Comparative Evolutionary Analysis of rDNA ITS Regions in *Drosophila*

Christian Schlötterer, * Marie-Theres Hauser,† Arndt von Haeseler,* and Diethard Tautz*

*Zoologisches Institut der Universität München and †Max Planck Institut für Biochemie

The internal transcribed spacer (ITS) of the ribosomal DNA is generally considered to be under low functional constraint, and it is therefore often treated as a typical nonfunctional spacer sequence. We have analyzed the ITS regions of five species from the *Drosophila melanogaster* subgroup, two *Drosophila* species from outside this group (*D. pseudoobscura* and *D. virilis*), as well as from the more distantly related dipteran fly *Musca domestica*. The sequence comparisons show a distinctive conservation/divergence pattern, indicating that some regions are more conserved than others. Moreover, secondary-structure calculations indicate several conserved structural elements within the ITS regions. On the other hand, a statistical test that allows us to estimate the fraction of sites that are not under selective constraint suggests that more than half of the spacer is apparently free to diverge and evolves with a rate that is close to the neutral rate of sequence evolution in *Drosophila*. The ITS sequences can be used to derive a molecular phylogeny for the species under study. We find that the ITS tree is largely in line with the so-far-known phylogeny of this group of species, with one difference. The species most distant within the *D. melanogaster* subgroup is *D. yakuba*, rather than *D. arena*, as is normally assumed.

Introduction

The eukaryotic ribosomal DNA (rDNA) genes are arranged in tandemly repeated clusters with each cluster containing the genes for the 18S, 5.8S, and 28S ribosomal RNA (rRNA) (reviewed in Gerbi 1985). The genes are separated by several spacers, namely, the IGS (intergenic spacer, formerly called the "nontranscribed spacer"), the ETS (external transcribed spacer), and the ITS (internal transcribed spacer). The IGS separates the different repeat units; the ETS lies between the promoter and the 18S gene; and the ITS lies between the 18S and 28S coding regions. The ITS can be further subdivided in ITS1, which is located between the 18S and 5.8S (5S and 2S in *Drosophila*) genes, and ITS2, which separates the 5.8S and 28S genes. The entire unit is transcribed by RNA polymerase I as a single 45S precursor molecule. Maturation of the rRNAs occurs by an ordered cleavage and removal of the spacer sequences (Gerbi 1985). In contrast to the regulation of rDNA transcription and termination (reviewed in Sollner-Webb and Mougey 1991), less is known about the processing of this large precursor RNA. In particular, the role of the spacers in the processing reaction is not clear. The spacers show generally a fast evolutionary divergence and are therefore usually not considered to have a function. However, the functional analysis of the ITS in yeast has yielded clear evidence for the role that spacers play in the processing reactions (Musters et al. 1990; van der Sande et al. 1992). ITS1 is required for the processing of the 3' end of the 18S and the 5' end of the 5.8S molecules, while ITS2 is required for the processing of the 3' end of the 5.8S and the 5' end of the 28S molecules. The yeast results indicate furthermore that the secondary-structure of these spacers is important for the processing reactions (van der Sande et al. 1992).

ITS sequence comparisons are becoming an increasingly popular tool for phylogenetic analysis (Lee and Taylor 1991; Baldwin 1992; Pleyte et al. 1992; Wesson et al. 1992) and for the differentiation of populations (Nazar et al. 1991; Bakker et al. 1992; Kooistra et al. 1992; O'Donnell 1992; Gardes and Bruns 1993; Vogler and DeSalle 1994 [in this issue]). It is therefore important to get an insight into the potential functional constraints that may effect the evolutionary divergence of this region. Furthermore, it is highly desirable to obtain an estimate for the average rate of divergence for these spacers. Sequence phylogenies of closely related species can only be resolved by comparing fast-evolving DNA regions, preferably regions that evolve with a neutral
rate. We show here that a large proportion of the ITS in *Drosophila* evolves indeed with a rate that is close to the neutral rate.

