Order Number 9325030

Molecular evolution, organization and expression of the actin gene family in Tephritid fruit flies

He, Mei, Ph.D.

University of Hawaii, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

MOLECULAR EVOLUTION, ORGANIZATION AND EXPRESSION OF THE ACTIN GENE FAMILY IN TEPHRITID FRUIT FLIES

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN BIOMEDICAL SCIENCES
(Genetics)

MAY 1993

BY

Mei He

Dissertation Committee:

David Haymer, Chairman
John A. Hunt
Rebecca J. Cann
Samuel S. M. Sun
Gert DeCouet

ACKNOWLEDGEMENTS

The author wishes to express her thanks to Dr. David Haymer for his advice and help, and to the other four committee members, Dr. John A Hunt, Dr. Rebecca L. Cann, Dr. Samuel S. M. Sun and Dr Gert De Couet. This work was supported by the Genetics Department at the University of Hawaii and by USDA Grant #88-34135-3583.

ABSTRACT

The actin multigene family has been a model system for molecular studies in eukaryotes. Studies of this gene family have led to new understanding of phenomena such as gene duplication, regulation of gene expression and evolution within multigene families. This work describes molecular genetic studies of actin genes conducted in the Oriental fruit fly, *Bactrocera dorsalis* and the Mediterranean fruit fly (Medfly), *Ceratitis capitata*. Four actin genes from *B. dorsalis* and two from *C. capitata* have been isolated. These genes, which are the first to have been cloned from the Oriental fruit fly genome and among the first from the Medfly genome, have been characterized in terms of nucleotide sequence, RNA expression patterns and genomic distribution.

The nucleotide sequences and their derived protein sequences have been analyzed and compared with published data from other Diptera as well as more distantly related organisms. At the protein level, a high degree of similarity is observed among all the genes compared. At DNA level more divergence is observed. The differences between the DNA and protein sequence comparison suggests a high level of silent substitutions and selective constraints at the protein level. Among the silent substitutions, strong preferences for different synonymous codons can also be seen for the different genes. A single intervening sequence has been found in four of the six actin genes. Phylogenetic relationships have been surveyed using tree construction.

Beyond the coding regions, the 3' and 5' flanking regions of these genes exhibit virtually no detectable sequence homology. This feature has been used to construct gene specific probes to detect RNA expression patterns of individual genes in a temporal and spatial manner. The patterns indicate that five genes are encoding muscle specific actins, and one gene is encoding a cytoskeletal actin. This is consistent with other comparisons made between these actin genes and the actin genes from *D. melanogaster*. The gene

specific probes have also been used to localize individual genes on chromosomes by *in situ* hybridization.

These studies have resulted in new understanding of fundamental similarities and differences in the molecular biology of Dipteran species.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

I. ENZYMES

B: *Bgl*II

Ba: *Bam*HI

C: *Cla*I

E: *Eco*RI

H: *Hin*dIII

Hc: *Hinc*II

K: *Kpn*I

RV: *Eco*RV

P: *Pst*I

S: *Sal*I

Sa: *Sau*3A

Sp: *Sph*I

Ss: *Sst*I

T: *Taq*I

X: *Xba*I

II. CHEMICALS AND SOLUTIONS

Amp: ampicillin

$CaCl_2$: calcium chloride

cDNA: complementary DNA

CHISM: chloroform and isoamyl alcohol in the ratio of 24:1.

dATP: deoxyadenosine triphosphate

dCTP: deoxycytidine triphosphate

ddH_2O: double distilled water

dGTP: deoxyguanosine triphosphate

dH_2O: distilled water

dTTP: deoxythymidine triphosphate

dUTP: deoxyuridine triphosphate

EDTA: ethylenediamine tetraacetic acid, disodium salt

KH_2PO_4: potassium phosphate monobasic

LB: Luria-Bertani bacterial medium

LiCl: lithium chloride

$MgCl_2$: magnesium chloride

mRNA: messenger RNA

NaCl: sodium chloride

NaH_2PO_4: sodium phosphate monobasic

NaOH: sodium hydroxide

NZCYM: bacterial medium

SDS: sodium dodecyl sulfate

SM: medium for dilution and storage of bacteriophage

SSC: Sodium chloride and Citric acid buffer

TE: Tris-HCl and EDTA buffer

Tris: tris (hydroxymethyl) aminomethane

tRNA: transfer RNA

X-gal: 5-bromo-4-chloro-3-indolyl-b-D-galactoside

III. MEASURMENTS

A: angstrom

bp: base pair

kb: kilobase pair

KD: kilodalton

Ci: curie

mCi: millicurie

cpm: counts per minute

g: gram

mg: milligram

ug: microgram

ng: nanogram

hr(s): hour(s)

l: liter

ml: milliliter

ul: microliter

M: molar

mM: millimolar

O.D.: optical density

pfu: plaque forming unit

RFLP: restriction fragment length polymorphism

rpm: revolution per minute

Sod: superoxide dismutase

u: unit of enzyme activity

UTR: untranslated region

xg: relative centrifugal force (times gravity)

CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

The actin protein, found in all eukaryotes, constitutes one of the major cellular components in both the cytoskeleton and the muscle sarcomere. It is associated with many other proteins and plays multiple functions in cells. The muscle specific actin is involved in myofibrillar construction in the muscle cells (Korn, 1978; Vandekerckhove et al. 1986, Karlik et al. 1984). The cytoskeleton actin is involved in a variety of cellular events, such as cell motility, cytokinesis, chromosome segregation, organelle transportation, neurulation, exo- or endo-cytosis and cell shape change (Schroeder 1976, Mollenhaner & Morre 1976, Brawley & Quatrano 1979, Fowler & Pollard 1982, Pollard et al. 1984). Actin has also been found to be abundant in the cell nucleus (Clark & Merriam, 1977; Goldstein et al., 1977), and plays a role as an initiation factor for RNA polymerase II (Egly et al., 1984).

Actin is encoded by a multigene family in most eukaryotes (Kindle & Fintel 1978, Cleveland et al. 1980, Engel et al. 1981) with the exception of a single actin gene in some yeast and protozoan species (Ng & Abelson 1980, Gallwitz & Seidel 1980; Cupples & Pearlman 1986; Fidel et al., 1988; Dudler, 1990; Kim et al., 1992). It has been suggested that the multigene families are likely to have arisen from gene duplications followed by divergence of the duplicated copies during evolution (Ohta 1982, Maeda & Smithies 1986). The actin protein has been highly conserved during evolution compared to some other multigene families such as globins (Hightower and Meagher, 1986). However, higher divergence is seen at the DNA sequence level not only in the coding region, but also to a greater extent between the immediate 3' and 5' flanking regions of different gene family members. In addition, the position of

intervening sequences in actin genes is highly variable within as well as between species compared to other multigene families (Davidson, 1982). In many organisms, different members of the gene family have also been shown to have differential patterns of expression and perform different functions in the cell.

By comparing the similarities (or differences) in the gene sequences, especially in 5', 3' flanking and untranslated regulatory-regions as well as intron regions and positions, such a multigene family provides a good system for the study of the relationship between differences in sequences and in gene expression within as well as between species. It also provides a good system for evolutionary studies of genetic mechanisms that promote conservation versus divergence and the occurrence of multigene families.

The actin gene family has been well studied in many species such as yeast, sea urchin, *Drosophila melanogaster,* some mammals, and some higher plants. My studies reported here for the completion of PhD degree are on the "Molecular evolution, organization and expression the actin gene family from Tephritid fruit flies"

1.2 GENERAL LITERATURE REVIEW

1.2.1 Actin Gene Family

In general, actin genes exist in all eukaryotic organisms and mostly in multiple copies per genome. The number of actin genes varies over a wide range. Single copy actin genes have been found in yeast *Saccharomyces cerevisiae* (Ng & Abelson 1980, Gallwitz & Seidel 1980), in *Tetrahymena thermophila* (Cupples & Pearlman 1986), in *Asperggillus nidulans* (Fidel et al., 1988), in the oomycete *Phytophthora megasperma* (Dudler, 1990), and in *Cryptosporidium parvum* (Kim et al., 1992). In most animal, plant and protozoan species, actin genes are found in multiple copies such as two in the carrot *Daucus carota* (Stranathan et al., 1989), three in *Arabidopsis thaliana* (Nairn et al., 1988), four in the nematode *Caenorhabditis elegans* (Scott et al., 1989), five in the potato *Solanum tuberosum* (Drouin and Dover, 1987), six in *Drosophila melanogaster* (Fyrberg et al., 1980), six to eight in the soybean *Glycine max*, maize *Zea mays* (Shah et al., 1982, Shan et al., 1983), eight in the rice *Oryza sativa* (McElroy et al., 1990a), ten in the tomato *Lycopersicon esculentum* (Bernatsky and Tanksley, 1986), fifteen in the sea urchin *Strongylocentrotus purpuratus* (Davidson et al., 1982), and seventeen in *Dictyostelium discoideum* (McKeown et al., 1978). At least six genes are found in most mammal and bird species (Vanderkerckhove & Weber, 1979b). Higher copy numbers have also been found in some species. For example, over 20 actin related sequences have been detected in the human genome, the majority of which have been found to be pseudogenes (Engel et al., 1982). The genome of *Ascaris suum* consists of 50-75 members of actin gene sequences (Winrow and Sodja, 1991). In *Petunia hybrida*, the genome contains a superfamily of actin genes with as many as 200 members which can be divided into at least six highly divergent subfamilies (Baired and Meagher, 1987; McLean et al., 1988; McLean et al., 1990b). Also, amplified numbers of actin-related sequences have been reported in organisms such as the mouse (Minty et al., 1983). It is

suggested that the multigene families are likely to have arisen from gene duplications followed by divergence of the duplicated copies during evolution (Ohta, 1982; Maeda & Smithies, 1986). This leads logically to the prediction that not all copies of the actin gene will be functional, and that some copies will be a process of transformation between functional and non-functional states.

Actin gene organization in a genome is highly variable compared to some other multigene families such as globin genes. In many species, including the fruit fly (Tobin et al., 1980; Fyrberg et al., 1980), chicken (Cleveland et al., 1981), and mammals (Czosnek et al., 1983; Minty et al., 1983; Gunning et al., 1984b; Robert et al., 1985), the actin genes are dispersed throughout the genome. However, in some species such as the sea urchin (Scheller et al., 1981) and the nematode *C. elegans* (Files et al., 1983; Krause & Hirsh, 1986), actin genes are found in clusters.

Studies indicate that there is higher homology between actin genes in clusters than those scattered over the chromosomes. A good example comes from the studies on the superfamily of actin genes in *Petunia* . The six actin gene subfamilies in the *Petunia* genome are located on five different chromosomes. The members of at least three subfamilies (PAc3, PAc4 and PAc7) exist in clusters (McLean et al., 1990b). The predicted amino acid sequences are divergent between members of different subfamilies (Baird and Meagher, 1987). However, at least in the PAc4 cluster, members of this subfamily are more closely related to each other in their DNA sequences as compared to the members of other subfamilies (McLean et al., 1990b). The similarities seen among the members within a cluster may result from nonreciprocal exchanges such as unequal crossing-over and gene conversions which would promote homogenization.

One implication of close linkage is that the genes found as a cluster may be regulated together, meaning they are expressed at the same time and place in the life cycle. Distant linked members may be expressed in different regulatory modes

(Davidson et al., 1982). For example, in the chorion genes of the silk moth, several distinct clusters of genes are used at different stages of choriogenesis, and genes found in the same cluster are used at the same stage of choriogenesis (Jones & Kafatos, 1980; Iatrou et al., 1980). However, this is not always the case. For example, the genes within the β-globin cluster in the human genome are utilized at different developmental stages (Fritsch et al., 1980). In addition, examples exist such as *Drosophila*, where a cluster of muscle-related genes has been found in the 88F polytene chromosome subdivision. This includes one actin gene, two tropomyosin genes, and two unidentified myofibrillar protein coding regions (Karlik et al., 1984). In *C. elegans*, a myosin gene (myo-3) and its specific suppressor gene (sup-3) are adjacent to a three actin gene cluster (Krause & Hirsh, 1986).

The localization of actin genes can also be useful in inferring phylogenetic relationships and the evolutionary history of genes and chromosomes. Actin genes have been mapped on chromosomes by in *situ* hybridization for seven distantly related *Drosophila* species: *D. melanogaster; D. madeirensis, D. subobscura, D. hydei; D. virilis; D. gimshawi* and *D. robusta.* Six dispersed actin loci were observed in these species with similar distributions on four polytene chromosomes (Loukas & Kafatos, 1986). This reinforces the hypothesis that chromosomes have maintained their essential identity throughout *Drosophila* evolution. This notion may be true in even a broader realm. Similarity in linkage of biochemical markers has been observed among distantly related Dipteran species (Foster et al., 1981; Malacrida et al., 1984), which suggests that fundamental elements of chromosomes may have been conserved during Dipteran evolution.

1.2.2 Structure of Actin Proteins

Actin can exist in a monomeric form called globular actin (G-actin) or as polymers called microfilaments (F-actin). The monomeric actin is able to self-assemble, which results in the forming of linear polymers. It also contains binding sites for a variety of proteins and different cations. This property enables actin monomers or actin filaments to interact with, and to form a variety of three dimensional supramolecular organizations with other molecules.

G-actin consists of a single polypeptide chain of 375-377 (before modification) amino acids which has a molecular weight of about 42KD. Actin proteins are modified post-translationally by N-terminal acetylation and methylation of histidine in position 73 (Soloman and Rubenstein, 1985; Soloman and Rubenstein, 1987). The acetylation process is different between muscle and nonmuscle actins. The N-terminal methionine in both actins is removed after acetylation. However, in muscle actins the aspartic acid residue is then acetylated to complete the process, while in nonmuscle actins, the next cysteine is removed again after acetylation, followed by acetylation of the exposed aspartic or glutamic acid residue to complete the process (Soloman and Rubenstein, 1985).

The three dimensional structure of an actin:DNase I complex has been resolved recently by Kabsch et al. (1990) at 2.8 A resolution. The use of DNase I in the complex is to prevent the self-polymerization of actins. The structure without actin has already been resolved in atomic detail (Suck et al., 1984; Oefner and Suck, 1986). Figure 1.2.1 shows a schematic representation of the actin protein secondary structure (from Kabsch et al., 1990). This structure reveals two domains, and each of these domain consists of two subdomains. The small domain contains both N and C termini and consists of subdomain 1 (residues 1-32, 70-144, and 338-372) and subdomain 2 (residues 33-69). The large domain consists of subdomain 3 (residues 145-180 and 270-337) and

subdomain 4 (residues 181-269). Because the sheet topology is identical between subdomain 1 and subdomain 3, it has been suggested that the two domains of actin evolved from a common ancestral protein by gene duplication, although the amino acid sequence does not show significant similarity (Kabsch et al., 1990). The atomic structure of actin suggests common features between the actin and hexokinase proteins, and structural similarities between actin and the N-terminal ATPase fragment of the 70KD bovine heat shock cognate protein.

Each actin monomer has a defined polarity, and the monomers polymerize head to tail into F actin. The maximum diameter is between 90-95A with the large domain at a small radius and the small domain at large radius from the F-actin helix axis. All four subdomains are involved in contacts with neighboring monomers (Holmes et al., 1990). The structure of F-actin, determined by electron microscopy (Egelman, 1985), shows a helix with 13 actin monomers per 6 turns (left-handed) and a repeat of about 360 A. The orientation is determined by the complex of F-actin with the myosin subfragment 1 (S1), in which the "pointed" and "barbed" ends of the S1-decorated filaments correspond to the "+" and "-" ends of F-actin, respectively (Woodrum et al., 1975).

The polymerization process involves four steps: activation, nucleation, elongation, and possibly annealing (reviewed by Pollard, 1986; 1990) and the process at all steps is reversible. In many cells, the equilibrium for the polymerized actin and monomeric forms of actin is constant. The polymerization reaction is affected by multiple parameters such as the temperature, ion (especially Mg^{2+}) and ATP concentrations, as well as actin binding proteins. The temperature, ion and ATP concentrations in cells usually strongly favors polymerization. Therefore, the equilibrium appears to be in fact regulated by a variety of actin binding proteins which can bind selectively to the monomer or polymer of actin.

Actin contains a single high affinity site for divalent cations and several moderate and low affinity sites for mono-, di-, or trivalent cations. The high affinity site binds a single divalent cation with a K_d of 2nmol/liter for Ca^{2+} and 10nmol/liter for Mg^{2+} (Gershman et al., 1986). Other sites have moderate or weak affinity with K_ds in the micromolar or millimolar range (Zimmerle et al., 1987). The occupancy of these cation binding sites reduces the net negative charge of the monomers and promotes polymerization. Actin also contains a nucleotide binding site for ATP and ADP. The affinity of binding can be moderate to high depending on the divalent cations bound to the actin molecules at both the high and moderate affinity sites (Frieden and Patane, 1988). The nucleotide binding sites and cation binding sites may reside in the cleft between the two domains of actin (see Figure 1.2.1).

Actin also contains specific binding sites for a number of actin associated or binding proteins. Figure 1.2.2 shows the schematic representation of the actin molecule structure and regions involving in actin:actin and/or actin:actin-binding protein interactions (modified from McElroy et al., 1990b). The domains corresponding to the three dimensional structure shown in Figure 1.2.1 are indicated. The nine regions (A-J) are highly conserved regions which coincide with the actin binding sites.

It has been suggested that many actin-binding proteins mimic the actin molecule around the sites which are involved in actin-actin interaction (Tellam, 1989). For example, a DE(S/A)G tetrapeptide motif present in the C-terminus of actin (residues 363-366, in subdomain 1, see Figure 1.2.2) has been found in several actin capping/severing proteins and other proteins. Another LTDYL pentapeptide in actin (residues 185-189, in subdomain 4, see Figure 2) also has homologs in a number of proteins which bind to actin, including profilin, gelsolin, fragmin, severin, villin, DNase I, vitamin D-binding protein and cofilin. In addition, the LTDYL region in severing/capping proteins is

Figure 1.2.1: Schematic Representation of the Three Dimensional Structure of Actin Protein.

First and last amino acid residues in the helices and sheet strands are specified. (From Kabsch, 1990).

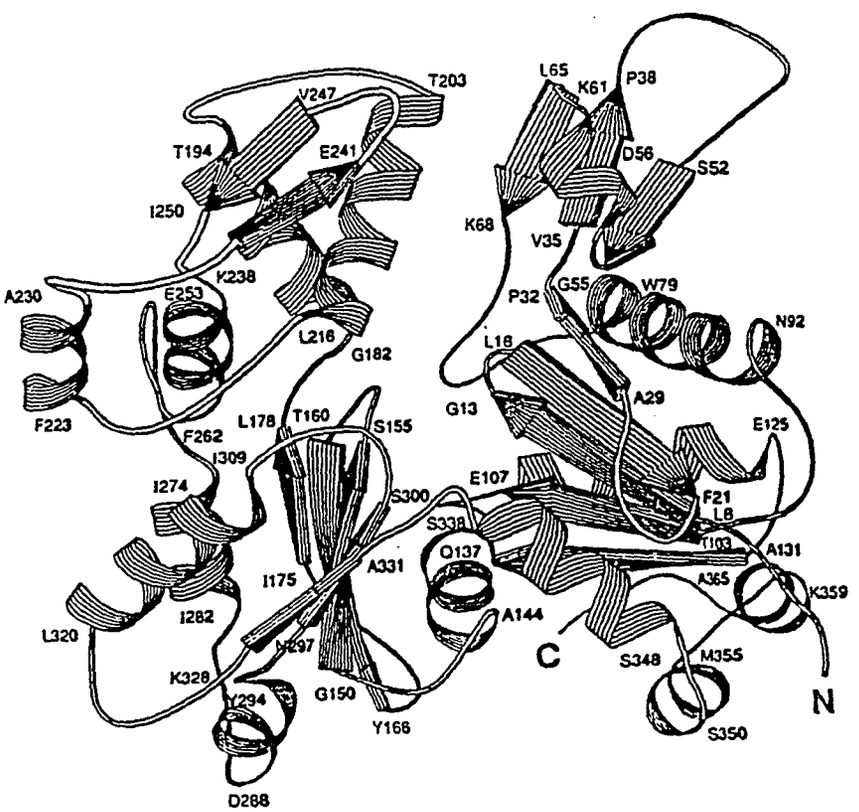
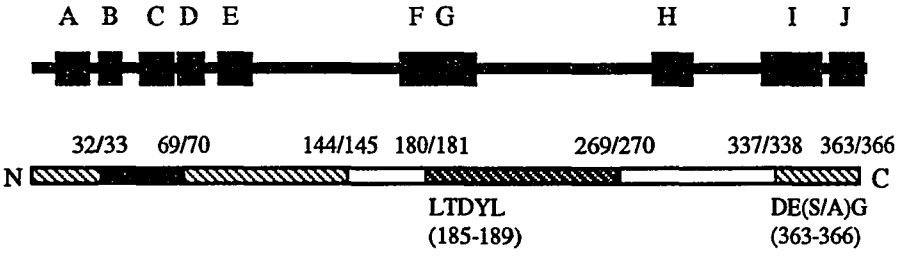

T203
V247
T194
E241
I250
K238
A230
E253
L216
G182
F223
F262
L178
T160
S155
I309
I274
S300
E107
S338
Q137
A331
I175
A144
I282
L320
N297
K328
Y294
G150
Y166
D288
L65
K61
P38
D56
S52
K68
V35
P32
G55
W79
N92
L18
G13
A29
E125
F21
L8
T103
A131
A365
K359
S348
M355
S350
C
N

Figure 1.2.2: Actin Structure and Actin Binding Sites.

The top schematic diagram is modified from McElroy (1990) indicating the position of nine conserved amino acid regions (black box) within animal and plant actin proteins. The sites of actin:actin and actin:actin-binding protein interactions are indicated below. The schematic diagram below is drawn based on the three dimentional structure of actin protein resolved by Kabsch (1990). The four subdomains indicated are explained below. LTDYL and DE(S/A)G are two sequences which have also been found in some actin-binding proteins.



A (11-22): Provilin, Vinculin, Myosin heavy chain

B (27-38): Vinculin, Proilin

C (47-60): Vinculin, Actin-actin, Profilin

D (64-76): Actin-actin, ATP exchange, Vinculin, Profilin

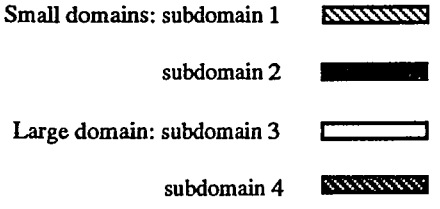
E (83-98): Actin-actin

F G (173-200): Actin-actin, Profilin, Gelsolin, Fragmin, Severin, Villin

H (280-298): Actin-actin

I (331-357): Depectin, Profilin, Gelsolin, Villin, Fragmin, Severin

J (363-377): Actin-actin, Myosin light chain



located immediately after the DE(S/A)G motif. These two regions are believed to be involved in actin-actin interactions.

1.2.3 Structure of Actin Genes

Knowledge of actin gene structure has come from the analysis of actin gene sequences. A huge collection of actin sequence data from a variety of organisms is now available. Most of the sequence data includes not only the coding region of the gene, but also the 3', 5' untranslated and flanking regions as well as the intervening sequences. The coding sequences of actin genes encode proteins of 375-377 amino acids, except the macronuclear actin gene in *Oxytricha fallax* which has only 356 amino acids due to a large deletion between position 65 and 82 (Kaine and Spear, 1982). All of the one or two amino acid deletions occur within first 10 amino acid positions at the N-terminus. Actins with 375 amino acids include the two vertebrate cytoplasmic actins (Erba et al., 1986; Ponte et al., 1984), the *Acanthamoeba* actin (Nellen and Gallwitz, 1982), the *Aspergillus* actin (Fidel et al., 1988), the *Phytophthora* actin (Dudler, 1990) and the yeast actin (Ng and Abelson, 1980). All of the *Drosophila* actins (Fyrberg et al., 1981; Sanchez, 1983), the *Bombyx mori* actins (Mounier and Prudhomme, 1986; Mounier et al., 1987), the nematode *C. elegans* actins (Krause et al., 1989), the Human smooth muscle γ-actin (Miwa et al., 1991), the two *Taenia* actins (Campos et al., 1990), the *Tetrahymena* actin (Hirono et al., 1987), the sea urchin *Strongylocentrotus purpuratus* actins (Durica et al., 1980; Scheller et al., 1981) and a few of plant actins (Meagher, 1991) have 376 amino acids. The other three vertebrate muscle actins (Hamada et al.,1982; Hanauer et al, 1983; Kamada and Kakunaga, 1989), the *Dictyostelium* actin (Vandekerckhove and Weber, 1980) and most plant actins (Meagher, 1991) have 377 amino acids.

In addition to the differences in the number of amino acids, actin gene products also occur as different isoforms or isovariants. Each of them may be encoded by one or more members of the actin gene family. For example, in mammals there are at least six actin isoforms, two are cytoplasmic actins (β- and γ-actins) and four are muscle-type actins. The muscle actins are subdivided into two striated muscle actins (α-skeletal and α-cardiac actins) and two smooth muscle actins (α-aortic type and γ-enteric type actins) (Vanderkerckhove and Weber, 1978a; 1979a, b). In *D. melanogaster*, there are at least three different isoforms (Storti et al., 1978; Fyrberg and Donady, 1979; Horovitch et al., 1979). Plant actin isovariants occur as well, and these tend to have higher variation in isoelectric points than animal (Meagher, 1991).

The position and number of intervening sequences in actin genes varies to a great extent within as well as between species, especially compared to many other multigene families. More than 26 different intron positions have been found in all eukaryotic actin sequences studied so far. Animal actin genes show greater variability in their intron positions, especially invertebrates, as apposed to plant actin genes. In the six actin genes of *D. melanogaster*, two share a common intron within codon 307, one has a intron within codon 13, and one has an intron in the 5'UTR region and two have no intron (Fyrberg et al., 1981). In *B. mori*, one of the three actin genes has a intron within codon 116 and the remaining two have no intron (Mounier and Prudhomme, 1986; Mounier et al., 1987). In vertebrates, the intron positions seems to be isoform specific. For example, in human the two cytoplasmic actin genes, two striated muscle actin genes and two smooth muscle actin genes share intron positions, respectively. Some intron positions are shared by all six actin genes, such as position 41/42 and position 327/328. All six actin genes have a intron in 5'UTR regions (Ueyama et al., 1984; Hamada et al., 1982; Zakut et al., 1982; Hu et al, 1986; Fornwald et al., 1982; Chang et al., 1985; Mohun et al., 1986; Nakajima-Iijima et al., 1985; Nudel et al., 1983; Kost et al., 1983;

Miwa et al., 1991). In contrast, intron positions in plant actin genes appear to be more highly conserved (Shah et al., 1983). Three identical intron positions (19/20, 150 and 355/356) are found in all plant actin genes studied (Shah et al., 1983; Baird and Meagher, 1987; McElroy et al., 1990a; Reece et al., 1990; Nairn et al., 1988). An intron in 5'UTR regions may also be conserved in plant actins (McElroy et al., 1990b, c; Pearson and Meagher, 1990; McKnight, 1983).

The 5', 3' untranslated region from different members of actin genes in a family vary in length. For the six *D. melanogaster* actin genes, mRNA sizes range from 1.6kb to 2.2kb in length. This is much larger than 1131bp protein coding region and it is due primarily to transcription of sequences in the 3' region, ranging in size from 300 to 700bp, which are not translated. These 3' UTRs show less homology between different gene members. This property has been used to construct gene specific or isoform specific probes (Fyrberg et al., 1983). The same gene may also produce more than one transcript during development by differential processing events. For example, it has been found that three different sized messages 2.2kb, 1.95kb and 1.7kb are complementary with the act5C specific probe during different portions of the life cycle (Fyrberg et al., 1983). Other studies indicate that the act5C gene may generate at least 9 and perhaps 15 different transcripts through alternative processing pathways, such as differential splicing, differential start points or termination points, etc (Burn et al., 1989). On the other hand, different genes may generate transcripts of the same size. The act42A, act87E and act88F genes are all found to be complementary to only a 1.7kb transcript.

1.2.4 The Multiple Functions of Actin Protein

The actin protein has been conserved during evolution, more so than hemoglobins but less than histones. The high degree of conservation seen in actin

proteins implies vital functions which are common to all eukaryotic cells. It is believed that the actins, together with associated or binding proteins, play many crucial roles in eukaryotic cells. The best understood role of actin is in muscle contraction. In mammals and birds, for example, there are four typical muscle actin isoforms found in muscle tissues. Of these, the α-skeletal actin is a dominant sarcomere form found in skeletal muscle, α-cardiac actin is found mostly in heart tissues, α-vascular actin is mostly present in vascular tissues and γ-enteric actin appear in gastrointestinal and genital tracts (Vanderkerckhove et al., 1986). They show extreme tissue specific and species-independent structural conservation (Vanderkerckhove & Weber, 1984). In *D. melanogaster* , a series of nonsense and missense mutations in the act88F actin gene have been identified based on the flight-impaired or flightless phenotype of mutant flies (Mogami and Hotta, 1981; Karlik et al., 1984, 1987; Okamoto et al., 1986; Hiromi and Hotta, 1985; Hiromi et al., 1986; Sakai et al., 1990; Drummond et al., 1991a,b; Sparrow et al., 1991). Expression of the act88F gene is restricted to the indirect flight muscle (IFMs), and it is the only actin expressed in these muscles (Fyrberg et al., 1983; Geyer and Fyrberg, 1986; Hiromi et al., 1986; Ball et al., 1987). Many of the act88F mutations cause a disruption in IFM structure (Mogami and Hotta, 1981; Hiromi and Hotta, 1985; Drummond et al., 1991b). Some of them form myofibrils that lack Z discs of the muscle sarcomere (Ball et al., 1987; Drummond et al., 1991b), and some of them have near-normal myofibrillar structure but alterations in the kinetics of the crossbridge (Drummond et al., 1990), or effects on thin filament assembly and protein-protein interactions. Others affect actin monomer stability in vitro (Drummond, et al., 1991a, b). The studies on these mutants demonstrate the role of this one specific actin gene plays in the development of indirect flight musculature.

Evidence has also indicated the role of actin in subcellular movements, such as cytokinesis, chromosome segregation and organelle transportation (Schroeder, 1976;

Mollenhaner & Morre, 1976; Brawley & Quatrano, 1979; Fowler & Pollard, 1982; Pollard et al, 1984). With the advantage of having only single actin gene in the genome (Shortle et al., 1982), yeast actin mutations have provided more direct evidence for studies of the essential roles of actin as well as for the relationships between the actin amino acid sequence structure and the function of the protein. Two temperature-sensitive lethal actin mutants have been described (Shortle et al., 1984). Both mutant actins differ from wild type by only one amino acid substitution, in which act1-1 has a change from proline to leucine at position 32, while act1-2 has a change from alanine to threonine at position 58. These two mutants show phenotypic abnormalities under the restrictive temperature, including defects in actin assembly, chitin deposition, protein secretion, vesicle accumulations, viability of the cells as well as osmotic sensitivity (Novick & Botstein, 1985). Recently Johannes and Gallwitz (1991) have reported the construction of a number of actin mutations in yeast. In this work, they made substitutions or deletions at eleven residues which are conserved from yeast to man. Five out of sixteen actin mutants had phenotypic alterations, in which the substitutions of Asp11 to either Lys or Gln resulted in dominant lethality. The C-terminal deletions of Phe375 or Cys374 and Phe375 resulted in temperature sensitivity, while the deletions of Lys373, Cys374 and Phe375 resulted in lethality. However, substitutions of Lys373 or Cys374 did not alter the phenotype.

