights). The number of covariations/pair was calculated as the number of covariations per length of the core stem, *i.e.*, the stem which is preserved across species. Four pairings were not included for reasons indicated in Table 1 (last column).

up (particularly in noncoding regions) if linkage disequilibrium data based on intraspecific DNA sequence variation are available. Although the observational window of the phylogenetic method is limited, this has only little effect on the main analysis given in the next section.

Distance effect for compensatory mutations in Drosophila: In Figure 2, the number of covariations per base pair (of a helix) is plotted against their physical distance (in base pairs). To put the evolutionary process onto a comparable time scale, only helices that are sufficiently conserved across subgenera were included. We omitted pairings that exist only in Sophophora. Whereas nearly all pairings that occur in coding regions are conserved across all subgenera, the pairings involving noncoding pre-mRNA are not conserved. Therefore, for noncoding helices, the numbers of covariations per base pair are likely to be underestimates. Despite this, Figure 2 indicates that the rate of coevolution in noncoding pairing regions is higher than in coding ones. This is in accordance with the observation that substitution rates in noncoding DNA are generally higher, which is usually explained by differences in selective constraints.

The number of pairings from coding regions is sufficiently large to analyze KIMURA's (1985) model. KIMURA computed the expected time to fixation of double mutants, $E\{T(p_1, p_2)\}$, under continuous mutation pressure, recombination and random genetic drift, where p_i ($i = 1, 2$) are the initial frequencies of these mutants. Single mutations are assumed to be deleterious but neutral in appropriate combinations that, in the context of RNA, are Watson-Crick pairs. From $E\{T(p_1, p_2)\}$, the expected number of fixation of double mutants per generation, or the rate of coevolution, is obtained as the inverse of $E\{T(0, 0)\}$. We follow here the usual definition of a rate. For single mutants, it is consistent with the definition of the rate of molecular evolution, which can be seen as follows. The rate of molecular evolution, k, is defined as the product of the per generation mutation rate times the probability of fixation of a mutant. This is not a correct definition of a rate in that it neglects the time to fixation. The definition of the fixation probability only requires that the mutant goes eventually to fixation. For single mutants, it can be shown that $E\{T(0)\}^{-1}$ is equal to the rate of molecular evolution, k, if $4N\mu << 1$, where N is the effective population size and μ the per nucleotide mutation rate. For instance, in the neutral case $E\{T(0)\} = 4N[\Psi(4N\mu) + \gamma]/(4N\mu - 1)$, where Ψ is the digamma function and $\gamma = 0.577$ (KIMURA 1980b). For $4N\mu << 1$, this reduces to the well known formula for the rate of molecular evolution $E\{T(0)\}^{-1} = k = \mu$.

Figure 2 shows the general trend for the number of covariations per base pair (in helices from coding regions) to decrease with distance. This property is qualitatively in agreement with KIMURA's (1985) model, which predicts that $E\{T(0, 0)\}$ increases with the distance between base-pairing nucleotides. However, it is at present not clear whether this model can also be used for quantitative predictions. It is noteworthy that the decay in Figure 2 is much more dramatic for short-range pairings than for longer-range ones. Average recombination rates per bp in Drosophila are about 10^{-8} or higher (CHOVNIK, GELBART and MCCARRON 1977; JUDD 1987). Given these relatively high recombination rates and an effective population size, N, in the order of 10^6, the model of compensatory neutral mutations (KIMURA 1985) may predict a significant drop of the rate of coevolution of base-pairing residues over short distances, which may explain the steep decay for short-range pairings. However, for this to happen $4N\mu$ has to be in the order of 1. The large number of covariations we observed in Drosophila (Table 1) indicates that the value of $4N\mu$ is sufficiently high to allow the spread of compensatory mutations. This may be partly due to the fact that synonymous substitution rates in Drosophila are rather high (10^{-8} per site per year or higher; MORIYAMA and GOJOBORI 1992). In addition, in pairing regions the rate for secondary (compensatory) mutations may be increased by "templated mutation." In this process, the frequency of nucleotide substitutions is influenced by the repetitive DNA structure, which, in the case of pairing regions, is palindromic. An appreciable portion of mutations at *Adh* appear to be templated by palindromes (GOLDING

