Nucleotide Sequence Comparison of the rp49 Gene Region between Drosophila subobscura and D. melanogaster

Montserrat Aguadé
Department de Genètica, Facultat de Biologia, Universitat de Barcelona

A 1.6-kb fragment encompassing the rp49 gene, which codes for a ribosomal protein, has been cloned and sequenced in Drosophila subobscura. The rp49 coding region has accumulated 46 nucleotide differences out of 402 bp since D. subobscura diverged from D. melanogaster. Forty-three percent of the effectively silent sites have changed since both species diverged. Both silent and replacement differences are distributed at random between the two exons of the gene. The frequency of silent differences in exons does not differ from that observed in the 5′ leader sequence and in the intron. The frequency of silent differences in exon and intron sites is much greater than the number of amino acid replacement differences. This observation indicates strong purifying selection against amino acid replacements.

Introduction

The rp49 gene, which codes for a ribosomal protein, is, like other ribosomal protein genes cloned in Drosophila melanogaster (see Kay and Jacobs-Lorena 1987 for references), a single-copy gene. It is located on the right arm of the third chromosome and consists of two exons (31 and 103 codons long, respectively) separated by a small intron.

This gene was chosen for its cloning and sequencing in D. subobscura for the following two main reasons: (1) According to chromosome homologies, rp49 should be located in D. subobscura on segment I of the O chromosome, where alternative gene arrangements are present. In future studies, rp49 might become a marker of those gene arrangements. (2) The structure of the rp49 gene makes it suitable for analysis of nucleotide variation in different functional regions.

In the present study, the region including the rp49 gene from D. subobscura was cloned and sequenced and the sequence obtained was compared with that of D. melanogaster (O'Connell and Rosbash 1984; Cadic-Jacquier and M. Rosbash, personal communication).

Material and Methods

Total genomic DNA was isolated according to the method of Bingham et al. (1981) from a strain of Drosophila subobscura (H27) homozygous for the standard gene arrangement of its five acrocentric chromosomes and sib-mated for 27 generations (strain provided by R. de Frutos). This DNA was partially digested with Sau3A, and fragments of 15–20 kb were selected by sucrose gradient fractionation. A representative random library was constructed by ligation of the Sau3A fragments into the BamHI

1. Key words: molecular evolution, sequence comparison, rp49.

Address for correspondence and reprints: Montserrat Aguadé, Department de Genètica, Facultat de Biologia, Universitat de Barcelona Diagonal 645, 08028 Barcelona, Spain; phone—343—3309951, x. 148.

© 1988 by The University of Chicago. All rights reserved.
0737-4038/88/0504-0008$02.00

433
site of the EMBL4 phage vector (Frischauf et al. 1983). Approximately $10^5$ recombinant plaques were screened (Benton and Davis 1977) for sequences hybridizing to a *D. melanogaster* 600-bp fragment that includes most of the *rp49* transcriptional unit (clone HR0.6 provided by M. Rosbash). The nick-translation probe was used at a concentration of $1.0 \times 10^6$ dpm/ml in the hybridization solution. Hybridization to phage DNA on nitrocellulose filters was performed at 65 C for 14–16 h with gentle rocking. Filters were washed at 45 C in low-stringency conditions (0.15 M NaCl, 0.015 M sodium citrate, pH 7.0, 0.1% sodium dodecyl sulfate), and autoradiographs were obtained after exposure of Kodak XAR film at -80 C.

Four phages positive for *rp49* were detected. Positives were plaque purified, and DNA was prepared according to the method of Maniatis et al. (1982).

DNA of each positive phage was digested with *AvaI*, *EcoRI*, *PstI*, *SalI*, and *SmaI*, and the digests were transferred to nylon membranes (Gene Screen®; DuPont) after size fractionation in agarose gels (Southern [1975], with modifications). The same 600-bp *D. melanogaster* fragment was used to probe the filters, with the conditions for hybridization and washes being identical to those used to screen the library. In all cases, digestion with *AvaI* gave a single 3.45-kb fragment that would hybridize with the probe. This 3.45-kb fragment from one of the positives was electroeluted on Na45 membrane (from Schleicher and Schuell), the ends were filled in with Klenow, and the fragment subsequently was subcloned in the *SmaI* site of pUC8. Digestion of the recombinant plasmid (pUC8.38) with *EcoRI* and *BamHI* (sites immediately contiguous to the *SmaI* site in the pUC8 polylinker) gave four different fragments—1.6 kb, 0.95 kb, 0.55 kb, and 0.35 kb—only one of which (the 1.6-kb fragment) would hybridize to the *D. melanogaster* probe.

