# Interspecific Comparisons of the Structure and Regulation of the Drosophila Ecdysone-Inducible Gene *E74*

**C. Weldon Jones, Mark W. Dalton and Laurie H. Townley**

*Department of Biology, Bethel College, St. Paul, Minnesota 55112*

Manuscript received October 2, 1990
Accepted for publication November 21, 1990

ABSTRACT

The *Drosophila melanogaster E74* gene is induced directly by the steroid hormone ecdysone and is a member of a small set of "early" genes that appear to trigger the onset of metamorphosis. The gene consists of three overlapping transcription units encoding two proteins, E74A and E74B, which possess a common C terminus. According to the Ashburner model for ecdysone's action, an *E74* protein product potentially functions as a transcriptional activator of "late" genes as well as a repressor of early genes. We have taken an evolutionary approach to understand the function and regulation of *E74* by isolating the homologous genes from *Drosophila pseudoobscura* and *Drosophila virilis* and comparing them to *D. melanogaster E74* sequences. Conserved characteristics of the *E74* genes include ecdysone inducibility, localization to ecdysone-induced polytene chromosome puffs, and gene size. Amino acid sequence comparisons of the E74A protein reveal a highly conserved C-terminal region that is rich in basic amino acid residues and which has been proposed to possess sequence-specific DNA binding activity. The moderately conserved N-terminal region has maintained its overall acidic character and is a potential transcriptional activator domain. The central region contains conserved glutamine and alanine homopolymeric repeats of variable lengths. Nucleotide sequence comparisons of the E74A promoter region fail to reveal ecdysone-response elements but do identify conserved sequences that may function in E74A regulation.

THE process of cell differentiation and ultimately of animal development is dependent upon the coordinated action of a large number of individual genes. An ideal system for studying such a process is the regulation of gene expression by the steroid hormone 20-OH ecdysone (hereafter referred to as ecdysone) in *Drosophila melanogaster*. Ecdysone is the primary steroid hormone utilized by insects and is responsible for initiating such developmental events as larval molting and metamorphosis to the adult fly.

Early studies by ASHBURNER and colleagues on the effects of ecdysone on polytene chromosomes of late third instar larval salivary glands led to a working model for ecdysone's action at the molecular level (ASHBURNER *et al.* 1974). It was observed that shortly after salivary gland exposure to ecdysone, about six "early" transcriptionally active puffs appear independently of protein synthesis and regress after a few hours. The early puffs are followed by the appearance of over 100 late puffs which are dependent upon prior protein synthesis and whose protein products are presumably responsible for the tissue's morphological changes observed *in vivo*. According to the Ashburner model, ecdysone acts at the cellular level by first binding to a protein receptor which then assists in activating a small number of early genes (those genes at the early puff sites). The early gene products are then responsible for turning on the larger number of late genes and also for turning off the transcription of their own genes. Recently, an extension to Ashburner's model has been proposed by BURTIS *et al.* (1990), who suggest that ecdysone's tissue specificity is the result of the hormone inducing small overlapping sets of regulatory genes in all tissues. The products of the regulatory genes would serve to control the expression of larger sets of genes which are responsible for the observed developmental changes in the tissues.

At the top of this genetic regulatory hierarchy in Drosophila are the early genes. Given their apparent function, the early genes would be expected to encode proteins with multiple domains, *i.e.*, DNA binding domain, transcriptional activator domain, and/or a region with transcriptional repressor activity. Three early genes have been isolated and partially characterized, those at polytene chromosome position 2B5 (CHAO and GUILD 1976), 74EF (BURTIS *et al.* 1990), and 75B (SEGRAVES and HOGNESS 1990). All three genes are large (>60 kb) and complex, with multiple overlapping transcription units.

The early gene at 74EF, designated *E74*, contains three transcription units which share C-terminal exons but which have unique promoters. The E74A unit is 60 kb, generates a 6.0-kb mRNA, and is ecdysone inducible in several larval and imaginal tissues. In addition, it is induced during each of the six Drosophila developmental stages that are marked by a pulse

of ecdysone (BURTIS *et al.* 1990; THUMMEL, BURTIS and HOGNESS 1990). Within one of the E74A introns are two E74B promoters which are used to produce two 4.8- and 5.1-kb mRNAs but a single protein. The E74B transcripts are expressed according to their own distinct developmental pattern (BURTIS *et al.* 1990; THUMMEL, BURTIS and HOGNESS 1990). Because of its clear ecdysone inducibility, the E74A transcription unit has been the focus of attention for understanding the genetic regulatory mechanisms underscored here.

The E74A protein product of *D. melanogaster* is 829 amino acids long and may be subdivided into three regions (domains) according to amino acid sequence and possible function. The N-terminal domain is relatively rich in acidic amino acids and is a potential transcriptional activator domain (BURTIS *et al.* 1990; PTASHNE 1988). The C-terminal domain, rich in basic amino acids, is a potential DNA-binding domain with sequence similarity to the proposed DNA-binding domain of the *ets* proto-oncogene protein product (WASYLYK *et al.* 1990; KARIM *et al.* 1990). The central region of the E74A protein contains multiple homopolymeric sequences of specific amino acids, especially glutamine and alanine. These repeats may be simply spacer sequences between the two outer domains or may possess an as yet unidentified function.

If the early genes of Drosophila represent a small set of developmentally important regulatory genes utilized by many tissues, the genes would be expected to be conserved through insect evolution. By isolating *E74*-homologous genes from other Drosophila species and comparing them to *E74* from *D. melanogaster*, it should be possible to identify evolutionarily constrained sequences important for *E74* regulation and function. This approach has been used successfully to study other Drosophila genes, *e.g.*, heat shock protein genes (BLACKMAN and MESELSON 1986) and chorion genes (MARTINEZ-CRUZADO *et al.* 1988).

This paper describes the isolation and characterization of *E74*-homologous genes from *D. pseudoobscura* and *D. virilis*. *D. melanogaster* and *D. pseudoobscura* belong to the same subgenus (*Sophophora*) but different species groups and their estimated divergence time is approximately 20–50 million years. *D. virilis* belongs to a different subgenus (*Drosophila*) which separated from the *Sophophora* subgenus approximately 50–80 million years ago (THROCKMORTON 1975; BEVERLEY and WILSON 1984). Here we show that both *D. pseudoobscura* and *D. virilis* *E74*-homologous genes are ecdysone inducible and map to ecdysone-inducible puffs on their respective polytene chromosomes. A comparison of DNA and protein sequences from these genes identifies presumptive *E74* functional elements that may be important in defining the gene's role in regulating Drosophila development.