1987). However, it is unlikely that $4N\mu$ is much greater than 0.1. As OHTA (1989) pointed out, KIMURA's approximate analysis of his model leads to an overestimation of the expected time to fixation of a double mutant, when $4N\mu << 1$. It is therefore of limited use. An analysis of a model of compensatory neutral mutations that was constructed specifically for RNA secondary structure gives a more satisfactory result, even when $4N\mu$ is low (W. STEPHAN, unpublished data). This model that describes the covariation process from one Watson-Crick pair to another one by considering only the most common intermediate state G·U seems to account for the rapid decay of the rate of coevolution for base-pairing residues in short-range helices. However, more work on this model is required.

For longer range interactions, the rate of coevolution decreases slowly. This may be due to selection that counteracts the retarding effects of recombination. In larger helices, selective constraints on single Watson-Crick pairs appear to be relaxed. This is suggested by the highly significant correlation between length of pairing regions and physical distance (Spearman correlation coefficient $r_s = 0.90$, $P < 0.001$).

DISCUSSION

Distance effect: Our analysis of covariations in larval *Adh* pre-mRNA of Drosophila indicates that the rate of compensatory substitutions decreases with increasing physical distance (in base pairs) between base-pairing residues. Assuming that the recombination rate between sites increases with their physical distance, this result may be interpreted based on KIMURA's (1985) model of compensatory fitness interactions. This model predicts that the rate of joint fixation of coupled mutations decreases with increasing recombination distance. In this model, single mutations remain in low frequency, because they are assumed to be deleterious in that they may destabilize functionally important structures. Most double mutants (new Watson-Crick pairs) are formed by a second mutation on a singly mutant chromosome. As long as the double mutants are in low frequency, the primary role of recombination is to decrease their frequency by crossing over with the wild type (KIMURA 1985). Thus, recombination retards the fixation of double mutants.

More data on the *Adh* locus are needed to put the conclusions of this study on a firmer basis. It is also necessary to examine whether a distance effect can be found in other RNAs, such as ribosomal RNAs. So far, we analyzed the secondary structure of ribonuclease P RNA in bacteria which was constructed by phylogenetic analysis of 23 species from five of the 11 major phylogenetic branches of bacteria (JAMES *et al.* 1988; BROWN and PACE 1992). This RNA is approximately 360 nucleotides long. Eleven pairing regions are conserved across phyla. In contrast to Drosophila *Adh*, these data show no distance effect (D. A. KIRBY and W. STEPHAN, unpublished data). It is noteworthy that most covariations/pair occur in the terminal pairing region. If we eliminate this longest range pairing as an outlier, there seems to be a weak decay. However, the nonparametric Spearman test of rank order for the reduced data set is also not significant ($r_s = -0.53$; $P > 0.1$). Furthermore, unlike at *Adh*, we found no correlation between stem length and distance ($r_s = 0.29$; $P > 0.2$). These apparent differences between Drosophila *Adh* pre-mRNA and bacterial ribonuclease P RNA may be attributed to differences in recombination rates and/or selective constraints.

Functional significance of RNA secondary structure: A functional role of RNA secondary structure in splicing (SOLNICK 1985; EPERON *et al.* 1988; ENG and WARNER 1991) and gene expression (LIEBHABER, CASH and ESHLEMAN 1992) has been suggested. Pre-mRNA secondary structure is thought to play an important role in splice site selection. The access of the splicing machinery to splice sites appears to be blocked if these sites are placed within pairing regions of hairpins (SOLNICK 1985; EPERON *et al.* 1988). We found that both 5′- and 3′-splice sites in the *Adh* introns are not locked into long stable pairing regions. However, it is still possible that the very conserved 5′- and 3′-splice sites are parts of short helices, which are too short to be detected by our analysis of covariations. Species outside the subfamily Drosophilinae have to be used to increase the depth of our phylogenetic analysis.

