The Evolution of Duplicate Glyceraldehyde-3-Phosphate Dehydrogenase Genes in Drosophila

Kathy M. Wojtas, Laurence von Kalm,[1] Jonna R. Weaver and David T. Sullivan

Department of Biology, Syracuse University, Syracuse, New York 13244

Manuscript received March 6, 1992
Accepted for publication August 3, 1992

ABSTRACT

In *Drosophila melanogaster* there are two genes which encode the enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH), *Gapdh-43E* and *Gapdh-13F*. We have shown that *Gapdh-43E* codes for the GAPDH subunit with an apparently larger molecular weight while *Gapdh-13F* encodes the GAPDH subunit having an apparently smaller molecular weight. Immunoblots of sodium dodecyl sulfate gels were used to survey species from throughout the genus and results indicated that two classes of GAPDH subunits are present only in Drosophila species of the melanogaster and takahashi subgroups of the melanogaster group. Only the smaller subunit is found in species of the obscura group while all other species have only a large subunit. *Drosophila hydei* was analyzed at the DNA level as a representative species of the subgenus Drosophila. The genome of this species has a single *Gapdh* gene which is localized at a cytogenetic position likely to be homologous to *Gapdh-43E* of *D. melanogaster*. Comparison of its sequence with the sequence of the *D. melanogaster Gapdh* genes indicates that the two genes of *D. melanogaster* are more similar to one another than either is to the gene from *D. hydei*. The *Gapdh* gene from *D. hydei* contains an intron following codon 29. Neither *Gapdh* gene of *D. melanogaster* has an intron within the coding region. Southern blots of genomic DNA were used to determine which species have duplicate *Gapdh* genomic sequences. Gene amplification was used to determine which species have a *Gapdh* gene that is interrupted by an intron. Species of the subgenus Drosophila have a single *Gapdh* gene with an intron. Species of the willistoni and saltans groups have a single *Gapdh* gene that does not contain an intron. Species of the obscura and melanogaster groups have two *Gapdh* genes neither of which have an intron. In *Drosophila pseudoobscura* these are located at cytogenetic positions homologous to those of *D. melanogaster*. Therefore, the simplest model for the evolution of the *Gapdh* genes proposes that the intron in the *Gapdh* gene was lost early in the Sophophoran lineage. Later in the Sophophoran lineage, at a point leading to the obscura and melanogaster groups, a duplication of the *Gapdh* gene occurred.

THE genome of *Drosophila melanogaster* has been shown to contain two genes which encode isozymes of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). In contrast to other well characterized duplications in Drosophila, the two *Gapdh* genes are unlinked. They have been localized to cytogenetic positions 13F and 43E, respectively (SULLIVAN *et al.* 1985). A comparison of the sequence divergence of the two genes is consistent with the duplication having occurred about 60 million years ago (T'SO, SUN and WU 1985). This would place the time of duplication at about the time that has been estimated for the origin of the genus *Drosophila* by THROCKMORTON (1975). This suggests that the *Gapdh* duplication could be a useful marker for studying the divergence of species groups which occurred early during the evolution of the genus.

We have shown previously that two subunit forms of GAPDH from *D. melanogaster* can be detected on immunoblots of extracts prepared from larvae or adults. These forms copurify and appear to differ by about 1–2 kD in molecular mass. However, the basis of this size difference is unclear, since translation of the nucleotide sequence does not predict a significant size difference between the gene products. In addition, *in vitro* translation of *D. melanogaster* mRNA in rabbit reticulocyte lysates yields two GAPDH polypeptides having the same sizes as the forms immunologically detected on western blots of whole animal extracts, thereby indicating that post-translational modification is not a likely basis for the size isoforms (SULLIVAN *et al.* 1985). Consequently, it is likely that the apparent size differences of the GAPDH subunits derive from some structural property that results in anomalous mobility in denaturing gels. The first goal of the work reported here has been to associate each subunit with a specific *Gapdh* gene.

Our second goal has been to describe the evolutionary history and events of significance during the evolution of the *Gapdh* loci. We report here the analysis of representative species from the major groups of the

[1] Present address, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720.

genus *Drosophila* at the protein and DNA level. Immunoblots of extracts from a number of species from throughout the genus suggest that the *Gapdh* duplication occurred early in the Sophophoran radiation of the genus *Drosophila*. *Drosophila hydei* was selected as a representative species of the subgenus Drosophila for detailed study. A single *Gapdh* gene was found in *D. hydei* and this gene was discovered to have an intron which interrupts its coding region. Neither *Gapdh* gene of *D. melanogaster* has any introns in the coding region. Consequently, we have focused our studies on the possibility of simultaneous gene duplication and intron loss during evolution.

At present the nomenclature with respect to the Drosophila *Gapdh* genes is in a confused state. The genes have been investigated by a group at Syracuse (SULLIVAN *et al.* 1985) and another at Cornell (T'SO, SUN and WU 1985; SUN, LIS and WU 1988; SUN *et al.* 1988). Each group independently named these genes according to their own clone isolation history. In consultation with the Cornell laboratory and with their agreement, we propose to name each gene according to its cytogenetic localization. Hence the names *Gapdh-43E* and *Gapdh-13F* will be used for the genes located on the second and *X* chromosomes, respectively.

MATERIALS AND METHODS

Animals: *D. melanogaster* used in these studies were from our standard laboratory strain which was originally inbred from strain Oregon-R. *D. hydei* were from our standard laboratory strain which was inbred from a population originally collected in Mexico City by M. WASSERMAN. *Drosophila pseudoobscura* was an inbred line, Apple Hill 69. All other species were obtained from the species stock center at Bowling Green, Ohio.