**Material and Methods**

**Polymerase Chain Reaction (PCR) and Sequencing**

The *Drosophila melanogaster* sequence was taken from Tautz et al. (1988). The other sequences were obtained from cloned PCR fragments. Primers located in highly conserved regions of the 18S and 28S genes were used to amplify the ITS regions of the different species. The following primers were used (all primers in 5'-3' direction): 18S primer (CS249) TCGTAACAAAGGTTTCCG and 28S primer (CS250) GTT(A/G)GTTTCTTTCCTC. PCR reactions were set up according to the recommendations of the supplier of the Taq polymerase (Cetus) using 50 ng genomic DNA. The typical thermocycling profile consisted of 25 cycles with 1 min at 94°C, 2 min at 45°C, and 2.5 min at 72°C. The PCR products were treated with Klenow polymerase to make them blunt ended, were purified on an agarose gel, and were cloned into the Smal site of M13 mp18 by using XL1 (Stratagene) as host. Several clones were obtained for each species, and both directions were fully sequenced, in most cases. Sequencing was done by the dideoxy method using Sequenase 2.0 (USB) according to the supplier's protocols. In addition to the universal M13 primer, we have used the following specific primers to obtain the full sequences: (DT421) GCTGCATTTCCATCG, (CS632) CGATGAAGAACGCAGC, (CS633) GAAAGTGGAAGTCGA, (CS262) GCCACGATGAAGAACGCAGC, (CS263) ATTATCTCACATTTG, and (CS264) GCTAGACATTTCTCA. DT421 and CS632 are complementary primers located in a highly conserved region of the 5S gene. For most of the sequences, both orientations of the ITS were sequenced. In the ITS region in *Drosophila melanogaster* will be published elsewhere (C. Schlötterer and D. Tautz, unpublished data). The spacers have a very high A-T content, as mosquitoes (Wesson et al. 1992). The high A-T content is not generally typical for ITS sequences (reviewed in Torres et al. 1990) and is not found in more distantly related dipterans, such as mosquitos (Wesson et al. 1992). The high A-T content found for the ITS in *Drosophila* might be related to the fact that the rDNA clusters are located in the heterochromatin, since not only the ITS, but also the other rDNA spacers (Tautz et al. 1987, 1988), as well as *Drosophila* satellite sequences (Strachan et al. 1985), are A-T rich.

### Results

**Cloning and Sequence Analysis**

The ITS sequences of the different species were PCR amplified using primers within the 18S and 28S genes. A single unique length class was found for each species. The PCR fragments were cloned, and several independent clones of each species were sequenced. These revealed a few within-species polymorphisms at the sequence level. These were, however, rare (<0.05%) and can be neglected for the analysis presented in this paper.

An extensive survey of within-species polymorphisms in the ITS region in *Drosophila melanogaster* will be published elsewhere (C. Schlötterer and D. Tautz, unpublished data). The spacers have a very high A-T content, with >70% in ITS1 and >75% in ITS2 (table 1). This exceeds the average A-T content of noncoding DNA in *Drosophila*, which is ~60% (Moriyama and Hartl 1993). Furthermore, a high A-T content is not generally typical for ITS sequences (reviewed in Torres et al. 1990) and is not found in more distantly related dipterans, such as mosquitos (Wesson et al. 1992). The high A-T content found for the ITS in *Drosophila* might be related to the fact that the rDNA clusters are located in the heterochromatin, since not only the ITS, but also the other rDNA spacers (Tautz et al. 1987, 1988), as well as *Drosophila* satellite sequences (Strachan et al. 1985), are A-T rich.

### Table 1

<table>
<thead>
<tr>
<th>Species</th>
<th>ITS1 A-T (%)</th>
<th>Length (nt)</th>
<th>ITS2 A-T (%)</th>
<th>Length (nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. sechelia</em></td>
<td>73.4</td>
<td>686</td>
<td>78.5</td>
<td>382</td>
</tr>
<tr>
<td><em>D. simulans</em></td>
<td>73.5</td>
<td>690</td>
<td>79.4</td>
<td>383</td>
</tr>
<tr>
<td><em>D. mauritiana</em></td>
<td>73.7</td>
<td>686</td>
<td>78.9</td>
<td>384</td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td>73.1</td>
<td>729</td>
<td>80.0</td>
<td>385</td>
</tr>
<tr>
<td><em>D. orena</em></td>
<td>76.3</td>
<td>848</td>
<td>78.4</td>
<td>380</td>
</tr>
<tr>
<td><em>D. yakuba</em></td>
<td>74.4</td>
<td>820</td>
<td>79.1</td>
<td>378</td>
</tr>
<tr>
<td><em>D. pseudoobscura</em></td>
<td>77.5</td>
<td>600</td>
<td>79.1</td>
<td>408</td>
</tr>
<tr>
<td><em>D. virilis</em></td>
<td>77.7</td>
<td>525</td>
<td>77.0</td>
<td>353</td>
</tr>
<tr>
<td><em>M. domestica</em></td>
<td>72.1</td>
<td>511</td>
<td>74.9</td>
<td>349</td>
</tr>
</tbody>
</table>
The sequence alignments are shown in figure 1. Though there is a substantial divergence between the spacers, including several insertion/deletion differences, the alignments among the Drosophila species are fairly unequivocal, because of the presence of some more highly conserved regions. This is particularly evident for the 3' half of ITS1 and most of the ITS2 region. On the other hand, the Musca ITS region is so highly diverged from the Drosophila sequences that a general alignment was not possible. The two short regions included in the alignments were only detected after a closer analysis of the sequences (see below).