The 1.6-kb fragment was subcloned in M13mp19 (Bankier 1984) in both orientations. A series of nested deletions was generated according to the method of Dale et al. (1985). These clones were sequenced according to the method of Sanger et al. (1977) by using 35S-thio-α-dATP and TBE (50 mM Tris borate, 10 mM EDTA, pH 8.3) wedged 5% acrylamide gels (Biggen et al. 1983). The average number of bases sequenced per clone was 250–300. Staden’s (1982, 1984) programs were used to obtain the definite sequence. Each site was sequenced an average of 3.82 times. More than 95% of the nucleotide positions in the transcriptional unit and 50% of those outside the locus itself were sequenced on both strands. The *rp49* sequence of *D. subobscura* was compared with that of *D. melanogaster* by using Staden’s programs. Hydropathy plots for the predicted *rp49* proteins of both species were obtained using the algorithm of Hopp and Woods (1981). The chromosomal location of *rp49* in *D. subobscura* was determined by in situ hybridization using biotinylated probes (Langer et al. 1981; E. Montgomery, personal communication).

Results

Clone pUC8.38, which includes the *rp49* gene, hybridizes to chromosomal band 91C on chromosome O of *Drosophila subobscura* (fig. 1).

A 1.6-kb fragment encompassing the *rp49* gene has been sequenced in *D. subobscura* (fig. 2). As in *D. melanogaster*, the gene consists of two exons and one small intron. The sequence ATCAGT starting at position 820 is in perfect agreement with the consensus cap sequence proposed by Hultmark et al. (1986) for several other *Drosophila* genes. At position 788, 32 nucleotides upstream from the putative cap sequence, there is a putative TATA box—AATAAATA—and at position 745 there is a putative CAAT box—GCCAATC. Both the good agreement with consensus se-
Fig. 1.—An in situ hybridization of clone pUC8.38 to *Drosophila subobscura* polytene chromosomes. Hybridization (arrow) is on band 91C of the O chromosome according to the cytological map of the species (Künze-Mühl and Muller 1958).

quences (six of eight nucleotides for the TATA box and six of seven for the CAAT box) and the relative distance both between them (43 bp apart) and from the presumptive cap sequence (32 bp from the TATA box) make it highly probable that these are the correct signal sequences. At the 3' end of the gene starting at position 1381 there is a putative polyadenylation signal, AATACA, which has been found in other *Drosophila* genes. The relative position of the intron in *D. subobscura* is the same as that in *D. melanogaster*, although its length is not the same—62 bp in *D. subobscura* as opposed to 59 bp in *D. melanogaster*.

The *D. subobscura* and *D. melanogaster* *rp49* coding regions align perfectly, but the alignment is uncertain in the flanking regions. Therefore, only divergence in the transcriptional unit will be considered here (table 1).

The *rp49* region has accumulated 46 nucleotide differences in 402 bp since both species diverged. The observed ratio of transition:transversion differences is 26:20, which differs significantly from the 1:2 expectation when equal mutation rates and no differential selection are assumed. T-C transitions are more frequent than A-G transitions. The frequency of both A-C and G-C transversions is higher than would be expected if equal frequency of the four kinds of transversions were assumed. The number of substitutions in the three codon positions is significantly different, most of them having occurred at the third position (fig. 2).

Twenty-three percent (88 of 402) of all sites in the coding region of the *rp49* are effectively silent or synonymous. Thirty-eight changes in silent sites (eight in the first exon and 30 in the second exon) occur between *D. subobscura* and *D. melanogaster* sequences (table 1). Silent differences are distributed at random between the two exons. The number of silent differences per effectively silent site in the exons is not
FIG. 2.—Sequence of the rp49 region from Drosophila subobscura. An arrow shows the position of the presumptive capping site. The promoter and polyadenylation recognition sequences are underlined. The amino acid sequence of the rp49 gene is presented below the nucleotide sequence, beginning at +867 and ending at +1330, the gap in the amino acid sequence representing the intron. Nucleotide differences in the coding region of D. melanogaster as compared with that of D. subobscura are shown above the nucleotide sequence. CC} represents an insertion in the D. melanogaster sequence at position }. In the amino acid sequence, those amino acids that have changed between D. subobscura and D. melanogaster have been underlined.

Eight replacement differences have been observed between the two rp49 sequences.
Comparison of the \textit{rp49} Nucleotide Sequence in \textit{Drosophila} 437

These changes are distributed at random between both exons (table 1). Six of the eight replacement differences cause conservative amino acid replacements, while the remaining two give amino acids different in charge. Codon usage is similar to that observed in other \textit{Drosophila} genes (O'Connell and Rosbash 1984).