Blotting procedures: Immunoblots were performed as previously described (SULLIVAN *et al.* 1985). Southern blots were conducted as described (SULLIVAN *et al.* 1985). For *D. melanogaster* blots a probe from *Gapdh-43E*, an approximately 3000-bp fragment from recombinant phage, λ G3, which extends from position 207 bp 5′ of the *Gapdh* translation start to 1600 3′ of the translation stop, was used. The probe used for interspecific analysis was a mixture of fragments from *Gapdh-13F* and *Gapdh-43E*. An 1100-bp fragment from the recombinant phage, λ G2, which extends from a *Xho*I site at position 218 bp 5′ of the translation start to a *Rsa*I site 930 bp 3′ to the translation start was mixed with an 1100-bp fragment of λ phage G3 including 207 bp 5′ to the translation start and 34 bp 3′ to the translation stop signal.

Cytogenetic localization: *In situ* hybridization to salivary chromosome was conducted using a procedure based on that of LANGER-SAFER, LEVINE and WARD (1982) with the following modifications. Biotinylated dUTP (15.6 mM) was used as a substrate for DNA polymerase (Klenow fragment) using random primers. Since we planned to conduct *in situ* hybridization to several species using a *D. melanogaster* probe, we adjusted hybridization and washing conditions such that a probe from *Gapdh-13F* gave signals of equal intensity at positions 13F and 43E of *D. melanogaster* salivary chromosomes. These were: hybridization at 37° in 2 × SSC and 50% formamide for 12–16 hr and washing at 55° for 10 min in 2 × SSC.

DNA sequencing: A 1.5-kb *Hin*dIII fragment from a *D. hydei* genomic clone obtained from an EMBL-4 genomic library was subcloned into M13mp18. Nucleotide sequencing was performed by the chain termination method (HONG 1982) using [^{35}S]dATP. Buffer gradient gels of BIGGIN, GIBSON and HONG (1983) were used. Gels were read and sequences ordered using a digitizer and computer programs from DNASTAR, Madison Wisconsin. Sequences were determined on one strand by synthesizing oligonucleotide primers spaced at about 350 to 400 bp. Each nucleotide was sequenced at least three times using both the Klenow fragment of *Escherichia coli* DNA polymerase and T7 DNA polymerase (Sequenase; U.S. Biochemical Corp.) in parallel lanes. This proved useful in resolving ambiguities since the two enzymes generate different sequence reading difficulties.

Gene amplification: The nucleotide sequences of the two *Gapdh* genes of *D. melanogaster* and the single *Gapdh* gene of *D. hydei* were used to design primers that were likely to form hybrids with a *Gapdh* gene of any Drosophila species. The 5′ primer is 23 nucleotides long and begins at the first nucleotide of codon 1. The 3′ primer is 20 nucleotides long and ends at the second nucleotide of codon 85. Amplification was conducted using 20 ng of genomic DNA of each species, and a gene amplification kit from U.S. Biochemical Corp. used according to the instructions supplied by the manufacturer.

RESULTS

Analysis of *D. melanogaster* extracts for GAPDH using immunoblots from sodium dodecyl sulfate gels reveals two subunit forms which differ in size. Since the two forms are also found in the immunoprecipitated products from *in vitro* translation of Drosophila mRNA (SULLIVAN *et al.* 1985) it is likely that each protein subunit is the product of one gene. However, no formal connection between a specific gene and protein has yet been established. *Gapdh-43E* is closely linked to the cinnabar (*cn*) locus. A. HOWELLS and W. WARREN of Australian National University (personal communication) have shown that *Gapdh-43E* and *cn* are located within a few thousand base pairs of each other, using the recombinant λ phage, G3, (SULLIVAN *et al.* 1985). HOWELLS and WARREN have used the series of small deletions, described by ALEXANDROV (1984) to localize the *cn* DNA region within the genome. Their success suggested to us that these deletions might be useful for associating the *Gapdh-43E* gene with a specific isoform.

The strains 79b9, 79b13 and 74d2 each have deletions in the 43E region. Each is homozygous lethal but heterozygotes between them are viable while deficient for a small region of overlap in 43E (ALEXANDROV 1984). Figure 1A shows Southern blots of DNA from wild type flies and deletion strains. These blots were probed with a fragment from the *Gapdh-43E* gene. Two *Gapdh* bands are found in wild-type DNA (Figure 1A, lane 1). The stronger hybridizing 5.1-kb fragment is known to be derived from *Gapdh-43E*.

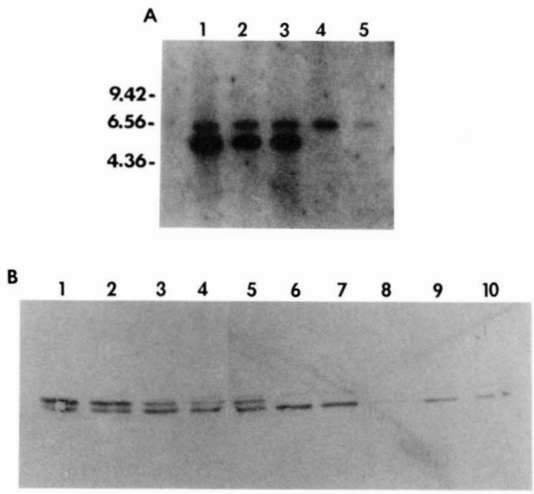

FIGURE 1.—(A) Southern blots of genomic DNA probed with a *Gapdh*-43E probe. Lane 1, $^{+}$/+; lane 2, $^{+}$/79b9; lane 3, $^{+}$/74d2; lane 4, 79b13/74d2; lane 5, 79b9/74d2. (B) Immunoblots reacted with GAPDH antisera. Lane 1, $^{+}$/+ females; lane 2, $^{+}$/+ males; lane 3, $^{+}$/74d2 females; lane 4, +/79b13 female; lane 5, $^{+}$/79b9 male; lane 6, 74d2/79b13 female; lane 7, 79b13/74d2 female; lane 8, 74d2/79b9 female; lane 9, 74d2/79b13 male; lane 10, 79b13/74d2 male.