Phylogeny

The PAUP package was used for a phylogenetic analysis of the data. Only a single optimal tree was found with the exhaustive search option (fig. 2). The bootstrap analysis (using the branch and bound option) showed that all branches are highly significant (98% support for the D. sechellia/D. mauritiana branch, and 100% support for all the other ones). The same tree was also obtained using tree reconstruction methods that are not based on parsimony criteria, such as neighbor joining and maximum likelihood. Nonetheless, the tree found is not fully in line with previously obtained trees using morphologic, cytogenetic, and DNA-DNA hybridization data (Lemeunier et al. 1986; Caccone et al. 1988; reviewed in DeSalle and Grimaldi 1991). In the ITS tree, D. orena is more closely related to the rest of the group than D. yakuba, while it is normally assumed that D. orena is the most distant species within this group. In an attempt to resolve this conflict, we have sequenced the ITS2 of D. erecta, which is supposed to be a sister species to D. orena. The ITS2 sequence obtained for D. erecta supports the close relationship between D. erecta and D. orena (not shown), but we still find that D. yakuba is the most distant species. Furthermore, two other trees based on DNA sequence data support our findings. The distance tree of alcohol dehydrogenase (ADH) sequences included in the publication of Moriyama and Gojobori (1992) shows the same grouping of D. orena and D. yakuba as the ITS tree. Similarly, a recently published molecular phylogeny of the genus Drosophila, based on 28S ribosomal RNA sequences, supports this grouping (Pélandakis and Solignac 1993). On the other hand, both of these publications partially deviate from our tree with respect to the grouping of the other species. The ADH tree proposes a different grouping of the more closely related species of D. simulans, D. mauritiana, and D. melanogaster. These relationships are, on the other hand, supported by the analysis of shared polymorphisms at the period locus between these species (Kliman and Hey 1993). The 28S tree by Pélandakis and Solignac (1993) proposes an even closer relationship of D. melanogaster and D. orena than the other trees. However, the sequence data in this case do not provide a good support of this node. We believe therefore that our tree currently has the best statistical support for this species group.

Nucleotide Substitution Rates

We have used the alignment to calculate the average number of nucleotide substitutions per site for the whole spacer sequences. This was done in pairwise sequence comparisons by using the method of Kimura (1980), which compensates for multiple hits and biases in transition/transversion rates (see below). Table 2A shows the respective distance matrix. The average number of substitutions is apparently correlated with the evolutionary distances of the species within the D. melanogaster species group. However, the numbers between the D. melanogaster group and D. pseudoobscura/D. virilis are very similar, even though the D. melanogaster/D. virilis split is much older than the D. melanogaster/D. pseudoobscura one. This suggests a saturation effect, which is indicative of regions within the spacer that are under selective constraint and that are therefore not free to diverge. To get an estimate for the substitution rate in regions that are free of such constraints, it is necessary to omit the apparently invariable regions from the analysis.