The branch point sequence, which base-pairs with U2 snRNA, is another major determinant for the assembly of the splicing apparatus (PARKER, SICILIANO and GUTHRIE 1987). We found that the branch point is not sequestered in a pairing region. In both *Adh* introns, the branch point sequences (at coordinates 918 and 1392 in Intron 2 and 3, respectively) are unpaired. In Intron 2, the branch point is immediately downstream of the hairpin detected by phylogenetic comparison. In Intron 3, we did not detect a secondary structural element that was sufficiently conserved across all 10 species. However, in the weakly conserved structure detected in *D. pseudoobscura* and its close relatives (discussed below), the branch point is also in the loop of a hairpin. To our knowledge, the only case in which the structure of RNA containing the branch point has been analyzed experimentally reports a similar observation. The branch point of the first intron of human β-globin pre-mRNA appears to be in the loop of a stem/loop structure (HALL, GREEN and REDFIELD 1988).

There is some evidence that RNA secondary structure may be important for *Adh* expression. LAURIE,

BRIDGHAM and CHOUDHARY (1991) have suggested that an insertion, ∇1, at position 448 in *D. melanogaster Adh-Fast* alleles may cause the difference in ADH protein level associated with the Fast/Slow allozyme polymorphism. This insertion occurs in the adult intron. We have used phylogenetic comparisons to analyze the region around the insertion site. Because the physical organization of the *Adh* locus is very different in the subgenus *Drosophila* (see above), only species from the subgenera *Sophophora* and *Scaptodrosophila* were included in the comparison. We were able to align the sequences from these species with the *Slow* allele of *D. melanogaster*. We found that the region around the insertion site is one of the few regions in the adult intron that is conserved across the two subgenera. Furthermore, we detected a putative pairing region 425–430/457–462 (coordinates according to KREITMAN 1983) that is supported by two covariations in the two subgenera (D. A. KIRBY and W. STEPHAN, unpublished data). Fast alleles carrying ∇1 have the sequence 448–476 removed (which encompasses the pairing region) and replaced by a different larger sequence, such that a hairpin structure can no longer be formed. This indicates a variation in secondary structure between *Fast* and *Slow* alleles. It is, however, uncertain whether this structural difference within the adult intron can affect the level of ADH protein. In particular, LAURIE and STAM (1988) have not found differences in mature RNA levels between *Slow* and *Fast*. To our knowledge, a mechanism whereby structural variation in intron pre-mRNA influences translation efficiency has not been described.

Interpretation of linkage disequilibrium patterns: For population geneticists, analysis of RNA secondary structure and compensatory fitness interactions may become very important for the interpretation of patterns of linkage disequilibrium. Extensive linkage disequilibrium has been observed in recent DNA polymorphism studies in Drosophila. For instance, in a survey of DNA sequence variation in the *Adh* gene region of *D. pseudoobscura*, clustering of significant linkage disequilibria was found in two small regions: one in Intron 1 and the other one in Intron 3 (S. W. SCHAEFFER and E. L. MILLER, unpublished data). Outside these regions significant linkage disequilibria appear to be randomly scattered. To interpret this pattern, pre-mRNA secondary structure can be used. So far, we have examined linkage disequilibrium in these two regions, where the clustering occurs. In the case of Intron 3, we used only sequences from *D. pseudoobscura* and its sibling species *Drosophila persimilis* and *Drosophila miranda* in our phylogenetic analysis, since this intron is very diverged (see above). We found evidence for a hairpin in this region within this species complex. One stemloop structure is found in *D. miranda*, an alternative one that differs by three covariations in *D. persimilis*. In *D. pseudoobscura* we found a polymorphism of these two stemloop forms and various transient intermediates (D. A. KIRBY and W. STEPHAN, unpublished data). The hairpin structure ranges from coordinate 1463 to coordinate 1501 (the coordinate system is from SCHAEFFER and AQUADRO 1987). This result is noteworthy. It means that the covariation process from one Watson-Crick pair to another one can be observed within a single species. A similar transition with the same "endpoints" of the evolutionary trajectory is found in the population (subspecies) from Bogotá (SCHAEFFER and MILLER 1991). Since there are no intraspecific data available from *D. persimilis* and *D. miranda*, it is not known whether this polymorphism exists also in these two sibling species of *D. pseudoobscura*. Outside the *D. pseudoobscura* complex these structures do not exist or are different.