The weaker hybridizing fragment at 6.3 kb is known to be from *Gapdh-13F* (SULLIVAN *et al.* 1985). The relative intensity of the *Gapdh-43E* band in Figure 1A, lanes 2 through 5, is proportional to the dosage of wild-type alleles. Flies heterozygous for a single deletion show approximately one half the level of hybridization as compared to wild-type flies. This fragment is absent in DNA from flies heterozygous for overlapping deletions (Figure 1A, lanes 4 and 5). No additional hybridizing fragments are present in lanes 4 and 5 indicating that all or most of the *Gapdh-43E* gene is absent in these strains. Immunoblots of protein extracts of these strains reveal that the larger GAPDH is reduced in amount in deletion heterozygotes (Figure 1B, lanes 3, 4 and 5) and absent in deletion homozygotes (Figure 1B, lanes 6 through 10). The amount of the small isoform of GAPDH is similar in each of the flies. Therefore, we conclude that *Gapdh-43E* encodes the larger GAPDH isoform and *Gapdh-13F* encodes the smaller GAPDH isoform.

We surveyed a number of species from throughout the genus *Drosophila* using immunoblots as a preliminary strategy to determine the evolutionary history of the *Gapdh* duplication A summary of all species analyzed is presented in Table 1. Species having two resolvable protein isoforms are found in the melanogaster and takahashi subgroups. Species of other subgroups from the melanogaster group have only a single GAPDH band. Species from the obscura group have one predominant band with mobility equivalent to the small GAPDH isoforms. All species of the subgenus Drosophila have a single large GAPDH isoform.

The number of protein bands revealed by Western blots is not a definitive guide to evolutionary history since it is conceivable that two protein isoforms might occupy the same position following electrophoresis. In addition, the molecular basis for the separation of the two isoforms of *D. melanogaster* is unknown because their conceptual translation does not predict isoforms of different subunit molecular weight. In all likelihood the separation of the GAPDH subunits on denaturing gels is due to retention of some secondary structure. We have reported earlier that two forms of GAPDH are immunoprecipitable from *in vitro* translation of *D. melanogaster* mRNA (SULLIVAN *et al.* 1985). Since the isoform pattern is phylogenetically coherent we have used these observations as a guide for further study of the evolutionary history of the *Gapdh* duplication at the DNA level.

We selected *D. hydei* as a representative species of the subgenus Drosophila for detailed study. Southern blots of *D. hydei* genomic DNA using a *D. melanogaster Gapdh-43E* probe yield a series of hybridizing fragments which are best interpreted as a single gene because none of eight enzymes generated more than a single hybridizing fragment. Five of these are shown in Figure 2. The others, *Eco*RI, *Bam*HI and *Sal*I all generated large single *Gapdh* fragments (data not shown). Therefore, in *D. hydei* all *Gapdh* genes probably reside on a single fragment and one of these fragments, that generated by *Hin*d III digestion is only 1.5 kb which is too small to contain more than one *Gapdh* gene. A genomic library of *D. hydei* DNA was prepared as a partial *Mbo*1 digest in the lambda vector EMBL-4 (MENOTTI-RAYMOND, STARMER and SULLIVAN 1991). Nine recombinant lambda clones were isolated using the *Gapdh-43E* probe. All clones were judged by their restriction maps to overlap and come from the same chromosomal region (data not shown). A single *Hin*dIII fragment of 1.5 kb was found to hybridize to a *Gapdh* probe. This fragment was isolated and its nucleotide sequence determined. The sequence and its translation is shown in Figure 3. This sequence starts 346 bp 5′ to the translation start codon and ends 165 bp 3′ to the termination signal. A striking aspect of the *D. hydei* gene is the presence of a 69-bp intron (nucleotides 434–502) which interrupts the coding region following amino acid 29. Neither *D. melanogaster* gene has any introns within the coding region. Comparison of the *D. hydei* and *D. melanogaster Gapdh* genes reveals that the two *D. melanogaster* genes are more similar to each other than either is to the *D. hydei* gene (Table 2). This is the case when comparisons of the nucleotide sequences or the encoded amino acid sequences are performed.

In situ hybridization of a *Gapdh* probe to *D. hydei* salivary chromosomes results in a single site of hybridization on chromosome *5* (Figure 4). This is equivalent

TABLE 1

GAPDH isozyme distribution in the genus *Drosophila*

Subgenus	Group	Subgroup	Species	GAPDH forms
Drosophila	Virilis	Virilis	*virilis*	1 large
	Picture Winged	Hawaiiensis	*grimshawi*	1 large
	Repleta	Hydei	*hydei*	1 large
		Mulleri	*mojavensis*	1 large
			mulleri	1 large
			arizona	1 large
Dorsilopha			*busckii*	1 large
Sophophora	Willistoni		*willistoni*	1 large
	Saltans		*saltans*	1 large
	Obscura	Obscura	*persimilis*	1 small
			pseudoobscura	1 small
			subobscura	1 small
		Affinis	*affinis*	1 small
	Melanogaster	Melanogaster	*melanogaster*	Both
			simulans	Both
			yakuba	Both
			mauritiana	Both
		Takahashi	*takahashi*	Both
			lutescens	Both
		Susuki	*mimetica*	1 large
			lucipennis	1 large
		Ananassae	*ananassae*	1 large
			parabipetinalis	1 large
		Ficusphila	*ficusphila*	1 large
		Eugracilis	*eugracilis*	1 large
		Elegans	*elegans*	1 large
		Montium	*bamali*	1 large
			auraria	1 large
			kanaplae	1 large
Scaptodrosophila	Victoria	Victoria	*lebanonesis*	1 small