The most simple way to do this is to exclude all positions that are conserved between all species in the alignment. However, this would be a suboptimal procedure for at least two reasons. First, some of the positions might have been "conserved" by chance rather than by selection, and, second, the procedure would depend heavily on the number of the species in the alignment and their relatedness. Hence, it is necessary to develop a correction procedure for estimating the sites that are not free to vary. For this correction procedure, we start with an estimate of the number of sites within the given tree topology and for the given branch length that should be "conserved" by chance alone, i.e., in the absence of constraints. If this number deviates from the observed counts, the counts are corrected toward the expected numbers. This estimate of variable sites is then used to reconstruct a new tree. The above procedure is repeated until the expected and observed counts of constant sites are approximately the same. The fraction of invariable sites is then simply the difference between the observed constant sites and the variable sites at the end of the cycle. We have developed the appropriate algorithms for this procedure and have done studies to show the robustness of the procedure. The results of this analysis will be published elsewhere (A. von Haeseler, C.-M. Ordonfer, and D. Tautz, unpublished data).

In order to use the procedure for the ITS sequences, we have excluded all positions that are not fully aligned,
i.e., all positions that contain a gap in one of the sequences. In this way, we completely ignore all mutational effects that lead to insertions/deletions even though these may account for almost half of the mutational events (compare also Tautz et al. 1987). Moreover, the mutational events that occur in those sequences that would be alignable in the respective regions are also ignored. Thus, we certainly underestimate the true mutation rates in this way, but we believe that this is still the only way to make the data comparable to those obtained for the divergence rates at third codon positions.

For the 763 positions that could be analyzed in this way, we find that 324 are invariable and that 439 are variable. The substitution rates that are obtained for this new data set are listed in table 2B. The numbers are now of course somewhat higher than those for the whole spacer comparisons, since we have omitted a considerable portion of constant positions. More important, however, the above-mentioned saturation effect is not seen any more. Instead, the degree of divergence correlates better with the assumed evolutionary distances of these species, indicating that the sequences evolve in a reasonably constant manner.

The divergence rates obtained under these conditions can be compared with those found previously for neutral sequences in this group of species. Cacccone et
al. (1988) have used DNA-DNA hybridization studies to calculate a rate of at least 0.85% substitutions/Myr (note that these authors cite a value of 1.7% / Myr, but have calculated this value along both lineages). Sharp and Li (1989) have analyzed the substitution rate at third codon positions and find a range of 0.8%-1.6% / Myr, whereby the higher rate was found for genes with a lower codon-usage bias. A similar analysis was performed by Moriyama and Gojobori (1992) who find a range of 0.7%-3.3% substitutions/Myr for third codon positions. When the same divergence times for the different species as those assumed by Moriyama and Gojobori (1992) are taken into account, we calculate the substitution rates from our data as follows: the average substitutions per site for the *D. melanogaster* group (*D. sechellia, D. simulans, D. mauritiana, and D. melanogaster*) versus *D. pseudoobscura* is 0.73. These species have split about 30 Mya, that is, they had 60 Myr of independent evolutionary history. We calculate thus an average substitution rate of 1.2%/Myr. A similar calculation of the rate between the *D. melanogaster* group species and *D. virilis* (split 40 Mya) yields a substitution rate of 1.1%/Myr. These values fall well within the range of values cited above. However, they are not as high as the values found for third codon positions that are under low codon-usage constraint (Sharp and Li 1989; Moriyama and Gojobori 1992). It must remain open, at this point, whether this is a significant difference or whether this is simply due to the different procedures taken to estimate these rates. Still, it seems safe to conclude that the unconstrained positions in the ITS evolve with a rate that is close to the neutral rate, in this species group.

### Transition/Transversion Ratio

For many sequence comparisons, a bias in the transition-transversion ratio is frequently observed, with

#### Table 2

Matrices for the Average Numbers of Substitutions per Site and for the Transition/Transversion Ratios for the *Drosophila* ITS Sequence Comparisons