The clustering of linkage disequilibria in Intron 1 can also be explained by compensatory interactions on the pre-mRNA level. In this case, phylogenetic comparison of all five species of the subgenus *Sophophora* revealed a hairpin structure that is supported by two covariations. One of the covariations was found within the *obscura* group, the other one between the *melanogaster* and *obscura* groups. This indicates that this secondary structure is more conserved than that in Intron 3. Consistent with this explanation is the observation that the coevolutionary process taking place in *D. pseudoobscura* does not feature the whole transition from one Watson-Crick pair to another one. Only one Watson-Crick pair and various intermediates were observed in the survey of *D. pseudoobscura*. The hairpin structure ranges from coordinate 317 to coordinate 355 (according to the coordinates of SCHAEFFER and AQUADRO 1987).

We thank LIN CHAO for stimulating discussions and comments on this manuscript; and STEVEN SCHAEFFER for sharing his unpublished DNA sequence data. We also thank ANDREW CLARK and two anonymous reviewers for valuable suggestions which improved this manuscript. This work was supported by a Shannon Award (GM 46233) from the National Institutes of Health and a Biomedical Research Award from the University of Maryland.

LITERATURE CITED

ASHBURNER, M., 1992 GenBank, accession number X54116.

BENYAJATI, C., N. SPOEREL, H. HAYMERLE and M. ASHBURNER, 1983 The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5′ end in different developmental stages. Cell **33:** 125–133.

BROWN, J. W., and N. R. PACE, 1992 Ribonuclease P RNA and protein subunits from bacteria. Nucleic Acids Res. **20:** 1451–1456.

CHOVNICK, A., W. GELBART and M. MCCARRON, 1977 Organization of the *rosy* locus in *Drosophila melanogaster*. Cell **11:** 1–10.

DEVEREUX, J., P. MAEDERLI and O. SMITHIES, 1984 A compre-

hensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 384–395.

ENG, F. J., and J. R. WARNER, 1991 Structural basis for the regulation of splicing of a yeast messenger RNA. Cell **65:** 797–804.

EPERON, L. P., I. R. GRAHAM, A. D. GRIFFITHS and I. C. EPERON, 1988 Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? Cell **54:** 393–401.

FISCHER, J. A., and T. MANIATIS, 1985 Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. Nucleic Acids Res. **13:** 6899–6917.

FOX, G. E., and C. R. WOESE, 1975 5S rRNA secondary structure. Nature **256:** 505–507.

GILBERT, D. G., 1992 LOOPDLOOP available via anonymous ftp to ftp.bio.indiana.edu.

GOLDING, G. B., 1987 Nonrandom patterns of mutation are reflected in evolutionary divergence and may cause some of the unusual patterns observed in sequences, pp. 151–172 in *Genetic Constraints on Adaptive Evolution*, edited by V. LOESCHCKE. Springer-Verlag, Berlin.

HAKEN, H., 1977 *Synergetics—An Introduction*. Springer-Verlag, Berlin.

HALDANE, J. B. S., 1931 A mathematical theory of natural selection. VIII. Stable metapopulations. Proc. Cambridge Philos. Soc. **27:** 137–142.

HALL, K. B., M. R. GREEN and A. G. REDFIELD, 1988 Structure of a pre-mRNA branch point/3′ splice site region. Proc. Natl. Acad. Sci. USA **85:** 704–708.

JAMES, B. D., G. J. OLSEN, J. LIU and N. R. PACE, 1988 The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. Cell **52:** 19–26.

JEFFS, P., and M. ASHBURNER, 1991 Processed pseudogenes in Drosophila. Proc. R. Soc. Lond. Ser. B **244:** 151–159.

JUAN, E., M. PAPACEIT and A. QUINTANA, 1990 Nucleotide sequence of the *Adh* gene of *Drosophila lebanonensis*. Nucleic Acids Res. **18:** 6420.