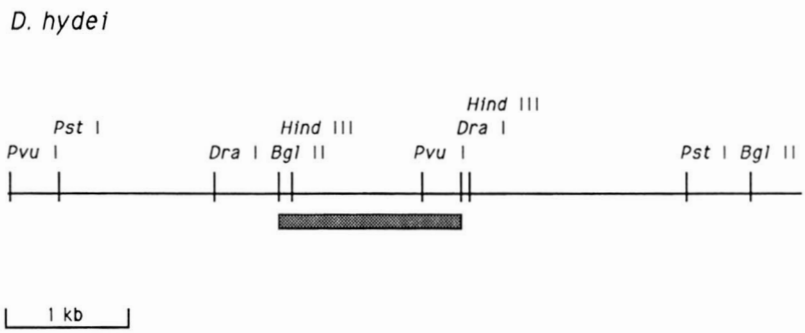

FIGURE 2.—Restriction map of *D. hydei* genomic DNA determined by probing with a *D. melanogaster* *Gapdh* gene.

to genetic element C of the virilis-repleta radiation and chromosome 5 of *Drosophila virilis* (WASSERMAN 1982). WHITING *et al.* (1989) have conducted an extensive series of *in situ* hybridizations to *D. virilis* salivary chromosomes using probes of *D. melanogaster* origin. Five of these probes are localized to chromosome 5 of *D. virilis*. Their probe 555 hybridizes to 43A-B of *D. melanogaster* and at a position near the tip of the *D. virilis* chromosome. Another probe, 514, hybridizes to 51B of *D. melanogaster* and at a position about 1/3 the distance from the *D. virilis* centromere. Another probe DT83^{b} which hybridizes to 56C of the *D. melanogaster* chromosome is distal to 514 on the *D. virilis* chromosome. Therefore, one might expect that a probe from 43E of *D. melanogaster* might hybridize either near 555 at the distal end of the chromosome or to a site proximal to the sites of 514 and DT83^{b} or within the proximal third of chromosome 5 depending on how the cytogenetic events which have occurred during evolution separated positions 43A-B, 43E and 51B. The site of hybridization of the *Gapdh* probe to *D. hydei* chromosomes is clearly within the first third of chromosome 5. Therefore, within the limits of cytogenetic comparisons possible, we conclude that the *D. melanogaster* 43E site and *D. hydei* sites are homologous and that the 43E site in *D. melanogaster* is homologous to the the ancestral position of the single *Gapdh* gene.

FIGURE 3.—Nucleotide sequence of the *Gapdh* region of *D. hydei*.

TABLE 2

Comparison of *D. melanogaster* and *D. hydei Gapdh* genes

Gene	Amino acid differences	Percent nucleotide similarity[a]
Gapdh-43E/Gapdh-Hy	13	95.8
Gapdh-13F/Gapdh-Hy	12	96.4
Gapdh-43E/Gapdh-13F	8	97.6

[a] Percent nucleotide similarity is calculated for nucleotides within the coding region only.

This comparison of the *Gapdh* genes of *D. melanogaster* and of *D. hydei* supports the hypothesis that the original *Gapdh* duplication occurred early in the evolution of the Sophophoran radiation and raises the possibility that the mechanism of duplication was through a retrotransposition event thereby accounting for the loss of the intron. We have proceeded to survey the genomes of other species with respect to the number of *Gapdh* genes and whether an intron at or near codon 29 is present. Since it is likely that the duplication occurred early in the evolution of the genus we have selected for study species from the major groups of the two major subgenera, Sophophora (melanogaster, obscura, willistoni and saltans groups) and Drosophila (virilis, repleta and immigrans groups) and the subgenera Dorsilopha and Scaptodrosophila.

DNA from this representative group of species were digested using a number of restriction enzymes and Southern blots were probed with a mixture of *Gapdh-43E* and *Gapdh-13F* derived probes. Representative enzyme digests of the DNA of these species which most clearly demonstrate the minimum number of genes present are shown in Figure 4. It is evident that the genomes of *Drosophila busckii* (Figure 5A, lanes 1, 2 and 3) and *Drosophila immigrans* (Figure 5A, lanes 4, 5 and 6) yield single *Gapdh* restriction fragments following digestion with either of two enzymes and a single fragment after digestion with both enzymes. Given that the duplicate *Gapdh* genes when present are not arranged in tandem, this pattern of restriction digestion indicates with high probability that the genomes of these species contain a single *Gapdh* gene. *Drosophila auraria* (Figure 5A, lanes 7, 8 and 9) on the other hand has a restriction digest pattern that has three bands. Since the smallest of the three bands