Actin may also act as a regulator and play a crucial role in the control of cell growth and differentiation either by directly acting on the nucleus, or by indirectly acting through the microfilaments. Maness and Walsh (1982) demonstrated that when the actin structure of 3T3 fibroblasts was disrupted by cytocholasin D, the quiescent cells were prevented from entering S phase even though the serum growth factors were present. It has also been found that a reduction in actin synthesis in 3T3 cells greatly inhibits the overall synthesis of other proteins (Farmer et al., 1983). Increased expression of actin is

associated with the G_0-G_1 transition (Campisi et al., 1984; Elder et al., 1984; Greenberg & Ziff, 1984; McCairns et al., 1984). Such a pattern of gene expression has been seen among a number of proto oncogenes as well, such as c-myc (Kelly et al., 1983; Campisi et al., 1984) and c-fos (Greenberg & Ziff, 1984).

Actin found in the nucleus is also a major protein associated with the nuclear matrix (Capco et al., 1982). The nuclear matrix appears to be involved in many metabolic processes such as DNA replication, RNA transcription and processing, etc.. Nuclear actin has been implicated as a factor required for accurate transcription by RNA polymerase II. When actin antibodies or actin binding proteins are injected into amphibian germinal vesicles, transcription mediated by RNA polymerase II is dramatically inhibited. Chromosome morphology is affected as well on lampbrush chromosomes (Scheer et al., 1984). Ankenbauer and his colleagues (1989) have recently identified a Ca^{2+}-independent nuclear actin binding protein which exists in a wide range of species.

1.2.5 Actin Gene Expression and Regulation

1.2.5.1 Differential expression of actin genes

A multigene family consists of homologous genes with related functions. Different gene members of many multigene families often exhibit differential patterns of expression in temporal and spatial manners. This is the case for the actin gene family as documented for a variety of organisms including the nsea urchin (Garcia et al., 1984; Cox et al., 1986); *Drosophila* (Zulauf et al., 1984; Fyrberg et al., 1983; Sanchez et al., 1983; Courchesne-Smith and Tobbin 1989; Tobin et al., 1990), Ascidians (Beach and Jeffery, 1990), *Xenopus* (Mohun et al., 1984), the chicken (Schwartz and Rothblum, 1981), mammals (Vandekerckhove and Weber, 1981; Minty et al., 1982; Buckingham,

1985), and several plant species (Hightower and Meagher, 1985; McElroy et al., 1990b; McLean et al., 1990a).

In the six *Drosophila* actin genes, two encode cytoplasmic actins and four encode muscle specific actins. In general, the cytoplasmic actins are found abundantly expressed in early embryogenesis and at early pupation. According to Fyrberg et al. (1983), the act5C and act42A genes represent cytoplasmic actins. They appear to peak in expression during early to mid embryogenesis and early pupal development, and they are expressed in all cells and tissues of the developing embryos. They show no correlation with muscle differentiation. In addition, they have no expression in differentiated muscle rich tissues, but transcripts from these genes do appear in undifferentiated tissues such as ovary and undifferentiated cell lines where cell division is occurring. However, the transcripts of these two genes are differentially localized (Burn et al., 1989; Tobin et al., 1990).

The other four genes act57A, act87E, act79B, and act88F, are muscle specific actin genes. They are found to be in abundance corresponding with muscle differentiation, but transcripts from each gene are predominant at different stages and in different tissues as well. For instance, the transcripts of both act57A and act87E accumulate in larval intersegmental muscle, but at different levels and in different expression patterns. Abundant transcripts are detectable at larval stages for act57A gene but not for act87E gene. Transcripts of act87E are found in more abundant than that act57A gene at late stages such as pupal stage (Fyrberg et al., 1983; Tobin et al., 1990). This larval tissue later on becomes the origin of the adult abdominal muscles. These two genes encode larval-and-adult muscle actins. Expression of the act88F gene is correlated with adult musculature and only generates the myofibrils for the construction of indirect flight muscles in the thorax (Mahaffey et al., 1985; Hiromi & Hotta, 1985). The act79B gene shows a somewhat similar pattern of expression as that of act88F. Both of them

encode adult muscle actins and are expressed at about the same time during development in the thorax (Sanchez et al., 1983; Fyrberg et al., 1983). However, the expression of act79B gene expression is more broad and is most abundant in such tissues as leg muscles, direct flight muscles in thorax and the muscles which support the abdomen and head, and the male-specific muscles as well (Ball et al., 1987; Courchesne-Smith and Tobin, 1989).

In mammals and birds the α-skeletal actin gene encodes the major actin species in adult skeletal muscle actin and the α-cardiac actin gene encodes the major actin species in adult heart as well as in embryonic and fetal skeletal muscle. These two actin isoforms differ by only four amino acids, two of which are located at the N-terminus. They are found to be coexpressed in some tissues, but each of them is predominant in different tissues. For instance, in normal mice hearts, the percentages of α-cardiac mRNA versus α-skeletal mRNA are 95.8% to 4.2% while in skeletal muscles, the cardiac mRNA becomes the minor component. However, in the adult mice which possess a 9.5kb duplication of the promoter and first three exons in the 5' upstream region of the α-cardiac actin, the percentage of cardiac mRNA is decreased to 52.4% while the percentage of α-skeletal mRNA is elevated to 47.6% (Garner et al., 1989). It has been shown that the elevated expression of the skeletal muscle actin gene is due to the reduced expression of the cardiac actin gene which has a trans effect on the expression of the skeletal muscle actin gene (Einat et al., 1990).

The coexpression of different actin isoforms is a rather common situation. In addition to the striated actins (α-skeletal and α-cardiac), β- and γ-cytoplasmic actins are coexpressed in developing avian heart and limb tissues as well (Bains et al., 1984; Hayward & Schwardz, 1986). The α-smooth muscle actin is also expressed in both cardiac and skeletal muscles during the development of these two fetal tissues (Woodcock-Mitchell et al., 1988). In fact, it appears that neither muscle nor nonmuscle

cells can distinguish different actin isoforms. Both muscle actin and brain actin were able to be incorporated into all types of microfilament structures, including cortical filaments of the leading edge, stress fibers, and myofibrils (Mckenna et al., 1985).

However, there are examples of actin isoform switching as a genetic program during the early stages of smooth muscle and sarcomeric muscle development. This involves the partial replacement of nonmuscle β- and γ-actins with muscle-specific actins (Saborio et al., 1979; Garrels & Gibson, 1976, Whalen et al., 1976), as well as replacement between muscle-specific isoforms. In chicken and mouse, the initially expressed cardiac muscle actin mRNA is gradually and almost completely replaced by skeletal α-actin mRNA in the adult skeletal muscles (Hayward & Schwartz, 1986; Minty et al., 1982) while in humans, the α-cardiac actin mRNA (5% of total) still remains in the adult skeletal muscles (Gunning et al., 1983). In the case of tail regeneration of the amphibian *Pleurodeles waltlii*, the embryonic expression patterns seems to reoccur and the skeletal actin isoforms revert back to one similar to that seen in embryonic terminal differentiation. This is likely to be the α-cardiac actin isoform (Khrestchatisky et al., 1988). This differs from the regeneration event found in invertebrates, where the specific muscle actin mRNA still accumulates during the dedifferentiation phase (Fontes et al., 1983).

The replacement of actin isoforms during muscle development is accompanied by the accumulation of several muscle-specific actin-binding protein isoforms, such as tropomyosin (Hallauer et al., 1987), troponin T, C and I (Breitbart et al., 1985; Toyota & Shimada, 1983), and myosin heavy and light chains (Whalen et al., 1979). These actin associated proteins are necessary for the assembly and regulation of contractile and cytoskeletal supremolecular structures in the cell. One important question is what is the significance of actin isoform switching? It could be related to the affinities between actin-binding protein isoforms and the binding region of the actin isoforms (Strauch &

Reeser, 1989). Alternatively, the expression of the different isoforms is developmentally regulated which is coupled with the expression patterns of the actin-binding proteins. In fact, despite of the high conservation at protein level, the 3' and 5' UTRs as well as flanking regions of different isoforms are remarkably different. These regions may be responsible for the differential expression of different actin genes.

1.2.5.2 Regulation of actin gene expression

The expression of a particular member of the actin gene family is differentially regulated during development, as has been discussed. How the regulatory machinery works so that each member of the gene family is turned on or off precisely during development is a complicated but interesting question. Studies have shown that the 3' and 5' flanking and untranslated regions may carry the information necessary for regulating expression in tissue specific and temporal specific ways. Mechanisms of regulation using structures in flanking regions have already been studied extensively in some genes such as histones, lymphokines, some oncogenes and the transferin receptor. The 3' UTRs have been shown to have an important role in gene regulation (Koeller et al., 1989; Luscher et al., 1985; Shaw and Kamen, 1988). Thus, studies of actin gene regulation may help elucidate general developmental genetic principles.

In the sea urchin, for example, the 11 actin genes identified so far are used at different times and in different tissues during development. They are distinguished as three classes by their different 3' untranslated sequences. The 3'C class includes six actin genes A, B, C, D, E and J; the 3'G class includes four actin genes F, G, H and I; the K gene forms the third class by itself (Davidson et al., 1982). The 3' untranslated sequences within the same group show more homology to each other than do the intervening sequences, implying that they might have been conserved in evolution. In terms of expression patterns, a 1.8kb actin mRNA generated from the 3'G group genes is found in embryos during the blastula stage of embryogenesis. At least two of the five

3'C genes are active during oogenesis and in early embryos, producing a 2.2kb mRNA. This 2.2kb mRNA is also present in adult cells such as coelomocytes or intestines.

The different transcripts also show different rates of turnover. The 1.8kb species has shorter life than the 2.2kb transcript. These types of 3' untranslated sequences associated with particular actin mRNAs have also been reported in other organisms, such as in chick skeletal muscle α-actin mRNA (Ordahl et al., 1980), β- and γ-cytoplasmic actin of chicken brain (Cleveland et al., 1980), In *Dictyostelium* actin genes (Mckeown & Firtel, 1981) and in *Drosophila melanogaster* (Fyrberg et al., 1981; Zulauf et al., 1981).

In addition, in mammals and birds, the mRNAs of various actin genes have been found to contain domains within their 3' UTRs which are isoform specific. Fragments corresponding to cDNAs of 3' UTRs have already been identified for the α-skeletal, α-cardiac, β-, γ-cytoplasmic actins (Cleveland et al., 1980; Erba et al., 1986; Gunning et al., 1983; Ponte et al., 1983; Hsu & Frankel, 1987; Kocher & Gabbiani, 1987). These 3' UTR domains show high conservation in sequences across widely divergent species (Gunning et al., 1984a; Ordahl & Cooper, 1983; Ponte et al., 1983; Ponte et al., 1984; Shani et al., 1981; Yaffe et al., 1985; Zafar & Sodja, 1983), which suggests their possible roles in the regulation of expression of the genes (Gunning et al., 1984a; Yaffe et al., 1985; Zafar & Sodja, 1983). As an example, a 40 bp sequence in the 3' UTR of the β-actin, which is absent in the γ-actin, has been found responsible for regulating the expression of this gene during myogenesis (Deponti-Zilli et al., 1988). However, the 3' UTR for the smooth muscle isoactins seems to be more complex and species specific than that of the other actin isoforms (Kamada et al., 1989; Miwa et al., 1991). This suggests that the expression of smooth muscle actins may be controlled by mechanisms different from those operating on other actin isoforms during development (Mchugh and Lessard, 1988).

From the studies on the expression pattern of several actin genes in hybrid embryos of *sea urchin* species (hybrids of *S. purpuratus* and *L. pictus*), it has been suggested that trans-acting factors are necessary to regulate the expression of these genes in a temporal and spatial manner. The factors are present in the hybrid embryos (Bullock et al., 1988; Nisson et al., 1989), and suggest that given the evolutionary distance of these species from each other, conservation of trans-regulatory sequences has coevolved with the regions they regulate.

The 5' untranslated regions and upstream promoter regions may also contain the element(s) responsible for the differential patterns of expression in actin genes. In *D. melanogaster*, a DNA fragment extending approximately 2kb upstream of the ATG codon has been shown to contain all the sequences necessary for correct expression of the 88F actin gene. This includes an enhancer type element (Hiromi et al., 1986). In rice, the 5' region of the Act1 actin gene has been characterized (McElroy et al., 1990c). The 5' upstream region up to 1.3 kb from the translation initiation codon appears to contain all of the regulatory elements required for high level of constitutive transcription of the gene under this promoter. This region includes the 5' leader intron. It has found that this intron is essential for the efficient expression which seems to be associated with efficient mRNA splicing.

Another example comes from the studies on the differential patterns of expression of the two striated muscle specific actin genes. A report by Carroll (1986) indicates that the 5' transcription regulatory sequences surrounding the CCAAT box of cardiac, skeletal, vascular α-actin genes in chicken appear to be conserved. This sequence, named as CArG element or box which has the sequence of CC(A+T-rich)$_6$GG, has also been found in the α-cardiac actin genes of mouse, frog, human (Minty et al., 1986; Gustafson et al., 1988; Mohun et al., 1989), in the α-skeletal actin genes of rat, as well as in a number of other muscle-specific genes (Minty et al., 1986a).

By analyzing a variety of mutations consisting of internal deletions and linker scanning mutations, four CArG box motifs from -80 to -220 upstream were located in the *Xenopus* α-cardiac actin gene promoter.

Only the most proximal copy of the CArG box is essential for muscle-specific expression of the gene (Mohun et al., 1989). In humans, the CArG box is repeated in the upstream regulatory regions of the α-cardiac and α-skeletal actin genes. Again, the proximal CArG box between -113 to -103 upstream of transcriptional start site appears to be essential for the gene specific expression. Another copy of the CArG box, one located at around -150bp, causes 40% to 50% loss of activation when deleted.

A specific nuclear factor appears to be able to interact with the CArG motif in vitro. This factor is present not only in muscle cells but also in nonmuscle cells. It can also interact with the human c-fos serum responsive element, which contains a CArG box as well.

Recent studies on the muscle-specific expression of the cardiac α-actin gene by Sartorelli and colleagues (1990) provided a clearer picture on gene regulation in multigene family and the interaction of DNA-protein complexes in gene activation. The promoter region of human cardiac α-actin contains numerous cis acting elements upstream of the TATA box and is expressed exclusively in myogenic cells (Gustafson and Kedes, 1989; Minty and Kedes, 1986). At least three transcription factors corresponding to three different cis elements are required for the muscle specific expression of this gene. One is the SRF (serum response factor) related protein which recognizes the CArG box [$CC(AT)_6GG$] located at -100bp from the transcription start site. A second factor is the Sp1 protein (zinc finger protein) which selectively binds to GC box cis element (GGGCGG) located at -75bp. These two proteins are ubiquitous transcription factors and are expressed in all cell types (Boxer, 1989a, b). The third factor involves one of the myogenic determinants MyoD1 (a helix-loop-helix DNA

binding protein) which binds to the CANNTG consensus sequence located at -50bp. The MyoD1 protein is a specific regulator in muscle cells. All three DNA-protein complexes are required for the muscle specific expression of the cardiac α-actin gene.

1.2.6 The Molecular Evolution of Actin Gene Family

1.2.6.1 Sequence diversity of actin genes

Duplication with subsequent divergence appears to be a common mechanism for acquiring new members in a multigene family. The divergence may occur in different regions of the gene (coding region, regulatory region etc.) depending on what the functions of the gene family are. Members of the gene family, such as the rRNA and histone gene families may remain uniform because large amounts of specific gene products are needed in the organism. For other gene families, differentiated functions may promote aquisition of new adaptive traits in the organism. Divergence may occur in the regulatory regions of different gene members (e. g. lactate dehydrogenase), or possibly in both regulatory and coding regions. Such cases have been seen in many multigene families such as hemoglobin and chorion genes (Markert, 1987; Perutz, 1983; Ohta, 1991). In some cases, genes become less similar over time and only a portion of the gene shows homology. These historically related but functionally distinct genes constitute supergene families. The homeotic genes which all contain the homeodomain in their gene product belong to such a supergene family (Hartl and Clark, 1989).

Members of actin gene family, like globin genes, have diverged in both regulatory regions as well as coding regions within a species. However, the divergence among the members of actin gene family has its own characteristics. Actin proteins are often highly conserved over a wide spectrum of species during evolution. For example, there is 95% amino-acid homology between two muscle specific actin genes Act79B and Act88F in *Drosophila melanogaster.* Each of them shows 91% to 92% amino-acid

homology with rabbit skeletal muscle actin. Most variants occur in the first ten amino acids at the amino-terminal of the proteins (Sanchez et al., 1983). Within an isoform family in mammals and birds (e.g., cytoplasmic, striated and smooth muscle actins), 99% amino acid homology is typically observed (Vandekerckhove and Weber, 1978b; Cross et al., 1988). The comparison between actins from yeast and mammals show 89% amino acid identity (Gallwitz & Sures, 1980; Ng & Abelson, 1980). Plant actins, however, show a higher degree of divergence from each other as compared to animal actins (Meagher, 1991). Up to 20% - 30% amino acid differences are observed for some actins from Protist or Fungi relative to all other organisms, such as *Oxytricha*, *Tetrahymena* and *Phytophthora* (Meagher, 1991; Kaine et al., 1982; Hirono, 1987; Dudler, 1990).

It has been estimated that most actin sequences have evolved at an approximately linear rate of 1% nonsynonymous substitutions per 110(±50) million years (Hightower and Meagher, 1986). However, different mutation rates are observed in different lineages when rDNA sequences are used as comparative measurement. Actins in animals and some of amoebae show much lower mutation rates than those in higher green plants, higher fungi and oomycetes (Bhattacharya et al., 1991).

Although overall protein sequence does appear to be conserved, there are many synonymous substitutions, especially at third codon positions in the coding region, producing divergence among different members of actin gene families (Minty et al., 1983). In addition, there is little homology in 3' and 5' UTRs between the six actin genes in *D. melanogaster* (Fryberg, 1981; Zulauf, 1981). As has been discussed, the highly variable 5', 3' untranslated and flanking regions may be responsible for the differential expression of each gene member.

Intron sequences and positions are also highly variable in actin gene families. The positions of intron sequences in many multigene families are highly conserved. In

the case of the globin genes (Efstradiatis et al., 1980), the ovalbulin genes and their homologues (Royal et al., 1979; Heilig et al., 1980), the vitellogenin genes (Wahli et al., 1980), and the immunoglobulin genes (Ellison et al., 1981; Yamawaki-Kataoka et al., 1981), all members of the gene family within a species contain the intron sequences at exactly the same positions. In globin genes, the sizes of the intervening sequences are also very similar. In contrast, the positions of intron sequences of actin genes vary greatly among different members within as well as between species, as has been discussed in an earlier section.

The reason why there is high variability of intron positions in the actin gene family is not quite clear. It has been hypothesized that domains in a gene, divided by introns might long ago have been separately located in the genome until an event (e. g. recombination) brought them together into one gene, so called "exon shuffling" (Gilbert, 1978). According to this idea each exon corresponds to a functional unit. For instance, in globin genes, the two introns subdivided the gene into exons encoding distinct functional regions of the protein. The central exon encodes the entire heme-binding region, whereas the two terminal exons encode non-heme-binding regions.

However, such "intron-domain" relations are not obvious in the case of actin genes, where over 26 different intron positions have been found in a variety of organisms (Davidson, 1982; Mounier and Prudhomme, 1986; Dibb and Newman, 1989; Bhattacharya et al., 1991). Dibb and Newman (1989) proposed a model based on the distributions of a total of 60 intron positions in tubulin genes and actin genes. They concluded that the introns found in these two multigene families were gained rather than lost during eukaryotic evolution in a nonrandom fashion between G and R of the proto-splice site (C/AAGR), which is a consensus sequence flanking most introns. Thus, it is possible that intron position is also related to regulatory charges in actin expression.

1.2.6.2 Phylogeny of the actin gene family

In vertebrates, two main types of actins have been identified based on the amino acid sequences and the tissuler localization. The four muscle specific actin isoforms including two striated muscle actins (α-skeletal and α-cardiac) and two smooth muscle actins (α-vascular or aortic and γ-enteric) are distinctly different from the two cytoplasmic actin isoforms (β-actin and γ-actin) (Vandekerckhove and Weber, 1978b; Buckingham and Minty, 1983). In invertebrates, the distinction between muscle specific actins and cytoplasmic actins has also been found (Mounier et al., 1992). However, all invertebrate actins including both muscle and cytoplasmic actins are found to be relatively more similar to the two vertebrate cytoplasmic actin isoforms (β-, γ-actins) than to vertebrate muscle actins, based on the comparison of amino acid sequences. In addition, the invertebrate cytoplasmic actins are more closely related to vertebrate cytoplasmic actins as opposed to invertebrate muscle actins (Vandekerckhove & Weber, 1984; Schuler et al., 1983; Crain et al., 1987; Krause et al., 1989; DesGroseillers et al., 1990; Fyrberg et al., 1981; Bhattacharya et al., 1991; Mounier et al., 1992).

In other words, the cytoplasmic actins of all metazoan organisms and the muscle actins of invertebrates could be descended directly from the cytoplasmic actins found in protozoa and fungi, while the four muscle actin isoforms found in vertebrates may have diverged and subdivided more than twice to yield a distinguishable family. The invertebrate muscle actins and vertebrate muscle actins may be an example of convergent evolution. They perform similar functions but with different solutions in their molecular sequences.

Plant actins are well separated from actins of animal, fungal and protist (Meagher, 1991). It has been suggested that the actin genes within higher green plants might have undergone gene duplication before the separation of monocots and dicots

(Drouin and Dover, 1990). This is also consistent with the fact that actin genes of all the higher plants studied so far have the same intron structure.

However, in the animal lineage, the comparison based on the intron positions suggest a contrary conclusion to that from the comparison of amino acid sequences. In most vertebrates, some intron positions are shared by both cytoplasmic and muscle specific actin genes, such as positions 41/42 and 327/328. In addition, all six actin genes have a intron in the 5'UTR region (Ueyama et al., 1984; Hamada et al., 1982; Zakut et al., 1982; Hu et al, 1986; Fornwald et al., 1982; Chang et al., 1985; Mohun et al., 1986; Nakajima-Iijima et al., Nudel et al., 1982; Kost et al., 1983; Miwa et al., 1991). Most invertebrate actin genes have intron-exon organizations. Base on intron positions, the vertebrate cytoplasmic actins show a closer relationship to vertebrate muscle actins rather than to the primitive actins in the muscle and nonmuscle cells of invertebrates (Buckingham et al., 1985). This possibly suggests different selection pressures on the protein sequence and the positioning of introns (Pollard and Cooper, 1986).

1.2.6.3 Concerted evolution in a gene family

A concept of "concerted evolution" or "molecular drive" has been used to described the phenomenon often observed in multigene families, in which a higher degree of sequence homogeneity is found between members of a gene family within a species than between homologous genes of different species (Hood et al., 1975; Zimmer et al., 1980; Dover, 1982). Several mechanisms have been described to account for concerted evolution in the multigene families, such as unequal crossing over, gene conversion, transposition, slippage-replication, and RNA-mediated exchanges (Hood et al., 1975; Dover, 1982; Dover and Tautz, 1986). Evidence for or against this idea is largely lacking, although a gene conversion event between a muscle actin gene and a cytoskeletal actin gene in the sea urchin has been reported (Crain et al., 1987).

However, despite extensive studies using coding region sequences, no evidence of nonreciprocal exchanges was found between the potato actin gene family (Drouin and Dover, 1990). The actin genes in this species appear to evolve independently from each other.

Concerted evolution could have profound effects on the construction of phylogenetic trees using gene families because the similarity or homology seen may not be due solely to shared ancestry in the pedigrees (Dover, 1987). Genes from the same species will tend to be grouped together if there is concerted evolution within species.

1.3 STUDIES OF THE MEDITERRANEAN FRUIT FLY AND THE ORIENTAL FRUIT FLY

1.3.1 Background and Rationale

The Mediterranean fruit fly (medfly) *Ceratitis. capitata* and the Oriental fruit fly *Bactrocera dorsalis* are higher dipteran species which are ideal organisms for general studies as well as for comparative molecular studies with *D. melanogaster*, one of the genetically best studied organisms. Both *C. capitata* and *B. dorsalis* belong to the family *Tephritidae*. The *Tephritidae* appears to be a younger family than the *Drosophilidae,* which contains *D. melanogaster* (Oldroyd, 1964). These two families are distant relatives, all belonging to the section of Acalyptrate in the order Diptera. Beverly and Wilson (1984) placed the time of divergence between these two families at 120 million years. However, recent studies on the Cu, Zn Sod (superoxide dismutase) gene by Kwiatowski et al. (1992) indicated that the divergence of these two families is dated about 80 million years, based on the assumption that the rate of Sod gene evolution in Dipterans is the same as in mammals, being 30 amino acid relacements per 100 residues per 100 million. The sequencing of the actin genes from these two Tephritid species plus the available sequence data from *D. melanogaster* may provide more information on the phylogenetic relationships between these species.

In addition to the evolutionary interest in these species, both of these Tephritid species are insects of economic importance amenable to biological control. The Mediterranean fruit fly, *C. capitata* is an important agricultural pest of over 250 species of fruits, nuts and vegetables in temperate, subtropical and tropical regions of the world, such as southern Africa, the Mediterranean region, and south-western Australia. It was first found in Hawaii in 1910 and soon became a very serious pest (Prokopy & Roitberg, 1984). The absence of these flies in the mainland of the United States causes restrictions on fruit as well as other plant exportation from Hawaii in order to prevent

their spreading. At least eight occasions of invasion by medflies on the mainland USA have been recorded (Prokopy & Roitberg, 1984). Because of this, extra expense in treatment has to be performed to guard against infestation before exportation. *B. dorsalis* is also a major pest species. Like many other pest species in this genus it is polyphagous, has a large distributional area, wide climatic tolerance, high reproductive potential and high mobility. *B. dorsalis* infests a number of fruits including apricot, plum, mango, peach, pear, and guava. *B. dorsalis* was first discovered in Hawaii in 1945 (Van Zwaluwenburg, 1947), some 35 years later than the invasion of *C. capitata*. It soon replaced *C. capitata* in the more tropic lowland areas, but decreased in abundance in the more temperate upland areas (Bess, 1953).

Considerable effort has been made in order to control the pest population so as to reduce the economic damage caused by these flies. One of the methods used to control the pest insects is the sterile insect technique (SIT), a form of biological control. This technique has been shown to be very effective against certain pest species. Although SIT can be an effective tool for biological control of pest species, in the case of *C. capitata* and *B. dorsalis* the technique must be made more efficient before it can be applied on a large scale. In particular the released females, although sterile, inflict extensive damage to fruit. The ability to produce a "male only" strain through genetic engineering may provide the means to make SIT more efficient and acceptable. Knowledge at the molecular level about these insect species will therefore be useful for the development of genetic engineering techniques.

In addition to general knowledge, the study of actin genes in these two species, especially the 5' flanking promoter region, may provide an efficient promoter for transgenic flies, which can be used in biological control. Recently McElroy and his colleagues have reported the use of rice actin 1 gene (Act1) 5' region as an efficient promoter for regulating the constitutive expression of a foreign gene in transgenic rice

(McElroy et. al. 1990c). Thus, biotechnology may benefit from a broder understanding of actin gene regulation in these species.

What we know about these species is very limited. In *C. capitata,* the polytene chromosomes have been well characterized (Zacharopoulou, 1987). There are five pairs of autosomes and one pair of sex chromosomes in this species (Radu et al., 1975). A few biochemical and morphological markers have been identified, and some linkage relationships have been established (Gasperi et al., 1986). Cloning of two chorion protein genes, one actin gene, four vitellogenin genes, and a Cu, Zn Sod gene was reported recently (Konsolaki et al. ,1990; Tolias et al., 1990; Haymer et al., 1990; Kwiatowski et. al., 1992). However in *B. dorsalis*, virtually nothing has been done in terms of contemporary molecular genetic studies except the cloning of two putative actin genes (He and Haymer, 1991). In addition, knowledge of the cytogenetics of this species is lacking. Studies of actin genes in *C. capitata* and *B. dorsalis* can not only provide the general knowledge about these two species, but also provide more information and comparison for the studies of actin genes, especially within the order of Diptera. Furthermore, the localization of the actin genes on chromosomes by means of in situ hybridization may provide chromosome markers which can be used as a basis for developing linkage or physical maps, especially of the *B. dorsalis* genome.

Although actin genes have been studied in detail in many organisms, including all the six actin genes in *D. melanogaster*, there is a lack of information about any other Dipteran species. In addition, even with these studies, there are still questions to be answered. For example, it is clear that actin genes are highly conserved in coding region sequences, however, they also display a high degree of diversity in regions immediately flanking the coding region, as well as in terms of intervening sequences and positions. What are the mechanisms which promote the conservation at the protein level but allow other regions to change? If selective constraints have played roles in terms of protein

sequences, what is the relationship between protein sequence and functions for different members of the gene family? Finally, for multigene families such as actin, it remains to be resolved how phylogenetic relationships can best be defined.

The fundamental question addressed in this study is: What can studies of the actin multigene family in related species , such as these two Tephritid species, reveal about the molecular biology of those species and about evolution in multigene families? First of all, the studies of actin gene in these two Tephritid species should help to fill the information gap for the order Diptera. Second, with the knowledge from *D. melanogaster*, it should be possible to nake detailed comparisons with two species from the same family and one from a distantly related family. This may provide information on the phylogenetic relationships between these species, the evolutionary origins of gene families and, the relationship between gene duplication and divergence within a multigene family. Last but not least, the studies should be beneficial in terms of providing basic molecular genetic knowledge of these two species.

1.3.2 Previous Work on Actin Genes in These Flies

The previous work on actin genes in *C. capitata* and *B. dorsalis* included the isolation and characterization of actin homologous subclones from *C. capitata* (pmed21, pmedC1) (Haymer et al., 1990) and from *B. dorsalis* (pDdA1 and pDdA5, the name was given based on the previous species name *Dacus dorsalis*) (He and Haymer, 1991). The pmed21 was originally isolated by screening a Charon 4 genomic library first with an actin gene probe from *sea urchin* and then with the 1.9kb DmA2 probe from the *Drosophila melanogaster* 5C actin gene. The pmedC1 was isolated from a λgt10 female late pupa cDNA library using the pmed21 as a probe. The two *Bactrocera* subclones were isolated from a EMBL3 genomic DNA library using the DmA2 as a probe. Cross hybridization has been carried out between pDdA1, pDdA5, pmed21, pmedC1, and the

DmA2. All of them show cross hybridization except pmedC1 which shows no homology with any of the subclones other than pmed21. The pmedC1 most likely represents the 3' untranslated region of the gene in pmed21

The genomic representations of actin genes in both species have also been surveyed. Both of them contain multiple copies of actin homologous sequences in their genome. Up to six homologous fragments are detectable in *Eco*RI digests of *C. capitata* genomic DNA while seven homologous fragments are detectable in *Eco*RI digests of *B. dorsalis* genomic DNA. The pattern of expression of actin gene(s) in *C. capitata* has been examined in both temporal and spatial manner by means of Northern blot using pmed21 as a probe. The result indicates that the gene in pmed21 encodes most likely a muscle specific actin. Due to the fact that pmed21 contains the coding region of the actin gene and we know that the actin genes are highly conserved in the coding sequence, the transcripts detected are not necessarily only from the gene in pmed21.