## MATERIALS AND METHODS

**General methods:** Restriction enzyme digests, agarose and polyacrylamide gel electrophoresis, DNA purification from gels, nick translation, Southern blot hybridization, subcloning into plasmids, bacterial transformation, and λ phage library screening were carried out using methods essentially as described by MANIATIS, FRISCH and SAMBROOK (1982). For chromosome walking the procedure of BURTIS *et al.* (1990) was followed. Phage purification and growth were performed according to DAVIS, BOTSTEIN and ROTH (1980). The *D. virilis* and *D. pseudoobscura* λ (EMBL) libraries were kindly provided by R. BLACKMAN (Harvard) and S. SCHAEFER (Pennsylvania State University), respectively.

**Drosophila culture methods:** *D. pseudoobscura* and *D. virilis* flies were obtained from stocks maintained at the Harvard Biological Laboratories and were grown on instant Drosophila medium (Carolina Biological Supply Co.) at 25°. For culturing salivary glands *in vitro*, late third instar larvae were dissected in Robb's saline (ROBB 1969) and the glands placed in Petri dishes containing oxygenated Robb's saline. When used, ecdysone (Sigma) was added to a final concentration of $4 \times 10^{-6}$ M and cycloheximide (Sigma) was added to a final concentration of $7 \times 10^{-5}$ M. Glands were incubated for about 3 hr before they were harvested for use.

**RNA isolation:** For RNA isolation from salivary glands, following their incubation in Robb's saline the glands were transferred with medium to 1.5-ml microfuge tubes and centrifuged briefly. The supernatant was removed and the glands homogenized in a 1:1 mixture of phenol and extraction buffer (0.3 M NaCl, 0.1 M Tris base, 20 mM EDTA, 1% Sarkosyl). After centrifugation the aqueous phase was reextracted with phenol and then ether extracted. The RNA was precipitated with ethanol, the pellet washed and resuspended in distilled water.

***In situ* hybridization to polytene chromosomes:** For polytene chromosome slide preparations salivary glands were squashed as described by ENGELS *et al.* (1986). The DNA probes used for the *in situ* hybridization experiments shown in Figure 3 were genomic restriction fragments containing exons 2–5 from pE74A or vE74A. These fragments were nick translated using biotin-14-ATP and the BRL Bionick Labeling System (BRL). The *in situ* hybridization procedure and subsequent DNA detection were carried out according to ENGELS *et al.* (1986) using the streptavidin-alkaline phosphatase (BRL) staining method.

**Northern blot hybridization:** RNA suspended in distilled water was glyoxylated (MCMASTER and CARMICHAEL 1977) and fractionated on 1.0% agarose gels in 0.01 M sodium phosphate (pH 7.2). A total of 10 μg RNA was electrophoresed per lane. Following electrophoresis the gels were stained with ethidium bromide to verify equal nucleic acid quantities per lane. The RNA was blotted to Zeta-Probe membrane (Bio-Rad) which was then hybridized to primer-extended single-stranded DNA probes obtained from M13 clones with inserts of exons 3–5 from pE74A or vE74A.

**DNA sequencing:** DNA sequencing of end-labeled restriction fragments from pE74A and vE74A was carried out using the method of MAXAM and GILBERT (1977). Additional sequencing utilized the Sanger method (SANGER, NICKLEN and COULSON 1977). Here DNA fragments to be sequenced were inserted into the M13 vectors, M13mp18 and M13mp19 (YANISCH-PERRON, VIERA and MESSING 1985), which were then transformed into *Escherichia coli* strain JM101. Single-stranded template DNA was prepared as described by MESSING (1983) and the inserts sequenced using Sequenase (modified T7 DNA polymerase, U.S. Biochemical).

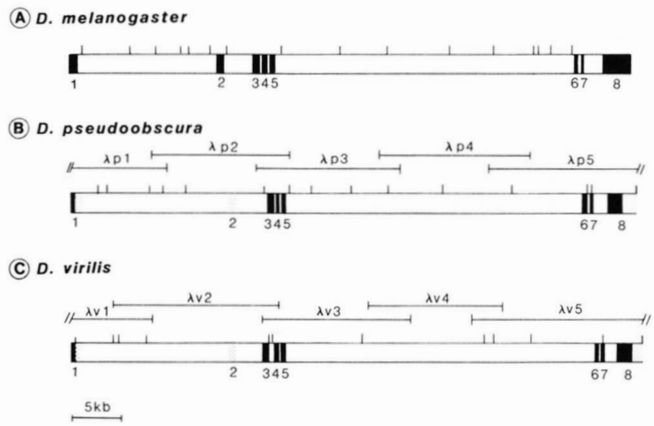


FIGURE 1.—Structural maps of *E74*. Maps of the approximately 60-kb *E74* genes from (**A**) *D. melanogaster* (from BURTIS *et al.* 1990), (**B**) *D. pseudoobscura*, and (**C**) *D. virilis* are shown. The locations and names of the genomic clones that span the *D. pseudoobscura* and *D. virilis* genes are indicated by the horizontal lines above the maps. The short vertical lines along each map represent *Eco*RI sites. The eight exons of the E74A transcription units are numbered, with those exons or portions of exons that have been sequenced shown in solid black and those that have not been sequenced shown in stipple. The positions for the 3′ ends of exons 1 and 8 in *D. pseudoobscura* and *D. virilis* are indicated with the assumption that both exons are approximately the same size as those of *D. melanogaster*. The positions of exon 2 in *D. pseudoobscura* and *D. virilis* were determined by hybridization to exon 2 from *D. melanogaster*.