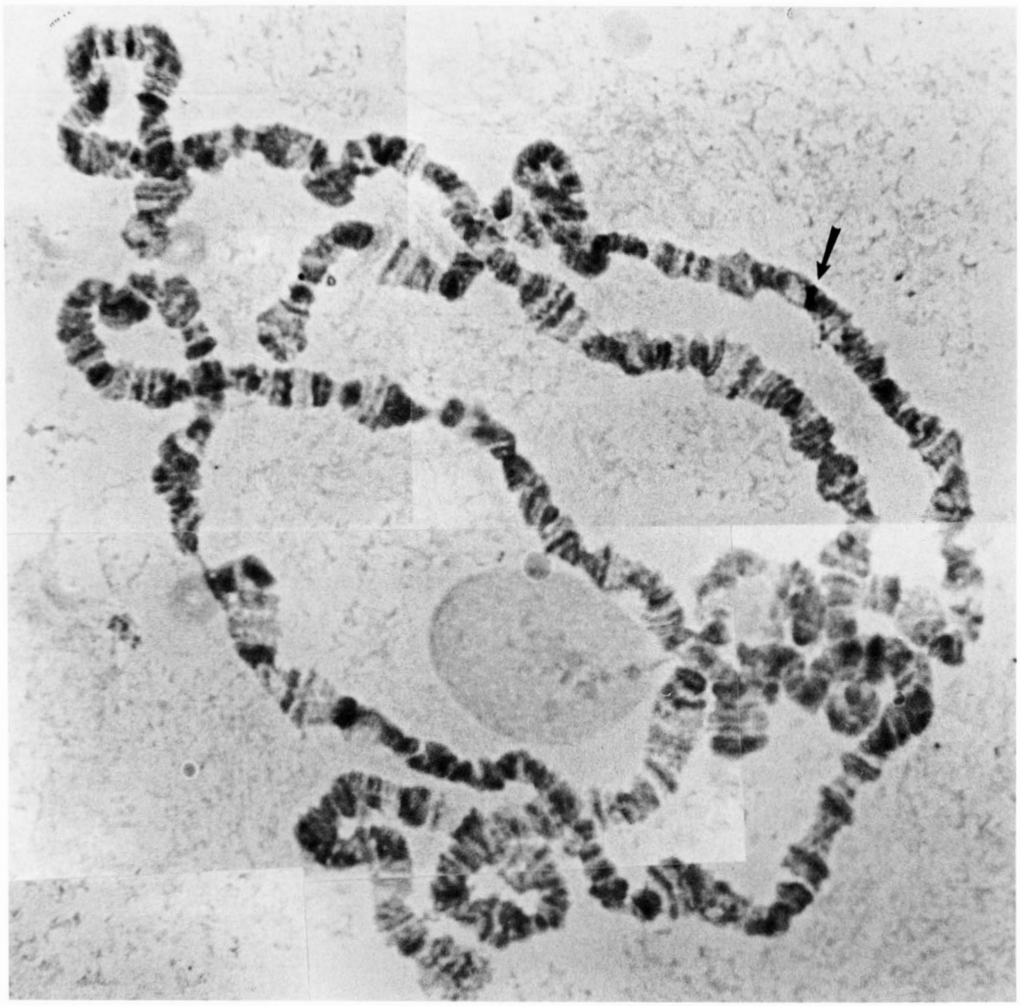

FIGURE 4.—*In situ* hybridization of a *D. melanogaster Gapdh* probe to polytene chromosomes of *D. hydei*.

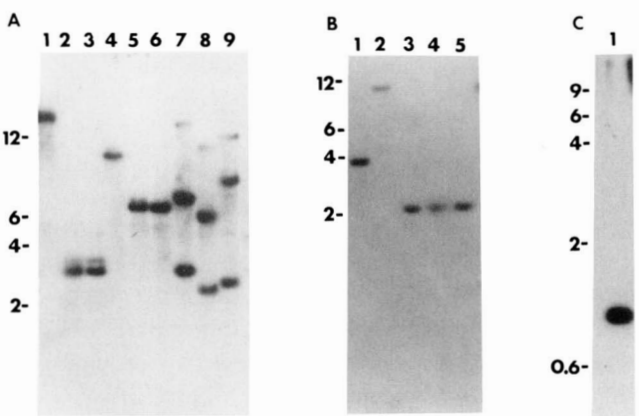

FIGURE 5.—Southern blots of DNA from various species probed with a mixture of *Gapdh 13F* and *Gapdh 43E* fragments. (A) Lanes 1, 2 and 3 are *D. busckii* DNA; lane 1, *Eco*RI; lane 2, *Eco*RI-*Pst*I; lane 3, *Pst*I. Lanes 4, 5 and 6 are *D. immigrans* DNA; lane 4, *Hin*dIII; lane 5, *Hin*dIII-*Pst*I; lane 6, *Hin*dIII. Lanes 7, 8 and 9 are *D. auraria* DNA; lane 7, *Nde*I; lane 8, *Nde*I-*Hin*dIII; lane 9, *Hin*dIII. (B) *D. saltans* DNA; lane 1, *Bam*HI; lane 2, *Eco*RI; lane 3, *Hin*dIII-*Eco*RI; lane 4, *Hin*dIII; lane 5, *Hin*dIII-*Bam*HI. (C) *D. willistoni* DNA digested with *Cla*I.

in each digest is larger than any known *Gapdh* gene, these data require that the genome contains more than one gene or alternatively the existence of a 2-kb or greater intron positioned so as to occupy most of the smallest restriction fragment seen in the digests. Furthermore, this would have to be a newly created intron since such an intron is not found in other species which have common ancestors with *D. auraria*. This combination of unlikely events makes it improbable that *D. auraria* has a single *Gapdh* gene. The genomes of *Drosophila saltans* (Figure 5B) and *Drosophila willistoni* (Figure 5C) generate only single *Gapdh* genomic restriction fragments and in the case of *D. willistoni* this band is less than 1.5 kb. Since it is unlikely that two nontandemly arranged genes would each have two *Cla*I sites identically positioned on such a small fragment, it is likely that the genomes of these related species contains a single *Gapdh* gene.