CHAPTER 2
MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Fly Strains

Ceratitis, capitata: a laboratory strain which has been in continuous culture for about 30 years. Originally collected in Kula, Maui.

Bactrocera, dorsalis: a laboratory strain in culture for several years and was originally collected in Puna, Hawaii.

2.1.2 Actin Homologous Subclones

DmA2: 1.9kb *Hin*dIII fragment subcloned in pBR322, it represents the complete coding region of *D. melanogaster* act5C actin gene.

pmed21: 4.7kb *Eco*RI fragment isolated from *C. capitata*, containing the complete coding region of the CcA1 actin gene. Subcloned in pUC9.

pmed5: 2.0kb *Eco*RI fragment isolated from *C. capitata*, containing part of the coding region of the CcA2 actin gene. Subcloned in pUC9.

pDdA1: 1.8kb *Eco*RI fragment isolated from *B. dorsalis* containing part of the coding region of the DdA1 actin gene. Subcloned in pUC9, pUC19 and pUC18.

pDdA1.2: 620bp *Eco*RI fragment isolated from *B. dorsalis*, containing part of the coding region of the DdA1 actin gene. Subcloned in pUC18.

pDdA2: 4.0kb *Eco*RI fragment isolated from *B. dorsalis*, containing most of the coding region of the DdA2 actin gene. Subcloned in pUC9, pUC19 and pUC18.

pDdA2.1: 6kb *Eco*RI fragment isolated from *B. dorsalis*, containing the 5' exon of the DdA2 actin gene. Subcloned in pUC19 vector.

pDdA3: 4.1kb *Eco*RI fragment isolated from *B. dorsalis*, containing the complete coding region of the DdA3 actin gene. Subcloned in pUC9, pUC19 and pUC18.

pDdA5: 5.6kb *Eco*RI fragment isolated from *B. dorsalis*, containing the complete coding region of the DdA5 actin gene. Subcloned in pUC9, pUC19 and pUC18.

2.1.3 Gene Specific Probes

pmedC1: Approximate 500bp cDNA subclone, isolated from a *C. capitata* female late pupal cDNA library using pmed21 as probe. It represents the 3' untranslated region of the CcA1 actin gene.

pmed5(5'): Approximate 1.3kb fragment containing the 5' flanking region of the CcA2 gene, 150bp before the start codon.

pDdA1(3'): Approximate 310bp fragment containing the 3' region of the DdA1 gene, 80bp after the stop codon.

pDdA2(3'): Approximate 570bp fragment containing the 3' region of the DdA2 gene, 100bp after the stop codon.

pDdA3(3'): Approximate 530bp fragment containing the 3' region of the DdA3 gene, 60bp after the stop codon.

pDdA5(3')S: 248bp fragment containing the 3' region of the DdA5 gene, 8bp after the stop codon.

pDdA5(3')L: Approximate 740bp fragment containing the 3' region of the DdA5 gene, 8bp after the stop codon.

2.1.4 Inverse PCR Products

pDdTaqIPCR: Inverse PCR product generated from *B. dorsalis Taq*I digested genomic DNA sample.

pDdHincIIPCR: Inverse PCR product generated from *B. dorsalis Hinc*II digested geneomic DNA sample.

pDdHincIIPCR(R1-BglII): A subclone isolated from pBdHincIIPCR, containing the 5' region outside of pDdA2 and about 50bp coding region of the DdA2 gene.

pDdSau3APCR: Inverse PCR product generated from *B. dorsalis Sau*3A digested genomic DNA sample.

2.1.5 Reagents and Systems

Enzymes: all enzymes were purchased from either Gibco/BRL or Boehringer Mannheim Biochemicals.

Chemical reagents: purchased from variety of resources. In all cases, molecular biology grade reagents were used if available. For electrophoresis, electrophoresis grade agarose was used.

System products:

1. Random primer DNA labeling kit: purchased from Boehringer Mannheim.
2. Genius^{TM} nonradioactive DNA-labeling and detection kit: purchased from Boehringer Mannheim.
3. EMBL lambda cloning vectors: purchased from Promega Biotec.
4. Protoclone^{TM} lamda gt10 system: purchased from Promega Biotec.
5. Packagene lamda DNA packaging system: purchased from Promega Biotec.
6. Subcloning efficiency^{TM} *E. coli* DH5α competent cells: purchased from BRL.

7. Geneclean^{TM} kit: purchased from Bio101 Inc.
8. ExoIII/mung bean nuclease deletion kit: purchased from Stratagene, Inc..
9. Sequenase version 2.0 kits: purchased from US Biochemical, Inc..
10. GeneAmp PCR reagent kit: purchased from Perkin Elmer Cetus.
11. Nick translation kit: purchased from Boehringer Mannheim.

2.2 METHODS

2.2.1 Extraction of High Molecular Weight DNA

Genomic DNA was isolated according by the following two methods.

Blin and Stafford method:

Genomic DNA from both *C. capitata* and *B. dorsalis* were isolated following the procedure described by Jowett (1986), a modification of the method of Blin and Stafford (1976). All solutions were autoclaved if possible. Glassware was baked at 250°C for four hours and covered with aluminium foil until use. Plastic items were sterilized by autoclaving.

Procedure:

1. 1 gram adult flies from -70°C stock ground to a powder in a prechilled mortar and pestle are resuspended in 15ml lysis buffer (100mM Tris-HCl pH 8.0, 50mM NaCl, 1% SDS, 0.15mM spermine, 0.5mM spermidine), and transferred to a 30 ml corex tube.

2. Add Proteinase K to a final concentration of 100 ug per ml, incubate for 2 hours at 37oC.

3. Extract two times with phenol, one time with phenol/CHISM (chloroform : isoamyl alcohol = 24:1 in volume), one time with CHISM only.

Phenol extraction: Add 1 volume phenol, which has been saturated with TE (10mM Tris-HCl pH8.0, 1mM EDTA), to the sample, invert gently and spin at low speed for 5 minutes. Save the top aqueous layer into a fresh tube. Repeat one more time.

Phenol/ CHISM extraction: Add 1/2 volume Phenol and 1/2 volume CHISM to the aqueous phase, mix well and spin at low speed for 5 minutes. Transfer the top phase to a fresh tube.

CHISM extraction: Add 1 volume CHISM to the aqueous phase, mix gently and spin at low speed for 5 minutes. save the top phase.

4. Transfer the clear top aqueous into dialysis tubing (which should be rinsed with dH_2O and handled with gloves) and dialyse against buffer 1 (50mM Tris-HCl pH 8.0, 10mM EDTA, 10mM NaCl) for at least two days at 4°C with gentle stirring.

5. Add DNase-free RNaseA to a final concentration of 100ug /ml and incubate at 37°C for one hour.

6. Extract one time with phenol, one time with phenol/CHISM, and one time with CHISM only.

7. Transfer the top aqueous phase from the final phenol extraction into dialysis tubing and dialyse again against buffer 2 (TE, pH8.0) for at least two days at 4°C with gentle stirring.

8. Measure the density and determine the quality of the DNA sample according to O.D. (optical density) value of the UV absorption at 260nm and 280nm. The density is calculated by the formula:

Density (ug/ul) = $O.D._{260nm}$ X dilution factor X 50 (ug/ml)

The $O.D._{260nm}$ / $O.D._{280nm}$ should be around 1.8.

9. Check the DNA sample on 0.7% agarose gel (described in 2-2-6). Ideally, the DNA should appear as a single high molecular weight band without smearing

10. Precipitate DNA with 0.1 volume of 3M Sodium acetate and 2 volumes of absolute ethanol, keep at -20°C for at least 2 hours. Spin at 10,000rpm (Sorvall, type SS-34 rotor) at 4°C for 15 minutes, wash the pellet with 70% ethanol and spin for 5 minutes at same speed. Dry the pellet in vacuum and resuspend in small volume of TE.

The DNA sample can be stored in TE after second dialysis at 4°C or after the ethanol precipitation stage at -20°C.

Lifton preparation:

1. Grind 0.3g flies (fresh or from -70°C) in 5ml of grind buffer (0.2M Sucrose, 50mM EDTA, 100mM Tris HCl pH9.0, 0.5% SDS).

2. Prepare a strainer by inserting a cotton ball-sized piece of sterile polyester fiberfil into a 10cc syringe. Strain the homogenate through the fiberfil into a 15ml polypropylene tube, set on ice.

3. Add 50ul of 20ml/ml proteinase K to a final concentration of 0.2mg/ml. Incubate at 65°C for one hour.

4. Add 750ul of 8M potassium acetate to the sample and incubate on ice for at least one hour.

5. Centrifuge for 15 minutes at 10,000rpm at 4°C (Sorvall, type SS-34 rotor), save the supernatant in a new tube.

6. Add 2vol. of room temperature 95% ethanol, mix well. Centrifuge again at the same condition as step 5, but save the pellet.

7. Resuspend pellet in 500ul of TE, pH8.0.

8. Add 2.5ul of 10mg/ml RNase A for a final concentration of 50ug/ml. Incubate at room temperature for 15-30 minutes. Transfer the aliquot into an eppendorf tube.

9. Extract one time with phenol, one time with phenol/CHISM, and one time with CHISM only.

10 Precipitate DNA with 0.1 volume of 3M sodium acetate and 2 volumes of absolute ethanol, keep at -20°C for at least 2 hours. Spin at 10,000rpm (Sorvall, type SS-34 rotor) at 4°C for 15 minutes, wash the pellet with 70% ethanol and spin for 5 minutes at same speed. Dry the pellet in vacuum and resuspend in 100 - 200ul of TE.

11. Measure the density and determine the quality of the DNA sample according to O.D. (optical density) value of the UV absorption at 260nm and 280nm as described.

2.2.2 Extraction of Total Cellular RNA

The procedure for RNA extraction is similar to that described in LeMeur et al. (1981). Special precautions and sterile conditions are required for RNA work.

1. Specimens must be kept at -70°C for at least 30 minutes, but less than 6 months before use.

2. Glasswear should be new or used for RNA work only, and baked at 250°C for 4 hours.

3. Disposable plastic tips and eppendorf tubes should be sterilized by autoclaving.

4. Solutions should be prepared using baked glassware, filtered plus autoclaved dH_2O, and RNA reserved dry chemicals.

5. Gloves should be worn through all the steps during RNA work.

6. Work as quickly as possible and keep as cool as possible.

Procedure (for 100mg of specimen):

1. Samples from each stage are stored at -70°C in a kontes microhomogenizer. The samples representing the four body parts (head, thorax, abdomen and leg) are dissected using forceps on dry ice and stored at -70°C in kontes microhomogenizers.

2. Remove sample from -70°C stock, homogenize on dry ice in the same kontes tube with 0.5ml extraction buffer (3M LiCl, 6M Urea, 10mM Sodium acetate pH5.2, 0.2mg/ml Heparin, 0.1% of SDS).

3. Wash the tube with another 0.5ml extraction buffer, mix well, keep at -20°C overnight.

4. Centrifuge the sample in a microcentrifuge at 10,000 xg for 15 minutes at 4°C, save the pellet and keep on ice.

5. Add 0.5ml precooled wash buffer (4M LiCl, 8M Urea) to resuspend pellet.

5. Centrifuge at 10,000 xg for 15 minutes at 4°C, save the pellet and keep on ice.

6. Add 0.3ml dissolving buffer (0.1M sodium acetate pH 5.2, 0.1% SDS), vortex briefly and gently shake for 5 minutes on shaker at room temperature.

7. Add 1 volume of TE pH 8.0 saturated phenol and 1 volume of CHISM, vortex and shake for 20 minutes on shaker at room temperature.

8. Spin at 10,000 xg for 5 minutes, transfer the top aqueous phase and interface to a new eppendorf tube.

9. Add 1 volume of CHISM, vortex, and spin at 10,000xg for 5 minutes.

10. Transfer the top aqueous phase to a new eppendorf tube. Add 1/20 volume of 3M sodium acetate pH5.2 and 2.5 volumes of absolute ethanol. Keep at -20°C overnight.

11. Centrifuge at 10,000 xg for 20 minutes at 4°C, wash the pellet with 70% ethanol twice, dry the pellet in air, and resuspend the pellet in 100-200ul ddH2O.

12. Measure O.D. of 260nm, calculate the density as formula:

Density = $O.D._{260nm}$ x dilution factor x 40.

13. Store the RNA samples at -70°C.

2.2.3 Genomic DNA Library Construction

Three of the genomic subclones pmed21, pDdA1, pDdA5 and one cDNA subclone were isolated in previous work as described (He, 1989, Master's Degree thesis). The phage clone for the subclone pmed5 was originally isolated by L. Arcangeli by screening an EMBL4 genomic DNA library. The methods used to isolate pDdA2 and pDdA3 are described as follow:

Ligation of genomic DNA fragments into phage vectors:

The restriction fragment library construction method described by Pirrota (1986) was used. The genomic DNA from *B. dorsalis* was completely digested with *Eco*RI. The vector system was λgt10 using *Eco*RI digested, phosphatase treated λgt10 arm DNA as well as host strains C600 and C600 Hfl. The inserts can range from 0 up to 7.0kb. C600 strain is used for titering and propagation of λgt10. Recombinant phages (cI^-) and nonrecombinant phages (cI^+) can be distinguished on C600 versus C600hfl because of the different plaque morphologies (clear plaques versus turbid plaques) produced. The C600Hfl strain carries a mutation Hfl A150 and it is an ideal host for recombinant selection and library amplification. The ligation reaction mixture, including 1ug of *Eco*RI digested genomic DNA, 2ug of λgt10 arms, 1 unit of T4 ligase and ligation buffer (5x ligation buffer: 250mM Tris-HCl, pH7.8, 50mM $MgCl_2$, 100mM DTT, 5mM ATP, 250ug/ml BSA) in a total volume of 10ul was incubated at room temperature for 4 hours.

(For *B. dorsalis* , the genomic DNA library construction was done generally as described in Frischauf et al. (1983). The genomic DNA was partially digested with *Sau*3A (0.1 unit of Sau3A for 3 minutes at 37°C) and dephosphorylated with phosphatase (2.5 units of phosphatase for 15 minutes at 37°C) to generate fragments unable to self ligate. The phage vector used was EMBL3 which is derived from 1059 (Frischauf et al., 1983). The EMBL3 vector DNA was double digested with *Bam*HI and

*Eco*RI to prevent religation of the internal "stuffer" fragment. The phosphatased genomic DNA and digested vector DNA were coprecipitated, pelleted and resuspended directly into a ligation reaction as described for the restriction fragment library. Approximately 2ug of vector DNA were ligated to 0.5ug of genomic DNA. Although the genomic DNA was not size selected, only fragments in the range of 9-20kb will generate recombinant phage which can be packaged and propagated.)

Packaging reaction in vitro:

The packaging system contains 50ul extraction mixture per eppendorf tube stored at -70°C. For use, take out one tube and thaw on ice. Add the ligated DNA into the mixture, mix gently and incubate at 22°C for 2 hours. Add 0.5 ml of SM buffer (0.1M NaCl, 0.02M Tris-HCl pH 7.4, 0.01M $MgSO_4$, 0.01% gelatin) and 25ul chloroform, vortex 10 seconds and spin for 2 minutes. Store the constructed library at 4°C.

Titering the library:

To titer the genomic library, a series of dilutions are made. An absorption mixture consisting of 100ul diluted phage DNA from the library, 100ul of 10mM $CaCl_2/MgCl_2$, and about 7.5×10^7 fresh host bacterial cells in a total volume of 400ul in SM is incubated at 37°C for 20 minutes. Add 2.5ml melted top agarose (0.7% agarose) and pour the mixture into prewarmed small NZCYM or LB medium agar plates (100mm). Incubate the plates at 37°C about 6-8 hours until the phage plaques become visible and in proper size. Count the number of phage plaques in each dilution, and calculate the titer as follow:

Titer (pfu/ml) = number of plaques x 10 x dilution factor

Amplification of the library:

According to the titer, prepare two large (150mm) plates with a total of about 25,000pfu bacteriophage, 2.1×10^8 bacterial cells, 100ul of 10mM $CaCl_2/MgCl_2$, and 7.5ml top agarose. Incubate the plates at 37°C until the phage plaques show up. Stop

incubation before it becomes confluent. Add 12ml SM to each plate and keep at 4°C overnight to allow phage particles to diffuse into the liquid SM. Harvest the SM into one sterile plastic tube, wash plates with an additional 4ml of SM and combine with the previous solution. Add chloroform to final concentration of 5%, mix well by gently vortexing, let sit for 15 minutes and spin at about 2800 xg for 5 minutes, transfer the top aqueous phase into a sterile glass container. Store at 4°C. Titer the amplified library as described before.

2.2.4 Identification of Recombinant Clones

The method used to screen the recombinant phage library (genomic and cDNA) is based on the plaque hybridization procedure of Benton and Davis (1977) described by Maniatis et al. (1982). The basic idea is to transfer phage DNA from master plates onto filters (nitrocellulose or nylon). After denaturation and neutralization, the DNA is fixed on the filter by baking. These filters can then be hybridized to a radioactive or nonradioactive probe (described in 2.2.12). Plaques corresponding to positive hybridization signals can be recovered from the master plates.

Procedure:

1. Pour two large petri dish plates (150mm) with approximately 40,000pfu per plate from the original library or amplified library, incubate at 37°C until the plaques reach proper size but are not yet confluent. Chill the plates at 4°C for at least an hour or overnight.

2. Wearing gloves, mark the filters (nitrocellulose or nylon circles) with numbers. Place the first filter onto the surface of the plate, mark both the filter and the plate with a needle to make asymmetric holes around the edge of the plate. Let the filter stay on the plate for 2 minutes and then gently remove using flat forceps. Place second

filter onto the same plate, mark the filter using the same holes on the plate which marked the first filter. Let the second filter sit for 5 minutes and then remove.

3. Allow the filters from the plate to dry DNA side up on a piece of blot paper for a few minutes. Transfer to a stack of blot paper which has been saturated with denaturing solution (1.5M NaCl, 0.5M MaOH) for 5 minutes, again DNA side up.

4. Transfer the filter to third stack of blot paper which has been saturated with neutralizing solution (1.5M NaCl, 0.5M Tris-HCl pH 8.0) for 5 minutes, DNA side up.

5. Rinse the filter in 2xSSC (20xSSC stock: 175.3g NaCl, 88.2g Sodium citrate per liter in dH_2O, sterilize by autoclaving).

6. Bake the filters at 80°C for 2 hours under vacuum.

7. Prehybridize and hybridize the filters with a probe (described in 2.2.12).

After hybridization, washing and either autoradiography or a nonradioactive color reaction, the filters can be realigned with master plates using the orientation marks. The plaques corresponding to the positive signals are picked up from the master plates and suspended in 1ml SM with a drop of chloroform and stored at 4°C.

2.2.5 Analysis of Recombinant DNA

DNA can be purified from the recombinant in bacteriophage for further analysis by restriction endonuclease digestion, electrophoresis, Southern blotting, and subcloning into plasmid vectors.

Isolation of phage DNA:

The procedure for phage DNA isolation was introduced by J. Hunt (personal communication).

1. The phage with inserts of interest are plated out into two small petri dishes with about 10^5 pfu per plate and grown until confluent (but not beyond).

2. Add 5ml of SM to the plates and keep at 4°C overnight.

3. Harvest the SM from the plates into a 50ml Falcon tube, add 1ml more SM to the plates and leave it in tilted position for 15 minutes, transfer the SM into the same tube.

4. Add 120ul chloroform to the tube, vortex gently to mix, spin at about 1200 xg for 10 minutes to separate bacterial debris from the phage particles. Transfer the top aqueous phase into a 30ml corex tube.

5. Add 50ul of DNase (10mg/ml stock) and 50ul of RNase A (10mg/ml), incubate at 37°C for one hour.

6. Add 2ml 70°C prewarmed lysis buffer (0.5M Tris-HCl. 0.25M EDTA, 2.5% SDS), incubate at 70°C for 35min.

7. Add 2.5ml of 8M Potassium acetate, keep on ice for 30 minutes.

8. Centrifuge at 10,000rpm (Sorvall, type SS-34 rotor), 4°C for 20 minutes, transfer the supernatant into a new corex tube. Precipitate with 8ml isopropanol for at least 10 minutes at room temperature.

9. Centrifuge at 10,000rpm for 10 minutes at 20°C, save the pellets.

10. Resuspend the pellets in 900ul of 3M Sodium acetate, transfer the solution into a eppendorf tube, add 600ul isopropanol and keep at room temperature for at least 10 minutes.

11. Spin at 10,000rpm for 10 minutes, pour off the top aqueous phase and save the sticky pellet containing phage particles at the bottom. Resuspend the pellet in 500ul of 0.3M Sodium acetate.

12. Extract two times with phenol and CHISM.

13. Transfer the last top aqueous phase into a new eppendorf tube, add 300ul isopropanol and keep at room temperature for at least 10 minutes.

14. Spin 10 minutes at 10,000rpm to pellet, wash the pellets with 70% ethanol once, spin again.

15. Dry the pellets in a vacuum, resuspend in 50ul TE.

16. Measure $O.D._{260nm}$ and $O.D._{280nm}$ to determine the concentration and the quality of the phage DNA, check on agarose gel.

Subclone interested fragments into plasmid vectors:

After verifying the presence of insert fragments in the recombinant phage DNA by restriction enzyme digestion, electrophoresis, or Southern blotting (discussed in 2.2.9), the fragments can be subcloned into plasmid vectors.

Two methods have been used to subclone fragments of interest into plasmid vectors. In one way, the insert DNA in λgt10 was released from phage arms by an *Eco*RI digest. After phenol extraction and precipitation, the total DNA (arms and insert) were combined with *Eco*RI digested and dephosphorylated pUC9 vector DNA in a ligation reaction. In this type of "shotgun" subcloning, the phage arms tend to ligate at lower efficiency. Even if they do ligate, these recombinant product are generally not recovered after transformation again because of their size and because of incompatibilities between the phage and plasmid genes.

A second method was used to subclone the fragments from the EMBL4 phage clones. The EMBL4 recombinant phages carry a larger amount of insert DNA than λgt10 does. Also, unlike λgt10 which carries only one *Eco*RI fragment, there are usually several *Eco*RI fragments in the insert DNA of the EMBL4 phage. In this second method, the desired internal *Eco*RI fragments were identified by probing Southern blots of restriction digested phage DNA. The fragments which gave positive signals were then eluted from the agarose gel by the Geneclean technique (described in 2.2.8). Eluted fragments were ligated to plasmid vectors as described previously.

Transformation procedure:

Add 100ul of thawed competent cells into prechilled tubes (with blunt bottom). Different transformations are set up using 3ul and 5ul from each ligation mixture and 5ul from each control reaction mixture. For each reaction, volume is made up to 100ul with 0.1M Tris-HCl pH 7.2, and then added and mixed gently with the competent cells. A transformation control is done using 10ng of uncut plasmid. Incubate on ice for 30 minutes, heat shock 45 seconds at 42°C, add 0.8ml LB broth, incubate at 37°C for at least 20 minutes but not longer than 60 minutes.

From each transformation, first take out 100ul of the solution and plate onto prewarmed LB + X-gal + Amp selection plates. Spread the aliquous with a glass rod which is sterilized by dipping in ethanol and flaming. Next, concentrate the remaining cells by spinning the tube for one minute to pellet cells. Pour off most of the supernatant. Resuspend pellet and plate out 50ul of concentrated cells as before. For the controls, make only one plate with 100ul of cells. Incubate the plates upside down at 37°C overnight.

The transformants are selected by Ampicillin in the medium and grow into blue colonies for the normal plasmids and white colonies for the recombinant plasmid which contain insert DNA.

Restreak the white colonies on new LB + X-gal + Amp plates to obtain well isolated single colonies. At this point plates can be stored at 4°C.

Isolation of plasmid DNA:

I. Alkaline plasmid preparation:

Before conducting large scale plasmid DNA preparation, a simplified mini plasmid preparation procedure usually is necessary for the purpose of checking if the insert DNA is present in the plasmid, and if it is the correct size. After confirmation, a large scale preparation can be carried out using a procedure such as the alkaline plasmid preparation method.

Mini preparation procedure:

1. Pick a single white colony, inoculate 5ml LB broth with 5ul of 100mg/ml ampicillin, incubate at 37°C overnight with shaking.

2. Transfer 1ml of the overnight culture to an eppendorf tube, spin 10-15 second to pellet cells in microcentrifuge. Pour off supernatant and save pellet.

3. Freeze pellets in dry-ice ethanol bath for 30 minutes.

4. Resuspend the pellet in 10ul of solution I (50mM Glucose, 10mM EDTA, 25mM Tris-HCl, pH 8.0) with lysozyme to a final concentration of 10mg/ml. Incubate on ice for 15 minutes.

5. Add 20ul of solution II (0.2N NaOH, 1% SDS), incubate on ice for 10 minutes.

6. Add 10ul of solution III (3M Potassium Acetate, adjust pH to 4.8 by adding glacial acetic acid, sterilize by autoclaving), incubate on ice for 15 minutes.

7. Centrifuge at 4°C for 15 minutes, transfer the supernatant to a new tube and add 0.6 volume of isopropanol. Precipitate at -20°C for at least 2 hours.

8. Centrifuge at 4°C for 15 minutes, pour off the supernatant, dry the pellets in vacuum.

9. Resuspend pellets in 20ul of TE, add 2ul of 10mg/ml RNase A, incubate at 37°C for 20-30 minutes.

10. Check for recovery of plasmid on a mini gel and/or restriction digest to characterize plasmid DNA recovered.

Large scale alkaline plasmid preparation:

1. Start a cell culture as in the mini preparation, grow overnight.

2. Make 2 flasks of 250ml LB broth, autoclave and add 250ul of 100mg/ml ampicillin. Add 10ul of overnight cell culture to each flask, incubate at 37oC waterbath overnight with vigorous shaking.

3. Transfer the culture into plastic centrifuge tubes, centrifuge at about 2000 xg for 15 minutes to pellet cells. Pour off supernatant and freeze pellet in ethanol-dry ice bath for 1 hour. (The supernatant should be treated as biohazard material.)

4. Resuspend cell pellet in 2.5ml of solution I with 5mg of lysozyme powder added, keep on ice for 15 minutes. Mix occasionally, transfer into one 50ml OakRidge tube.

5. Add 10ml of solution II (from mini plasmid prep.), mix well and keep on ice for 10 minutes.

6. Add 5ml of solution III, mix well and keep on ice for 15 minutes.

7. Centrifuge at 10,000rpm (Sorvall, type SS-34 rotor) for 15 minutes at 4°C, transfer the supernatant into a 30ml high speed corex tube, add 0.6 volume of cold isopropanol, incubate at -20°C for at least 2 hours to precipitate.

8. Centrifuge at 12,000rpm for 15 minutes at 4°C, save pellets and resuspend in 1ml TE, pH 8.0.

9. Add DNase-free RNase A to the final concentration of 50-100ug/ml, incubate at 37°C for 30-60 minutes.

10. Extract two times with Phenol/CHISM, one time with CHISM.

11. Ethanol precipitate with 0.1 volume of 3M Sodium acetate and 2 volume of absolute ethanol, incubate at -20°C for at least 2 hours.

12. Centrifuge at 12,000rpm for 15 minutes at 4°C, wash with 70% ethanol once, centrifuge for another 10 minutes, save the pellets. Dry the pellets in vacuum.

13. Resuspend the pellets in total of 1.8ml TE pH 8.0 in 1M NaCl.

14. Prepare pZ523 column by spinning at about 1100xg for 1 minutes to remove storage buffer from the column.

15. Add the 1.8ml sample to the column, spin at the same speed for 24 minutes, save the eluted volume.

16. Precipitate the DNA sample from the elution with 0.6 volume of isopropanol. Keep at -20°C until use (can divide the sample into four eppendorf tubes).

17. Precipitate one tube of the sample by spinning for 15 minutes at 4°C, wash with 70% ethanol once and dry the pellets in vacuum. Resuspend the pellets in 100ul TE, pH 8.0.

18. Measure O.D. of 260nm and 280nm and check on the mini gel as described before.

II. Rapid boiling method: The rapid boiling method was first described by Holmes and Quigley (1981). This method has been used mostly in subclone selection and plasmid DNA preparation for DNA sequencing (described in 2.2.14).

1. Pick a single white colony, inoculate 2ml 2xYT medium (16g of select peptone, 10g of select yeast extract, 5g of NaCl in 1000ml dH_2O, sterilize by autoclaving) with 2ul of 100mg/ml ampicillin, incubate at 37°C waterbath overnight with shaking.

2. Overnight bacterial cultures are pelleted by centrifugation in an Eppendorf microcentrifuge for 30 seconds.

3. The pellets are resuspended in 300ul of STET buffer (8% sucrose, 50 mM Tris pH 8.0, 50 mM EDTA, 5% Triton X-100, filter sterilized).

4. Add 15ul of freshly made lysozyme (20mg/ml in 0.01M Tris, pH 8.0) and gently vortex, set on ice for 10 minutes.

5. Boil the mixture for 45 seconds, centrifuge at room temperature for 15 minutes to pellet the cellular debris and chromosomal DNA.

6. Removed the pellets with toothpicks. Add 300ul of isopropanol into the supernatant to precipitate the plasmid DNA. Incubate at room temperature for 10 minutes.

7. Centrifuge at 4°C for 15 minutes, wash with 70% ethanol two times. Dry the pellets in vaccum.

8. Resuspend the pellets in 50ul ddH_2O.

2.2.6 Restriction Endonuclease Digestion

The information about the concentration and the buffer conditions for each enzyme is supplied by manufacturer. Usually the buffers differ only in terms of the concentration of salt or Tris-HCl.

Normally, 2-3 units of enzyme per 1ug DNA sample is used in the reaction. All the restriction enzyme digestions are performed at 37°C. The incubation time varies according to the DNA being digested. For example, in the case of phage DNA, more units of enzyme and longer incubation may be required for complete digestion. For genomic DNA, normal amounts of enzyme, and 2 hour incubations usually work well. For single enzyme digestions of plasmid DNA, 1 hour incubations almost always produce complete digestion. In the case of double or more enzyme digestions, another hour is often added. If the enzymes share the same buffer, they can be put together in one reaction. If the enzymes require different buffers, the digestion could be carry out by doing one enzyme reaction first, then using phenol extraction or the Geneclean technique (described in 2.2.8) to reisolate the DNA for the second digest. Sometimes, the buffer condition can be adjusted to meet the requirements of the second enzyme. The results of digestion can be examined on agarose gels.

2.2.7 Agarose Gel Electrophoresis

The agarose gel can vary in many ways, including the type and concentration of the agarose, the size of the gel, the thickness of the gel and the voltage applied. For large fragments, such as genomic DNA, phage DNA or plasmid DNA, 0.7% gels have

been used. For fine scale restriction mapping or smaller fragments, 1% - 1.5% agarose gels have been used.