## RESULTS

**Isolation and initial characterization of *E74*-homologous genes from *D. pseudoobscura* and *D. virilis*:** The *D. melanogaster E74* gene is approximately 60 kb and its E74A transcription unit (also 60 kb) has eight exons (BURTIS *et al.* 1990). Genomic λ phage libraries from *D. pseudoobscura* and *D. virilis* were screened with exon-containing fragments from *D. melanogaster E74* (designated *mE74*). Once the corresponding exon-containing genomic segments were isolated from the two species, the remainder of each gene was obtained by chromosome walking. Sufficient clones were analyzed by restriction enzyme mapping to conclude that the entire *E74*-homologous gene had been isolated and that the gene was unique in these species. The *E74*-homologous genes were designated *pE74* (from *D. pseudoobscura*) and *vE74* (from *D. virilis*). The precise positioning of the E74A-homologous exons was determined by Southern analysis, restriction enzyme mapping, and for most exons, by DNA sequencing. Figure 1 compares the organization of the three *E74* genes. Only the E74A transcription unit exons are indicated. The *E74* gene length and the E74A transcription unit length of approximately 60 kb is conserved among the species despite over 85% of the gene length consisting of introns.

**Expression and chromosome location of *pE74* and *vE74*:** The *mE74* gene has been shown to be induced directly by ecdysone. This was demonstrated by incubating tissues in the presence of both ecdysone and the protein synthesis inhibitor cycloheximide and then analyzing the transcribed RNA (BURTIS *et al.* 1990). Similar experiments were carried out using tissues from *D. pseudoobscura* and *D. virilis*. Third instar larval salivary glands from each of the two species were incubated in the presence or absence of ecdysone. To test for primary gene induction cycloheximide was also added. Figure 2 shows the results of Northern analyses using the RNA isolated from these glands and species-specific E74A exon-containing probes. An approximately 6-kb mRNA, equivalent in size to that produced in *D. melanogaster*, is generated by each species upon ecdysone induction (BURTIS *et al.* 1990).

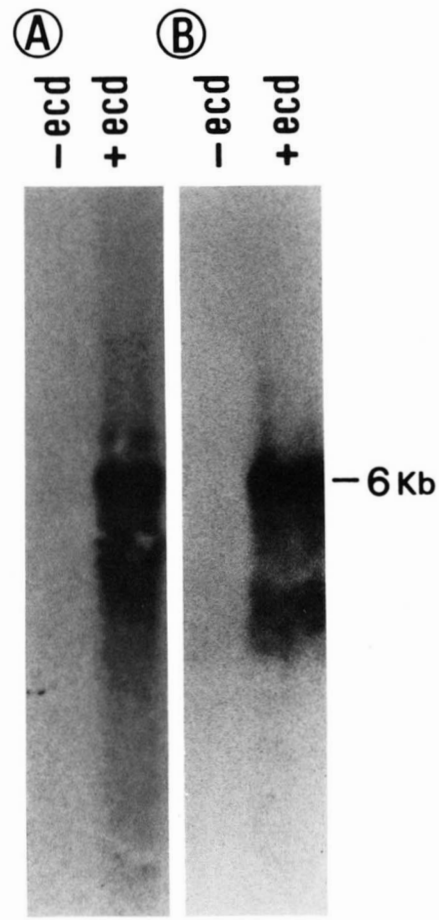


FIGURE 2.—Effect of ecdysone on E74A expression. Northern blot analyses of total RNA isolated from cycloheximide-treated third instar larval salivary glands incubated in the absence (− ecd) or presence (+ ecd) of ecdysone are shown. Radioactive single-stranded DNA from exons 3–5 of E74A from the respective species were used as probes. (**A**) *D. pseudoobscura* and (**B**) *D. virilis*.

The chromosome location of *pE74* and *vE74* was determined by *in situ* hybridization of biotin-labeled E74A DNA fragments to polytene chromosome preparations from the respective species. Figure 3 shows the results when using chromosomes from ecdysone-induced salivary glands. Both *pE74* and *vE74* genes are associated with ecdysone-inducible chromosome

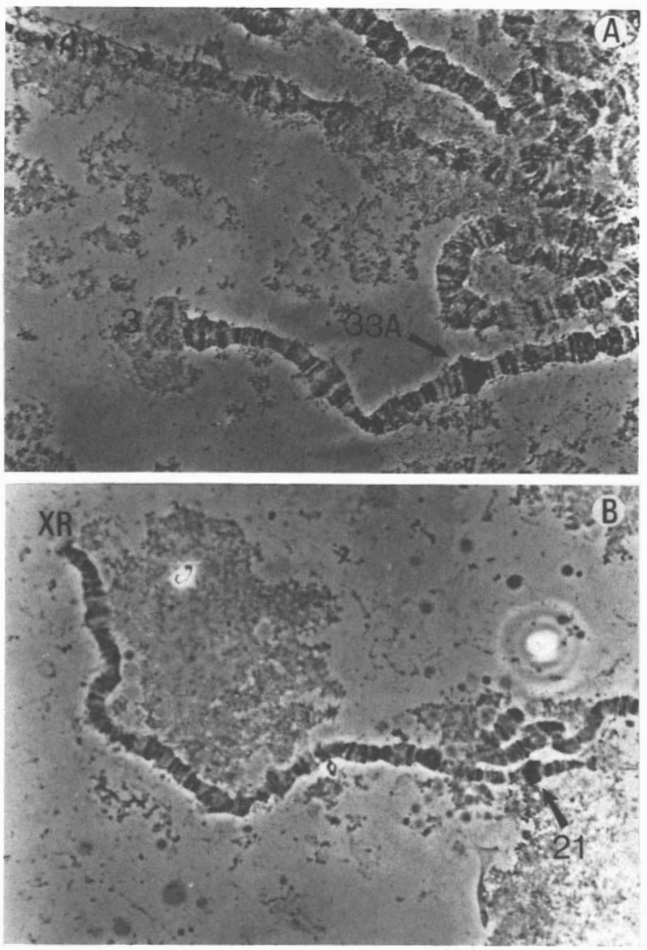


FIGURE 3.—*In situ* hybridization of *E74* genomic fragments to ecdysone-induced polytene chromosomes from (**A**) *D. pseudoobscura* and (**B**) *D. virilis*. Third instar larval salivary glands from *D. pseudoobscura* and *D. virilis* were incubated with ecdysone, squashed, and their polytene chromosomes hybridized to biotin-labeled E74A genomic fragments containing exons 2–5 from the respective species. Chromosome positions for the *E74* genes, indicated by the arrows, were determined by comparisons with the polytene chromosome maps of SORSA (1988).

puff sites, although *D. pseudoobscura* consistently gave a smaller puff than *D. virilis*. *pE74A* maps to site 21 on *D. pseudoobscura* chromosome *XR*, the homolog to chromosome arm 3L (location of *mE74*) of *D. melanogaster* (SORSA 1988). This *D. pseudoobscura* chromosome site had been shown previously to rapidly puff in response to ecdysone (STOCKER and KASTRITSIS 1973). Position 33A of *D. virilis* chromosome *3*, also homologous to chromosome arm 3L of *D. melanogaster*, is the site of *vE74* (SORSA 1988) and has been shown by others to produce an ecdysone-inducible puff (GUBENKO and EVGEN'EV 1984). Thus, it can be concluded that *E74* has remained ecdysone inducible and associated with the same chromosome arm and puff site for 50–80 million years.