In situ hybridization to *D. pseudoobscura* salivary gland chromosomes reveals several sites of hybridization. Two of these, marked A and B in Figure 6, are on chromosomes *XL* and *3*, respectively. These are likely to be homologous to the 13F and 43E positions of *D. melanogaster*. A third region of hybridization, C, which is in or near the heterochromatin of chromo-

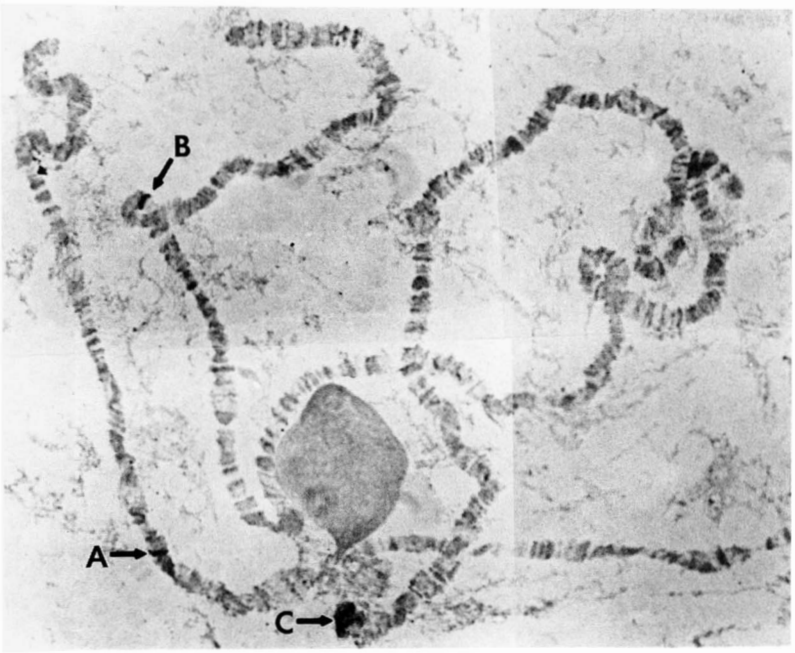

FIGURE 6.—*In situ* hybridization of a *D. melanogaster Gapdh* probe to polytene chromosomes of *D. pseudoobscura*.

some *XR* and sometimes appears as two separate regions can also be seen in Figure 6. The nature of these heterochromatic regions has not been further investigated. Southern blots of genomic DNA generate a pattern of fragments consistent with two *Gapdh* genes (not shown). This indicates that species in the obscura group contain duplicate *Gapdh* genes and that the presence of a single isoform in these species (Table 1) is best interpreted as resulting from two proteins occupying the same gel position or the translational silencing of one gene.

Since the single *Gapdh* gene of *D. hydei* has an intron, while neither of the two *Gapdh* genes of *D. melanogaster* has an intron, we have explored the possibility that duplication and intron loss happened coincidentally during evolution which would be the case if the *Gapdh* duplication occurred by means of a retrotransposition event. We have used a gene amplification technique to measure the distance from codon 1 to codon 85 of the *Gapdh* gene in genomic DNA of a number of species. The two primers were designed to complement conserved regions in *Gapdh* genes so as to hybridize equally well to both of the *D. melanogaster* genes and the *D. hydei* gene in order to be effective with any *Gapdh* from the genus. Since the amplified fragment spans the position of the intron (codon 29) found in the *D. hydei* gene, a *Gapdh* gene without the intron in its genomic DNA should result in a fragment of 254 bp, while a fragment larger than this would be found when the intron is present in genomic DNA. The exact size of the fragment might vary with variation of intron size but in the case of *D. hydei* the amplified fragment should be 323 bp. The results of the gene amplification analysis are shown in

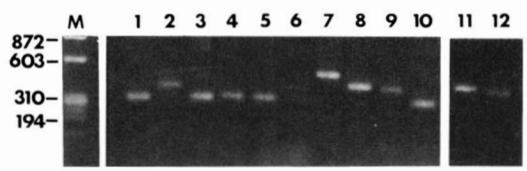

FIGURE 7.—Amplification of the *Gapdh* genes of several species across the region of the putative intron position. Lane 1, *D. melanogaster*; lane 2, *D. hydei*; lane 3, *D. saltans*; lane 4, *D. pseudoobscura*; lane 5, *D. willistoni*; lane 6, *D. lebanonensis*; lane 7, *D. busckii*; lane 8, *D. immigrans*; lane 9, *D. americana*; lane 10, D. *auraria*. Since *D. lebananensis* is of particular interest to these studies (see DISCUSSION) and lane 6 is poorly amplified there is a second lane of the products of amplification of *D. lebanonensis* in lane 12 with a simultaneously run *D. hydei* as a control, lane 11.

Figure 7. All species of the subgenus Sophophora, including *D. melanogaster*, *D. saltans*, *D. pseudoobscura*, *D. willistoni* and *D. auraria*, generate a 254-bp fragment indicating that there is no intron in the *Gapdh* gene(s) of these species. *D. immigrans*, (a member of the immigrans group of the quinaria section), *D. hydei* (a member of the repleta group) and *Drosophila americana* (a member of the virilis group), have an amplified fragment of about 320 bp. *D. busckii* of the subgenus Dorsilopha (lane 7) has an amplified fragment larger than 320 bp. *Drosophila lebanonensis* of the subgenus Scaptodrosophila has an amplified fragment a few nucleotides shorter, but close to the size of the amplified fragment of *D. hydei*, lanes 6 and 12. The initial amplification of *D. lebanonensis* DNA (lane 6) was only partially successful so a second DNA sample was prepared and amplified (lane 12) and compared to a second sample of *D. hydei* DNA (lane 11). Since GAPDH is an extremely conserved protein, flexibility in amino acid number internally in the protein is highly unlikely. Accordingly, species with

an amplified fragment larger than 254 bp must have an intron in their *Gapdh* gene. Since species from each of the four groups of the subgenus Sophophora have a *Gapdh* gene without an intron but only two of these species groups have a duplication, these results indicate that the loss of the intron in the *Gapdh* genes occurred very early in the evolution of the Sophophoran lineage and in all likelihood preceded the *Gapdh* duplication.