A mini gel or a medium gel is made in 25ml or 50ml of electrophoresis buffer 1xTBE (5xTBE stock: 54g Tris-base, 27.5g boric acid, 25ml of 0.5M EDTA pH 8.0 per liter in dH_2O, sterilize by autoclaving). The gel is run in a mini or medium electrophoresis tank (model: MGU-100 or MGU-200 purchased from C.B.S. Scientific Co.). The mini gel or the medium gel is good for quick checking small amounts of DNA, as little as 15 nanograms (ng). It can be run at up to 100 volts and it is usually completed in 1 hour.

A large gel gives better banding resolution and is used for restriction mapping, Southern etc.. The gel can be made in 100-250ml of 1xTBE and run in the large electrophoresis gel tank (model: HE99 purchased from Hoefer Scientific Instruments). The gel is run at low voltage of 20-45 volts overnight.

An even larger gel which can hold as much as 60 samples has been used mostly in subclone selection after deletion reaction (described in 2.2.13). This gel is made in 200ml of 1xTBE and run in the gel tank (model: HRH, made by International Biotechnologies, Inc.). It can be run at as high as 90 volts, and gives very good resolution.

Procedure:

1. Add the correct amount of agarose powder to 1xTBE, heat in a microwave oven until the agarose dissolves.

2. Cool the solution to 50-55°C, add ethidium bromide to a final concentration of 1ug/ml.

3. Seal the both edges of the gel plate with tape to form a mold. Place a comb near one end of the mold.

4. Pour the warm gel solution into the mold, let it set until solidifies.

5. Remove the comb and tape, put the gel in the electrophoresis tank.

6. Add 1xelectrophoresis buffer (TBE or TAE) to submerge the gel.

7. Mix DNA samples with loading buffer (10xstock: 50% glycerol, 0.4% bromophenol blue), load the sample carefully into the wells. Run the gel at desired voltage.

8. Stop electrophoresis, make a photograph of the gel under UV light with Polariod camera.

2.2.8 Geneclean Technique

Geneclean^{TM} kit is a product purchased from Bio 101 inc.. It is an aid in separating clean single or double stranded DNA from contaminants. For example, it can be used for elution of DNA fragments from agarose gels, for purification between enzymatic reactions, etc.

Procedure:

1. Excise DNA band from agarose gel under UV-light, and weigh the agarose.

2. Add 2.5-3.0 volume of sodium iodide (1g = 1ml), incubate at 50-55°C water bath for 5 minutes to dissolve the agarose. If the DNA sample does not contain agarose, after adding NaI, directly go to next step.

3. Add 5ul or less of glassmilk to solutions with 5ug or less DNA, for more than 5ug, add 1ul of glassmilk to each additional 0.5ug of DNA. Incubate on ice for 5 minutes.

4. Spin for 5 seconds to pellet, pour off the supernatant.

5. Wash the pellet with 200-700 volume of New Wash (Add the 7ml "New Concentrate" vial provided with the kit to 140ml of ddH_2O, mix well. Add 155ml of

100% ethanol, mix well. Store at -20°C), spin 5', pour off the supernatant. Wash three times. Make sure to remove all the liquid at the last wash.

6. Elute the DNA from glassmilk by resuspending the pellet in 5-10ul of TE or ddH2O, and incubate at 50°C - 60°C for 2-3 minutes. Spin for 30 seconds, transfer the supernatant which contains the eluted DNA into a new tube.

7. Repeat step 6, add the supernatant into the same tube.

2.2.9 Southern Blot

The Southern blot method developed by Southern (1975), is a technique to identify particular DNA sequences by transferring them, after size fractionation on an agarose gel, to a nitrocellulose or nylon filter for detection by hybridization.

Procedure (handle with gloves at each step):

1. After electrophoresis and photography, the DNA in the gel is denatured by soaking the gel in denaturing solution (1.5M NaCl, 0.5M NaOH) for one hour with gentle shaking.

2. The gel is then placed in a neutralizing solution (1M Tris-HCl pH 8.0, 1.5M NaCl) for one hour with gentle shaking.

3. In a large baking dish, a sponge slightly larger than the gel is saturated with 4xSSC to serve as a base support and buffer source for the blot transfer.

4. Wet two pieces of gel blotting paper in 2xSSC, lay on the soaked sponge one by one to cover the surface, remove all the bubbles.

5. Place the neutralized gel on the blotting paper, again avoiding air bubbles. Wet a piece of nylon filter in 2xSSC and lay carefully on the surface of the gel. Cut off any portions of gel uncovered, gently move out bubbles. Once in place, the filter should not be moved.

6. Use parafilm to surround the edges of the gel to direct buffer flow through gel only. Place two more pieces of blotting paper (prewet with 2xSSC) over the filter, carefully remove any bubbles.

7. Put a stack of paper towels on the blotting paper to wick buffer up.

8. Put a plate on the paper towel and a 500g weight on it.

9. Allow the transfer of DNA to continue for 16-20 hours.

10. After transfer, remove all the things above the filter and mark the orientation of the filter by cutting off a small piece at one corner.

11. Rinse the filter in 2xSSC, dry in air and bake for 2 hours at 80°C under vacuum.

The filter now is ready for prehybridization.

2.2.10 RNA slot blot

Procedure (handle with gloves at each step):

1. Wet nylon filter in ddH_2O, then soak in 20xSSC for one hour.

2. Clean slot blot manifold (purchased from Schleicher & Schuell) with 0.1N NaOH by submerging in 0.1N NaOH solution, rinse with ddH_2O.

3. Prepare RNA samples (1ug - 3ug) in 20ul ddH_2O, add 20ul of 100% formamide, 7ul of 37% formaldehyde, 2ul of 20xSSC.

4. Incubate the samples at 68°C for 15 minutes, set on ice. Add 2vol of 20xSSC, keep the samples on ice before loading.

5. Assemble the manifold as described by the manufactory.

6. Wash slots with 500ul of 10xSSC, apply gentle suction, turn off the vacuum after solution have gone through the filter.

7. Refill slots with 500ul of 10xSSC, apply gentle suction when samples are ready, then turn off the cacuum.

8. Load samples in each slot, apply gentle suction until all samples go through the slot, then turn off the vacuum.

9. Rinse each slot twice with 10xSSC, continue suction for 5 munites.

10. disassemble the manifold, dry the filter in air and bake the filter at 80°C for 30 munites with vacuum.

The filter now is ready for prehybridization.

2.2.11 DNA labeling

Radioactive DNA probe with ^{32}P-dCTP nucleotide is labeled by a random priming labeling method (Feinberg and Vogelstein, 1983; 1984), and has been used in RNA slot blot hybridization. Nonradioactive DNA probe is labeled by either random priming method or nick translation, in which a dUTP nucleotide with a Digoxigenin side group (Dig-dUTP) is incorporated into the newly synthesized DNA strands. Nonradioactive probes have been used in Southern blot hybridization and *in situ* hybridization (described in 2-2-16) as well.

Radioactive DNA probe labeling with ^{32}P-dCTP

1. Denature the linearized DNA in a volume of 9ul in ddH_2O by heating at 95°C for 10 minutes. Quick spin for 1 second and cool on ice.

2. Add 2ul of hexanucleotide mixture, 3ul of dATP, dGTP, dTTP mixture, 5ul of 50uCi ^{32}P-dCTP, and 1ul Klenow enzyme. Incubate for 30 minutes at 37°C.

3. Stop reaction by adding 2ul of 0.2M EDTA (pH 8.0).

4. Add 28ul ddH_2O, to make the total volume of 50ul.

5. Add 1ul spermine to a final concentration of 2mM. Incubate on ice at least for 5 minutes to precipitate DNA.

6. Centrifuge for 15 minutes at 4°C, pour supernatant off from pellet. Dry the pellet in air, and dissolve the pellet in 100ul TE, heat at 95°C for 7 minutes.

7. Measure the radioactivity as follows:

Add 3ml Scintilation cocktail into two samples, sample one contains 1ul of supernatant, sample 2 contains 1ul of pellet resuspension. Measure each sample in Scintilation counter.

Calculate by the formula:

Percentage of incorporation =

Total activity of pellet/total activity of pellet and supernatant

Total activity of labeled DNA (cpm) =

activity reading of pellet X total volume of pellet

specific activity (cpm/ug DNA)=

Total activity / The amount of DNA used

Nonradioactive DNA probe labeling with Dig-dUTP

I. Random priming DNA labeling:

1. Denature the DNA sample in a total volume of 15ul in ddH_2O by heating at 95°C for 10 minutes, quick spin for 1 second and set on ice.

2. Add 2ul of hexanucleotide mixture, 2ul of Dig-DNA mix (dATP, dCTP, dGTP, dTTP, Dig-dUTP) and 1ul Klenow enzyme into the denatured DNA sample, incubate at 37°C for 1-2 hours.

3. Stop the reaction by adding 2ul of 0.2M EDTA.

4. Collect the newly synthesized DNA by gene clean or ethanol precipitation as follows:

Precipitate DNA with 2ul of 10mg/ml tRNA, 3ul of 4M LiCl, and 100ul absolute ethanol. Incubate at -20°C for at least 2 hours. Spin in microcentrifuge for 15 minutes at 4°C, wash the pellets with 70% ethanol once. Dry the pellets in vacuum. Resuspend the labeled DNA pellets in 50ul TE with 0.1% of SDS, incubate at 37°C for 10 minutes.

The labeled DNA probe can be stored at -20°C.

II. Nick translation:

1. Make the reaction mixture in a total volume of 20ul with 5ul of Dig-DNA mix from genius kit, 2ul of 10x buffer and 2ul of enzyme mix from Nick translation kit, and the DNA sample.

2. Incubate the mixture at room temperature for 90 minutes.

3. Stop the reaction by adding 2ul of 0.2M EDTA, pH8.0.

4. Collect the newly synthesized DNA by gene clean or ethanol precipitation.

Stored the labeled DNA probe at -20°C.

2.2.12 Hybridization and Detection System

Two methods of hybridization and detection have been used to probe DNA or RNA which has been immobilized on filters by means of Southern blot or RNA slot blot. One is radioactive detection system, in which ^{32}P-labeled probe is used in a hybridization reaction and detected by autoradiographic exposure to X-ray film. Another method is the nonradioactive detection system, which is based on principles of immunology. In the nonradioactive method, after hybridization between the filter bound DNA or RNA and the probe, hybrids are detected by using an antibody-conjugate (anti-digoxigenin alkaline phosphatase conjugate, <Dig>AP) following by an enzyme catalyzed color reaction with 5-bromo-4-chloro-3-indolyl phosphate (X-phosphate) and nitro blue tetra zolium salt (NBT). Positive reactions produce a purple/blue coloration. This nonradioactive system with some modifications is also used in *in situ* hybridization for locolizing individual genes on polytene chromosomes as well as mitotic chromosomes (described in 2.2.17)

Radioactive hybridization and detection system: Procedures:

1. Put the filter which have the RNA fixed on in a sealable plastic bag, add 20ml prehybridization solution for per 100cm^2 filter. Incubate at 42°C for a few hours or overnight. The prehybridization solution consists of 50% (full stringency) or 43% (reduced stringency) formamide, 6xSSC, 10mM Na/K PO_4 pH 6.5 (for 100ml prehybridization solution, take 10.36ml of 67mM NaH_2PO_4 and 4.44ml of 67mM KH_2PO_4), 1xDenhardt's solution (50x: 5g Ficoll, 5g Polyvinylpyrrolidone, 5g BSA Penlax Fraction V, to 500ml in d H_2O) and 250ug/ml salmon sperm (Boil 5 minutes, chill on ice before adding)

2. Add ^{32}P-labeled DNA in fresh prehybridization solution (2.5ml for per 100cm^2 filter) with or without 10% Dextran Sulphate (Dextran Sulphate may increase the background in the case of RNA slot blot) to make hybridization solution, denature by heating for 10 minutes at 95°C.

3. Pour off the prehybridization solution from the bag, replace with hybridization solution containing denatured probe. Seal bag well and incubate at 42°C overnight.

4. Pour off the hybridization solution (saving for possible future use), wash filters in a container with following solutions at 65°C (full stringency) or 50°C (reduced stringency) with occasional shaking.

First wash: in 50% formamide, 2xSSC and 0.1% SDS, for 15' at RT;

Second washes: in 2xSSC and 0.1% SDS, for 15' at RT, twice;

Third washes: in 0.2xSSC and 0.1% SDS, for 15' at 50°C, twice.

5. Dry the filters in air, cover each filter with saran wrap, place in cassette with x-ray film with or without intensifying screen for autoradiography . Expose at -70°C appropriate length of time depending on the intensity of radioactivity.

6. Develop the films in dark room.

The positive results are given by black bands.

Nonradioactive hybridization and detection system: Procedures:

Hybridization

1. Put the filters which have DNA fixed on into a sealable plastic bag, add 20ml/100cm^2 filter of prehybridization solution into the bag, incubate at 42°C for at least 1 hour. The prehybridization solution consists of 50% (full stringency) or 43% (reduced stringency) formamide, 5xSSC, 5% of blocking reagent (supplied with the kit), 0.1% of N-lauroylsacosine Na-salt and 0.02% of SDS

2. Pour off the prehybridization solution from the bag, add denatured (heat at 95°C for 10 minutes) nonradioactive probe in about 2.5ml/100cm^2 of prehybridization solution into the bag. Seal bag and Incubate at 42°C overnight.

3. Remove hybridization solution, remove filters and wash as follows:

First and second washes: 2.0xSSC, 0.1% of SDS, at room temperature for 5 minutes.

Third and fourth washes: 0.1xSSC, 0.1% of SDS, at 55°C for 15 minutes or at 60°C for 15 minutes.

The filters are now ready for color reaction.

Color reaction (all steps are performed at room temperature)

1. Wash the filters with buffer 1 (0.1M Tris-HCl pH 7.5, 0.15M NaCl) for 1 minute.

2. Incubate the filters in buffer 2 (0.5% of blocking reagent in buffer 1) for 30 minutes.

3. Wash the filters with buffer 1 for 1 minute.

4. Dilute <Dig>Ap-conjugate in buffer 1 to 150u/ml, put the filters into a bag, pour into the diluted <Dig>Ap-conjugate solution, incubate for 30 minutes.

5. Wash filters with buffer 1 for 15 minutes with gently shaking, repeat once.

6. Equilibrate filters with buffer 3 (0.1M Tris-HCl pH 9.5, 0.1M NaCl, 0.05M $MgCl_2$) for 2 minutes.

7. Make up color solution with 45ul of NBT and 35ul of X-phosphate in per 10ml of buffer 3. Incubate the filters in the color solution for color reaction for hours to one day.

8. Stop the reaction with TE (10mM Tris-HCl pH 8.0, 1mM EDTA) for 5 minutes when the positive signals are detectable. Dry the filters in air.

9. Photocopy or photograph the filters for documentation.

2.2.13 Nested Deletions

The ExoIII/Mung bean deletion kit from Stratagene, Inc. is used to produce unidirectional nested deletions in a DNA insert. The double stranded DNA, purified by the rapid boiling preparation, is double digested with restriction enzymes.which (1) generate a unique 3' overhanging restriction site and (2) a unique 5' or blunt restriction site between the insert and the 3' site. If there is no proper 3' overhanging site, a 5' overhanging digest filled in using deoxy-thioderivatives and Klenow enzyme can be used. The double digested DNA is then treated with Exonuclease III to digest the DNA from the 5' overhanging site or blunt restriction site progressively. The remaining single stranded DNA is digested to blunt ends with Mung bean nuclease. The deletion subclone is then ligated and transformed into DH5α competent cells.

Procedures: (use only one tenth the materials of the original protocol provided by the manufactuer)

1. Prepare the plasmid DNA using rapid boiling method.

2. Digest the DNA (0.5ug for each time point) with appropriate enzyme as described. Make sure the digestion is complete. Skip to step d if a 3' overhanging

enzyme is used. Otherwise do the thioderivative filling if the 5' overhang enzyme is used as following:

(1). Heat the restriction digest (20-30ug in 500ul) to 75°C for 15 min.

(2). Add 2ul of the 1mM stock of thio-dNTP and 1ul of 5u/l Klenow polymerase, incubate at room temperature for 10 min.

(3). Phenol extraction by one time phenol, one time phenol/CHISM (chloroform : isoamyl alcohol = 24:1 in volume), one time CHISM.

(4). Ethanol precipitate with 0.1v of 3M sodium acetate, 2 vols of absolute ethanol. Incubate at -20°C at least 2 hours.

(5). Check fill-in by incubating 1ug DNA with 20U Exonuclease III for 15 min, at 37°C and run on an agarose gel. The sample should be protected from deletion.

3. Do second restriction enzyme digestion, followed by Phenol extraction and ethanol precipitation.

4. Exonuclease III/Mung Bean nuclease deletion reactions:

(1). Dilute 2ul of 10x Mung Bean Buffer into 15.5ul of ddH_2O in an eppendorf tube for each time point.

(2). Exonuclease III deletion reaction mixture contains:

Ingredient	one time point	N time points
double digested DNA	0.50ug	Nx0.50ug
2x ExoIII buffer	1.25ul	Nx1.25ul
100mM b-mercaptoethanol	0.25ul	Nx0.25ul
100U/ul Exonuclease III	0.10ul	Nx0.10ul
Add ddH_2O to total Vol	2.50ul	Nx2.50ul

Incubate the mixture at 30°C at which the Exonuclease III converts the double stranded DNA to single stranded DNA at the rate of approximately 230bp/minute.

Remove 2.5ul aliquot from the reaction mixture at each desired time point and directly add into the tube containing the 17.5ul of diluted Mung bean buffer. Place the tube on dry ice.

(3). Heat tubes at 68°C for 15 minutes after all aliquots have been removed and placed on dry ice. Place the tubes on ice.

(4). Dilute Mung Bean Nuclease stock (85U/ul) with 1x Mung Bean dilution buffer to 15U/ul. Add 1ul of the diluted nuclease into each time point tube and incubate for 30 minutes at 30°C.

5. Check 3-4ul of each time point deletion on mini gel.

6. Geneclean the deletion samples with 1ul of glassmilk and resuspend in 15-20ul of ddH_2O.

7 Do the ligation reaction and transformation at normal condition with 5ul of the geneclean sample.

8 Screen deletion subclones using rapid boiling method for plasmid identification. The subclones with proper deletions can be used for sequencing directly.

2.2.14 DNA Sequencing

Subcloning and plasmid preparation

Overlapping subclones for DNA sequencing were obtained either by utilizing restriction endonuclease sites or by generating nested deletions using the ExoIII/mung bean nuclease deletion kit from Stratagene, Inc. The cDNA subclone pmedC1 was sequenced directly. The rapid boiling method is used in the preparation of plasmid DNA for DNA sequencing.

Denaturation of double stranded DNA sample:

About 5ug of plasmid DNA is diluted in ddH_2O to a total of 36ul. The DNA is denatured with 4ul of denaturation buffer (2N NaOH, .2 mM EDTA, pH 8.0) for 5

minutes at room temperature, and then neutralized in 4ul of neutralization buffer (2M ammonium acetate, pH 4.5). The denatured DNA templates are precipitated with 100ul of absolute ethanol at -20°C for at least 2 hours (or overnight). The samples are then centrifuged at 4°C for 15 minutes and washed with 1ml of 70% ethanol once. The pellets are dried in a vacuum and are ready for sequencing.

Sequencing reaction:

Sequenase version 2.0 kits from US Biochemical, Inc. are used for the DNA sequencing reactions. The primers used are the M13/pUC sequencing primer and M13/pUC reverse sequencing primer from pUC sequencing kit of Boehringer Mannheim.

Procedures: (Sequenase version 2.0 kits from US Biochemical, Inc.)

1. Annealing reaction: the annealing mix includes the denatured DNA sample in pellets, 0.5pmol of primer, 2ul of sequenase reaction buffer, in a total of 10ul in ddH_2O. Heat the mixture to 65°C for 2 minutes, then slowly cool down to 30°C, set on ice.

2. Prepare 4 small eppendorf tubes for each sample, mark the tubes as T, C, G, A, and add 2.5ul of each dideoxynucleotide termination buffer (ddTTP, ddCTP, ddGTP and ddATP) respectively to the each marked tube. Warm the tubes at 37°C for at least 1 minute before use.

3. Dilute the sequenase version 2.0 with sequenase dilution buffer 1:8.

4. Make reaction mix stock as following:

Ingredient	one sample	N samples
ddH_2O	1.6ul	Nx1.6ul
DTT	1.0ul	Nx1.0ul
5xlabeling mix	0.4ul	Nx0.4ul
^{35}S-dATP	0.5ul	Nx0.5ul

diluted sequenase	2.0ul	Nx2.0ul

Transfer 5.5ul of the reaction mix into the annealing reaction tube for each sample, place the aliquout on the side of the tube. Quick spin to mix well simultaneously for all samples.

5. Immedately transfer 3.5ul of above mixture into the four small eppendorf tubes containing the dideoxynucleotide termination buffers, again placing the aliquout on the side of the tube. Quick spin to mix well simultaneously for all samples. Incubate at 37°C for 5 minutes.

6. Add 4ul of stop buffer to each tube. Keep at -20°C before polyacrylamide gel electrophoresis.

Polyacrylamide gel electrophoresis:

In most cases, the buffer gradient polyacrylamide sequencing gel (6%) are used (Sambrook et al. 1989). Make stock solutions as follows:

6% buffer gradient gel top solution:

40% acrylamide:bis (1:19) solution	75ml
5xTBE	50ml
Urea	230g

Adjust volume to 500ml with ddH_2O, filter and store at 4°C.

6% buffer gradient gel bottom solution:

40% acrylamide:bis (1:19) solution	25ml
5xTBE	50ml
Urea	46g
Sucrose	10g
Bromophenol blue	5mg

filter and store at 4°C.

Take 70ml of 6% buffer gradient gel top solution and 13ml of 6% buffer gradient gel bottom solution, put in seperate beakers. Add 240ul and 52ul of freshly made 10% Ammonium persulphate into the top solutions and the bottom solutions respectively, vortex gently to mix. Add 100ul of TEMED into the top solution, vortex gently to mix well. Take up about 50ml of the top solution with a syringe, keep it for later use. Add 19.5ul of TEMED into the bottom solution, vortex gently to mix well. Take up 12ml of the top solution with a 25ml pipet tube, and then take up 12ml of the bottom solution in the same pipet. Allow 3 to 4 bubbles to run through the pipet, then pour the solution from the pipet into the gel plates. Now pour the top solution from the syringe into the gel plates until it is full. Insert the flat edge of the comb inside the plates about 2mm from the top of the small plate. Clamp both sides of the gel plates at the top. Keep the plates in a tilted position, allow the gel polymerize for about 45 minutes. Cover the top with 1xTBE dampened kimwipe and cover with saran wrap. Leave it overnight.

Remove the comb and reinsert with just contacting gel surface and prerun the gel for 30 minutes to 60 minutes at 60 watts before loading the samples. After the prerun, wash individual wells with a syringe. Samples are heated at 90 - 100°C for about 3 minutes and set on ice. Load about 2.0ul - 3.0ul of each sample. Gels are typically run at 60 watts for 2 hours until the first dye runs off. After running, the gels are fixed for about 15 minutes in a 15% methanol, 5% acetic acid bath. After fixation, gels are dried on a gel dryer and autoradiographed using Fuji-RX X-ray film. Sequences from these films are read either manually or by using a hand held gel input reader from CBS Scientific, Inc. Results are then entered into a MacIntosh computer for analysis using either the DNA Inspector II (Textco,Inc.), or MacVector (IBI, Inc.), or other computer programs.

2.2.15 Inverse PCR

Inverse Polymerase Chain Reaction (IPCR) was first described by Ochman et al. (1988) which allows the amplification of upstream or downstream sequences flanking a region of interest. To do this, the PCR primers are oriented in the reverse direction of the usual orientation. DNA templates are digested with a restriction enzyme and self ligated to form a circle. Typical PCR reactions can then be applied.

Circularization of template DNA:

Genomic DNA of *B. dorsalis* is digested with a designed restriction enzyme. The ligation reaction for self circularization is set up under the conditions which favors the formation of monomeric circles (Collins and Weissman, 1984). The DNA fragments from the restriction enzyme digestion are diluted to a final concentration of 0.5ng/ul in a total volume of 200ul including 20ul of the 10x ligation buffer and 4ul of the T4 DNA ligase (Boehringe Mannheim). The reaction mixture is incubated at 15°C for 16 hrs. The samples are then treated with phenol extraction, followed by ethanol precipitation.

Primer preparation:

A computer program "Primer" was used to search for two primer sequences within the *Eco*RI-*Bgl*II fragment of the pDdA2 subclone. The primer 92058 with the sequence of 5' -CAAGTACCCCATTGAGCACGG- 3' is located in the coding region of pDdA2 before *Bgl*II restriction site. The primer 92059 with the sequence of 5' -CGGCGGGAAACGAA-GAAAC- 3' is located in the intron region of pDdA2. The melting temperature is 64.8°C for primer 92058 and 65°C for primer 92059.

PCR reaction:

The GeneAmp PCR reagent kit from Perkin Elmer Cetus is used to perform the PCR. The reaction is set up in a total of 100ul with 10-50ng of circularized DNA template, 20-50pmol of each primer, 200uM of dNTPs, 1.5mM of $MgCl_2$ and 2.5units of AmpliTaq DNA Polymerase. 75ul of mineral oil is laid on the top of the reaction

mixture. The conditions for the PCR cycles are: initial denaturation for 2 min at 94°C for one cycle, followed by 35 cycles of denaturation at 94°C for 1 min, primer annealing at 65°C for 1 min, and extension at 74°C for 2'.

A portion (15ul) of the amplified sample is checked on an agarose gel. The PCR products are purified by phenol extraction and ethanol precipitation or by excising from agarose gel by the Geneclean procedure. The products are incubated in a 20ul reaction mix with 1ul of dNTPs and 2ul of 10x buffer (from nick translation kit), 1ul of 2U/ul klenow enzyme at 22°C for 15-30 min to make double stranded blund ends before being subcloned into pUC vectors.

2.2.16 Mitotic Chromosome Preparation

For general staining:

1. Dissect brains from third instar larve of *C. capitata* or *B. dorsalis* in 0.7% NaCl.
2. Fix the brain in 1% Sodium Citrate for 10 minutes.
3. Stain the brain in a drop of 1% aceto orcein for 5 - 10 minutes on a coverslip. Add another drop of 2% lacto-aceto orcein and mix.
4. Pick up the coverslip with a slide. Squash to spread the chromosomes.
5. Check the chromosomes under microscope.

For *in situ* hybridization:

1. Dissect brains from third instar larve of *C. capitata* or *B. dorsslis* in 0.7% NaCl.
2. Fix the brains in 1% Na Citrate for 10 minutes.
3. Transfer the brain in the 8 -10ul of acetic acid: ddH_2O:lactic acid (3:2:1) on a siliconized coverslip (18mm^2).

4. Pick up the coverslip with a slide, leave one side of the coverslip overhanging the slide. Squash to spread the chromosomes.

5. Dip the slide with the coverslip end into liquid nitrogen for a few seconds until it is frozen.

6. Flick off the coverslip with a forceps. Immediately immerse in cold 95% ethanol for 10 minutes, three times.

7. Air-dry the slide in a dust free atmosphere. Check the chromosome under microscope (Olympus BH-2). Select the slides with good chromosome spreading for in situ hybridization.

*2% lecto-aceto orcein preparation: Dissolve 2g of orcein in 50cc of heated glacial acetic acid. Remove from heat and add 50cc of 85% lactic acid. Filter the solution.

2.2.17 *In Situ* Hybridization

The genius color reaction method is used for hybridization detection. Probes from both nick translation and random priming methods have been used.

Denaturation: several methods have been used.

I. Heat denature:

In this method, the chromosomes on the slide and the probe are denatured at the same time. Apply 10ul of hybridization solution with probe (undenatured) onto a siliconization coverslip ($22mm^2$), pick up the coverslip with the slide which has the chromosomes (undenatured). Seal the coverslip with generous rubber cement. Dry in air. Put the slides in a humid chamber. Incubate at 80°C for 15 minutes to heat denature. Incubate at 37°C ovenight.

II. Heat denature in 70% formamide:

As in method A, the chromosomes on the slide and the probe are denatured at the same time. Apply 10ul of hybridization solution with probe (undenatured) onto a siliconization coverslip ($22mm^2$), pick up the coverslip with the slide which has the chromosomes (undenatured). Seal the coverslip with generous rubber cement. Dry in air. Preheated the slides at 37°C. To denature, immerse the slides in 70°C preheated solutions with 70% formamide and 2xSSC for 2 minutes. Put the slides in a humid chamber. Incubate at 37°C ovenight.

III. Separately denature chromosomes and probes:

Denature the chromosomes on the slides as follows:

1. Immerse slides in 70% formamide, 2xSSC at 70°C for 2 minutes.
2. Dehydrate the slides in cold 70% Ethanol for 5 minutes.
3. Transfer the slides to cold 95% Ethanol for 5 minutes.
4. Transfer the slides to cold 100% Ethanol for 5 minutes.

Air dry the slides until the ethanol is invisible.

Hybridization:

1. Prepare the solutions for hybridization reactions:

2x Hybridization Base Mix: for 10ml

Ingredient	Amount
20x SSC	4ml
50x Denhart's solution	400ul
50% Dextran Sulphate	2ml
10mg/ml Salmon sperm	300ul (boil 5' before adding)
10mg/ml Yeast tRNA	150ul
ddH_2O	3.15ml

Hybridization solution: both nick translation and random priming labeling methods have been used to make probes.

Ingredient	Amount
2x Hybridization Base Mix	25ul
100% Formamide	20ul
Nonradioactive probe	10-50ng
ddH_2O	make to 50ul

2. Denature the hybridization solution by heating at 95°C for 10 minutes, and set on ice.

3. Apply 10ul of hybridization solution onto a siliconization coverslip ($22mm^2$), pick up the coverslip with the slide which has the denatured chromosomes. Seal the coverslip with rubber cement. Dry in air.

4. Incubate the slides in a humid chamber at 37°C, overnight.

Post hybridization:

1. Carefully peel off rubber cement with forceps and float the coverslip off in 2xSSC.

2. Wash the slides as follows:

First wash: in 50% formamide, 2xSSC, for 30' at 37°C;

Color detection (Boehring Mannhein Genius kit):

1. Place the slides in buffer 1 (0.1M Tris-HCl pH 7.5, 0.15M NaCl) for 1 minute.

2. Incubate the slides in buffer 1 with 2% of blocking reagent and 0.3% of triton X-100 for 30 minutes.

3. Make 1:500 dilution of <Dig>Ap-conjugate in buffer 1 with 1% of blocking reagent and 0.3% of Triton X-100. Apply 100ul onto each slide where the chromosomes are and cover with a siliconized coverslip ($22x40mm^2$). Incubate for 2 hours.

4. Wash the slides in buffer 1 for 15 minutes with gently shaking, repeat once.

5. Equilibrate each slide with 0.5ml of buffer 3 (0.1M Tris-HCl pH 9.5, 0.1M NaCl, 0.05M $MgCl_2$) for 2 minutes.

6. Make up color solution with 45ul of NBT and 35ul of X-phosphate, and 2.4mg of Levamisole in per 10ml of buffer 3. Apply 500ul onto each slide and cover with a unsiliconized coverslip (22x40mm^2). Keep the slides in a humid chamber and incubate in dark overnight.