<table>
<thead>
<tr>
<th></th>
<th><em>D. sechellia</em></th>
<th><em>D. simulans</em></th>
<th><em>D. mauritiana</em></th>
<th><em>D. melanogaster</em></th>
<th><em>D. orena</em></th>
<th><em>D. yakuba</em></th>
<th><em>D. pseudoobscura</em></th>
<th><em>D. virilis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. sechellia</em></td>
<td>0</td>
<td>0.014</td>
<td>0.016</td>
<td>0.072</td>
<td>0.146</td>
<td>0.230</td>
<td>0.401</td>
<td>0.399</td>
</tr>
<tr>
<td><em>D. simulans</em></td>
<td>0</td>
<td>0.021</td>
<td>0.077</td>
<td>0.143</td>
<td>0.217</td>
<td>0.402</td>
<td>0.409</td>
<td>0.397</td>
</tr>
<tr>
<td><em>D. mauritiana</em></td>
<td>0</td>
<td>0.078</td>
<td>0.150</td>
<td>0.226</td>
<td>0.411</td>
<td>0.426</td>
<td>0.424</td>
<td>0.424</td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td>0</td>
<td>0.150</td>
<td>0.248</td>
<td>0.707</td>
<td>0.878</td>
<td>0.701</td>
<td>0.871</td>
<td>0.871</td>
</tr>
<tr>
<td><em>D. orena</em></td>
<td>0</td>
<td>0.283</td>
<td>0.415</td>
<td>0.731</td>
<td>0.923</td>
<td>0.768</td>
<td>0.988</td>
<td>0.988</td>
</tr>
<tr>
<td><em>D. yakuba</em></td>
<td>0</td>
<td>0.410</td>
<td>0.425</td>
<td>0.745</td>
<td>0.962</td>
<td>0.788</td>
<td>0.969</td>
<td>0.969</td>
</tr>
<tr>
<td><em>D. pseudoobscura</em></td>
<td>0</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Table 2 Continued

<table>
<thead>
<tr>
<th></th>
<th><em>D. sechellia</em></th>
<th><em>D. simulans</em></th>
<th><em>D. mauritiana</em></th>
<th><em>D. melanogaster</em></th>
<th><em>D. orena</em></th>
<th><em>D. yakuba</em></th>
<th><em>D. pseudoobscura</em></th>
<th><em>D. virilis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. sechellia</em></td>
<td>0</td>
<td>0.021</td>
<td>0.221</td>
<td>0.110</td>
<td>0.206</td>
<td>0.281</td>
<td>0.707</td>
<td>0.878</td>
</tr>
<tr>
<td><em>D. simulans</em></td>
<td>0.80</td>
<td>0.028</td>
<td>0.113</td>
<td>0.19/</td>
<td>0.259</td>
<td>0.701</td>
<td>0.871</td>
<td>0.871</td>
</tr>
<tr>
<td><em>D. mauritiana</em></td>
<td>0.80</td>
<td>0.50</td>
<td>0</td>
<td>0.118</td>
<td>0.215</td>
<td>0.285</td>
<td>0.731</td>
<td>0.923</td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td>0.55</td>
<td>0.53</td>
<td>0.60</td>
<td>0.206</td>
<td>0.322</td>
<td>0.768</td>
<td>0.988</td>
<td>0.988</td>
</tr>
<tr>
<td><em>D. orena</em></td>
<td>0.46</td>
<td>0.44</td>
<td>0.39</td>
<td>0</td>
<td>0.291</td>
<td>0.788</td>
<td>0.969</td>
<td>0.969</td>
</tr>
<tr>
<td><em>D. yakuba</em></td>
<td>0.56</td>
<td>0.52</td>
<td>0.55</td>
<td>0.49</td>
<td>0.47</td>
<td>0.745</td>
<td>0.962</td>
<td>0.962</td>
</tr>
<tr>
<td><em>D. pseudoobscura</em></td>
<td>0.53</td>
<td>0.51</td>
<td>0.53</td>
<td>0.53</td>
<td>0.46</td>
<td>0.41</td>
<td>0</td>
<td>0.629</td>
</tr>
<tr>
<td><em>D. virilis</em></td>
<td>0.46</td>
<td>0.46</td>
<td>0.47</td>
<td>0.48</td>
<td>0.40</td>
<td>0.39</td>
<td>0.52</td>
<td>0</td>
</tr>
</tbody>
</table>
transitions being usually more frequent than transversions, in particular in mitochondrial sequence comparisons (DeSalle et al. 1987). It is interesting that we find the opposite effect in our sequences. Transversions are more frequent than transitions, with a transition/transversion ratio of ~0.5 in most of the pairwise comparisons (table 2B). Only the most closely related species show a slightly higher value. This is, however, likely to be a sampling error, since these differ only at a total of nine sites (four transitions and five transversions; note that this calculation is restricted to the alignment where all gaps have been excluded). We consider this difference therefore as nonsignificant. A transition/transversion ratio of 0.5 is exactly the rate that would be expected, if there were no bias at all, since the calculated chance for a given nucleotide to become a transversion is twice as large as that for becoming a transition. It is interesting that the same ratios were found for a comparison between satellite sequences in the D. melanogaster species group (Strachan et al. 1985). It appears, therefore, that in the nuclear genome of Drosophila, there are no biases in the propensity to mutate to any other nucleotide, not even steric biases that could be expected when two purines or two pyrimidines lie opposite each other in the DNA helix after a misincorporation event.