DISCUSSION

All available evidence is consistent with the *Gapdh* duplication having occurred at an early point during the evolution of the Sophophoran subgenus. There is no evidence of duplicate *Gapdh* genes in any species of the subgenus Drosophila. In *D. hydei* a number of independent lines of evidence clearly indicate a single gene. Therefore it appears that early in the evolution of the subgenus Sophophora two events occurred. One was the loss of the intron and the second was duplication of the intron-less *Gapdh* gene. We began these studies attracted to the hypothesis that the loss of the intron and the duplication might have occurred simultaneously by a single retrotransposition event. However, the simplest, most economical hypothesis based on the analysis of species presented here is that the intron was lost substantially earlier in the Sophophoran lineage than the time at which the duplication occurred. The species of the related groups, saltans and willistoni, clearly have a *Gapdh* gene without an intron. Yet Southern blots indicate that these species each have only a single *Gapdh* gene. More complicated hypotheses can be invoked that might connect the events of gene duplication and intron loss. Accordingly, the current data are consistent with a model in which an original duplication occurred through a retrotransposition event followed by the loss of one of the genes in the saltans-willistoni lineage. Alternatively, the initial duplication might have occurred through retrotransposition and the subsequent loss of the intron from the ancestral gene by means of gene conversion from the retrotransposally derived copy in ancestors of the obscura and melanogaster subgroup species. However, these hypotheses will be difficult to test since the presumed gene loss would have happened in the distant evolutionary past, early in the Sopohoran lineage.

Consideration of each of the relevant subgroups suggests the possibility of additional complications in the history of evolution of the *Gapdh* genes. Species of the melanogaster subgroup have two proteins and two functional genes. There is no evidence of additional genes or pseudogenes. Two protein isoforms are also found in species of the closely related takahashi subgroup. However, species of the melanogaster group from subgroups more distantly related to the melanogaster subgroup, *e.g.*, *D. auraria* of the montium subgroup and species of the obscura group apparently contain only a single protein form yet contain two *Gapdh* genes. Two explanations for the existence of only one isoform are plausible. It is possible that amino acid substitutions have resulted in the two protein forms having identical mobilities. However, we have not observed any compensating increase in the intensity of the remaining single protein form as compared to the large band of *D. melanogaster*. Alternatively, one of the coding genes may have become translationally silent and may now be a pseudogene. It seems possible that a species could survive with only a single functional *Gapdh* gene since deletions of one of the *D. melanogaster* genes (Figure 1) results in flies that while flightless are viable. The flightless phenotype is consistent with the importance of glycolysis in flight muscle energy generation and the observation of SUN *et al.* (1988) who showed that *Gapdh-43E* is abundantly expressed in thoracic muscle.

Our model regarding the origin of the *Gapdh* duplication assumes that intron loss occurred in the Sophophoran lineage rather than intron gain in the Drosophila lineage. The origin of introns remains the subject of debate (GREEN 1988). Several genes which encode enzymes that function as part of common metabolic pathways which are thought to be very old, such as glycolysis have been studied with respect to intron position (MARCHIONNI and GILBERT 1986). One view that has emerged is that the position of introns is conserved over long evolutionary times, therefore introns are thought to represent early events in protein evolution which are subject to loss over time (GILBERT, MARCHIONNI and MCKNIGHT 1986). Conserved intron positions in the *Gapdh* genes support this view and have been used to argue that intron positioning predates the prokaryotic-eukaryotic divergence (QUIGLEY, MARTIN and CERF 1988; SHIH, HEINRICH and GOODMAN 1988). The manner through which introns are lost has not been ascertained but it seems reasonable that the precision of intron loss requires reverse transcription. FINK (1987) has argued that the scarcity of introns in the yeast genome is the result of reverse transcription of a large fraction of yeast genes over evolutionary time. Accordingly, if one accepts the position that introns are old and only subject to loss, it seems most likely that the presence of an intron in the *Gapdh* gene of *D. hydei* and its absence in the genes of *D. melanogaster* is due to loss of the intron in the *D. melanogaster* lineage. There is no direct demonstration that the history of *Gapdh* evolution in the genus *Drosophila* could not have involved the gain of a new intron at codon 29 early in the evolution of the subgenus Drosophila. However, *D. lebanonensis*, a member of the subgenus Scaptodrosophila, has an intron in its *Gapdh* gene.

The early evolution of subgenera in the genus *Drosophila* is not unambiguously determined but the subgenus Scaptodrosophila is thought to have diverged before the divergence of the Sophophora and Drosophila subgenera (BEVERLEY and WILSON 1982). If this view is correct, then there would have been an intron in the *Gapdh* genes of the ancestors of the Sophophoran groups thereby making it highly likely that the absence of the intron in the *Gapdh* genes of species from these groups is due the loss of an old intron.