7. Float off the coverslip in ddH_2O. Check under microscope with a drop of dH_2O and a coverslip.

8. Stain the chromosomes in 5% of Giemsa (0.5ml of Giemsa stock, 0.5ml of 1M Sodium Phosphate, pH6.8, 49ml of ddH_2O) for 20-30 second.

9. Mount the slides with "CoverBond Mounting Media".

10. Take pictures with normal camera or store the image in a computer using a video camera for making slides.

CHAPTER 3
RESULTS AND DISCUSSION

3.1 ISOLATION OF ACTIN HOMOLOGOUS SUBCLONES

3.1.1 Screening Genomic DNA Library

Four actin homologous subclones, two from *Ceratitis capitata* and two from *Bactrocera dorsalis* , were isolated in previous studies (Haymer et al., 1990; He and Haymer, 1991). The clone containing the CcA1 actin gene (designated pmed21) was originally isolated by screening a Charon 4 *C. capitata* genomic DNA library with the *Drosophila* clone DmA2 (see Haymer et al. 1990). The DmA2 clone contains the entire protein coding region of the *D. melanogaster* actin gene 5C (Burn et al. 1989). The cDNA clone (designated pmedC1) was isolated from a *C. capitata* female pupal cDNA λgt10 library using pmed21 as probe (Haymer et. al. 1990). The clones for DdA1 and DdA5 actin genes (designated pDdA1 and pDdA5, respectively) were originally isolated by screening a *B. dorsalis* genomic DNA EMBL3 library with again the *Drosophila* clone DmA2 (He and Haymer, 1991). In addition to these subclones, several new actin homologous subclones have been isolated in this study. Two were isolated by screening an *Eco*RI λgt10 restriction fragment genomic library of *B. dorsalis*. About 80,000 pfu of the original λgt10 library were plated out in two large plates each time and screened with a DmA2 nonradioactive probe under reduced stringency (43% formamide, 50^0C-55^0C washes, see Chapter 2, section 2.2.4). A total of ten plaques from the screening which gave intense positive signals were picked and plaque purified. The phage DNA was then extracted and purified. Two of the insert fragments from the phage clones were new based on their sizes on agarose gel electrophoresis and were then subcloned into pUC9 vector. These two subclones were given the names pDdA2 and pDdA3 with

the sizes of 4.0kb and 4.1kb respectively. Both of them correspond to fragments identified in a genomic DNA Southern blot, showing all actin homologous sequences.

A subclone of a 620bp *Eco*RI fragment, which contains additional sequences of pDdA1 covering part of the coding region and 3' flanking region, has been isolated from an EMBL3 phage clone λDd1-1 the same as pDdA1. This subclone has been given the name pDdA1.2.

An actin homologous *Eco*RI fragment originally isolated from a *C. capitata* EMBL 3 genomic library by L. Arcangli was subcloned into pUC9 (named as pmed5 with size of 2.0kb). However, this subclone does not correspond to any fragments on the genomic DNA Southern blot.

3.1.2 Inverse Polymerase Chain Reaction

In order to recover the additional sequences of pDdA2 containing the 5'exon of this gene, the inverse polymerase chain reaction (IPCR) technique was employed. The two primers used here were designed based on the sequences within the *Eco*RI-*Bgl*II fragment of pDdA2 (see Figure 3.2.4) using a computer program "Primer" developed by Stephen et al. (version 1991). One primer (92058) is located in the coding region with the same orientation as the reading frame, while the other primer (92059) is located in the intron region with opposite orientation as the reading frame. The genomic DNA of *B. dorsalis* was digested separately with *Taq*I, *Sau*3A, *Bam*HI-*Bgl*II, *Eco*RV, and *Hinc*II. The digested genomic DNA was then circularized by using T_4 DNA Ligase. The PCR has been run using various parameters. At low annealing temperature (<=60^0C), more than one band was usually seen in each reaction. A major fragment from the *Bam*HI-*Bgl*II digested sample, when sequenced, turned out to be an unrelated sequence. To elimilate this kind of false product, a higher annealing temperature (65^0C) was used. In this case a single fragment was seen in *Taq*I, *Sau*3A and *Hinc*II digested samples, while

nothing was produced in *Bam*HI-*Bgl*II and *Eco*RV digested samples. The three PCR products recovered were subcloned into pUC19 and given the names pDdTaqIPCR, pDdSau3APCR, and pDdHincIIPCR, respectively.

pDdHincIIPCR contains the most 5' region among the three subclones. It extends about 800bp upstream of the pDdA2 subclone. However, it still does not cover the 5' exon of the DdA2 gene. In an attempts to recover the 5' exon, a *Eco*RI-*Bgl*II fragment from the pDdHincIIPCR clone was isolated. This fragment, when used as a probe, hybridized strongly to a single band of about 6kb in length in *Eco*RI digests of *B. dorsalis* genomic DNA. It was then used as a probe to screening a *B. dorsalis* λgt10 genomic DNA library. Two positive plaques from a total of 80,000pfu were obtained. One of them (λDdA-2-1) carried more than one *Eco*RI fragments, while another one (λDdA-1-1) carried one large *Eco*RI fragment. Both of them had the 6kb fragment as described above. This 6kb fragment was then subcloned into a pUC19 vector and given the name pDdA2.1.

3.2 SEQUENCE ANALYSIS

3.2.1 Sequencing Strategy and Gene Structure

Sequenase version 2.0 from U.S. Biochemical was employed in the sequencing reactions, performed based on the dideoxy chain termination method of Sanger et al. (1977). A total of six actin genes have been sequenced, including four actin genes (DdA1, DdA2, DdA3, DdA5) isolated from *B. dorsalis,* two actin genes (CcA1, CcA2) and one cDNA subclone (pmedC1) isolated from *C. capitata.* Most of the sequenced regions contain complete coding regions, intervening sequences (if present), and reasonable lengths of 5' and 3' untranslated region sequences of each gene. The total sequenced lengths in different regions for each gene are summarized in Table 3.2.1. Figures 3.2.1 to 3.2.6 show the sequencing strategy used in each case as well as the gene structures for CcA1, CcA2, DdA1, DdA2, DdA3 and DdA5 respectively. Restriction maps of each subclone are also shown in the figures. The relative size of each restriction fragment is indicated in the figures as kilobases of nucleotide pairs (kb) by the scale shown.

For the CcA1 gene, restriction sites were used to obtain smaller overlapping subclones. In order to sequence across one of the EcoRV sites, a deletion subclone was isolated by using Exonuclease III and Mung bean nuclease. The pmedC1 clone was sequenced directly from both ends without further subcloning. This cDNA cone shows almost complete alignment with 3' untranslated region of CcA1 gene, except two nucleotides are switched from CG in CcA1 to GC in pmedC1. The four genes from *B. dorsalis* DdA1, DdA2, DdA3 and DdA5 and the CcA2 gene from *C. capitata* were sequenced by using Exonuclease III deletions. Some of them were sequenced from both directions, such as DdA1 and DdA2, while others were sequenced from only one direction. For DdA5, most sequence data was obtained from deletions in one direction, except one that was obtained from the other direction in a *Kpn*I-*Pst*I subclone in order to

Table 3.2.1. Sequenced Regions for Each Gene or Subclone.

Name	5'region (bp)	Coding region Nt/AA	3'region (bp)	Intron position	Intron size
pmedC1	...	....	416		
CcA1	625	1131/377	499	307th codon	59
CcA2	1488	588/196	...	5'UTR?	
DdA1	804	1131/377	395	307th codon	99
DdA2	2514	1131/377	1119	13th codon	984
DdA3	903	1131/377	1797	None	
DdA5	2396	1131/377	1952	307th codon	74

Figure 3.2.1: Sequencing Strategy and Gene Structure for the CcA1 gene.

The top schematic diagram represents the subclone containing the CcA1 gene and the restriction sites. Arrows indicate sequenced fragments. The size in kb is indicated by the scale at right. The diagram below represents the gene structure of the CcA1 gene. Exons are shown as solid boxes. Introns are shown as a solid line between exons. Also shown are the localization of a putative TATA box and polyadenylation signal (AATAAA). The bottom diagram represents the sequencing strategy for the pmedC1 cDNA subclone and the position corresponding to sequences in the CcA1 gene.

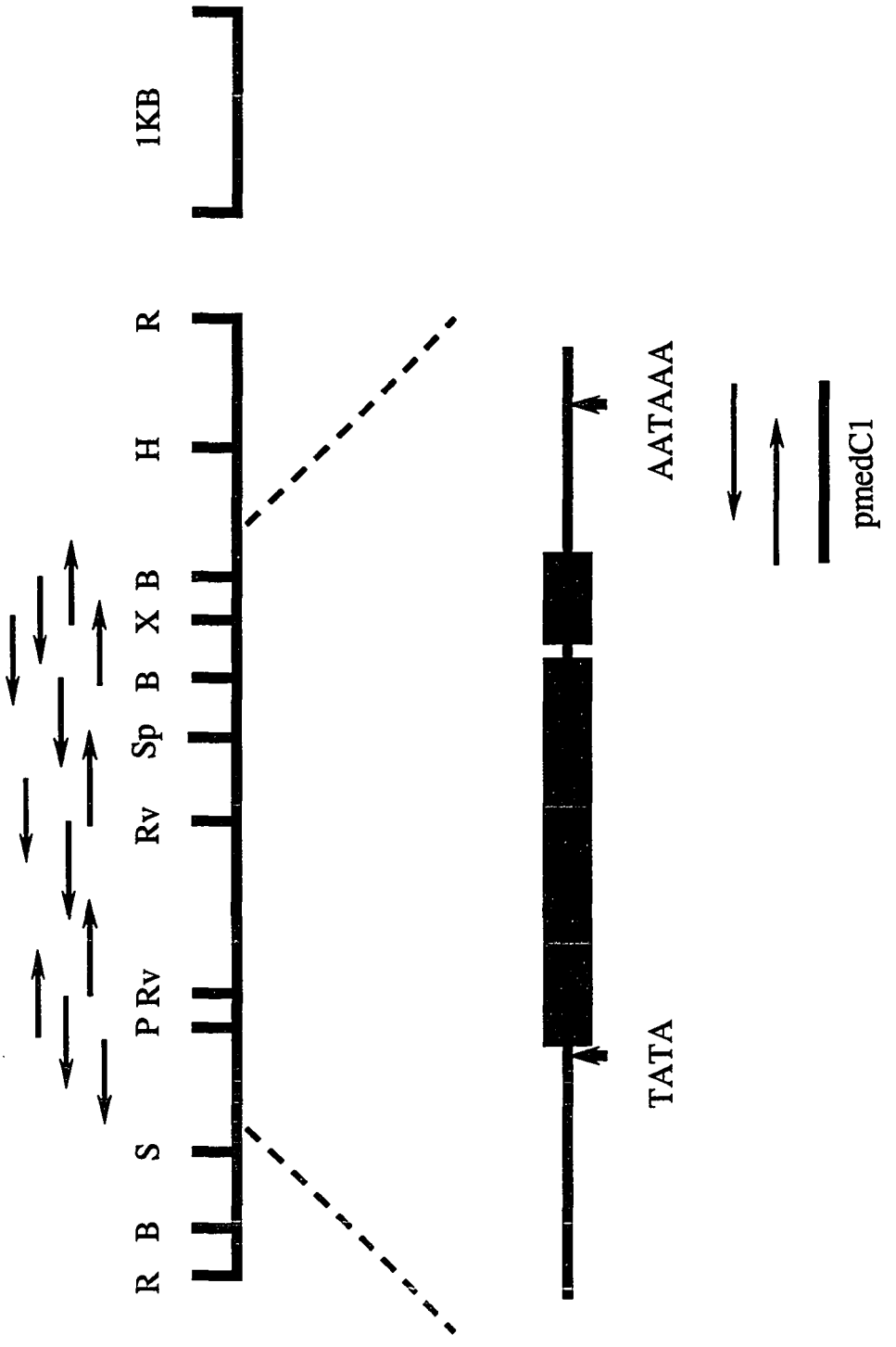

1KB
R B S P Rv Rv Sp B X B H R
TATA
AATAAA
pmedC1

Figure 3.2.2: Sequencing Strategy and Gene Structure for the CcA2 gene.

The top schematic diagram represents the subclone containing the CcA2 gene and the restriction sites. Arrows indicate sequenced fragments. The size in kb is indicated by the scale at right. The diagram below represents the gene structure of the CcA2 gene. Exons are shown as solid boxes. Also shown are the localization of two putative TATA boxes. The bottom diagram represents the subclone pmed(5') used as a gene specific probe and the position corresponding to sequences in the CcA2 gene.

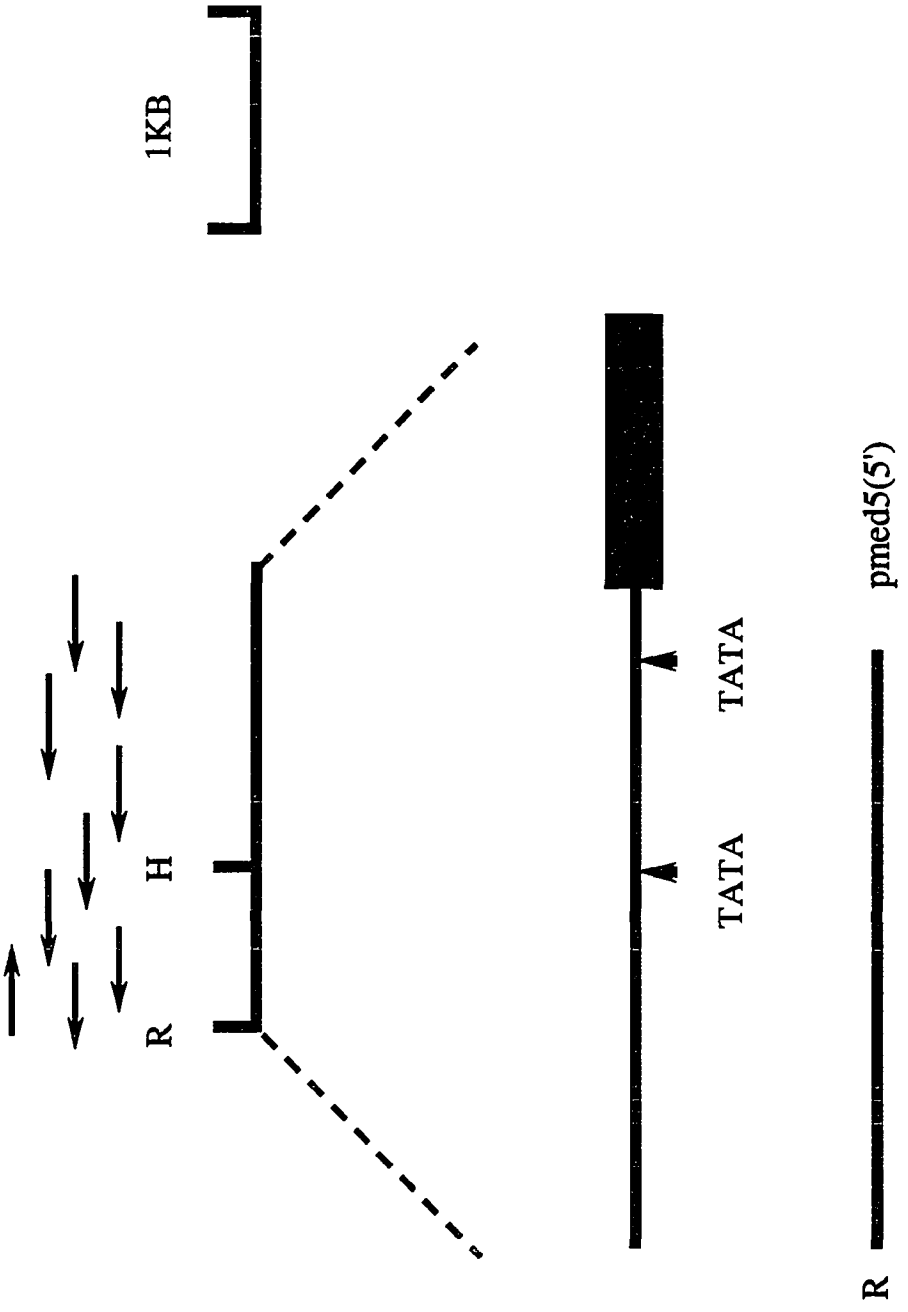

1KB
H
R
TATA
TATA
pmed5(5')
R

Figure 3.2.3: Sequencing Strategy and Gene Structure for the DdA1 gene.

The top schematic diagram represents the subclones containing the DdA1 gene and the restriction sites. Arrows indicate sequenced fragments. The size in kb is indicated by the scale at right. The diagram below represents the gene structure of the DdA1 gene. Exons are shown as solid boxes. Introns are shown as a solid line between exons. Also shown is the localization of a putative TATA box. The bottom diagram represents the subclone pDdA1(3') used as a gene specific probe and the position corresponding to sequences in the DdA1 gene.

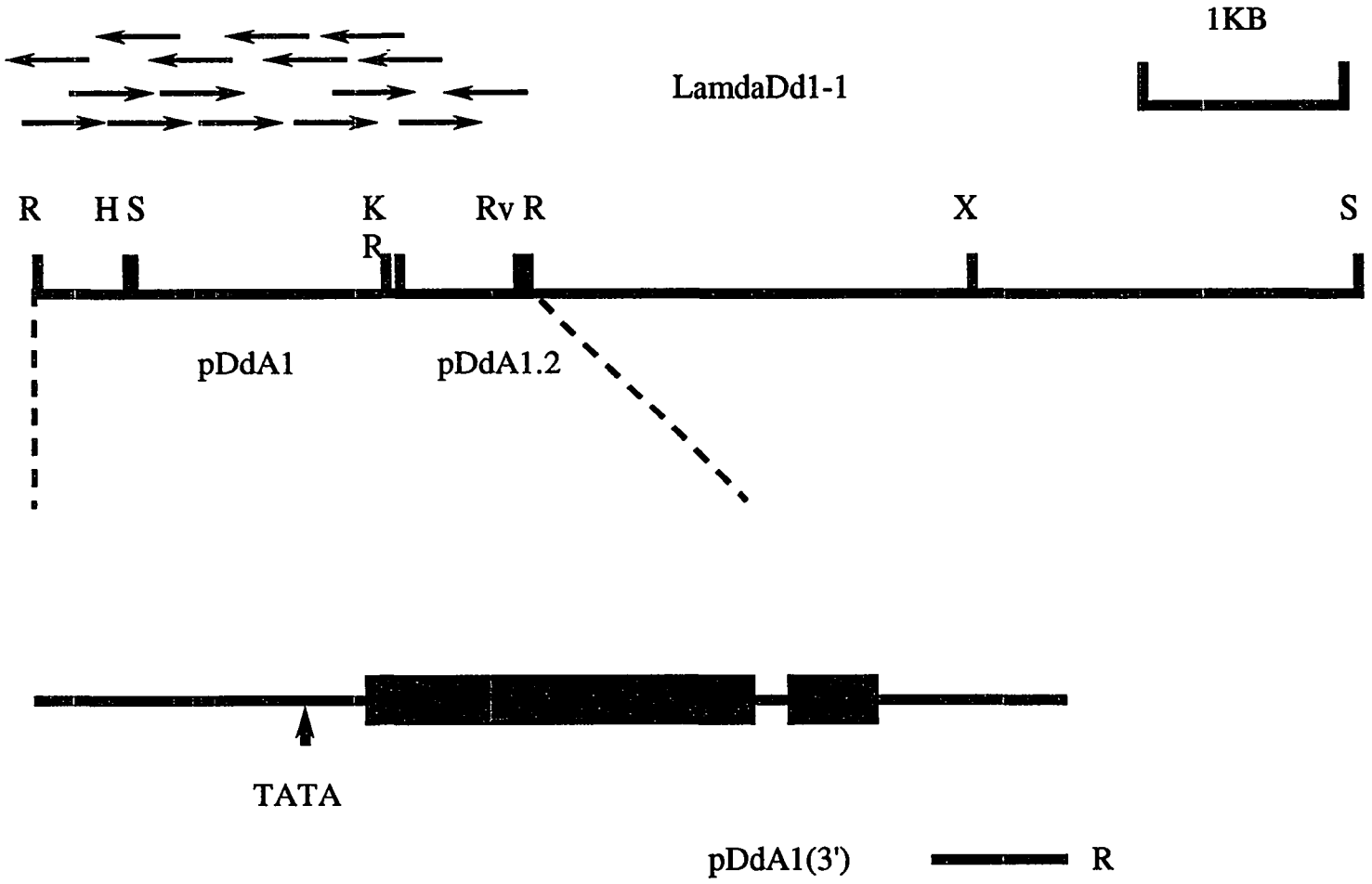

1KB
LamdaDd1-1
R
H S
K
R
Rv R
X
S
pDdA1
pDdA1.2
TATA
pDdA1(3')
R

Figure 3.2.4: Sequencing Strategy and Gene Structure for the DdA2 gene.

The top schematic diagram represents the subclones containing the DdA2 gene and the restriction sites. Arrows indicate sequenced fragments. The size in kb is indicated by the scale at up left side. The positions for PCR products of pDdTaqIPCR, pDdSau3APCR and pDdHincIIPCR are shown below. The diagram below represents the gene structure of the DdA2 gene. Exons are shown as solid boxes. Introns are shown as a solid line between exons. Also shown are the localization of a putative TATA box and polyadenylation signal (AATAAA). The bottom diagram represents the subclone pDdA2(3') used as a gene specific probe and the position corresponding to sequences in the DdA2 gene.

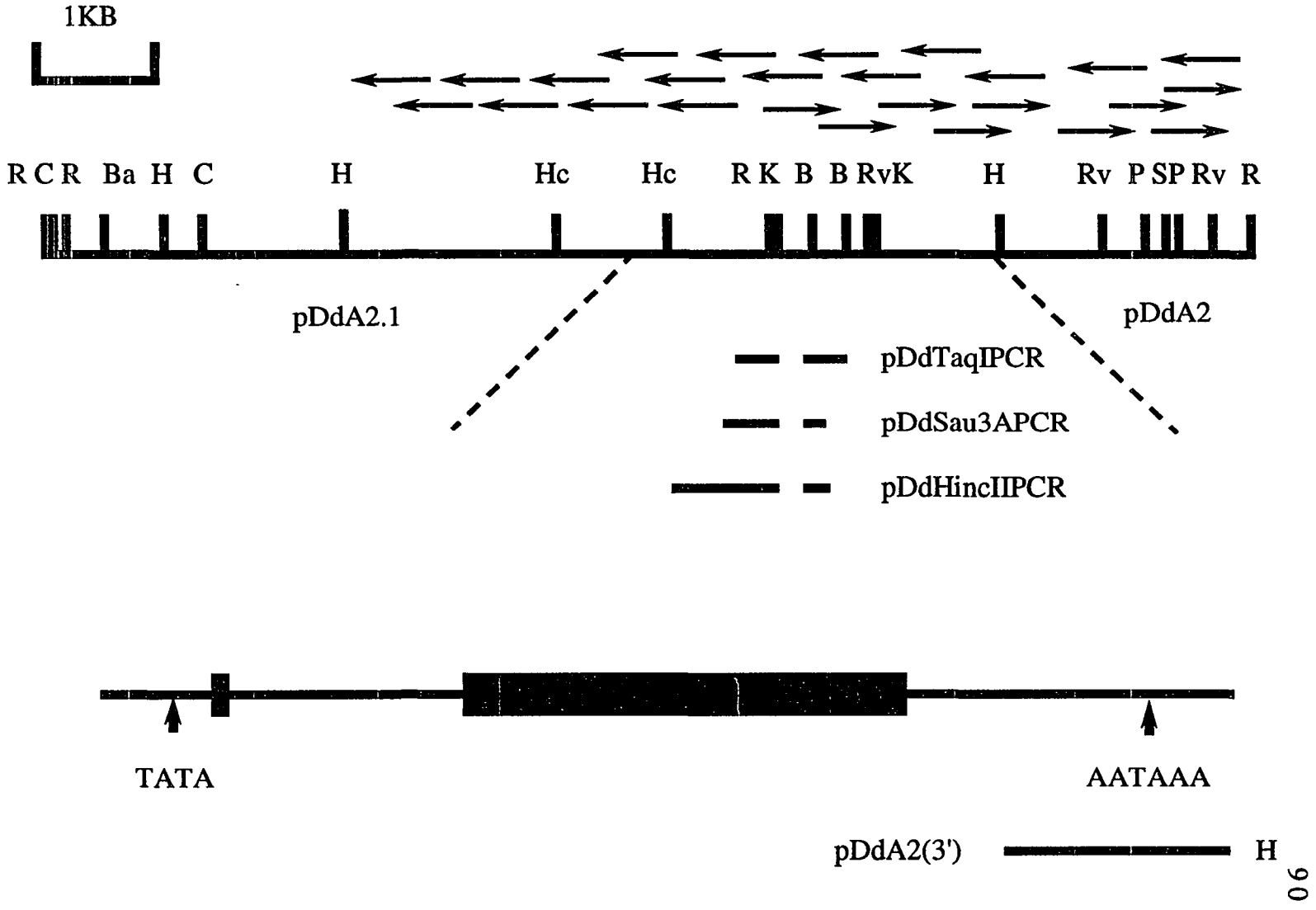

1KB
R C R Ba H C H Hc Hc R K B B RvK H Rv P SP Rv R
pDdA2.1
pDdA2
pDdTaqIPCR
pDdSau3APCR
pDdHincIIPCR
TATA
AATAAA
pDdA2(3')
H

Figure 3.2.5: Sequencing Strategy and Gene Structure for the DdA3 gene.

The top schematic diagram represents the subclone containing the DdA3 gene and the restriction sites. Arrows indicate sequenced fragments. The size in kb is indicated by the scale at right. The diagram below represents the gene structure of the DdA3 gene. Exons are shown as a solid box. Also shown are the localization of a putative TATA box and two polyadenylation signals (AATAAA). The bottom diagram represents the subclone pDdA3(3') used as a gene specific probe and the position corresponding to sequences in the DdA3 gene.

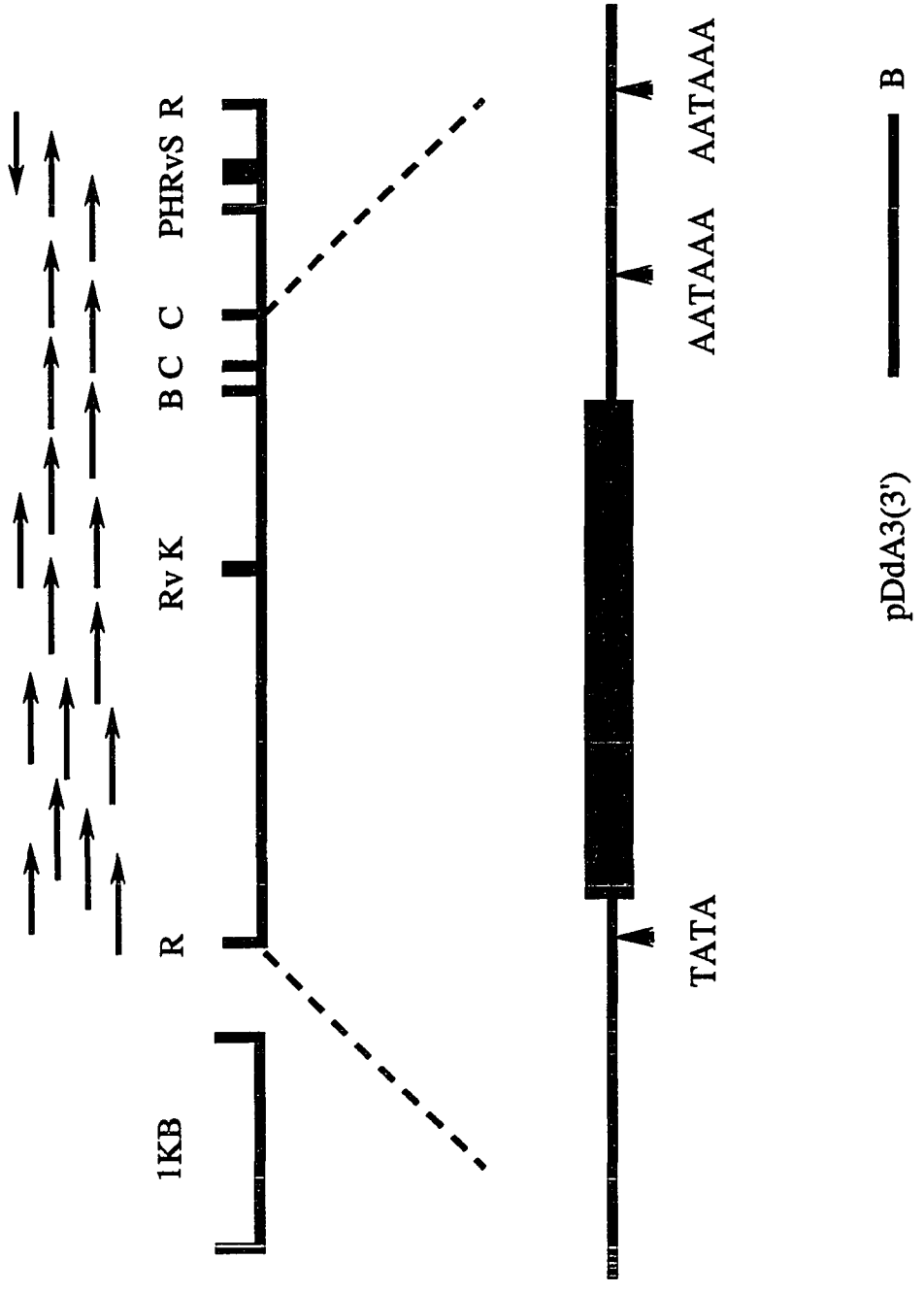
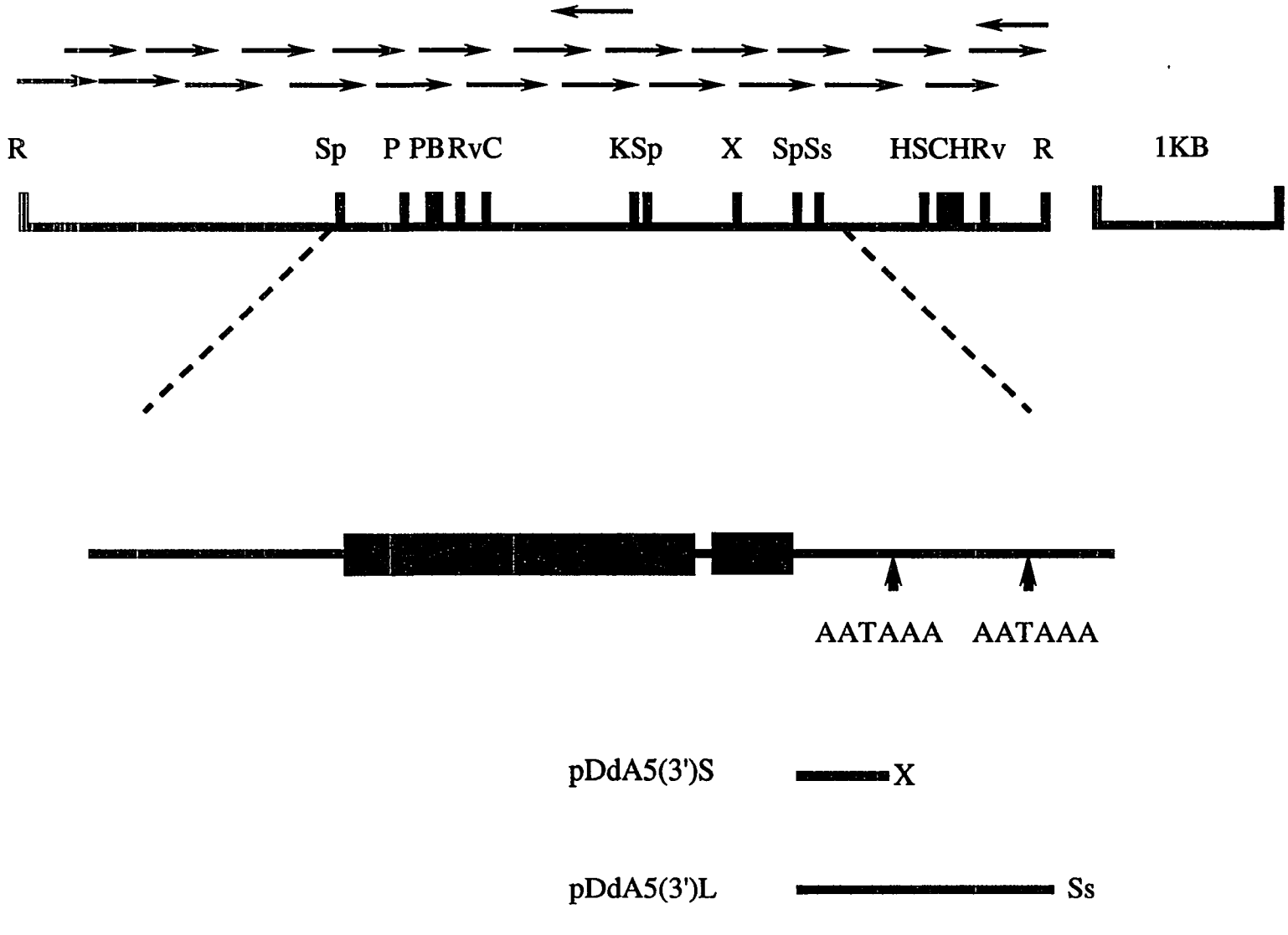

1KB
R
Rv K
B C C
PHRvS R
TATA
AATAAA
AATAAA
pDdA3(3')
B

Figure 3.2.6: Sequencing Strategy and Gene Structure for the DdA1 gene.