**E74A protein sequence comparison:** The mE74A protein consists of 829 amino acid residues. BURTIS *et al.* (1990) have divided the protein structure into two outer domains and a central region, each with its own presumed function (see Introduction). The coding portions of the pE74A and vE74A transcription units were sequenced and the assigned amino acid sequences are displayed in Figure 4. The sequences are aligned with the mE74A protein sequence to show maximum homology. The pE74A and vE74A proteins are 20 amino acid residues less than and 44 residues greater than the mE74A protein, respectively. Stretches of amino acids greater than five residues and identical among all three species have been boxed. The most striking homology is found in the C-terminal region (residues 688–829 of mE74A protein) where there is 96% sequence identity. Table 1 confirms that the basic amino acid content in this region is significantly greater than the remainder of the protein for all three species.

The presumptive N-terminal domain (residues 1–190 of mE74A protein) has two long (greater than 20 residues) sequences of identity along with four shorter conserved sequences. The overall percent identity for this domain among the three species is approximately 60%. The importance of these conserved sequences for E74A protein function is unknown. However, while the conserved sequences represent only 36–41% (depending on the species) of the entire domain, from 56–62% of the domain's acidic amino acids are located there (Table 1 and Figure 4). All three proteins are relatively abundant in acidic residues at this end of the protein (from 13.5% to 16.8%; Table 1).

The central regions of all three proteins contain variable lengths of homopolymeric sequences, especially glutamine (Q) and alanine (A) (Figure 4). Interspersed with the homopolymeric sequences are several conserved sequences, most notably those between residues 390–451 (62 residues) and 476–517 (42 residues) from the mE74A protein. The importance of these conserved regions to E74A protein function is not known.

**E74A promoter sequence comparison:** The promoter regions of the pE74A and vE74A transcription units were sequenced to identify conserved DNA elements that might play a role in the regulation of *E74* gene expression. The positions for the startpoint of each transcription unit were determined by comparing the sequences to the *D. melanogaster* gene (BURTIS *et al.* 1990). While nearly 600 nucleotides upstream from the E74A transcription startpoint were determined for each species, significant homology was obvious only within 250 nucleotides of the startpoint. Figure 5 shows the aligned promoter sequences. DNA sequences identical among the three species and at least five nucleotides long (with one exception) are boxed. Four of the conserved sequences (dashed boxes) contain elements identified by THUMMEL (1989) that bind the Drosophila transcription factor