JUDD, B. H., 1987 The *white* locus of *Drosophila melanogaster*, pp. 81–94, in *Results and Problems in Cell Differentiation 14—Structure and Function of Eukaryotic Chromosomes*, edited by W. HENNIG. Springer-Verlag, Berlin.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–132, in *Mammalian Protein Metabolism*, edited by H. N. MUNRO. Academic Press, New York.

KIMURA, M., 1980a A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

KIMURA, M., 1980b Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical, and pseudo-sampling methods. Proc. Natl. Acad. Sci. USA **77:** 522–526.

KIMURA, M., 1985 The role of compensatory neutral mutations in molecular evolution. J. Genet. **64:** 7–19.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.

LAURIE, C. C., J. T. BRIDGHAM and M. CHOUDHARY, 1991 Associations between DNA sequence variation and variation in expression of the *Adh* gene in natural populations of *Drosophila melanogaster*. Genetics **129:** 489–499.

LAURIE, C. C., and STAM, L. F., 1988 Quantitative analysis of RNA produced by Slow and Fast alleles of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **85:** 5161–5165

LIEBHABER, S. A., F. CASH and S. S. ESHLEMAN, 1992 Translation inhibition by an mRNA coding region secondary structure is determined by its proximity to the AUG initiation codon. J. Mol. Biol. **226:** 609–621.

MARFANY, G., and R. GONZALEZ-DUARTE, 1991 The *Adh* genomic region of *Drosophila ambigua*: evolutionary trends in different species. J. Mol. Evol. **32:** 454–462.

MENOTTI-RAYMOND, M., W. T. STARMER and D. T. SULLIVAN, 1991 Characterization of the structure and evolution of the *Adh* region of *Drosophila hydei*. Genetics **127:** 355–366.

MORIYAMA, E. N., and T. GOJOBORI, 1992 Rates of synonymous substitution and base composition of nuclear genes in Drosophila. Genetics **130:** 855–864.

OHTA, T. 1989 Time for spreading of compensatory mutations under gene duplication. Genetics **123:** 579–584.

PARKER, R., P. G. SICILIANO and C. GUTHRIE, 1987 Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell **49:** 229–239.

PLEIJ, C. W., K. RIETVELD and L. BOSCH, 1985 A new principle of RNA folding based on pseudoknotting. Nucleic Acids Res. **13:** 1717–1731

ROUSSET, F., M. PÉLANDAKIS and M. SOLIGNAC, 1991 Evolution of compensatory substitutions through G·U intermediate state in *Drosophila* rRNA. Proc. Natl. Acad. Sci. USA **88:** 10032–10036.

ROWAN, R. G., and W. J. DICKINSON, 1988 Nucleotide sequence of the genomic region encoding alcohol dehydrogenase in *Drosophila affinidisjuncta*. J. Mol. Evol. **28:** 43–54.

ROWAN, R. G., and J. A. HUNT, 1991 Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene in five closely related species of Hawaiian Drosophila. Mol. Evol. Biol. **8:** 49–70.

SCHAEFFER S. W., and C. F. AQUADRO, 1987 Nucleotide sequence of the *Adh* gene region of *Drosophila pseudoobscura*: evolutionary change and evidence for an ancient gene duplication. Genetics **117:** 61–73.

SCHAEFFER, S. W., and E. L. MILLER, 1991 Nucleotide sequence analysis of *Adh* genes estimates the time of geographic isolation of the Bogota population of *Drosophila pseudoobscura*. Proc. Natl. Acad. Sci. USA **88:** 6097–6101.

SOLNICK, D., 1985 Alternative splicing caused by RNA secondary structure. Cell **43:** 667–676.

THROCKMORTON, L. H., 1975 The phylogeny, ecology and geography of Drosophila, pp. 421–469 in *Handbook of Genetics*, Vol. III, edited by R. KING. Plenum Press, New York.

WRIGHT, S., 1932 The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proc. Sixth Int. Congr. Genet. **1:** 356–366.

Communicating editor: A. G. CLARK