The top schematic diagram represents the subclone containing the DdA5 gene and the restriction sites. Arrows indicate sequenced fragments. The size in kb is indicated by the scale at right. The diagram below represents the gene structure of the DdA5 gene. Exons are shown as solid boxes. Introns are shown as a solid line between exons. Also shown are the localization of two putative polyadenylation signals (AATAAA). The bottom diagram represents the two subclones pDdA5(3')S and pDdA5(3')L used as gene specific probes and the position corresponding to sequences in the DdA5 gene.


R
Sp
P PB RvC
KSp
X
SpSs
HSCHRv
R
1KB
AATAAA
AATAAA
pDdA5(3')S
X
pDdA5(3')L
Ss

resolve a false stop at one point of the sequence. The figures also indicate gene structures, including exons (in solid box), introns (in open box), as well as potential TATA box(es) and polyA tail signaling sequences where they have been identified. The positions of subclones used as gene specific probes and the PCR products in DdA2 gene are also shown.

3.2.2 Characterization and Comparison of Actin Gene Sequences

3.2.2.1 Coding region sequence analysis

Figure 3.2.7 represents the alignment of the coding region sequences of the six actin genes from the two Tephritid fruit flies, using a computer program "Eyeball Sequence Editor" developed by Eric Cabot (1990 version 1.04). Five of them (CcA1, DdA1, DdA2, DdA3 and DdA5) have completed coding region sequences, but one (CcA2) has only partial sequences. The complete coding region sequences of all the genes here consists of total 1131 base pairs, including a stop codon. These encode actin proteins with 376 amino acids in length, which is the same as the six actin genes from *D. melanogaster* and three actin genes from *B. mori*. In fact, all the invertebrate actin proteins studied so far consist of a total of 376 amino acids. The derived amino acid sequences of all the six genes are presented in Figure 3.2.8. These sequences are aligned with the derived amino acid sequences of the six actin genes from *D. melanogaster* as well as one actin gene from yeast. The rabbit β-actin is used as a standard for the alignment. Overall, a total of 42 positions have amino acid changes comparing the three Dipteran species to the rabbit β-actin. Twenty of the positions have changes in all actins from the Dipteran species. Of these, 14 have the same amino acids, while 6 substitutions have different amino acids. The majority of the substitutions are found in only a small group of genes or in a single gene. For example, several changes are unique in DdA1 and act88F, such as the Aspartic acid residue at position 4, the

Threonine residue at position 6, the Isoleucine residue at position 10, the Phenylalanine residue at position 69, and the Serine residue at position 297.

The nucleotide sequences of the five completed actin genes from the two Tephritid species, together with the derived protein sequences, have been compared in a pairwise manner with previously published actin sequences of *D. melanogaster* using a computer program developed by Bailey Kessing (personal communication) (Table 3.2.2). The compared regions include the complete coding region DNA sequences of actin gene (1131bp) which encode a protein with 376 amino acid sequences and 1 stop codon. The protein sequences were compared by percentage of similarity (upper portion), while the nucleotide sequences were compared by percent of differences (not corrected for multiple hits). Overall, the actin proteins show very high degree of similarity in a range from 95.0% to 100% among these three species. The nucleotide sequences, however, show a higher degree of diversity in a range from 9.0% to 22.0%. Within a species, *D. melanogaster* displays a range of 95.2% to 99.2% amino acid similarity and 9.2% to 16.4% nucleotide differences, while *B. dorsalis* displays 97.3% to 99.5% amino acid similarity and 9.0% to 16.8% nucleotide differences. The CcA1 actin gene from *C. capitata* shows 95.0% to 98.4% amino acid similarity and 15.5% to 19.6% nucleotide differences compared to the genes from *D. melanogaster*, and 96.0% to 98.7% amino acid similarity and 14.0% to 15.8% nucleotide differences compared to the genes from *B. dorsalis*. All the five genes from the Tephritid species show 95.0% to 100% amino acid similarity and 11.0% to 22.0% nucleotide differences compared to the genes from *D. melanogaster*. From these data it can be seen that overall there is clearly greater divergence at the DNA level between these genes. Also at the DNA level, there is the tendency that more divergence is seen when the comparison is made between as opposed to within species. At the protein level, there is essentially no difference for these same comparisons.

Figure 3.2.7: Coding Region Sequence Alignment of the Six Tephritid Actin Genes.

The DNA sequence of CcA1 gene is used as a standard. Characters not changed are shown as periods. Changed characters are shown by individual bases. The name for each gene is indicated on the left side. Numbers of nucleotides are indicated on the right side.

```
CcA1    ATGTGTGATGAAGAAGCATCAGCTCTGGTCGTGGACAATGGGTCCGGCATGTGTAAGGCTGGCTTCGCCG  70
CcA2    ........C.......TTG.T..CT....T..T........T.....T.....C..A.....T.......
DdA1    ........C..T..T..GGGT..C.....TA.T...........T..T.....C.................
DdA2    ........C..T..G.TTG.T...T.......T..T.....C...........C.....C..T.......
DdA3    ........C..C....TTG.T........T..T........C...........C.....C..T.......
DdA5    ........C..T........T........G..C..T.....C.....A.....C.................

CcA1    GTGACGATGCCCCGCGTGCTGTATTTCCCTCGATAGTGGGCCGCCCACGTCATCAAGGTGTGATGGTGGG 140
CCA2    ....T..C..A..C..C.....C..C..A..A..T.....T............................T..
DdA1    ....T.....G..C.....C.....C..T.....T..T.....G.....C..C..G..............
DdA2    ....T.....G..C.....C..C..C.....A..T..A..T............................T..
DdA3    ....T.....G..C........C..C.....A..T.....T..............G...........C..
DdA5    ....T........T.....G.....C..A.....T.....T..T.....C.....G..............

CcA1    TATGGGTCAGAAGGACGCTTATGTTGGTGATGAGGCGCAAAGCAAACGAGGTATACTCACATTGAAATAT 210
CCA2    .........A.....TT.A.....C........A..A..G........T.....CT.A..CC.......C
DdA1    ............A..TT.C.....A..C...........G......A................A......
DdA2    ......A..A......T.A..C..C........A..T...TC...GA.......CT.G..CC.C..G..C
DdA3    ......A.........T.G.....C...........T...TC...GA.......C..............C
DdA5    ............A...T.G.....C..C...........G........T.....C.....CC....G..C

CcA1    CCCATCGAGCACGGCATCATCACCAACTGGGATGACATGGAGAAGGTGTGGCATCATACATTCTACAATG 280
CcA2    ..A..T........T..TG.G..............T.........AA.C.........C.............
DdA1    .....T.....T...........A..T.........T.........AA.C...........G..........
DdA2    .....T........T.....................T.....A...A.C........C..T..........
DdA3    .....T........T.....................T.........AA.C........C..T..........
DdA5    ......................G........C...........A...........C.....T.......

CcA1    AGTTGCGTGTGGCACCTGAGGAGCACCCAGTGCTACTCACTGAAGCGCCACTGAATCCAAAAGCCAATCG 350
CCA2    .A........C..C..C.....A.....T..TT..T.G..C.....A..TT.............T.....
DdA1    .A..................A.....T.....AT..T.G..C........T..T.....C...........
DdA2    .A........C..C..A..............CT.GT.G........C...T....C.....G..T..C..
DdA3    ..........C........A...........AT.G.....C.....A...T...................
DdA5    ............G........A........T..G..G..C.....A.....T.....C.....G.....

CcA1    CGAGAAGATGACCCAAATCATGTTTGAGACCTTCAACTCGCCCGCCATGTATGTTGCCATTCAGGCTGTG 420
CCA2    ............T..G........C..A..T......A.A..............G..............A
DdA1    T..A........T.................A.....T.................G................
DdA2    T.........................A...........C..A...........C.....C........C
DdA3    T.........................A.........A.A..............C...........C..A
DdA5    ..................T.....C..A..A........A.....A........C..............A

CcA1    CTCTCTCTGTATGCTTCGGGTCGTACCACCGGTATTGTGCTAGACTCCGGTGATGGTGTTTCCGACACAG 490
CCA2    T.A..A.....C.....T...........T...........G..T..T...........A...C.......
DdA1    .....C...........C.....................TT.G.....T........C......C....T.
DdA2    .....C.....C.....C.................C..CT.G..T..T...........C...C....C.
DdA3    .....A.....C..C..C.................C..CT.G.................C..AC....C.
DdA5    T.G..A..T.....C..T..C.....T..A..C.....AT.G.....T..C........C...C......

CcA1    TGCCCATTTATGAAGGTTATGCACTGCCACACGCCATCTTGCGTTTAGACTTGGCTGGTCGTGATTTGAC 560
CcA2    .......C..............C..............C.C.....G..............C.........
DdA1    .A.....C..........TC..CT.......T.....TC.....C.T........C.....C.........
DdA2    .A..A..C..C............CT....C..T......C.C.....G..T...........C.........
DdA3    .A.....C..............T.....C..T.....TC.C.....G..T.....C.....C.........
DdA5    .C.....C..............TT.A..T..T..T.........C.G........C..C..C.........

CcA1    CGACTATCTTATGAAGATTTTGACCGAGCGCGGTTATTCGTTCACCACCACCGCTGAGCGTGAAATCGTC 630
CCA2    ......C..A........CC
DdA1    .......T.G........CC.C..............C..A...........T..................G
DdA2    ......CT.G........C.....T..A..T.....C..T............................T...
DdA3    T..T..C..C.........C....T..A.....C..C..............A..............T..A
DdA5    ......C..C.....A...C.A..T..A..T..C..C..A..T..A.....T..................G
```

```
CcA1        CGTGATATCAAGGAAAAGTTGTGCTATGTCGCCTTGGACTTTGAGCAGGAGATGGCTACTGCTGCTGCCT 700
CcA2
DdA1        A.G.....T.....G..A..........T..T.................A..........C.....A.
DdA2        .....C........G..A..........T..T........C..A..A..A........C..C.......
DdA3        .....C..T.....G..A..........T..A.............A..A.....C.....C.......
DdA5        ..G..C........G.....A..T........A.................A.....A..A..C.....A.

CcA1        CCTCGTCGTTGGAGAAGTCGTACGAGTTGCCTGATGGTCAAGTGATTACCATTGGCAACGAACGTTTCCG 770
CcA2
DdA1        ..A.C..T..........A..T..A...................C........T..T..G........
DdA2        ..A.C..CC.........T..T..AC.T..C.............C........T..T..........
DdA3        ..A.C..A........A..C..T..A...................C........T..T..........
DdA5        .TA.C..A..........T..T.................G........T.....T..T..........

CcA1        CACACCCGAAGCCCTCTTCCAGCCCTCATTCCTGGGCATGGAATCATGCGGTATTCACGAGACTGTCTAC 840
CcA2
DdA1        TTGT..A..G..G..G..............T.............G..T..C.....T............
DdA2        TTGC..A...T..........A.........T....T........CA.......C.....A..C......
DdA3        .TGC......T.A........A.........T....T........GA.......C..T............
DdA5        ........................G......T.A..T........C........C.....A.........

CcA1        CAGTCAATCATGAAGTGCGACGTGGACATCCGTAAGGACTTGTACGCCAACAACGTACTGTCCGGCGGCA 910
CcA2
DdA1        A.T..G..............T..C..T..A........TC....T......TCG..CT....T..T..T.
DdA2        A.C..C..............T..A..T.........................TT..CA.......T..T.
DdA3        A.T..G..............T..T..T..................T.:.....TT..CA....T..T..T.
DdA5        ..A..C..............T..A........A.....T..................T....T..T..T.

CcA1        CCACCATGTACCCAGGTATTGCCGATCGCATGCAAAAGGAAATCACTGCCTTGGCTCCATCGACCATCAA 980
CcA2
DdA1        ..........T..................T.....................C.C..A..T...........
DdA2        ........................T.....T..................C..............C........
DdA3        .............C........T.....T............................G.....C.....T..
DdA5        ....T..............C..T.....................T..............T........A..

CcA1        GATTAAGATCATTGCACCACCTGAGCGTAAATACTCCGTATGGATTGGTGGTTCTATCTTGCGTTCGCTC1050
CcA2
DdA1        ..........................A.A...........C.....................GC...TT.G
DdA2        ...C...........C.....A..A.....................C........C......GC...C...
DdA3        ...C...........C.....A.................................C......GCC..A...
DdA5        ..............T...........C.....T.....T.....................GCC..T..G

CcA1        TCCACCTTCCAGCAGATGTGGATCTCTAAGCAGGAGTACGACGAATCCGGCCCTGGCATTGTGCACCGCA1120
CcA2
DdA1        .....T.....A..............C........A..T..T.....T..A...................
DdA2        ...........A..............C...G.A..A..........................C.......
DdA3        ...........A..............G..A.....A.....T...........A........C.......
DdA5        ..........................C.....................T..T..C..T..C..........

CcA1        AGTGCTTCTAA                                                        1131
CcA2
DdA1        .A.........
DdA2        ...........
DdA3        .A.........
DdA5        ...........
```

Figure 3.2.8: Protein Sequence Alignment of the Six Tephritid Actin Genes.

The protein sequence of a rabbit β-actin is used as a standard. Only changed characters are shown by individual amino acids. Also shown are the six actins from *D. melanogaster* (act79B, act88F, act87E, act57A, act5C and act42A) and one actin from yeast *S. cerevisiae*. Numbers of amino acids are indicated on the left side.

R	ASP	GLU	ASP	GLU	SER	THR	ALA	LEU	VAL	CYS	ASP	ASN	GLY	SER	GLY	LEU	VAL	LYS	ALA	GLY
1.CcA1	cys	asp	glu		ala	ser				val						met	cys			
2.Act79B	cys	asp	glu		ala	ser				val						met	cys			
3.DdA5	cys	asp			ala	ser				val						met	cys			
4 DdA1	cys	asp		asp	ala	gly				ile						met	cys			
5.Act88F	cys	asp		asp	ala	gly				ile						met	cys			
6.DdA2	cys	asp			val	ala				val						met	cys			
7.Act57A	cys	asp			val	ala				val						met	cys			
8.Act87E	cys	asp			val	ala				val						met	cys			
9.DdA3	cys	asp			val	ala				val						met	cys			
10.Act41A	cys	asp	glu		val	ala				val						met	cys			
11.Act5C	cys	asp	glu		val	ala				val						met	cys			
12.CcA2	cys	asp	glu		val	ala				val						met	cys			
13.Yeast	met	asp	ser		val	ala				ile						met	cys			

21	PHE	ALA	GLY	ASP	ASP	ALA	PRO	ARG	ALA	VAL	PHE	PRO	SER	ILE	VAL	GLY	ARG	PRO	ARG	HIS
1.21																				
2.21																				
3.21																				
4.21																				
5.21																				
6.21																				
7.21																				
8.21																				
9.21																				
10.21																				
11.21																				
12.21																				
13.21																				

41	GLN	GLY	VAL	MET	VAL	GLY	MET	GLY	GLN	LYS	ASP	SER	TYR	VAL	GLY	ASP	GLU	ALA	GLN	SER
1.41												ala								
2.41												cys								
3.41																				
4.41																				
5.41																				
6.41																				
7.41																				
8.41																				
9.41																				
10.41																				
11.41																				
12.41																				
13.41			ile																	

61	LYS	ARG	GLY	ILE	LEU	SER	LEU	LYS	TYR	PRO	ILE	GLU	HIS	GLY	ILE	ILE	THR	ASN	TRP	ASP
1.61						thr														
2.61																				
3.61						thr														
4.61						thr														
5.61						thr														
6.61						thr														
7.61						thr														
8.61						thr														
9.61						thr														
10.61						thr										val				
11.61						thr										val				
12.61						thr										val				
13.61						thr										val				

81	ASP	MET	GLU	LYS	ILE	TRP	HIS	HIS	THR	PHE	TYR	ASN	GLU	LEU	ARG	VAL	ALA	PRO	GLU	GLU
1.81					val															
2.81					val															
3.81					val															
4.81																				
5.81																				
6.81																				
7.81																				
8.81																				
9.81																				
10.81																				
11.81																				
12.81																				
13.81																				

Continue

101	HIS	PRO	THR	LEU	LEU	THR	GLU	ALA	PRO	LEU	ASN	PRO	LYS	ALA	ASN	ARG	GLU	LYS	MET	THR
1.101			val																	
2.101			val																	
3.101			val																	
4.101			val																	
5.101			val																	
6.101			val																	
7.101			val																	
8.101			val																	
9.101			val																	
10.101			val																	
11.101			val																	
12.101			val																	
13.101			val							met				ser						

121	GLN	ILE	MET	PHE	GLU	THR	PHE	ASN	VAL	PRO	ALA	MET	TYR	VAL	ALA	ILE	GLN	ALA	VAL	LEU
1.121									ser											
2.121									ser											
3.121									ser											
4.121									ser											
5.121									ser											
6.121									ser											
7.121									ser											
8.121									ala											
9.121									thr											
10.121									thr											
11.121									thr											
12.121									thr											
13.121												phe			ser					

141	SER	LEU	TYR	ALA	SER	GLY	ARG	THR	THR	GLY	ILE	VAL	LEU	ASP	SER	GLY	ASP	GLY	VAL	THR
1.141																				ser
2.141																				ser
3.141																				ser
4.141																				ser
5.141																				ser
6.141																				ser
7.141																				ser
8.141																				ser
9.141																				ser
10.141																				ser
11.141																				ser
12.121																				ser
13.141				ser																

161	HIS	ASN	VAL	PRO	ILE	TYR	GLU	GLY	TYR	ALA	LEU	PRO	HIS	ALA	ILE	MET	ARG	LEU	ASP	LEU
1.161	asp	thr														leu				
2.161		thr														leu				
3.161		thr														leu				
4.161		thr							phe							leu				
5.161		thr							phe							leu				
6.161		thr														leu				
7.161		thr														leu				
8.161		thr														leu				
9.161		thr														leu				
10.161		thr														leu				
11.161		thr														leu				
12.161		thr														leu				
13.161		val					ala			phe	ser					leu				

181	ALA	GLY	ARG	ASP	LEU	THR	ASP	TYR	LEU	MET	LYS	ILE	LEU	THR	GLU	ARG	GLY	TYR	SER	PHE
1.181																				
2.181																				
3.181																				
4.181																				
5.181																				
6.181																				
7.181																				
8.181																				
9.181																				
10.181																				
11.181																				
12.181													end							
13.181														ser						

Continue

201	VAL	THR	THR	ALA	GLU	ARG	GLU	ILE	VAL	ARG	ASP	ILE	LYS	GLU	LYS	LEU	CYS	TYR	VAL	ALA
1.201	thr																			
2.201	thr																			
3.201	thr																			
4.201	thr																			
5.201	thr																			
6.201	thr																			
7.201	thr																			
8.201	thr																			
9.201	thr																			
10.201	thr																			
11.201	thr																			
13.201	ser																			

211	LEU	ASP	PHE	GLU	ASN	GLU	MET	ALA	THR	ALA	ALA	SER	SER	SER	SER	LEU	GLU	LYS	SER	TYR
1.221					gln							ala								
2.221					gln							ala		thr						
3.221					gln							ala		thr						
4.221					gln							ala		thr						
5.221					gln							ala		thr						
6.221					gln							ala		thr						
7.221					gln							ala		thr						
8.221					gln							ala		thr						
9.221					gln							ala		thr						
10.221					gln															
11.221					gln															
13.221					gln			gln				gln				ile				

241	GLU	LEU	PRO	ASP	GLY	GLN	VAL	ILE	THR	ILE	GLY	ASN	GLU	ARG	PHE	ARG	CYS	PRO	GLU	THR
1.241																	thr			ala
2.241																	thr			ala
3.241																	thr			ala
4.241																				ala
5.241																				ala
6.241																				ser
7.241																				ser
8.241																				ser
9.241																				ser
10.241																				ser
11.241																				ala
13.241																	ala			ala

261	LEU	PHE	GLN	PRO	SER	PHE	ILE	GLY	MET	GLU	SER	ALA	GLY	ILE	HIS	GLU	THR	THR	TYR	ASN
1.261							leu					cys						val		gln
2.261							leu					cys						val		gln
3.261							leu					cys						val		gln
4.261							leu					cys						val		
5.261							leu					cys						val		
6.261							leu					ser						val		
7.261							leu					cys						val		
8.261							leu					cys						val		
9.261							leu					ser						val		
10.261							leu				ala	cys		leu						
11.261			his				leu					cys								
13.261			his			val	leu		leu						asp	gln				

281	SER	ILE	MET	LYS	CYS	ASP	ILE	ASP	ILE	ARG	LYS	ASP	LEU	TYR	ALA	ASN	ASN	VAL	MET	SER
1.281							val												leu	
2.281							val												leu	
3.281							val												leu	
4.281							val										ser		leu	
5.281							val										ser		leu	
6.281							val										ile			
7.281							val										ile			
8.281							val										ile			
9.281							val										ile			
10.281							val										thr		leu	
11.281							val										thr		leu	
13.281							val		val			glu			gly		ile			

Continue

301	GLY	GLY	THR	THR	MET	TYR	PRO	GLY	ILE	ALA	ASP	ARG	MET	GLN	LYS	GLU	ILE	THR	ALA	LEU
1.301																				
2.301																				
3.301																				
4.301																				
5.301																				
6.301																				
7.301																			ser	
8.301																				
9.301																				
10.301																				
11.301																				
13.301						phe		ser			glu									

321	ALA	PRO	SER	THR	MET	LYS	ILE	LYS	ILE	ILE	ALA	PRO	PRO	GLU	ARG	LYS	TYR	SER	VAL	TRP
1.321					ile															
2.321																				
3.321					ile															
4.321					ile															
5.321					ile															
6.321					ile															
7.321					ile															
8.321					ile															
9.321					ile															
10.321																				
11.321																				
13.321				ser			val										val			

341	ILE	GLY	GLY	SER	ILE	LEU	ALA	SER	LEU	SER	THR	PHE	GLN	GLN	MET	TRP	ILE	THR	LYS	GLN
1.341							arg											ser		
2.341																		ser		
3.341																		ser		
4.341																		ser		
5.341																		ser		
6.341																		ser		glu
7.341																		ser		glu
8.341																		ser		
9.341																		ser		
10.341																		ser		
11.341									ser								thr	ser		
13.341										thr								ser		

361	GLU	TYR	ASP	GLU	ALA	GLY	PRO	SER	ILE	VAL	HIS	ARG	LYS	CYS	PHE
1.361					ser			gly							
2.361					ser			gly							
3.361					ser			gly							
4.361					ser			gly							
5.361					ser			gly							
6.361					ser			gly							
7.361					ser			gly							
8.361					ser			gly							
9.361					ser			gly							
9.361					ser										
10.361					ser										
11.361					ser										
13.361					ser										

Table 3.2.2. Coding Region Sequence Pairwise Comparisons. Upper Portion: Amino Acid Sequence Comparisons in % Similarity; Bottom Portion: Nucleotide Substitutions in % Diversity.

	CcA1	DdA5	79B	DdA1	88F	DdA3	87E	DdA2	57A	5C	42A
CcA1	--	98.7	98.4	96.6	96.6	96.0	96.3	96.0	96.0	95.2	95.0
DdA5	14.0	--	98.9	97.9	97.9	97.3	97.6	97.3	97.3	95.5	95.2
Act79B	17.2	18.6	--	96.8	96.8	96.3	96.6	96.3	96.3	95.5	95.2
DdA1	14.7	15.7	19.7	--	100.0	97.3	97.6	97.3	97.3	95.5	95.2
Act88F	16.4	18.4	11.2	16.4	--	97.3	97.6	97.3	97.3	95.5	95.2
DdA3	14.1	14.8	18.7	13.8	16.3	--	99.5	99.5	98.9	96.3	96.6
Act87E	16.2	17.3	12.5	17.2	11.4	14.9	--	99.2	99.2	96.3	96.6
DdA2	15.8	16.8	17.5	16.0	17.1	9.0	14.2	--	99.5	95.8	96.0
Act57A	15.5	17.2	12.2	17.6	11.6	13.1	9.2	11.0	--	95.8	96.0
Act5C	16.7	18.7	12.5	18.4	12.6	16.3	10.1	15.9	10.0	--	98.1
Act42A	19.6	20.4	15.2	22.0	16.0	19.5	16.4	18.4	16.0	15.0	--

Based on the aligned protein sequences (Figure 3.2.8) and the overall percentage of similarity at protein level (Table 3.2.2), it is clear that all four actin genes (DdA1, DdA2, DdA3, DdA5) from *B. dorsalis* and one (CcA1) from *C. capitata* show more homology to the muscle specific actins of *D. melanogaster* including act79B, act88F, act87E and act57A than to the cytoplasmic actins act5C and 42A. Previous work has indicated that CcA1 represents a muscle specific actin (Haymer et al., 1990). The CcA1 gene shows the highest similarity score with the act79B gene (98.4%); the DdA1 gene with the act88F gene (100%); the DdA2 gene with the act57A gene (99.5%); the DdA3 gene with the act87E gene (99.5%); and the DdA5 gene with the act79B gene (98.9%). Between the two Tephritid species, CcA1 shows the highest similarity score with DdA5 (98.7%). The genes with a high degree of similarity also share amino acid substitutions (Figure 3.2.8). The CcA2 gene, although only a partial sequence is known, appears to be more closely related with the two cytoplasmic actin genes (act5C and act42A) from *D. melanogaster*. Actin genes with higher similarity may imply that they are orthologous genes and perform similar functions in their respective organisms.

The nucleotide sequence comparisons between genes with the highest similarities at the protein level, however, are not always consistent in terms of lowest diversity of nucleotide substitutions. For example, the difference at the DNA level between the DdA1 gene and act88F is 16.4%, even though these two genes have 100% similarity in their protein sequences. This apparent inconsistency suggests the occurrence of a high level of silent substitutions and selective constraints at the protein level. For example, as shown in Table 3.2.3, in a comparison of CcA1 with act79B of *D. melanogaster*, only 5.7% of the nucleotide substitutions are missense mutations. The remaining 94.3% of the substitutions are silent mutations. Of these, 86.9% occur at the third codon position, 6% occur at the first codon position, 2.7% occur at both first and third positions, and 0.5% occur at all three positions. The same is true for comparisons between all the actin

genes described here, i.e. the great majority of the nucleotide changes seen are synonymous substitutions (data not shown).

Table 3.2.3. The Nucleotide Substitutions Between CcA1 and Act79B

Codon position	Silent mutation	AA substitution
1	11	1
2	0	0
3	159	1
1+2	0	1
1+3	5	2
1+2+3	1	1
Total	183	11
Percentage	94.3%	5.7%

Another aspect of the silent substitutions reported here is reflected in the preferences for synonymous codons of the different genes. For those amino acids with three or more potential codons, Table 3.2.4 shows codon usage for the five actin genes from Tephritid species compared to that of the *Drosophila* genes. The most dramatic difference between genes from Tephritid species and *Drosophila* genes is the preference for T in Tephritid genes versus C in *Drosophila* genes, particularly where silent substitutions have occurred. For example, among the silent substitutions in Table 3.2.3, approximately 45% and 38% respectively of the total substitutions are caused by changes from T in CcA1 to C in act79B and act88F. Along this same line it appears that the *Drosophila* actin genes tend to have a higher GC content than the Tephritid genes. The GC contents of the five Tephritid actin genes are between 49% to 52%, while the GC contents of the six *Drosophila* genes are between 55% to 61%. The overall ratio of transitions to transversions between these genes is summarized in Table 3.2.5. In general, it appears that the ratio is higher (about 2.0) when the comparison is made between Tephritid genes and *Drosophila* genes as opposed to the comparison made among Tephritid genes or among *Drosophila* genes (less than 2.0).

Table 3.2.4. Codon Usage Comparison.