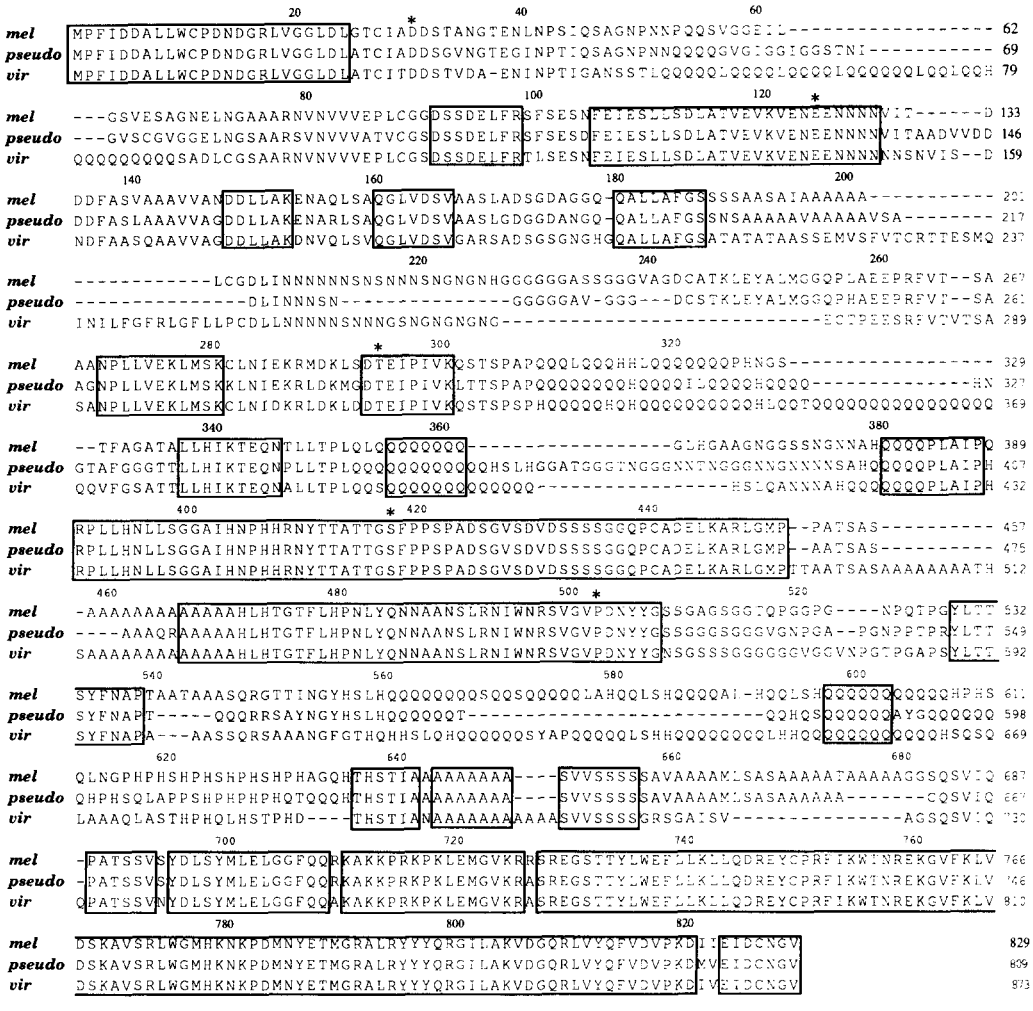

FIGURE 4.—Amino acid sequence comparisons of E74A-homologous proteins. Genomic fragments from *D. pseudoobscura* and *D. virilis* containing E74A protein-coding information were sequenced. Using the one-letter amino acid code, the E74A protein sequences from *D. melanogaster* (***mel***; from BURTIS *et al.* 1990), *D. pseudoobscura* (***pseudo***), and *D. virilis* (***vir***) are compared here. The sequences are aligned to demonstrate maximum homology. Dashes indicate gaps inserted to facilitate alignment. Amino acid sequences greater than five residues that are identical among the three species are boxed. Asterisks designate the first amino acids encoded following an intron-exon boundary. In the right margin are amino acid numbers for all three species while the *D. melanogaster* sequence has more detailed numbering above it for reference.

**TABLE 1**

**Percent acidic and basic amino acid residues for E74A protein subregions**

| | N-terminal domain | | | Central region | | | C-Terminal domain | | |
|---|---|---|---|---|---|---|---|---|---|
| Species | Residue Nos. | Percent Asp and Glu | Percent Arg and Lys | Residue Nos. | Percent Asp and Glu | Percent Arg and Lys | Residue Nos. | Percent Asp and Glu | Percent Arg and Lys |
| *D. melanogaster* | 1–190 | 16.8 | 2.6 | 191–687 | 3.4 | 3.2 | 688–829 | 11.3 | 19.7 |
| *D. pseudoobscura* | 1–202 | 15.8 | 2.9 | 203–667 | 3.4 | 4.3 | 668–809 | 11.3 | 19.0 |
| *D. virilis* | 1–215 | 13.5 | 2.8 | 216–731 | 3.5 | 3.5 | 732–873 | 11.3 | 18.3 |

Data are obtained from Figure 4. Sequence analysis of *D. melanogaster* E74A established three protein subregions (BURTIS *et al.* 1990). By aligning homologous protein sequences, the corresponding subregions were defined for *D. pseudoobscura* and *D. virilis* proteins and the percent acidic (aspartic acid and glutamic acid) and basic (arginine and lysine) amino acid residues for each subregion was calculated.

GAGA *in vitro*. Another conserved element, from −30 to −17, contains a presumptive TATA box but has been shown to be a potential repressor binding site instead (THUMMEL 1989). *In vitro* transcription experiments with mE74A suggest that the minimal sequences required for efficient transcription are from +3 to +43. Within this stretch of nucleotides is a conserved TATA sequence and a conserved 12-nucleotide sequence of unknown function. From −135 to −117 of *D. melanogaster* and −140 to −122 of *D.*

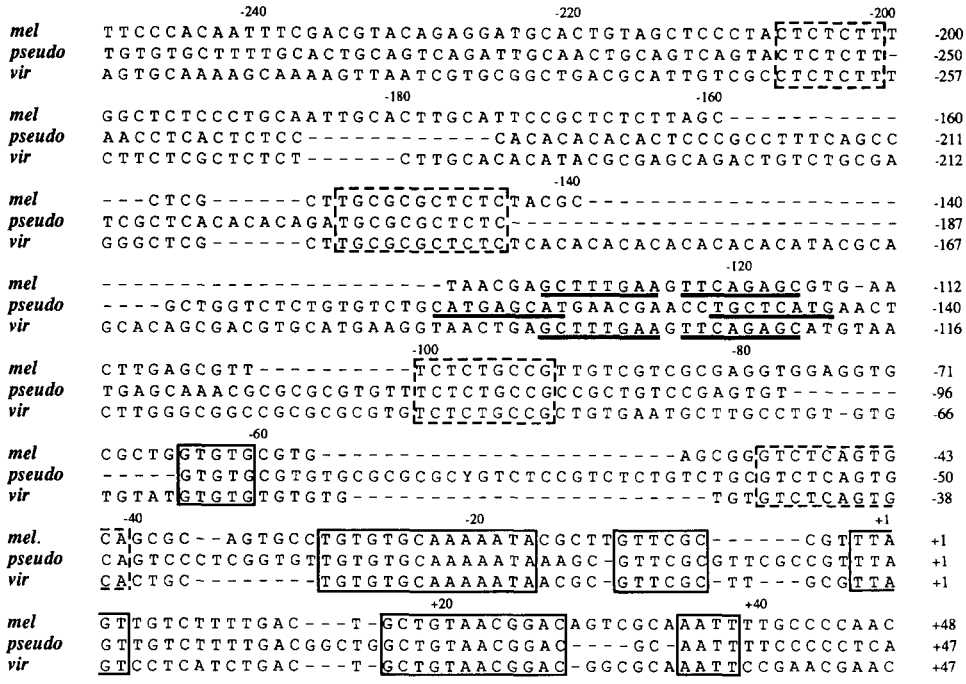

FIGURE 5.—Nucleotide sequence comparison of the E74A promoter region. The promoter regions of E74A from *D. pseudoobscura* (***pseudo***) and *D. virilis* (***vir***) were determined by nucleic acid hybridization and DNA sequence comparison to the same region in *D. melanogaster* (***mel***) (THUMMEL 1990; BURTIS *et al.* 1990). The three sequences are aligned here to maximize homology. The startpoint of transcription is indicated by +1. Sequences greater than five nucleotides that are identical among the three species are boxed. The one exception to this convention is the last box which has been proposed to contribute to the gene's TATA box (see text). Identical sequences that have been shown to bind the GAGA transcription factor (THUMMEL 1990) are enclosed by dashed boxes. Sequences with extensive dyad symmetry are underlined. The right margin indicates the nucleotide numbers for the three species while the more detailed numbering above the sequences is provided for *D. melanogaster* as a reference.

*virilis* is an identical 19-nucleotide sequence within which are 17 nucleotides with dyad symmetry (underlined in Figure 5). While this sequence is absent in *D. pseudoobscura,* a different sequence with dyad symmetry is located between −170 and −145. The two halves of the dyad are more widely separated in *D. pseudoobscura* than in the other two species (Figure 5). Whether these sequences with dyad symmetry play a role in E74A regulation is not known. Finally, careful examination of the promoter regions from the three species failed to identify any sequences similar to the ecdysone-response element defined by RIDDIHOUGH and PELHAM (1987).