Since GAPDH has been the object of extensive physical structural characterization the position of introns in *Gapdh* genes from several sources has been studied in order to further investigate the relation between intron position and the evolution of protein structure. For the most part the position of introns in *Gapdh* genes correlate well with the boundaries of proposed domains in GAPDH in the *Gapdh* genes of: chickens (STONE, ROTHBLUM and SCHWARTZ 1985); nematodes (YARBROUGH *et al.* 1987); humans (ERCOLAN *et al.* 1988); maize (QUIGLEY, MARTIN and CERF 1988); and nuclear and mitochondrial encoded GAPDH of arabidopsis (SHIH, HEINRICH and GOODMAN 1988). However, the relation between protein domains and intron position is not well correlated in the *Gapdh*/GAPDH of Aspergillus (PUNT *et al.* 1988). An intron at codon 29 as in *D. hydei Gapdh* has not been observed in other species. The detailed comparison of GAPDH domain organization and *Gapdh* intron position by SHIH, HEINRICH and GOODMAN (1988) reveals that introns have not yet been found at all potential domain delineating positions. Interestingly, these authors predict that an intron might be found in the interval between codons 20 and 30. The intron in the *D. hydei Gapdh* gene at codon 29 apparently confirms this prediction. Therefore, this intron is at a position that is located near the borders of two protein domains.

This research was supported by National Institutes of Health grant 26830 to D.T.S. We thank WYATT ANDERSON and MARVIN WASSERMAN for their assistance in the cytogenetic localization in *D. pseudoobscura* and *D. hydei*, respectively. The nucleotide sequence of the *Gapdh* gene of *D. hydei* has been submitted to the EMBL data base and has been assigned Acc. No. Z14144.

LITERATURE CITED

ALEXANDROV, I. D., 1984 Comparative genetics of neutron- and X-ray-induced *b*, *cn* and *vg* mutations in *D. melanogaster*. Drosophila Inform. Serv. **60:** 45–47.

BEVERLEY, S. M., and A. C. WILSON, 1982 Molecular evolution in *Drosophila* and higher Diptera. I. Micro-complement fixation studies of a larval hemolymph protein. J. Mol. Evol. **18:** 251–264.

BIGGIN, M. D., T. S. GIBSON and G. G. HONG, 1983 Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80:** 3963–3965.

ERCOLAN, L., B. FLORENCE, M. DENARO and M. ALEXANDER, 1988 Isolation and complete sequence of a functional human glyceraldehyde-3-phosphate dehydrogenase gene. J. Biol. Chem. **263:** 15335–15341.

FINK, G. R., 1987 Pseudogenes in yeast? Cell **49:** 5–6.

GILBERT, W. M., MARCHIONNI and G. MCKNIGHT, 1986 On the antiquity of introns. Cell **46:** 151–153.

GREEN, M. R., 1988 Mobile RNA catalysts. Nature **336:** 716–718.

HONG, G. F., 1982 A systematic DNA sequencing method. J. Mol. Biol **158:** 539–549.

LANGER-SAFER, P. R., M. LEVINE and D. C. WARD, 1982 Immunological method for mapping genes on *Drosophila* polytene chromosomes. Proc. Natl. Acad. Sci. USA **79:** 4381–4385.

MARCHIONNI, M., and W. GILBERT, 1986 The triosephosphate isomerase gene from maize: introns antedate the plant-animal divergence. Cell **46:** 133–141.

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

PUNT, P. J., M. A. DINGEMANSE, B. J.-M. JACOB-MEIJSING, P. H. POUWELS and C. A. M. J. J. VAN DER HONDEL, 1988 Isolation and characterization of the glyceraldehyde-3-phosphate dehydrogenase gene of *Aspergillus nidulans*. Gene **69:** 49–57.

QUIGLEY, F., W. F. MARTIN and R. CERF, 1988 Intron conservation across the prokaryote-eukaryote boundary: structure of the nuclease gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. Proc. Natl. Acad. Sci. USA **85:** 2672–2676.

SHIH, M.-C., P. HEINRICH and H. M. GOODMAN, 1988 Intron existence predated the divergence of eukaryotes and prokaryotes. Science **242:** 1164–1166.

STONE, E. M., K. N. ROTHBLUM and R. J. SCHWARTZ, 1985 Intron-dependent evolution of chicken glyceraldehyde phosphate dehydrogenase gene. Nature **313:** 498–500.

SULLIVAN, D. T., W. T. CARROLL, C. L. KANIK-ENNULAT, Y. HITTI, J. A. LOVETT and L. VON KALM, 1985 Glyceraldehyde-3-phosphate dehydrogenase from *Drosophila melanogaster*: Identification of two isozymic forms encoded by separate genes. J. Biol. Chem. **260:** 4345–4350.

SUN, X.-H., J. T. LIS and R. WU, 1988 The positive and negative transcriptional regulation of the Drosophlia *Gapdh-2* gene. Genes Dev. **2:** 743–753.

SUN, X.-H., J. Y. T'SO, J. LIS and R. WU, 1988 Differential regulation of the two glyceraldehyde-3-phosophate dehydrogenase genes during Drosophila development. Mol. Cell. Biol. **8:** 5200–5205.

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

T'SO, J. Y., X.-H. SUN and R. WU, 1985 Structure of two unlinked *Drosophila melanogaster* glyceraldehyde-3-phosphate dehydrogenase genes. J. Biol. Chem. **260:** 8220–8228.

WASSERMAN, M., 1982 Cytological evolution in the Drosophila repleta species group, pp. 49–64 in *Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila Model System*, edited by J. S. F. BARKER and. W. T. STARMER. Academic Press, Sydney, Australia.

WHITING, J. H. J., M. D. PLILEY, J. L. FARMER and D. E. JEFFEREY, 1989 *In situ* hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis*. Genetics **122:** 99–109.

YARBROUGH, P. O., M. A. HAYDEN, L. A. DUNN, P. S. VERMERSCH, M. R. KLASS and R. M. HECHT, 1987 The glyceraldehyde-3-phosphate dehydrogenase gene family in the nematode *Caenorhabditis elegans*: isolation and characterization of one of the genes. Biochim. Biophys. Acta **908:** 21–33.

Communicating editor: C. C. LAURIE