Figure 3 gives a hydropathy plot of \textit{rp49} protein for \textit{D. subobscura} and \textit{D. melanogaster}. This plot predicts the hydrophilic and hydrophobic domains of a protein according to the hydropathy values of its amino acids. The three amino acid replacements in the N-terminal region (among which are the two changes that result in a change of charge) occur in hydrophilic domains, while those in the C-terminal region occur in slightly hydrophobic domains.

Discussion

Comparison of the sequence of the \textit{rp49} gene of \textit{Drosophila subobscura} with that of \textit{D. melanogaster} leads to the following observations: (1) Differences are distributed at random in the two exons, both when all changes are considered and when only silent or replacement changes are. (2) The number of silent differences in exon and intron sites is much greater than the number of amino acid replacements.
Table 1
Changes in Nucleotide Sequence between
*Drosophila subobscura* and *D. melanogaster*
for the *rp49* Gene

<table>
<thead>
<tr>
<th>Region</th>
<th>No. of Sites*</th>
<th>No. of Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leader (5')</td>
<td>47/19</td>
<td></td>
</tr>
<tr>
<td>Exon 1:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silent</td>
<td>21/8</td>
<td></td>
</tr>
<tr>
<td>Replacement</td>
<td>72/2</td>
<td></td>
</tr>
<tr>
<td>Intron</td>
<td>59/31</td>
<td></td>
</tr>
<tr>
<td>Exon 2:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silent</td>
<td>67/30</td>
<td></td>
</tr>
<tr>
<td>Replacement</td>
<td>242/6</td>
<td></td>
</tr>
</tbody>
</table>

* For each codon, the proportion of nucleotide changes that would be silent among all nine possible substitutions has been calculated, excluding any substitution that changes a codon to a stop codon. The number of silent changes in each exon was counted considering the actual codons in the sequence.

The first observation indicates the absence of differential constraints in different regions of the protein, in contrast to data from the *Adh* locus. When the *Adh* sequences from *D. mauritiana* and *D. pseudoobscura* are compared (Schaeffer and Aquadro 1987), both replacement and silent differences are not distributed at random in the three axons, the third exon showing an excess and the second exon a deficiency of differences.

The second observation is similar to results of previous studies in *Drosophila* (Bodmer and Ashburner 1984; Cohn 1985; Blackman and Meselson 1986; Coyne and Kreitman 1986): an excess of silent vis-à-vis replacement differences points to the existence of strong purifying selection against amino acid replacements.

For both *Adh* and *hsp82*, sequences have been compared between one species of the *obscura* group and one species of the *melanogaster* group (Blackman and Meselson 1986; Schaeffer and Aquadro 1987, respectively). When the data presented here for *rp49* are compared with those previously reported, the level of silent substitutions (K_S) between *D. melanogaster* and *D. subobscura* (0.670) is comparable with that observed either between *D. melanogaster* and *D. pseudoobscura* for *hsp82* (0.645) or between *D. mauritiana* and *D. pseudoobscura* for *Adh* (0.702). If it is assumed that silent substitutions are subject to virtually no purifying selection, this observation would be in agreement with a constant rate of fixation of newly arisen neutral mutations—under that hypothesis the actual number of changes is dependent only on the time of divergence (in this case, between the *melanogaster* and *obscura* groups of *Drosophila*).

When the number of replacement substitutions per nucleotide site (K_A) between species of the *melanogaster* and *obscura* groups of *Drosophila* is considered, the divergence is highest for *Adh* (0.049 between *D. mauritiana* and *D. pseudoobscura*) and lowest for *hsp82* (0.016 between *D. melanogaster* and *D. pseudoobscura*), with *rp49* intermediate (0.025 between *D. melanogaster* and *D. subobscura*). If equal mutation rates are assumed, this would indicate a lower level of constraint, at the protein level, for *rp49* than for *hsp82—but a higher level than that for *Adh*.
Comparison of the \textit{rp49} Nucleotide Sequence in \textit{Drosophila} 439

Knowledge of the sequence of the \textit{rp49} gene in \textit{D. subobscura} will allow population studies of variation in this species by four-cutter analysis (Kreitman and Aguadé 1986) and will also be the starting point for subsequent sequence comparisons both within and between species of the same group.

Acknowledgments

I thank R. de Frutos, M. Rosbash, M. Kreitman, R. C. Lewontin, E. Montgomery, C. Laurie, and N. Miyashita. The first part of this work was done in R. C. Lewontin’s
laboratory while the author was on leave of absence. Travel funds were awarded to M.A. by Comissió Interdepartamental de Recerca i Innovació Tecnològica. This project was partly funded by a Research Grant awarded to M.A. by the Universitat de Barcelona.

LITERATURE CITED


MASATOSHI NEI, reviewing editor

Received November 23, 1987; revision received March 14, 1988