## DISCUSSION

We have compared the structure and regulation of *E74* genes isolated from *D. pseudoobscura* and *D. virilis* to the previously characterized *E74* gene of *D. melanogaster* (BURTIS *et al.* 1990; THUMMEL, BURTIS and HOGNESS 1990). Despite up to 50–80 million years separating the three species, several important general *E74* characteristics have been conserved, including the fundamental feature of ecdysone inducibility in the presence of protein synthesis inhibitors (Figure 2). We have also found that the *E74* genes are located at homologous polytene chromosome sites which puff in response to ecdysone (Figure 3), supporting the assumption that *E74* transcription and chromosome puffing are coupled events.

An additional conserved *E74* characteristic is gene length (approximately 60 kb; Figure 1). This may be surprising because nearly 90% of the E74A transcription unit consists of intron sequences (introns 1 and 5 alone constitute up to 75% of the unit), which as a rule diverge in both sequence and length at a much faster rate than protein-coding sequences (BLACKMAN and MESELSON 1986; KASSIS *et al.* 1986; C. W. JONES, unpublished results). Differences in length of greater than 35% have been observed when comparing homologous introns of genes between species (see, for example, KASSIS *et al.* 1986; this is also true for some of the smaller introns of E74A, C. W. JONES, unpublished results). One explanation for *E74* gene length conservation is that the large introns possess sequences required for gene regulation and function. Indeed, the E74A intron 5 contains the 5′ exons of the E74B transcription units (BURTIS *et al.* 1990). Ecdysone-response elements and additional as yet unidentified regulatory sequences may also be associated with the E74A introns. However, it is unlikely that these necessary components would account for a majority of the E74A intron lengths. An alternative explanation is derived from experiments by THUMMEL, BURTIS and HOGNESS (1990) who showed that approximately

one hour is required to transcribe the 60-kb E74A transcription unit *in vivo*. Because other early ecdysone-inducible genes will vary in length and therefore time required for transcription, it was proposed that gene length may be important to regulate when and for how long each of the gene products is available to the cell. Simply stated, in the absence of other regulatory mechanisms, a short gene protein product will be synthesized sooner than the product of a long gene. KORNFELD *et al.* (1989) have offered the same explanation for the regulation of timing of homeotic gene products during Drosophila early development. Our results suggest that length can be an important general characteristic of genes, especially those large genes that may be part of a regulatory hierarchy.

According to the Ashburner model (see Introduction), early gene products collectively activate late gene transcription and repress early gene transcription. However, the function of individual early gene products remains to be determined. Recently, URNESS and THUMMEL (1990) have demonstrated that the E74A protein binds both to specific sites in the *E74* gene and to other early and late ecdysone-inducible puffs. Thus, E74A protein could have a dual role in both transcription activation and repression. Such bifunctional proteins have been observed in Drosophila, for example, the *Ultrabithorax* (*Ubx*) protein, which binds to its own gene to stimulate transcription as well as to the *Antennapedia* gene to repress transcription (BEACHY *et al.* 1988).

By comparing E74A protein sequences from three Drosophila species we have been able to identify those features of the protein that are likely to be important for its function. The most highly conserved part of the protein is the basic C-terminal domain with over 95% identity among the three species. Others have shown that it is the DNA-binding domain of transcription factors that is the most conserved part of the protein (HAN, LEVINE and MANLEY 1989; SCOTT and WEINER 1984). The E74A protein C-terminal domain does not resemble other proposed DNA-binding domains, *e.g.*, the homeodomain and the zinc finger motif (for review, see MITCHELL and TJIAN 1989). Recently KARIM *et al.* (1990) have proposed that the Drosophila E74A protein belongs to a larger class of sequence-specific DNA-binding proteins with a similar DNA-binding motif, referred to as the ETS domain (named for the *ets* proto-oncogene protein, another member of this class). In *D. melanogaster* E74A protein this domain is 86 amino acid residues long (residues 730–815), all of which are identical in the *D. pseudoobscura* and *D. virilis* protein sequences (Figure 4).

The high acidic amino acid content of the E74A N-terminal region makes it a logical choice to function as a transcriptional activator domain (PTASHNE 1988). Despite the relatively lower sequence conservation in this region, all three species have maintained its overall acidic character (Table 1). This observation is consistent with other experiments that have demonstrated that acidic transcriptional activator domains often lack amino acid specificity but instead rely on their general acidic nature to function (PTASHNE 1988).

The long central region of the E74A protein (residues 191–687 in *D. melanogaster*) contains a comparatively low number of basic and acidic amino acids (Table 1). While there are several long stretches of amino acid sequence identity among the three species, perhaps the most striking features of this part of the protein are the glutamine-rich and alanine-rich segments, which often contain homopolymeric repeats of the respective amino acid. These simple repeating sequences belong to the *opa* family of structural elements that have been found in other nuclear regulatory proteins such as *Antennapedia* and *Ultrabithorax* (WHARTON *et al.* 1985; SCOTT and CARROLL 1987). Homopolymeric repeats of the *opa* type may be generated as a result of slippage during DNA replication (TAUTZ, TRICK and DOVER 1986) and may be tolerated as long as they are located within a part of the protein of nonessential function, such as a hinge region joining two protein domains. Within the central region there are three significant (greater than nine residues in all three species) conserved glutamine-rich sequences located at residues 308–324, 352–361, and 560–607 of *D. melanogaster* E74A protein. While no function can be assigned to them at present, glutamine-rich sequences have been suggested to play a role in transcriptional activation (COUREY and TJIAN 1988). The E74A protein may require multiple transcriptional activator regions to successfully stimulate the transcription of a variety of late genes, each with its own molecular environment. At residues 456–480 and 641–649 of *D. melanogaster* E74A protein are the two conserved alanine-rich regions. The latter region consists of a homopolymeric repeat of at least nine consecutive alanine residues in all three species. Recently, LICHT *et al.* (1990) reported that an alanine-rich region of the Drosophila *Kruppel* gene has transcriptional repressor activity associated with it. Other Drosophila proteins with proposed transcriptional repressor activity, including *engrailed*, *even-skipped*, and *Ultrabithorax* also possess alanine-rich regions (LICHT *et al.* 1990; KORNFELD *et al.* 1989). If E74A protein does have repressor activity, as its binding to *E74A* and other early genes suggests, these alanine-rich regions should be considered when attempting to determine which part of the protein is responsible for this function. If the E74A protein performs both in transcriptional activation and repression, the *in vivo* nuclear environment where the protein binds presum-

ably would determine which of the two functions is employed.