Secondary Structure

Previous comparisons in several species have shown that there is a high propensity to form secondary structures within the ITS region (Furlong and Maden 1983; Michot et al. 1983; Gonzalez et al. 1990; Torres et al. 1990; Yeh and Lee 1990). We have therefore calculated the secondary structures for each of the ITS1 sequences in this study, whereby the ITS1 and ITS2 regions were treated separately, since it has been shown that they are functionally independent (Musters et al. 1990). The structures found for the different species were compared with each other and analyzed for shared structural elements.

ITS1 Secondary Structure

The 5' half of the ITS1 is only poorly conserved and shows many insertions and deletions. Furthermore, because of the A-T richness, it allows many different folding structures, all with a similar energy. Accordingly, we could not detect any stable structures in this region, which could be compared between the species. This is different for the 3' half of the ITS1 sequences, where a stable structure was found. This structure forms a good hairpin, which is completely conserved between all Drosophila species examined (fig. 1a). This hairpin structure in fact is also found in the Musca ITS1 sequence, indicating that it is of high functional significance. It is interesting that the loop of the hairpin is more diverged than the stem in Musca. Similar comparisons in the coding parts of the rRNA frequently show the opposite effect, namely, that the loop regions are more highly conserved (Gerbi 1985). This is probably due to the fact that the loops interact with the ribosomal proteins, while the stems form structural elements. The same line of argument would suggest that the loop of the ITS hairpin does not interact with proteins and that it is only the secondary structure formed by the stem that is important. van der Sande et al. (1992) have reached a similar conclusion from their functional analysis of yeast ITS2 sequences.

ITS2 Secondary Structure

The ITS2 region is generally more conserved than the ITS1 region, making the secondary-structure predictions very consistent. The structures of the D. melanogaster group species are all very similar, showing four distinct domains (fig. 3a and b). In addition, the sequences at the 5' end and at the 3' end of the ITS2 can form a stable stem structure in these species. This stem structure is not found in the calculations for the D. pseudoobscura sequence, even though the four conserved domains found for the D. melanogaster species are still discernible (fig. 3c).

Two of the four conserved domains are also found in the D. virilis sequence, namely domain 1 and domain 3 (fig. 3d). Domain 3 encompasses the most highly conserved region in the whole ITS2 sequence and is again a domain that can still be found in the Musca ITS2 region (fig. 1). Furthermore, the predicted folding structure of this domain is supported by a compensatory mutation. The nucleotides at positions 218 and 266 (fig. 1b) form an A-U base pair in the D. melanogaster group sequences but form a G-C pair in D. pseudoobscura and D. virilis.

It was suggested that the sequence regions flanking the ITS2 sequence, namely the 3' end of the 5.8S gene and the 5' end of the 28S gene can pair with each other (Veldman et al. 1981; Subrahmanyan et al. 1982; Furlong and Maden 1983; Hindenach and Stafford 1983). This may be important for the correct processing of these genes and might therefore play a role in directing the folding of the ITS2 region. We have therefore repeated the ITS2 secondary-structure calculations by including the respective parts of the gene sequences. We found that the 3' end of the 2S gene (equivalent to the 5.8S in other species) and the 5' end of the 28S gene can indeed form a fairly stable stem structure but that this additional structural element does not change the predicted folding of the remainder of the ITS2. We conclude therefore that the formation of the ITS2 secondary structure is not affected by the coding region.
FIG. 3.—Secondary structures calculated for the ITS2 sequences, using the FOLD program of the UWGCG package. The domains indicated in fig. 1b are boxed and numbered. The Drosophila melanogaster structure (a) is representative for the structures found for D. sechellia, D. simulans, and D. mauritiana; the D. yakuba structure (b) is very similar to the D. arena structure. All four domains found in the D. melanogaster species structures can also be discerned in the D. pseudoobscura structure (c), while only D1 and D3 are found in the D. virilis structure (d).