Condon		CcA1	DdA1	DdA2	DdA3	DdA5	79B	88F	87E	57A	5c	42A
Phe	TTT	0.25	0.16	0.08	0.17	0.25	0.08	0.08	0.08	0.08	0.17	0.25
	TTC	0.75	0.85	0.92	0.83	0.75	0.92	0.92	0.92	0.92	0.83	0.75
Tyr	TAT	0.56	0.73	0.25	0.44	0.50	0.25	0.13	0.31	0.19	0.25	0.19
	TAC	0.44	0.27	0.75	0.56	0.50	0.75	0.87	0.69	0.81	0.75	0.81
Cys	TGT	0.33	0.43	0.17	0.17	0.33	0.14	0.14	0.14	0.14	0.29	0.43
	TGC	0.67	0.57	0.83	0.83	0.67	0.86	0.86	0.86	0.86	0.71	0.57
His	CAT	0.38	0.67	0.33	0.44	0.33	0.00	0.22	0.11	0.11	0.20	0.22
	CAC	0.63	0.33	0.67	0.56	0.67	1.00	0.78	0.89	0.89	0.80	0.78
Gln	CAA	0.39	0.33	0.82	0.58	0.23	0.23	0.08	0.00	0.09	0.00	0.33
	CAG	0.62	0.67	0.18	0.42	0.77	0.77	0.92	1.00	0.92	1.00	0.67
Asn	AAT	0.44	0.89	0.33	0.67	0.56	0.11	0.11	0.22	0.11	0.11	0.22
	AAc	0.56	0.11	0.67	0.33	0.44	0.89	0.89	0.78	0.89	0.89	0.78
Lys	AAA	0.21	0.42	0.11	0.42	0.32	0.00	0.11	0.16	0.11	0.00	0.11
	AAG	0.71	0.58	0.81	0.58	0.68	1.00	0.90	0.84	0.90	1.00	0.81
Asp	GAT	0.41	0.74	0.68	0.36	0.50	0.24	0.30	0.41	0.50	0.43	0.29
	GAC	0.59	0.26	0.32	0.64	0.50	0.76	0.70	0.59	0.50	0.57	0.71
Glu	GAA	0.39	0.46	0.71	0.56	0.52	0.14	0.00	0.15	0.11	0.11	0.25
	GAG	0.61	0.54	0.29	0.44	0.48	0.86	1.00	0.85	0.89	0.89	0.75
Ile	ATT	0.43	0.47	0.27	0.47	0.39	0.04	0.20	0.13	0.10	0.15	0.32
	ATC	0.50	0.47	0.73	0.53	0.57	0.93	0.80	0.87	0.90	0.85	0.68
	ATA	0.07	0.07	0.00	0.00	0.04	0.04	0.00	0.00	0.00	0.00	0.00
Pro	CCT	0.21	0.37	0.05	0.11	0.32	0.00	0.11	0.16	0.11	0.05	0.11
	CCC	0.32	0.32	0.32	0.47	0.32	0.63	0.63	0.63	0.74	0.68	0.37
	CCA	0.42	0.32	0.63	0.42	0.32	o.16	0.21	0.11	0.16	0.11	0.11
	CCG	0.05	0.00	0.00	0.00	0.05	0.21	0.05	0.11	0.00	0.16	0.42
Thr	ACT	0.17	0.30	0.13	0.25	0.29	0.09	0.04	0.04	0.04	0.00	0.32
	ACC	0.65	0.52	0.87	0.63	0.38	0.87	0.83	0.91	0.96	0.92	0.56
	ACA	0.17	0.13	0.00	0.13	0.39	0.00	0.00	0.04	0.00	0.04	0.04
	ACG	0.00	0.04	0.00	0.00	0.04	0.04	0.13	0.00	0.00	0.04	0.08

Table 3.2.4. Continued

Condon		CcA1	DdA1	DdA2	DdA3	DdA5	79B	88B	87E	57A	5c	42A
Val	GTT	0.13	0.24	0.17	0.22	0.09	0.04	0.10	0.17	0.17	0.17	0.25
	GTC	0.17	0.19	0.57	0.44	0.30	0.35	0.19	0.39	0.61	0.26	0.29
	GTA	0.13	0.19	0.17	0.22	0.22	0.13	0.05	0.04	0.00	0.00	0.08
	GTG	0.57	0.38	0.09	0.13	0.39	0.48	0.67	0.39	0.22	0.57	0.38
Ala	GCT	0.40	0.27	0.48	0.28	0.33	0.10	0.20	0.13	0.39	0.31	0.31
	GCC	0.40	0.43	0.48	0.55	0.33	0.77	0.67	0.70	0.57	0.59	0.48
	GCA	0.13	0.13	0.00	0.10	0.20	0.07	0.03	0.10	0.04	0.07	0.14
	GCG	0.07	0.17	0.03	0.07	0.13	0.07	0.10	0.07	0.00	0.03	0.07
Gly	GGT	0.62	0.67	0.88	0.79	0.66	0.35	0.47	0.52	0.59	0.43	0.25
	GGC	0.35	0.27	0.14	0.17	0.31	0.55	0.47	0.24	0.38	0.39	0.46
	GGA	0.00	0.03	0.03	0.03	0.03	0.07	0.07	0.24	0.03	0.14	0.25
	GGG	0.03	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.04	0.04
Ser	AGT	0.00	0.00	0.00	0.0	0.00	0.00	0.04	0.00	0.00	0.00	0.00
	AGC	0.04	0.04	0.04	0.04	0.04	0.08	0.04	0.04	0.00	0.11	0.12
	TCT	0.12	0.28	0.12	0.04	0.36	0.08	0.04	0.04	0.08	0.15	0.19
	TCC	0.31	0.32	0.73	0.44	0.28	0.64	0.40	0.58	0.77	0.41	0.35
	TCA	0.15	0.12	0.12	0.28	0.20	0.04	0.04	0.08	0.04	0.00	0.08
	TCG	0.35	0.24	0.00	0.20	0.12	0.16	0.44	0.25	0.12	0.33	0.27
Leu	TTA	0.02	0.07	0.00	0.00	0.11	0.00	0.07	0.00	0.00	0.00	0.07
	TTG	0.31	0.54	0.70	0.59	0.46	0.07	0.11	0.11	0.07	0.07	0.24
	CTT	0.02	0.07	0.04	0.00	0.07	0.07	0.00	0.04	0.07	0.04	0.21
	CTC	0.12	0.14	0.19	0.26	0.11	0.11	0.04	0.11	0.11	0.15	0.10
	CTA	0.05	0.00	0.00	0.00	0.04	0.11	0.00	0.00	0.04	0.00	0.07
	CTG	0.14	0.18	0.07	0.15	0.21	0.64	0.79	0.74	0.70	0.74	0.31
Arg	AGA	0.00	0.11	0.06	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	AGG	0.00	0.06	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
	CGT	0.63	0.56	0.78	0.67	0.50	0.33	0.33	0.50	0.67	0.67	0.50
	CGC	0.32	0.22	0.17	0.28	0.39	0.67	0.56	0.50	0.33	0.33	0.44
	CGA	0.05	0.00	0.00	0.00	0.06	0.00	0.06	0.00	0.00	0.00	0.06
	CGG	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00

Table 3.2.5. Overall Comparison of Transition and Transversion Ratios.

TS/TV	CcA1	DdA1	DdA2	DdA3	DdA5	79B	88F	87E	57A	5C	42A
CcA1	-										
DdA1	1.6										
DdA2	1.6	1.6									
DdA3	1.7	1.5	2.1								
DdA5	1.9	1.7	1.6	1.4							
Act79B	2.1	2.3	2.6	2.0	2.4						
Act88F	2.0	2.9	2.2	2.1	2.0	1.3					
Act87E	1.5	2.0	2.4	2.5	1.5	1.4	1.3				
Act57A	1.5	1.9	3.1	2.3	2.0	1.3	1.4	1.8			
Act5C	1.7	1.8	1.9	2.1	2.1	1.4	1.5	1.2	1.2		
Act42A	1.5	1.8	1.7	1.8	1.6	1.5	1.4	1.4	1.5	1.3	-

3.2.2.2 Intron sequences and positions

The intron and its position are determined by the canonical intron-exon junction sequence, and the conservation of amono acid sequence in the protein. Based on this, a single intron is found in the CcA1, DdA1, DdA2 and DdA5 genes. No intron is found in the DdA3 gene. It is not certain if there is any intron within the coding region of the CcA2 gene, because only part of this gene has been sequenced.

DdA1, DdA5 and CcA1 share a common intron position. The splice point occurs within the 307th codon (glycine). The same intron position is also present in two of the *D. melanogaster* actin genes act79B and act88F. The length of the intron is 59bp in CcA1, 99bp in DdA1, and 74bp in DdA5. The intron position in DdA2 is located within

the 13th codon (also glycine) with 984bp in length. The same intron position is also found in one of the *D. melanogaster* actin genes act57A. Figure 3.2.9 shows the intron-exon junctions and intron sequences in these actin genes from the two Tephritid species. The consensus junction sequences as well as the junction sequences of three actin genes from *D. melanogaster* are also shown in the figure.

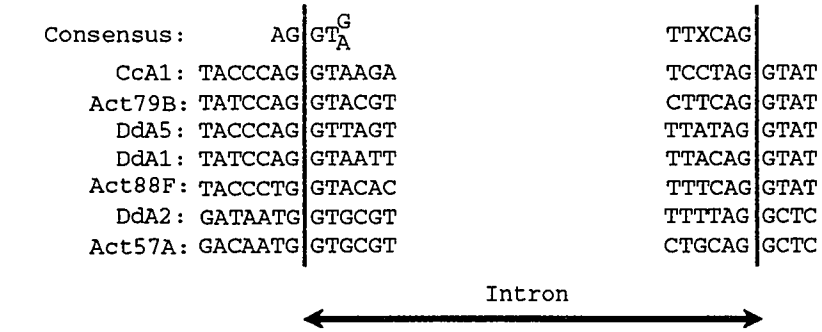
It is possible that there is a leader intron in the 5'UTR of this gene. The 5' leader intron has been found in many actin genes. For example, all six actin genes from mammals and birds have an intron in 5'UTR regions (Ueyama et al., 1984; Hamada et al., 1982; Zakut et al., 1982; Hu et al, 1986; Fornwald et al., 1982; Chang et al., 1985; Mohun et al., 1986; Nakajima-Iijima et al., Nudel et al., 1982; Kost et al., 1983; Miwa et al., 1991). The leader intron may also be conserved in plant actins (McElroy et al., 1990b, c; Pearson and Meagher, 1990; McKnight, 1983). In *D. melanogaster*, one of the actin genes act5C also has a 5' leader intron (Fyrberg, 1981). This will be discussed in the next section.

Because of the high variability in the intron positions of actin genes, especially among invertebrate species, the common intron positions shared by these three Dipteran species are of interest. Table 3.2.6 lists all the intron positions found so far from actin genes of a variety of organisms (Mounier and Prudhomme, 1986; Davidson, 1982; Krause and Hirsh, 1986; Dibb and Newman, 1989; Bhattacharya et al., 1991). The codon numbering is based on the β-cytoplasmic actin from mammals, which has 375 amino acids. The extra amino acids in the actins with 376 or 377 are numbered as -1 and -2. Actin genes from two species *A. nidulans* and *T. lanuginisus* also have an intron within the 13th codon, but these introns disrupt at the second position of the 13th codon rather than at the first position as in the case of the DdA2 and act57A genes.

Figure 3.2.9: Intron-Exon Junction Sequences and Intron Sequences for Four Tephritid Actin Genes.

(A). Alignment of intron-exon junctions of four Tephritid actin genes and three Drosophila actin genes (act79B, act88F and ac57A) with consensus sequence. (B). Sequences of the intron in the CcA1, DdA1, DdA2 and DdA5 actin genes.

A. Intron-Exon Junctions



B. Intron Sequences

CcA1:

GTAAGAGAAGCCAATTTCATACAATTACAGGGGCTTAATTCCTTTACTTACGTTCCTAG

DdA1:

GTAATTTAATTGTTTCCACAAAAAAGCCAGAAATAATGAATAAAGCCTTTGCGCTGTTTTCGATGTATTCAATG
AATTCTTTTTTTATATCTTTACAG

DdA2:

GTGCGTTAACATTGGATATTAGCTGTGGATATTGTTGTTGGTGGGCTGGATGAAACCCCAATCCACCAGCAGCAA
AGCACATTTGGATTACAGTTTGAAGGATTTACGGCGGAAGGAGGCGAAGATACGAGAATGGTGTAGTGAAAAGTG
TACAACACAAACATAGAATTGAAACGAAAAGGATAACAAAAGAGAGAAAAATNAAAAATAGTTTATTAAATCTCA
AGTGATTAAGGACAAGAGAGATACAGAAATATCTAGTTTCAAGAAATAGGGNGGTTGGCATTGGCCAGNATTGCG
TAGTGTGTGCCGTGAGACNCCCCTCAATGGGATCANGTAATATTTCATTCGAACTGTACCAGTTCCGTAATCTGA
AATAAAATGTGGTAAACATACCAAAAAGATATAGCAAACGCTGAAGGGCAAATGAGGAAAGGAAAAGAATCTCAT
TAAGTGATAGCAGAAGTTTCTAAATAATTTGGCTAATAATGAGAAATATATTTTATATTAAATTCAAAATTAAAT
TCAGCAGAAAGACTAAATGTTTTGAAAACTAAAAAAAAATACTCGAAAGATTTTGGAAATTTTTGTTACTAAAAA
TATCAGCACTTGGCAAAAATAGACGACACACTTTTCAATGAGATTAATTTTAGAACAATCATTTATTAAAAAAAA
GTAGGCCAATCAGGTTTCTAATTATATTCCTCAAACTGTTTTTTTTGCGGAAAGCAATTTTTTCAAAAATCAGGTT
TTTTGAGGTTAGCTGTTTCTAAGACGGGCCGCCAGATTTTGCGCCATTTTTCACNATACCCTCTCCGGAATTCGC
ACAATTTTGCGGTACCACATATCAAATGGCGCTGCGTCACTCACTTTTACCGCACGTTTCTTCGTTTCCCGCCGT
CCGTTTTTCCTACGACACATAAAGACCAGCGAAAGCATGGTGTATCACTTACTTACGTACGGTTTGTATTTTTTT
ATTTTTTAG

DdA5:

GTTAGTTTAGCACATCTGTAGCTGTTTCTGTTTTTTCTGTAGAAAAGAATTAAACATATTTGTACGTGTTATAG

Table 3.2.6. Intron Positions in Actin Genes from Various Sources.

Intron positions / Actin genes	5'UT	3	4	13	19 / 20	25	32	41 / 42	45	64	84 / 85	93	105 / 106	116	121 / 122	150	204	213 / 214	263	264 / 265	267 / 268	298 / 299	307	319 / 320	324	327 / 328	355 / 356	365 / 366
α-skeletal	+	.	.	.	.	.	.	+	.	.	.	.	.	.	.	+	+	.	.	.	+	.	.	.	.	+	.	.
α–cardiac	+	.	.	.	.	.	.	+	.	.	.	.	.	.	.	+	+	.	.	.	+	.	.	.	.	+	.	.
α–aortic	+	.	.	.	.	.	.	+	.	.	+	.	.	.	+	+	+	.	.	.	+	.	.	.	.	+	.	.
γ-enteric	+	.	.	.	.	.	.	+	.	.	+	.	.	.	+	+	+	.	.	.	+	.	.	.	.	+	.	.
β-cytoplasmic	+	.	.	.	.	.	.	+	.	.	.	.	.	.	+	.	.	.	.	.	+	.	.	.	.	+	.	.
γ-cytoplasmic	+	.	.	.	.	.	.	+	.	.	.	.	.	.	+	.	.	.	.	.	+	.	.	.	.	+	.	.
Sea urchin I, II	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	+	.	.	.	.	.	.	.	.	.	.	.
Sea urchin Gene J	.	.	.	.	.	.	.	+	.	.	.	.	.	.	+	.	+	.	.	+	.	.	.	.	.	.	.	.
D. melanogaster 79B, 88F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.
D. melanogaster 57A	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. melanogaster 5C	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. melanogaster 87E, 42A	no intron																											
C. capitata A1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.
C. capitata A2	undetermined																											
B. dorsalis A1, A5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.
B. dorsalis A2	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B. dorsalis. A3	no intron																											
B. mori. A3	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B. mori A1, A2	no intron																											
C. elegans I, II,III	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.
C. elegans IV	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	+	.	.	.
A. castellani	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P. polycephalum	.	.	.	.	.	+	.	.	.	.	.	+	.	.	.	.	.	+	.	.	.	.	.	+	.	.	.	+
P. polycephalum	.	.	.	.	.	+	.	.	.	.	.	+	.	.	.	.	.	+	.	.	.	.	.	+	.	.	.	.
A. nidulans	.	+	.	+	.	.	+	+	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.
T. lanuginosus	.	+	.	+	.	.	+	+	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.
C. costata	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.
Yeast	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Dictyostelium	no intron																											
Oxytricha	no intron																											
plants	+	.	.	.	+	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	+	.

3.2.2.3 5', 3' flanking and untranslated regions

The 5', 3' flanking and untranslated regions are also areas of interest. These regions may play a role in regulating temporal and spatial specific expression patterns as reviewed in Chapter 1.

The 5' region which has been sequenced covers about 600bp for CcA1;1.5kb for CcA2; 800bp for DdA1; 2.5kb for DdA2; 900bp for DdA3; and 2.4kb for DdA5 (See Table 3.2.1). The 5' region sequences are shown in Figure 3.2.10. Potential TATA box(es)are found in CcA1, CcA2, DdA1, DdA2, and DdA3. The position of the TATA box is indicated for each gene. TATA boxes are not present in the two *Drosophila* actin genes act79B and act88F. Because the transcriptional initiation site is not known for any of the six genes, it is not certain if these potential TATA boxes are the actual TATA boxes. In addition, a number of direct repeats, reverse complements ranging in length from eight to twelve bases are identified in this region of the CcA1, DdA1 and DdA5 genes. Figure 3.2.11 shows these repeats in the 5' region of the CcA1 gene. These 5' region repeats have also been found in act79B and act88F. It has been suggested that these repeats might be involved in transcriptional regulation or genome rearrangements (Sanchez, et. al., 1983).

In the CcA2 gene, there might be an intron in the 5' UTR located between -8 and -569 from the ATG start codon. The potential 5' splicing site around -569 has the sequence ATAGGT. AGGT is a typical of intron splicing site. In the *Drosophila* act5C gene, the 5' splicing site also located at -569 with the sequence AGGT (Vigoreaux and Tobin, 1987). The 3' splicing site in the act5C gene is located at -9 with the acceptor sequence AG. In all functional plant actin genes which potentially have the 5' leader intron, this sequence (AG) has also been found located at approximately -9 to -11 (McKnight, 1983; Pearson and Meagher, 1990). The AG sequence is found in the CcA2 gene at -8. Therefore it is possible that there is 5' leader intron(s) in the CcA2

Figure 3.2.10: 5' Untranslated and Flanking Region Sequences of the Six Tephritid Actin Genes.

5' UTR or flanking region sequences of the six Tephritid actin genes are presented. Putative TATA boxes are underlined. Unknown sequences are shown as ".....". For the CcA2 gene, the putative intron-exon splicing sites are also indicated with "^".