A central question regarding the *E74* gene is how is it activated by ecdysone? Steroid hormone-inducible genes may have binding sites for hormone receptor complexes (*i.e.*, hormone responsive elements) located both upstream and downstream from the gene's transcription startpoint (see, for example, PAYVAR *et al.* 1983; HAM *et al.* 1988). RIDDIHOUGH and PELHAM (1987) have identified a 23-bp ecdysone-responsive element, a homolog of which we have been unable to find in the E74A promoter regions presented here. This negative result is consistent with *P* element transformation experiments utilizing *D. melanogaster* DNA from this region (BURTIS *et al.* 1990). Indeed, *in vitro* DNA binding experiments with ecdysone receptor protein and the *E74* gene have identified potential hormone responsive elements within intron 1 of the E74A transcription unit (BURTIS *et al.* 1990).

In addition to the hormone receptor, activation of genes by steroid hormones may require additional transcription factors that interact synergistically with the receptor (SCHULE *et al.* 1988). Footprinting analysis by THUMMEL (1989) showed that within the E74A promoter region of *D. melanogaster* are six binding sites for GAGA and two for *zeste* transcription factors. Both of these factors will activate the Drosophila *Ubx* promoter *in vitro* (BIGGIN *et al.* 1988). We have found that four of the GAGA binding sites are conserved in the other Drosophila species examined here (Figure 5). The *zeste* binding sites do not appear to be as well conserved among the three species, although there are sequences that share significant homology with the *zeste* consensus sequence (C/T G A G C/T G; BENSON and PIROTTA 1989) within the promoter segment with extensive dyad symmetry (Figure 5). Thus, we would expect that these or similar transcription factors are required for *E74* gene regulation. Other conserved sequences in the E74A promoter region suggest that additional transcription factors may also be used. These could include a repressor that acts at the traditional location for the gene's TATA box. THUMMEL (1989) has suggested that the functional E74A TATA box may instead lie downstream from the transcription startpoint. Downstream *cis*-acting positive elements have also been shown to regulate the Drosophila *Antennapedia* and *engrailed* genes (PERKINS, DAILEY and TJIAN 1988; SOELLER, POOLE and KORNBERG 1988). While the proposed downstream TATA box is not part of a sequence with extensive identity among the three species, the homology is significant to warrant further examination of the role of this element in E74A regulation.

The identification of evolutionarily conserved protein sequences and cis-regulatory DNA sequences is a powerful tool for understanding the function and regulation of a gene. Of course, this analysis allows us only to infer function from sequence. However, we now are in a position to focus on specific components of the *E74* gene which we believe are functionally important. This will lead to experimental attempts to support our hypotheses regarding the variety of roles *E74* may play during Drosophila development.

We thank RON BLACKMAN and STEVE SCHAEFFER for Drosophila genomic libraries, FOTIS KAFATOS (Harvard) for providing encouragement and facilities for a part of this work, and CARL THUMMEL (Utah) for comments on the manuscript and for sharing information. We also thank colleagues and fellow students in our department for their support during this research. This work was supported by a grant from the National Science Foundation (DCB-8502670).

## LITERATURE CITED

ASHBURNER, M., C. CHIHARA, P. MELTZER and G. RICHARDS, 1974 On the temporal control of puffing activity in polytene chromosomes. Cold Spring Harbor Symp. Quant. Biol. **38:** 655–662.

BEACHY, P. A., M. A. KRASNOW, E. R. GAVIS and D. S. HOGNESS, 1988 An *Ultrabithorax* protein binds sequences near its own and the *Antennapedia* P1 promoters. Cell **55:** 1069–1081.

BENSON, M., and V. PIROTTA, 1988 The Drosophila *zeste* protein binds cooperatively to sites in many gene regulatory regions: implications for transvection and gene regulation. EMBO J. **7:** 3907–3915.

BEVERLEY, S. M., and A. C. WILSON, 1984 Molecular evolution in Drosophila and the higher diptera. II. A time scale for fly evolution. J. Mol. Evol. **21:** 1–13.

BIGGIN, M. D., S. BICKEL, M. BENSON, V. PIROTTA and R. TJIAN, 1988 *Zeste* encodes a sequence-specific transcription factor that activates the *Ultrabithorax* promoter *in vitro*. Cell **53:** 713–722.

BLACKMAN, R. K., and M. MESELSON, 1986 Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the Drosophila hsp72 gene. J. Mol. Biol. **188:** 499–515.

BURTIS, K. C., C. S. THUMMEL, C. W. JONES, F. D. KARIM and D. S. HOGNESS, 1990 The Drosophila 74EF early puff contains *E74*, a complex ecdysone-inducible gene that encodes two *ets*-related proteins. Cell **61:** 85–99.

CHAO, A. T., and G. M. GUILD, 1986 Molecular analysis of the ecdysterone-inducible 2B5 "early" puff in *Drosophila melanogaster*. EMBO J. **5:** 143–150.

COUREY, A. J., and R. TJIAN, 1988 Analysis of Sp1 *in vivo* reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell **55:** 887–898.

DAVIS, R. W., D. BOTSTEIN and J. R. ROTH, 1980 *Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

ENGELS, W. R., C. R. PRESTON, P. THOMPSON and W. B. EGGLESTON, 1986 *In situ* hybridization to Drosophila salivary chromosomes with biotinylated DNA probes and alkaline phosphatase. Focus **8:** 6–8.

GUBENKO, I. S., and M. B. EVGEN'EV, 1984 Cytological and linkage maps of *Drosophila virilis* chromosomes. Genetica **65:** 127–139.

HAM, J., A. THOMSON, M. NEDDHAM, P. WEBB and M. PARKER, 1988 Characterization of response elements for androgens, glucocorticoids and progestins in mouse mammary tumor virus. Nucleic Acids Res. **16:** 5263–5277.