A recent study of the ITS2 region in mosquito species has also revealed partially conserved secondary structural elements (Wesson et al. 1992). However, none of the structures defined by these authors resembles any of the structures that were identified in this study. This is in line with the observation that most of the Drosophila structural elements are not even found in Musca.

Discussion

Molecular phylogenies of closely related species can only be obtained with fast-evolving DNA regions. It is customary to use mitochondrial DNA regions for this purpose (Simon 1991). However, it is of great interest to find also a suitable universal region in the genomic DNA. The rDNA ITS region is such a candidate. It is present in all eukaryotes; it can easily be amplified with universal primers in the highly conserved flanking regions; and it shows a high rate of divergence, at least in comparison with the coding regions of the rDNA. However, little attention has so far been paid to the question of whether it is still under some selective constraints or whether it evolves with a rate that is close to the neutral rate of the remainder of the genome. Evidence for potential functional constraint comes from in vivo mutational analysis in yeast, where it was shown that the ITS regions are actively involved in the processing of the primary transcript (van der Sande et al. 1992). However, only a minor fraction of the nucleotides in the spacer might be involved in such a function, and it is therefore necessary to make an assessment on the general degree of conservation within the ITS. We have performed such an assessment for the Drosophila ITS. This analysis suggests that ~40% of the spacer sequences are not free to diverge (this estimate relates only to the alignable portions of the ITS). In addition, we have identified some more highly conserved regions in the ITS that form stable secondary structures and that may be of functional significance. Two of these regions were also found in the distantly related species Musca domestica. This shows that the naive assumption of a typical spacer sequence
that would be completely free to diverge is not fulfilled for the ITS.

We have employed a statistical procedure to estimate the proportion of sequences that are free to diverge within the spacer. Though this statistical procedure does not allow us to infer which particular nucleotide positions are under selective constraints, it yields a more reliable basis for assessing the rate of evolution in the nonconstrained part of the spacer. We find that this proportion of the spacer evolves with a rate close to the neutral rate in *Drosophila*. This suggests that these parts of the ITS region behave like neutral or almost-neutral DNA sequences. It is interesting that this rate is also close to that found for the mitochondrial genome in *Drosophila* (DeSalle et al. 1987), suggesting that the large rate differences found for nuclear and mitochondrial genomes of vertebrates may not be a general phenomenon, as was also noted elsewhere (Powell et al. 1986; Vawter and Brown, 1986; Sharp and Li 1989).

A potential problem in using the ITS for phylogenetic purposes comes from the fact that it is tandemly repeated. To keep all copies homogeneous, it is necessary that a new mutation in one of the units spreads first to all other units before a second new mutation arises. The process that accomplishes this is called "molecular drive" (Dover 1982), and the mechanisms are probably unequal crossing-over and gene conversion. The degree of homogeneity found in a tandemly repetitive sequence depends on the balance between the rate of homogenization and the rate of new mutations (Ohta and Dover 1983). If the rate of homogenization is low in comparison with the rate of new mutations, one would expect to find multiple ITS variants in a single species. We have therefore sequenced several independent copies of the ITS for each species. We found that the within-species polymorphism of the ITS is very low in our case and does not present a problem for phylogenetic analysis. However, a recent study on the ITS region in mosquitoes has shown a somewhat higher degree of polymorphism (Wesson et al. 1992). In addition, we have since identified some vertebrate species groups that show a degree of polymorphism in the ITS that will be of concern in a phylogenetic evaluation of the sequences. This suggests that the rate of homogenization may be different in different species groups. Within-species polymorphisms should therefore always be analyzed when one intends to use the ITS for phylogenetic purposes.

Acknowledgments

Flies and DNA samples used in this study were provided by A. Klarenberg, L. Nigro, and R. Sommer. We thank H. L. Sänger for his support, as well as Rob DeSalle and Paul Sharp for very helpful suggestions on the manuscript.

LITERATURE CITED


FURLONG, J. C., and B. E. H. MADEN. 1983. Patterns of major divergence between the internal transcribed spacers of ribosomal DNA in *Xenopus borealis* and *Xenopus laevis*, and of minimal divergence within ribosomal coding regions. EMBO J. 2:443–448.


Paul Sharp, reviewing editor

Received April 6, 1993

Accepted January 19, 1994