HAN, K., M. S. LEVINE and J. L. MANLEY, 1989 Synergistic

activation and repression of transcription by Drosophila homeobox proteins. Cell **56:** 573–583.

KARIM, F. D., L. D. URNESS, C. S. THUMMEL, M. J. KLEMSZ, S. R. MCKERCHER, A. CELADA, C. VAN BEVERAN, R. A. MAKI, C. V. GUNTHER, J. A. NYE and B. J. GRAVES, 1990 The ETS-domain: a new DNA binding motif that recognizes a purine-rich core DNA sequence. Genes Dev. **4:** 1451–1453.

KASSIS, J. A., S. J. POOLE, D. K. WRIGHT and P. H. O'FARRELL, 1986 Sequence conservation in the protein coding and intron regions of the engrailed transcription unit. EMBO J. **5:** 3583–3589.

KORNFELD, K., R. B. SAINT, P. A. BEACHY, P. J. HARTE, D. A. PEATTIE and D. S. HOGNESS, 1989 Structure and expression of a family of *Ultrabithorax* mRNAs generated by alternative splicing and polyadenylation in Drosophila. Genes Dev. **3:** 243–258.

LICHT, J. D., M. J. GROSSEL, J. FIGGE and U. M. HANSEN, 1990 Drosophila *Kruppel* protein is a transcriptional repressor. Nature **346:** 76–79.

MANIATIS, T., E. F. FRISCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

MARTINEZ-CRUZADO, J. C., C. SWIMMER, M. G. FENERJIAN and F. C. KAFATOS, 1988 Evolution of the autosomal chorion locus in Drosophila. I. General organization of the locus and sequence comparisons of genes *s15* and *s19* in evolutionarily distant species Genetics **119:** 663–677.

MAXAM, A. M., and W. GILBERT, 1977 A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74:** 560–564.

MCMASTER, G. K., and G. G. CARMICHAEL, 1977 Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA **74:** 4835–4838.

MESSING, J., 1983 New M13 vectors for cloning. Methods Enzymol. **101:** 20–78.

MITCHELL, P. J., and R. TJIAN, 1989 Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245:** 371–378.

PAYVAR, F., D. DEFRANCO, G. L. FIRESTONE, B. EDGAR, O. WRANGE, S. OKRET, J.-A. GUSTAFSON and K. R. YAMAMOTO, 1983 Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. Cell **35:** 381–392.

PERKINS, K. K., G. M. DAILEY and R. TJIAN, 1988 *In vitro* analysis of the *Antennapedia* P2 promoter: identification of a new Drosophila transcription factor. Cell **35:** 381–392.

PTASHNE, M., 1988 How eukaryotic transcriptional activators work. Nature **335:** 683–689.

RIDDIHOUGH, G., and H. R. B. PELHAM, 1987 An ecdysone response element in the Drosophila *hsp27* promoter. EMBO J. **6:** 3729–3734.

ROBB, J. A., 1969 Maintenance of imaginal discs of *Drosophila melanogaster* in chemically defined media. J. Cell Biol. **41:** 876–885.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHULE, R., M. MULLER, C. KALTSCHNIDT and R. RANKOWITZ, 1988 Many transcription factors interact synergistically with steroid receptors. Science **242:** 1418–1420.

SCOTT, M. P., and S. B. CARROLL, 1987 The segmentation and homeotic gene network in early Drosophila development. Cell **51:** 689–698.

SCOTT, M. P., and A. J. WEINER, 1984 Structural relationships among genes that control development: sequence homology between *Antennapedia, Ultrabithorax,* and *fushi tarazu* loci of Drosophila. Proc. Natl. Acad. Sci. USA **81:** 4115–4119.

SEGRAVES, W. A., and D. S. HOGNESS, 1990 The *E75* ecdysone-inducible gene responsible for the 75B early puff in Drosophila encodes two new members of the steroid receptor superfamily. Genes Dev. **4:** 204–219.

SOELLER, W. C., S. J. POOLE and T. KORNBERG, 1988 *In vitro* transcription of the Drosophila *engrailed* gene. Genes Dev. **2:** 68–81.

SORSA, V., 1988 Polytene chromosome maps of Drosophila, pp. 155–194 in *Chromosome Maps of Drosophila*, Vol. I. CRC Press, Boca Raton.

STOCKER, A. J., and C. D. KASTRITSIS, 1973 Developmental studies in Drosophila. VII. The influence of ecdysterone on the salivary gland puffing pattern of *D. pseudoobscura* larvae and prepupae. Differentiation **1:** 225–239.

TAUTZ, D., M. TRICK and G. A. DOVER, 1986 Cryptic simplicity in DNA is a major source of genetic variation. Nature **322:** 652–656.

THROCKMORTON, L. H., 1975 The phylogeny, ecology and geography of *Drosophila*, pp. 421–469 in *Handbook of Genetics*, Vol. 3, edited by R. C. KING. Plenum Press, New York.

THUMMEL, C. S., 1989 The Drosophila *E74* promoter contains essential sequences downstream from the start site of transcription. Genes Dev. **3:** 782–792.

THUMMEL, C. S., K. C. BURTIS and D. S. HOGNESS, 1990 Spatial and temporal patterns of *E74* transcription during Drosophila development. Cell **61:** 101–111.

URNESS, L. D., and C. S. THUMMEL, 1990 Molecular interactions within the ecdysone regulatory hierarchy: DNA binding properties of the *Drosophila* early ecdysone-inducible E74A protein. Cell **63:** 47–61.

WASYLYK, B., C. WASYLYK, P. FLORES, A. BEGUE, D. LEPRINCE and D. STEHELIN, 1990 The *c-ets* proto-oncogenes encode transcription factors that cooperate with c-*Fos* and c-*Jun* for transcriptional activation. Nature **346:** 191–193.

WHARTON, K. A., K. M. JOHANSEN, T. XU and S. ARTAVANIS-TSAKONAS, 1985 Nucleotide sequence from the neurogenic locus *Notch* implies a gene product that shares homology with proteins containing EGF-like repeats. Cell **43:** 567–581.

YANNISCH-PERRON, C., J. VIERA and J. MESSING, 1985 Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:** 103–119.

Communicating editor: P. CHERBAS