DdA1 5'region

GAATTCTTACGCAAGCGGTGTGTAGAAATAGCAACCGACCGAAGTACTTTCACGAGTGTATTTTTGTGTCCGACAATTAAGGAT

CTATGAAATTAAGACTGAATGAAAGTGAAAGGGGCCAATAAAAAGCGCTTGCTAATTATAGCAGCCATTTTATATGCTAAAATA

AAGCAAAAAGCTGATATATTCAATATAATTTATAATATTTTTTTGATAAAAGAAACTCCACAAAGTTTCATTCAATAAATATGT

TTCATTAATTTAGACTATTTTTAACAAAATTCAGTAATTTACTAAAAAATATTTAATCGAATTTGTAAAACGCTTTTTTGCTTT

CTTATTATAGAAAATCAAGCGAAAACATATTTAACTCAACTTTTCGATCCCCAAAATTATCAAAATTCTATATTTTTTTGCGCA

AAAGAATTTTACTTAAATATAGAAATAAGCTTGTCGACCGATGTTTCCAATTTAAGAAAAATATCACACCCGGCAATTTTTTTG

TCAACTCAAACTTTTTTCACGAAAAGAAAACAAATAATTATTTCATTTATTCAAAATTCGAAACCGTGTGAAATTTTCAATTGA

AATGGTTTATTGCATTTCATTTCTATAGTTATATATATATTGCATTTCGGATAAGTCAGTGCAAAATTATTAAAGTTGTAGGCA

TTTCGGAAGTAAAAGAATAAATTGCAATTCTCTCAATATCACAAAAAAAATGCATGTGAAAAACTATTAAGGCGATAAATTGTCA

TCACTTGCATATAATTCATTTGCAGAACTCACTTAACTAACTGCCAAA

DdA2 5'region

ATTCTCACTAATAGCAATAATCTGAATTATACAGATTTTCCAGAGTTGTAGAATTAAGCGGTCAGTCTGTAAACAATTGAATTT

GTTCTTTGGAAATAAGTCCATTTCACTTGTTCAAATTAAATCCAACAACAGCTGGAATACGTGAAAACTTTCCATGGCATTTCT

ATTTATCCGAATGCATTGATATATTTCCTGAAAAGCATTTAAATGTAGTGAAATACAAGATTTCTCCCAGTATGATGGAAAAGT

TACTGCCTAACATTGCAGAACATAACTATAGCTTATATGTAGCTGTCATACAAACTGACCGATCAAAATCAAGTTCTTGTATAG

GCCTTTTTTATTTTTTTCGGTGCAGCAGAAGTTTATATATTTTCTTGTTTAGAATACGGTTGGTTTTTTATGGCACATATTTGT

ATTAAATTTATCTTAATTTCGTTAATATACCAAAAATTCTCAATTTATGCTAAAATTATATTCCCATTGCATTCTACTTAATAT

TTCAGAAATACTTAACATTTTCAAATACACATACAAGTTCATACATCGTTCCCTGCTTGAATATATGTATGTTTGCGGCTATTT

CGAGCTGAACTAGCTCCACATTTAGAGCTGATCATTTTAATTATAACCTATGCCAGCAGAAATAGTTCGAGAATGCACCAATTC

TATTTACATAAAACTGCTGGAATCATTCCATTTGCTCGATTACAGAATCAAAATCACTTAGCATTCCAAAAATACTTGCGCTCC

ATGTACAAGTGTGGGCCGAGACATGTATGCTAGCATAGATATGTACATACATATGTTGGTGCATAGAATATTATATCCGATTTT

AATGGAACTTGAGGGTGTATGTGTGATAAAGATTTATTTCATATTTTATTGAGCTAAAACTCCAGAATTTTTACAGGAACAAAC

CCAAAATGTTTGTGTAAAGAAATCATGAACTTACTGTCCTGCCAACTGAAGAGCTCGCTAGAGCGACATTCAACATGTTGGTGC

AACAGTGTGTCGCCATGTAAACAAATGTGATTATGTATGTGTGTGTGTGTATATGTATAAATTTGTTTTGGTTCATGAGCTTAA

CATTTTATTAAAGCAAACTCAAATGTAGATATGTATATGAAAGTCCAAAAAATCGAGGTGGCAACACCGAAAAAATGCACAGAG

GGTGGCAACATGGAAATTACCAAAAATAGAAATGTACAACACAGAAAAGGCGCCAATAAATAAAATCAACAACGTAAGCGCTGA

CCACCTTGCTTCGCCTATCCAACGAATATTCATATTCACATACATACAGACATACACACATATGTATATGCATACAAAAATACG

AAAAATTCCTACATAATGATACACATGTACCATACCAATTTGCTCGGCCATACAAATGCAAATTTAGACGCCCGTGCGTGGGGG

CCACAACGGCTAACGAGAGCGATGCTCGTACGCAACGTGAGAGCCACAACACAACCACACAGCCTTACAGCCATACAACCAGAA

AATCGA.....ATAAAACAGAGTAAGCGAATTGAAACGGAGAAATTGAAGAAAACAAAAAGTGCCACACATTTTAGATTGAGAA

TAAATTTTGTTAAAAAAATACCAGCGCCGCACCGCACAGCACCCGCCAACCACGAACTCGCAACGGCAGGCGTTGTGGCGTTGA

CGCGGCCGCCTTGCCTTATACGGGCTTTTGGAGCGCTTTCCGTGGAGAGCCGAACCAATTACCCATGAAATATGCACACACATT

CGCACAATCATAAACAATCACACTCGATCAAGCCGAACGAACGCGAGCATACGCGGCGATCAACAACGAACAGCGAGCACACAA

ACACACGAAGTAAACGTGTGCACTGCGTGCCGCTATGACGGCACAACACGTTCGCTAGCAGGCCGAACCAACCTATACACCCAC

ACGCATACGATCAGCCGAAGATAGTGAGTGGGAGTGGGCAGTGCGCAGGAGCACGATCGCTATTGCTAACTCGCTCGCTGCGCG

CACATCGCTTGCCATCAAGCCTCCGCAATGCATGGCGTACATGTGTCGGTGTGCCGCTGATGACGGTGAGCGCGGCGGTATGAT

CGGTCTATTTTTAGCGACGTTGAAAAATCGTAAACACACCATATTCCTCGTTGCATGAGCTGGTGTTGGTGTTGCTGCTTGTGG

CGAAGCACCTGTATGTGTGCCTGTGTGTGGNGCGGTGTATCGTGTATGCCGAGTTGGTGGTGGAGGTGAGGTTGCAGCCGCCGC

CGTTGATGGAGTGTTGCGGGTGTTGCGACGCTGAGCGGCGTCGCGGCTGACGATCGCTGTTGCCGAAGAACTTTTTTGGATTTC

CCATACGATGAACATGAAAATTGTAGTTTTATGTTGATTGGTATAAAATCGACTGGCAGTGGCCAAGGCACTTTAGTCAGTGAG

TAGTGTTCCAAAGCGTACACATTTGCCTTATTATTTGATTATTTCTTCTTGCTCGTGGTTTTAAAACACAAAACACAAACAAA

DdA3 5'region

GAATTCCGGCATTCTATTACTTTGAAGTAGCTTAAAATAATTATAGAACTTTTTTGAAGAATGTGAAAAATCAGATTATGTTAT

TTTAAACTTAAACATTTTCAAGATGGATTATTTTTTTAATTAAAAAAAAAAAATAATAATAATTTTCAAAATTATTTATGCATTT

TAAAAAATTATTACTTTTTATATTTTTTAATTAATAATAATCAACACAAGTATAGACAAAAAATGATTATTGTATAAAACAATG

GGTTTTATATAAAAAAATATTTTTTTGTTTTAAAAATGTAGAATATTTTTTTGAATTGTAGAAATAAGAAAATAATTAATTTTC

AAAAATACCATTAGTTTTGCGTTAAAAATTTTTTTTTTGATTTAAATTTTAAAAATAAAAATTAAATATCTTTTTTAGTACTGA

ATAGAGAAAAATAAATTTTAAAAACTTTCATTTTATTAAAATTTTTAGAGGAAATTTACCATTAAAAAAAATTTTTTTTTTGTA

AATTTTTCAAGTTTTTTAATTTTTATTTTTTTAACATTTTATAAGTTTAAGAAAAGAATTAAAAAATAAAAATAAATATTTTTG

CAATTTTTTAAATACCATTTTTTAAATTATTAATTAACTTATATCAAAAAACACTCAGTTGAAACATATTTAGCAACTCGAAAA

TGTTAATAAATATTCATGAGTTAGACAGTTGTTAGGAATATTATAGTCAACAAAAAAAAAACAGTTTTGTTGTTTACTGCTAAA

ATTAATAAACAAAAAAAAAACGTAAAAAATAAAAATACAATTTATAAAGTCACAAAGCAACAGACACTTACCGTCTCTTTTCTT

TTACCTTCTCTCTCTTCCACGCACTTGCAATCGCCTAGTATTAGTATATAAAAATCAAACAAA

DdA5 5'region

GAATTCAACTTTTGTTTTCTTGTAGACCAGTGCAATCGTTCTTGTGATTAGGTCAAAGGTCTTAAATTCATATATTCTTCAGAC

GAATTGAGACAGATTCCTGATAATTAATAGTAGTGTTTATTTCTTTTGTTTTCAACATTCAGAAATGGGCAAAAACTATTTGTT

CTTCGTTCTTCCATTCATAAACTTCAAAACAAAAATTTCGAATGTTTTCCGCCAAATTTGCAGCGAAGTCATTCAACCAATCAC

TCATTCATAAAATAACGCAATAATTCACAATTCTTATTTCATATTTCATATCTAAACACACTACTGTATGATATTCAGTATGTA

GTCTTCAAGTCAGCAAGGCAACAAGTGCTAATCGTTGAATGTGCAACGCCAACTTCATAAATTGTCCGCGGAACTGAGAAATTT

CAAGCGATTCAACAATTCTTCTACGCGTGCAAGGGCCTCTTATACACGCATTTACCCACACCCAACAAGAAACAAAAATCCACC

AAAAATCATACCATCAGTGCCACTAAGAAAAGCCAAGCGAATTTTCTTAGCAATTGGTTGAAGGCGTGCCGCAATGCTTGACAT

AGAACGGGACCAGTGTTGCAGCTTATAAATAACATCAGCCACTTGTCAGAATTGCCAATGAATCAGCTCCCCACCCCGCTGTGA

TCGTTCTCGACGCTTTTACAACTGGGAGTGGCGAAATGGGGAGATTTTCAAAATATTTTCCGTTTTCATTTTCGCTAAGGAATT

TCGCTAAATTTATAGTGCATTCGAGTGTGTATATATGTATGTGTGCACATATCATTTCAATTGCGCTTTGTTACCGCAATCTGG

GCGTATTAAGTTGACAAGTTAATACTAATGTATGTATGTATGTGTGGATGTGTATGTGAGTTGCAGAAAAGTGCGGCGGGTTAT

AAATACACACCTCATTTTATGATGTTAATGATCTTCGGGCTGAGGTTCGCAAGACTTCACATTAATGTGAGCACATATACACTT

ACATACATACACATATATGTATTTGTAAGGCTTTCTTCTTCTGTGTGACGCGGAGAGGCACGTGGTATTTATTTCAAAAATTTG

TTGCCCACATACTCCTTGAACTTCAGTTAGTCATTGTACCCTTAATAAATTAAAATGAGAGATTTTTGTAAGCAAAAATAACTG

ATTACTATAAGGTCCTTGCAAACAAATTTAAAATTAGAAAACTTGGTACAAACTGAAATAGACGCTTAATAAAAGATAATATAC

ATGGAATGAACCCAATACATAAATATGGTAACCTATATGGTAACCTATTGTTGTTGAAAAGAATAACAAGGCCGGTTTAGGTAG

AATTAACCATTCTTTGGAAGATTAAATACCAATAAGGCCGCAAATTGCACACCGTTGCGTTTGCAGTAACATAGTTAGGTCCAC

ATAACCGGAACGGTACGGATTTCTGCTCAGTCAAAACTACTATTTTAATATTGTTCAACAATAACAACAATGAATCTCCTATGC

ACAATATTCACAGCGATTCCCTACCTTTTTTTCTACTTAATGCAACGTAAACATTAAAGCCCACTGTGGCAAAAATTTAGAAGG

CAAATGATCATTGGGTAAAGCGAGGGTGCAAAAAGACTTGCCCGTCATCACTTTATAAAATGCTTGTCGCACAATGTGGTGTGG

CGCAACAAGTGATTACTATGCATGCCACACCATCTATGGCGATCGTTACACTTCCAAGAGCGTTCTCTCAATACAGTACCGTTA

CAACACGTCCCTTCCCCATTTATATTTATAAACAAGCGAAAACAAAGAATGACGGCAAGCAAAAATTATTCTTGATCCTGCGCG

CCTACTAGTAGCGCCCATTCTCTGGGACTTTAGTTGCAAACCACACACATGTATGAATAGTACACTAACAACAATTGACAAAAC

AAAAATTTCATTACTTCTTGGCAGCAGAATTGTGCCTCACTTCGAAGGCAAGCAATGCATCCGAATAGAAATTCATCTCATTTT

GACGCCCTTTCTACGACTACATGATAAGCAAATCAACTGCAGATGAGAAAGTATTGATATGTGTCCGATATTGTGGAGAGCACC

CAAGCATCAAAATGAGATTTCATTTTTGAGAGCACTGTGTCAATTTAAAAGGCGCTGCGTCCAACATAACCGAATCACTTTGGT

TGCTGTCACGTCGGAAATCACACGTTACTGCAGATCTTCACGTCCATTGTAAATTGGATTTTGTACTTAATCAGTTGCTGTCGT

AGTTTGTGTAAGTTTACTCAGCATTACGTGGCGACAATTCGAATTTGGGTGCATTACACCAGTGATATCAGATTTCTCCATTAA

TTTGTATCTACTGTCTTTGTAGTGAAAAATCATAAAACTTCAAG

CcA1 5'region

TACACACAAGATGAATGACAAAGCTCACAGGTCCGCGCCTACATTGGCGCCTTCTGGTACACAGCACACACACCTACACTTGTA

CAGAGACAACCAAGAAATTGCATTACTTCTTGTGCAACAGACTCATTTGCAGGCAATCACAATTCATCCGAGAAGACACTCATC

TCACTTGAACGCCCATCCAAACGTCTTCTCTGCGACAAAATCATTTGTTTGTAGTGCTGCTTTATTGTTGGTGTAGTAAGATTA

CAAAAGAATAATAGTTGTTGTTGCTGATGAAGAGAGAATATTGATATGTCTCTGGCGGTGTGTGTGCGCTGAGATTCGCATTTG

ACCGAAGCGACCGGACTCATTGGATTTAAAAGCCGCCACGTCCAACATCGCCGAATCACTTTGGTTGCTGCCACATCGGATTCA

CACGTTACTGCAGGTCCTTCCGTCCACTTGTCGTTGGATATTGTGTGTTTAATTTGTGTTACTGTGTAAGTGTCACCACAGAGT

GATCTCCTTCTGTCACCCTTCAAGTCCTCTGTGATATCAAATTTCCCTTCTCAATTGTAATAAGTATTTATTTGTGTTATTGTT

GTTGT<u>TATA</u>GTGAAAGACTAAATAACATAACTCCAAA

CcA2 5'region

AAGTAGCATTAGACCGTTAATTCGTTTCTGTTGCCTTTTTGGTCAAACGAAAATACATATACAAATCTTTTTGTATGTACATAC

ATATGTATGTATGTATCTACAAAATGTATTTGTTAATAAGGCATATTATAAGAAAAATGCCGGTTTTGCCGAAGATGCTCTTTT

ATGGCATCTGAGTTTGGTTCAATTTTCACCCCCTGCTCTTTCTTAGCTTTAACAGTTTCTAGTTCTCCTACACTGACTGGCTCA

CTCACACTCATCGCACCATGTACACGCACAACGAACGTGATGGCTGAAAAAAAAGAAGTTACTTCTTTTTTTGTCGTTCGCAAG

GGTAATTCGAGAAATCGCGTATGACGATTTTTCATTGCAGCACAGCAAACATGACCATATATGGTCACGAAAATGGTATCGCGA

TACTGCTTGAATATCTACTCAGAGAGTCGCAAGTGTGGATTAAACGAAACGTGTTCGTTTGGCAAGTTGGTAGCGCCTAGCATA

TTGCCTTGCACTCTCTCTTTTGCTTAGTAGGCAAGTTACGAGCCAATAAGAAAGAGTTACATAGATTTTGATTCCAATAGAGCG

CTCTCACCATTCGGCGGGGAACAATTCGTTACTCGTTGAAATGATGACTCGTTGATAAGTAGAAATTTTAACTGGCATTTTTTG

CTTAACGGCGTCCCATTCGTTGTTGGACTGTCAAAGTGCTTACTTGAATTTCTCCGTTCAATCTTTAAGCTTAGATTTCATTTG

TGCCAAGTCAGCATCTCAAAGACCGTGAAATTTGTACCACACCTATAGTGCTACAATTAAAGTGTTTTAATTCAAGTTTATAAT

AAATTTAAAG^GTGAGAAGTTATTGTCTAACATTTCACTTTTATTTTATTTATTTTTTGTGGTTACCGGNTTTGTAATAGGTCT

ATTATGCACGCATTTATTGATCCGGCCGGCTTTCGGCCATGATTCTATTGGTCAAACGTTTTTTCTTTTATTCGTGATTACTAG

TCATCCATGCTATTAATGTCGGGCGTTGCCCTTTTGAACGATTCTGATATTAACAATGATCTTTTTTTTTAAGCTCATAACATAA

TATTTAATTTTTTTTTATCGTCTAAGCGAAAATTTTTTTTATTTTTGTCAATTACTACTAGATTGCTTCGATTTTTCTTTCAAA

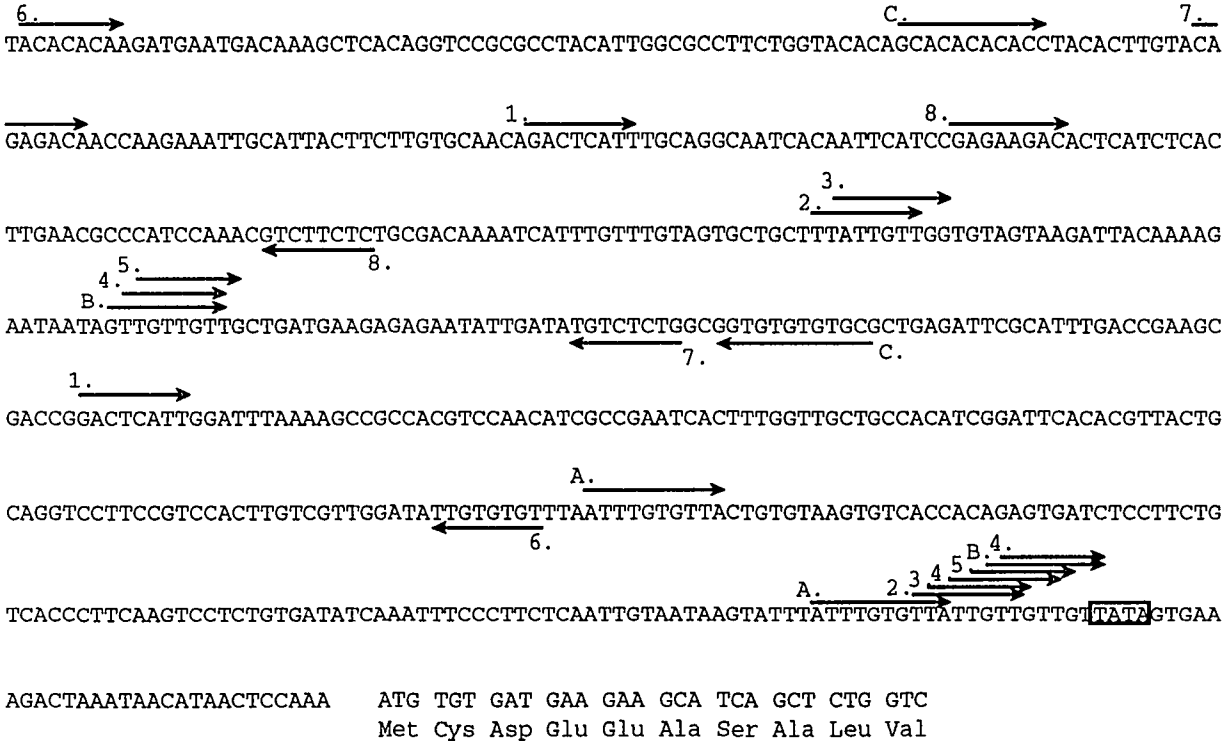
GGCTTGTTAAATTTTGCTAAATTATTGTGAAATTGTTGAAATTTTATGTTTTCTCCTATATTCGCGCGTTTACTTCCATATCGT

TTTACACAAATTTTGTACTGCTGCAACAAAAACTTTTGTTAAAATTGGAATTAAAAAGTCAAAACTACGTTCTAATAGCTATAA

TAAGGAAAAGAAGACCAACTACCGGAACTTGAAGTTTAGAAGTCAATACAAATCCAAATTAAATTTTTCAAACTAACGCATTCT

ATTTACATCGTATGTATGTGTATTAATTCGATGATTTCTTTGACTATTTTTAG^CAAACAAA

Figure 3.2.11: 5' Untranslated and Flanking Region Sequence of the CcA1 Gene. The putative TATA box is denoted by a rectangle. Also shown are several inverted repeats (letter designations) and direct repeats (number designations) found in this region.


6.
C.
7.
TACACACAAGATGAATGACAAAGCTCACAGGTCCGCGCCTACATTGGCGCCTTCTGGTACACAGCACACACACCTACACTTGTACA
1.
8.
GAGACAACCAAGAAATTGCATTACTTCTTGTGCAACAGACTCATTTGCAGGCAATCACAATTCATCCGAGAAGACACTCATCTCAC
3.
2.
TTGAACGCCCATCCAAACGTCTTCTCTGCGACAAAATCATTTGTTTGTAGTGCTGCTTTATTGTTGGTGTAGTAAGATTACAAAAG
8.
5.
4.
B.
AATAATAGTTGTTGTTGCTGATGAAGAGAGAATATTGATATGTCTCTGGCGGTGTGTGTGCGCTGAGATTCGCATTTGACCGAAGC
7.
C.
1.
GACCGGACTCATTGGATTTAAAAGCCGCCACGTCCAACATCGCCGAATCACTTTGGTTGCTGCCACATCGGATTCACACGTTACTG
A.
CAGGTCCTTCCGTCCACTTGTCGTTGGATATTGTGTGTTTAATTTGTGTTACTGTGTAAGTGTCACCACAGAGTGATCTCCTTCTG
6.
4.
B.
5
4
3
2.
A.
TCACCCTTCAAGTCCTCTGTGATATCAAATTTCCCTTCTCAATTGTAATAAGTATTTATTTGTGTTATTGTTGTTGTTATAGTGAA
AGACTAAATAACATAACTCCAAA
ATG TGT GAT GAA GAA GCA TCA GCT CTG GTC
Met Cys Asp Glu Glu Ala Ser Ala Leu Val

gene. If this is true, the CcA2 gene is most likely to be the orthologous gene of the 5C gene rather than the 42A gene, which does not have a 5' leader intron. The partial amino acid sequences of CcA2 gene are not sufficient to distinquish these two posibilities.

The 3' regions sequenced are 500 bp for CcA1; 395bp for DdA1; 1.8kb for DdA2, 2.04kb for DdA3, and 1.95kb for DdA5. The sequences of 3' region for each gene are presented in Figure 3.3.12. The AATAAA sequence typical of the polyA addition site is present in most genes except DdA1. This may due to the fact that the sequenced 3' region for this gene is too short. In *D. melanogaster*, the 3' untranslated region range from 300bp up to 700bp in size (Fyrberg et al., 1983). Alternatively, there might be no AATAAA site in this gene. In fact, no AATAAA sequence is found in the two *Drosophila* actin genes act79B and act88F (Sanchez et al., 1983). For the CcA1 gene, a potential AATAAA sequence is located between 370bp to 375bp after stop codon. In DdA2, it is at 521bp to 526bp after stop codon. For DdA3 and DdA5, there are each two potential AATAAA signals (indicated in each sequence).

The sequence of the cDNA clone pmedC1 aligns almost perfectly with the 3' untranslated region of CcA1 (See Figure 3.2.13), except for a change from CG in pmed21 to GC in pmedC1 at positions 95 and 96 after the stop codon. The complete sequence of pmedC1 covers a total of 416 bp from the stop codon to 38 bp after the potential polyA addition site.

By using the MacVector computer program to search for homologous sequences, it shows that there is almost no detectable homology in the 3' region between all the genes from *B. dorsalis*. Very low homology is detectable between CcA1 and three *B. dorsalis* genes DdA1, DdA2, DdA3. However, some homology is detectable in the 3' region between DdA5 and CcA1. This homology was also shown by the Southern blot (He, 1990, MS thesis). Almost no homology is detectable in the 3' region between these genes from the two Tephritid species and two of the *D. melanogaster* genes act79B

Figure 3.2.12: 3' Untranslated and Flanking Region Sequences of the Four Tephritid Actin Genes.

3' UTR or flanking region sequences of the six Tephritid actin genes are presented. Putative polyadenylation signals (AATAAA) are underlined.

DdA1 3'region

ATCTTTCGATTATGAAAGATTGCTGAATGCTTTGGCGCAAGAACACATACACAAAACTAGAATTAACAACAATAACAACATA

CCTACGTTTGGATACTTCGACTGCCAGCAAGACATCGTGCTGAAAGCGAATGTAACGACCAAAACCAAATTCATGGCGCAAT

TAAGAACTTCAATTACAACAACACAAAAGGCATTAGTAACAACAAGAACGAAAAGAAAGCTACAGTTATGTAGGGAAAGCAA

CAACAGTAAATTTAAGCAAATCAATAAGCACACAACGAATGCTCAAATTAATGGCGCCAAAAATAGTAACATTAAACACCAT

AACAAGAAGAAGAGCGCGCAGAACGGATATCATACGACAGTGACATGTGCGCCATCTGTAGGAATTC

DdA2 3'region

GCAGTTACTGTTACCAGCTAGTAATACGCTTTTACGTGTTCGCGCTGACGTGGCTATTGCATTGTTCGCCACATTATTGTGT

CTGATGATGCAGTATCACCATATTAGCTATAACAACAACAACATGCCTTCAGAGATGCCCGCCCATTCGTTTATTTATTTAT

TTGCATAAATGTTGTCAATACCAATTGATCACTGGTTCCAGATCGCAAATTCATAACGACTGAACAACACTCACAACTGTAA

ACACACAAACAACACTGAACCAGATTTATGTTTTATGTTTGTCATTTAACAATTTACCGTTTATGAATATGTTTCGTTTCAT

TTCGATTAACTGTTGAATTTGCTTATAGTTTTATGAAAGCATAATGAACGCATTATTGTTTTTTAAGGGTATTTTTGCCACA

TGTAAAATAATGTAATGATTTCTTATGTAATTAAATTAAAACAATTTATAATTATTTATAAATAAACGACTTTTTTAGACT

TAATTTTCCATATCAATGACTTTATAACGCAGTTATGGTCAAACAAGGGGGGTTGAGGGGCCGAGGGGCGGGTTGCCGCCAC

GCATAGCACTCAGCATTGGGTTGTATCTTCGCTATACTCCATAGGGCGCTTAGATCATACATGTTGTTGCAGTTGATGTTGT

TGCTAAGAAAAAGCTTTCCTAAGCAACAAAAGCATTGAAATTGTTTGACCCCCCTCCACTGTCGCCAAAAATATTACGAAAG

CTCCGATCTTATTTCCATGCGCAACGTCCGCATACACCGCCCATTTCTCCACTCGGCATTTTGAAACACACGAGCTTTCGGC

GTGTTGCATAACAGCTTCGATCATTCTTAAGTAATCTTGTAGAAATATAATATTTGTATACATACATACATATTAGCCGTTC

GAGTAAAAACATTGGCTGCCAGTTTGTCTATTTTTGGAAGCCAAAACGTTGCATTGGAACACAACCGAATGCTGCTGCTCGT

AACACCGCCAGCAGGTGCCAACGAGGCCACCGCCGAAAAAGCTCGCTGGCTCACATACATCAATGGCATTGCGTTGGCTGAA

ACACCGAAGCCACCATGCAGTTGTTGGCATTAGCCGTTCATGTATTGGGGCTG

DdA3 3'region

ACGCTCAATGAGAAACACAAAACTGAGGGCTTTTACGCTCCAAACCCCCCCTTTGCTATGTATTGTATTTGTTTTATTTTAA

TACTTTATTTTAAAATTAATTATGTATGTAAATATTTTTATAAGAACAAAAATGTAGCAAGAACAACGGCACTAGATATACC

AAAGAAAACGTTGATAACGCAAATGATGATAAAGATGATGATAGGAATTTGAAAATTGAAAAAGAATTATGAATAATGGATA

AAACGATTATATGCGAGAGAGCAAACACAAAGAGAAAAAACACTGCATTGAAAATCCCATGAAAGAAATAAACAAATTATTG

AAAAAATATACTATATACACCTACAATTTTATTGAAATCTACAAACACATAAATATCTGCGTTTTTATTGCCCTCAACTCAC

GAATATGCGAGTACATAGACTTTTGTGTATCCATAAATCATCTGGACATCTCCATTAAGCACGCATTCCGCTACACACCCCA

CACCGGGAAACGTCATTGCTGCCCGACTATTAGTCCATTAGTGGGCTAAATATAGCGGCGCTTCAATCGGCCACATGCAACA

AATCAAAATGCAGATCTCCTGCTACATTTCCATTTCGGGCGCTATTCCTTTGCACTTGACTCGCTGGCATATTATATATTTA

GACACCTGCACACAGGCAGCCAGCCGAGTGCGTCGATCGATGTTCGGTGCTCGCGCACTGAGCAACCCACCACACCAAGTTT

CTTCCAAATAAAGCCAGCAGTGTGCTGGGTGATTCACCCTTCTACGCGATGGTTTTGTTGATCTTCGTGTACGCAGTCCGAC

GGCATGTTCACCCACGATGCGCGCCGGATATATTTCGTTGGCTGCCACATACACACACAAACGACTGCATACATAAATTCAA

AAATATGCTTGACGTTTCTAAAAATATCCGCCATCGATATGTTTGTTTGCTGTCATAAATATACAGTTATAGTGTTGTGTTT

GTTGTTTGGTTTCGCAGCATTCCCATCTTTCAAAATTTCTTTTCCATTTTTGGCCGCGCGCAGATAAATTAATTATGGCATG

TACGCGGATGCATGAACTTAGGTGTAGATTTGTGTGTACATATATACATAGAATGATGGTTTTGCAAGCGTAATTGTTTATG

ACTTTTACTCATCGCTAAGGCTTTGATGTATTTACCGTTAGTTTCGTCGTAAAGTTTGGTGTTTTCATTTTCTNTTGAAGGA

AATAACATTGTAGATTCAACTAATTGAAAAAAGTTGGTTATTAGTTTTGATATTTCGTTTTAAAAATATTCTCCGTAAATTT

CGGCGATATGATCATATTTGTTGTAATTTTTTAACTTGTGTAATCGGCTTTAGACGAGTTTCTCTCCAAACAATATTTCCAG

AGTGAGTGAGGAAAATTTTTCCGGAATGGCTGCAGCAATAAATTATCGCACGAACTTCTTAGATATACTACTGTGAGCAAAA

AATTGTAAGACTTTATTCATAATTTTAAAATTCTTAAAATATTCTTCAAAAATTATGTCATCTCCCCCCGTCCCCGTTAGGA

TGTAAGGACGGTCAAAGCTTGTCGGAAGATTATTGCGCCCGAAGTTTGGTCAGCGTATTGTCTGATATCGTCGACTTAGTTG

AGGAATGACCTCTTAATGTTCCTAGGAGGCGGTTACGCCAAGCAGGAGACTGCGAGGATGATTTCTGAGGAAGCATCCCCGC

ATGAACGGCTTGGAGAGGAGTTTATAATGCACTATGCAGGTGTTTGATAGGAGACATCAAGAGGCATTCTGTCAT

DdA5 3'region

GCTCTAAATTCATCTGTTGCCAACCAAGGCGTGGTGATCCGAACTATCGGCTGATCACCATCCACCTACATCCCACAGTGTC

CGAACAGCTACCGAACCCACCCCAGGGTTGTAGGACCAATACGAACATCACATATATGATATGGAATATACAATGAAATACA

AAAAGGAACATTAAACTCAACCCATTCGAAATCTATCCTCAAGAATTTTTTGCTTTTTTCTCATTCGATTTGGTTTCCATTT

ACTTTCTAGATTATTGCAGTATTTGTTTGTATGAACAGAAATAAACTTTATTCGTTGGCCAGCAAACAACAGCGAAGTCTCT

TTACTTCCGCAGTTTCTGTTCAATTGGAGATGAGGGGCGAGTACGTGGTTCGAGAACTGTGGGACACCCGAACTACTAACGA

CGACTGCTTACTTAGTAAGCAGATTTCGAATTATTTCAAGTGATCCCACTTTAGTTGTAGGTTTTCATGGCACTTTTGTTGT

TGCGTCTCTGTCAAGTGGTTGTGCCTCACACGTCGCTCCACATTCGCACCCACCTCCTGGACACGGTACATTAAGGCATCGC

ATGCAGTGACATTTTAAAATTTACAATTGCAACTTTTAAGTGTTGCTTCTTATTTTTATCGGGCGAACGGGGTAATAAATGT

TGCCAAACTTTGTACTTGATGGCAGCTTTTATGTGTAGGTCCAAAATGGCTCGACTGCTCCTACAAGTGCGTGTGTACGTGC

ACAATGAGCTCGTAAAACGGTAATGTGCGTAATGTTGTTGCCATAGTTGTTGTTGTAGTTCTATCGCTATAATGCTTGATGT

TTGTTATTTGTTGTTTGGATGTTGTTATCCATGTGCGCCATTGATGGCTTCATAAATGCCACGGATGTGTTCTTTGCCACAT

TTATGTTGCTTTGTAAGCTCCTTGATGGATCTCGAGTGCTTTCTGCACAGCGAAGTGGCAGCGTTTCTGCTCTTCTCAGATG

TGGGTCGTCAAGTTCGCAAGTTTAATTGGTTGGACACACTTTACGAGGTTAACAGCGCTGTGGTGGCAGCCGAGGATTGGTC

GGGGATACTTTCGAAATTTCTAAGAATTAATGATGTCATTTACAACATTTACATTGCAGCAGAATGTTGTCATGCGTCGATT

TATTCGGCTACACTAAAATCTGGGTTTATTGGAATCATGTTAGGAATATTGCTAGCCTTAATTTAATCACTTGTCGTTGGCA

TAGTGGTACAGAGATGAAGGCGCTAGGCCAAGATCGTTGGACAAAGCTTACGTAAACCATGGCTGAAATGGTAAATGAATGT

CCAATCGCAGTCTTTATTATCCGAAAAATAAGAAAATCACTGTCGACACTGCCAACGATATGGTAAGAGAACCACTCTTCAT

ACTTGTGTTGAAGGCAATCGATGCAGATGGGTTGGGGTGTTTGGCTACAGTATTCTAAATCAGTGAAGTATAAGCTTTCGAG

ATCGCGATTTCGGTCCGAACCAAAACAGAGCTTTTAGTTGCGGACAAAGAATACACGAAGACCATAACAGATTTCATTGCTG

CTTTTCAAATAAGCGTTAAAACAAACTTTGAAGAAAACAAAGAAGATATCCAGAACGAGGTGTTTAAAAAACAGCGGAATAA

AAAGCATTTACATAAACATGTTTAAATTTTTTAAACATCTGTTATAAAATAATTGTAAGGAAAGGCATTACCGCATTAGCTT

CAATAGAAAAGGACAATAGCGTCGATTTAAATAATAACTTAATAGTCCCTATAGAATAATAGAAATTAGACAAATTGCATAT

GTTGCTAAGTATCTTTTCTAACGATAACTGGGAGATAGCGTGAGGAAAATTTGAGCTTACAGAACATCTTCTCTCTTCTTAC

TTCAATTTCCATAAATTTTTTGTTCTTTTTTTTCAAATAATAATTTGCCGAGCTGAAATCGAATTC

Figure 3.2.13: DNA Sequence of the 3' Untranslated and Flanking Region of CcA1 Aligned with the Sequence of pmedC1.

A dinucleotide which is reversed in pmedC1 is indicated by lower case letters and the putative polyadenylation signal is underlined in the CcA1 sequence.

```
  Phe
  TTCTAA     ACAGTCAATACTGACGTTAAAACAACAAAAACCCACACTATTAAAATCTGAAAAGATGCGGTCCAC
cDNA:TAA     ACAGTCAATACTGACGTTAAAACAACAAAAACCCACACTATTAAAATCTGAAAAGATGCGGTCCAC

ACGCTGTATTGGCCACCACTGAGCTACACGTACGCCCTGCCTGCCTTAAGATCTGCTGAGCATCCACCCCGCCCTCAC
ACGCTGTATTGGCCACCACTGAGCTACAgcTACGCCCTGCCTGCCTTAAGATCTGCTGAGCATCCACCCCGCCCTCAC

GGGAGTGTGTGTGTGTCTTAGCGGCATGTGGCACTCATACATATCAACAACACACATACATAAGAATGAAAATAATAA
GGGAGTGTGTGTGTGTCTTAGCGGCATGTGGCACTCATACATATCAACAACACACATACATAAGAATGAAAATAATAA

TATATAAAATAGTTATAGAAAAGGAACATTAAACTCAAGAACCCACTCGAAATCTCTTCCCTTCTATTTTCTCGCGAT
TATATAAAATAGTTATAGAAAAGGAACATTAAACTCAAGAACCCACTCGAAATCTCTTCCCTTCTATTTTCTCGCGAT

TCGATTCTTTTGCTTTGTTTTCGATTTATTTTCTAGATTAATGCAGTATTCGACGTTTGTATGAACAGGAATAAACTT
TCGATTCTTTTGCTTTGTTTTCGATTTATTTTCTAGATTAATGCAGTATTCGACGTTTGTATGAACAGGAATAAACTT

TATTTGTTCGCCAGAAAACAACAGCGAAGTCTCTTTACTTAATGCAGTTTCACTTAAAGGGGGTGTAGAGCGGGCCTC
TATTTGTTCGCCAGAAAACAACAGCGAAGTCTCTT

CTCGCTGGGAGGGTGGACAAATGAGCTATTAACCTCGAGATCT
```

and act88F. However, there is a seven nucleotide sequence (GCATCCA) found in the 3' untranslated region of both the CcA1 and 79B genes. In CcA1 it is located at 126bp after the stop codon while in 79B it is found immediately after the stop codon. This sequence is not found in any of the other genes.

In *D. melanogaster*, the lack of homology in 3' untranslated between the six actin genes has enabled the construction of gene specific probes. In this study, a similar lack of homology in the 3' regions between actin genes from the same species has also been used to generate gene specific probes.

3.2.3 Evolutionary Considerations

3.2.3.1 Phylogenetic analysis

Speciation is one of the most important natural processes related to gene divergence. For single copy genes, gene divergence from a common ancestor is reflected by the species evolution. For a multigene family, the situation is somewhat different. As duplication with subsequent divergence appears to be a common mechanism for acquiring new members or functions in a multigene family, gene evolution in a multigene family is, therefore, a consequence of both speciation as well as duplication processes. A duplication event can generate paralogous genes within a species while a speciation event generates orthologous genes between species.

To identify possible phylogenetic relationships including paralogous genes and orthologous genes between members of the actin gene family from these three Dipteran species, a phylogenetic tree based on amino acid sequences of actin proteins was constructed. The tree was constructed with a computer program "PAUP" (Phylogenetic Analysis Using Parsimony, version 3.0) developed by Swofford (1990). This method uses character states such as the nucleotide or amino acid at a particular site under the

principle of maximum parsimony, and it searches for the minimum-length trees based on these character states.

For this particular operation, a total of 12 actin sequences containing 377 amino acids each are included, of which 11 are from the three Dipteran species and one is from the yeast *Saccharomyces cerevisiae* (YSC). Unknown amino acids or gaps (one gap in the YSC gene) are replaced with "?". The YSC actin is used as an outside group. Numerical resampling by bootstrapping is applied with 100 replications using a "Branch and Bound" method to optimize trees (Swofford and Olsen, 1990). A total of 16 trees are saved after 100 replications of bootstrapping. The shortest tree found has 66 steps. A consensus tree with bootstrap 50% majority-rule is shown in Figure 3.2.14. Bootstrap values and the changes required for each branch are indicated on the tree. The consistency index (CI) is 0.866.

In this tree, several phylogenetic relationships appeared. The two cytoplasmic actins from *D. melanogaster* (Act5C and act42A) form a group. The DdA2 and DdA3 genes are grouped with the two larval-and-adult muscle actin genes of *D. melanogaster* act57A and act87E, within which DdA2 and act57A form an additional group. As discussed, the DdA2 and act57A genes share a common intron position while both DdA3 and act87E genes do not have any introns. The remaining three Tephritid actin genes, CcA1, DdA1 and DdA5, are grouped with the two *Drosophila* adult muscle actin gene act79B and act88F. The DdA1 gene is further grouped with the 88F gene. These five genes also share a common intron position. This tree appears to fit both the intron positions and amino acid sequence data very well. The genes with same intron makeup and higher amino acid similarity are grouped together.

To further extend the phylogenetic analysis, three more actin genes were added. These include three actin genes (BmA1, BmA2 and BmA3) from *Bombyx mori*. The BmA3 is a cytoplasmic actin gene while the BmA1 and BmA2 are muscle specific actin

genes. BmA1, as two of the *Drosophila* genes act57A and 87E, encodes a larval-and-adult muscle actin, while BmA2, like the other two *Drosophila* genes act79B and act88F, encodes a adult muscle actin (Mounier et al., 1992). The sequences of these three actin genes were extracted from Genebank. This time, a total of fifteen actin sequences containing 377 amino acids each are included, of which eleven from the three Dipteran species, three from *B. mori* and one from *Saccharomyces cerevisiae* (YSC). Unknown amino acids or gaps (one gap in the YSC gene and one unknown amino acid in the BmA3 gene) are replaced with "?". The YSC actin is again used as an outside group. Bootstrapping is applied here with 100 replications again using a "Branch and Bound" search. A total of 88 trees are saved after 100 replications of bootstrapping. The shortest tree found has 74 steps. A consensus tree with bootstrap 50% majority-rule is shown in Figure 3.2.15. The bootstrap values and the changes required for each branch are indicated on the tree. The consistency index (CI) is 0.828.

This tree retains the general pattern of relationships of the *Dipteran* genes. The BmA3 actin gene is grouped with the cytoplasmic actin genes of *Drosophila.* The BmA1 actin gene is placed in the group with the DdA2, DdA3, act57A and act87E actins, while the BmA2 actin is placed in the same group with DdA1 and act88F actins seperated from the other three genes the CcA1, DdA5 and act79B. Although the root is not placed, it should be somewhere between the YSC gene and all other genes.

This tree demonstrates several points. First, all five Tephritid actin genes are most likely to be muscle specific actin genes. The DdA2 and DdA3 represent the larval-and-adult muscle actins while the CcA1, DdA1 and DdA5 represent the adult muscle actins. Second, all the muscle actins here form a distinct clade, separate from the cytoplasmic actins. This is consistent with the result from Mounier et al. (1992), where 27 actin sequences from various aminal phyla were compared, including the six *Drosophila* actins and the three *Bombyx* actins.

Figure 3.2.14: Phylogenetic Tree of 12 Actin Protein Sequences Using the "PAUP" Method.

This tree was constructed by the "PAUP" method using amino acid sequences. The naked numbers are percentages of 100 bootstrap replicates where the same internal branch is recovered. The numbers in () are changes required for each branch. The tree was rooted by the yeast actin of *Saccharomyces cerevisiea.* Cc: *Ceratitis capitata*, A1: actin A1 gene; Dd: *Bactrocera dorsalis*; Dm: *Drosophila melanogaster*, Act79B: actin Act79B gene; YSC: yeast *Saccharomyces cerevisiea.*

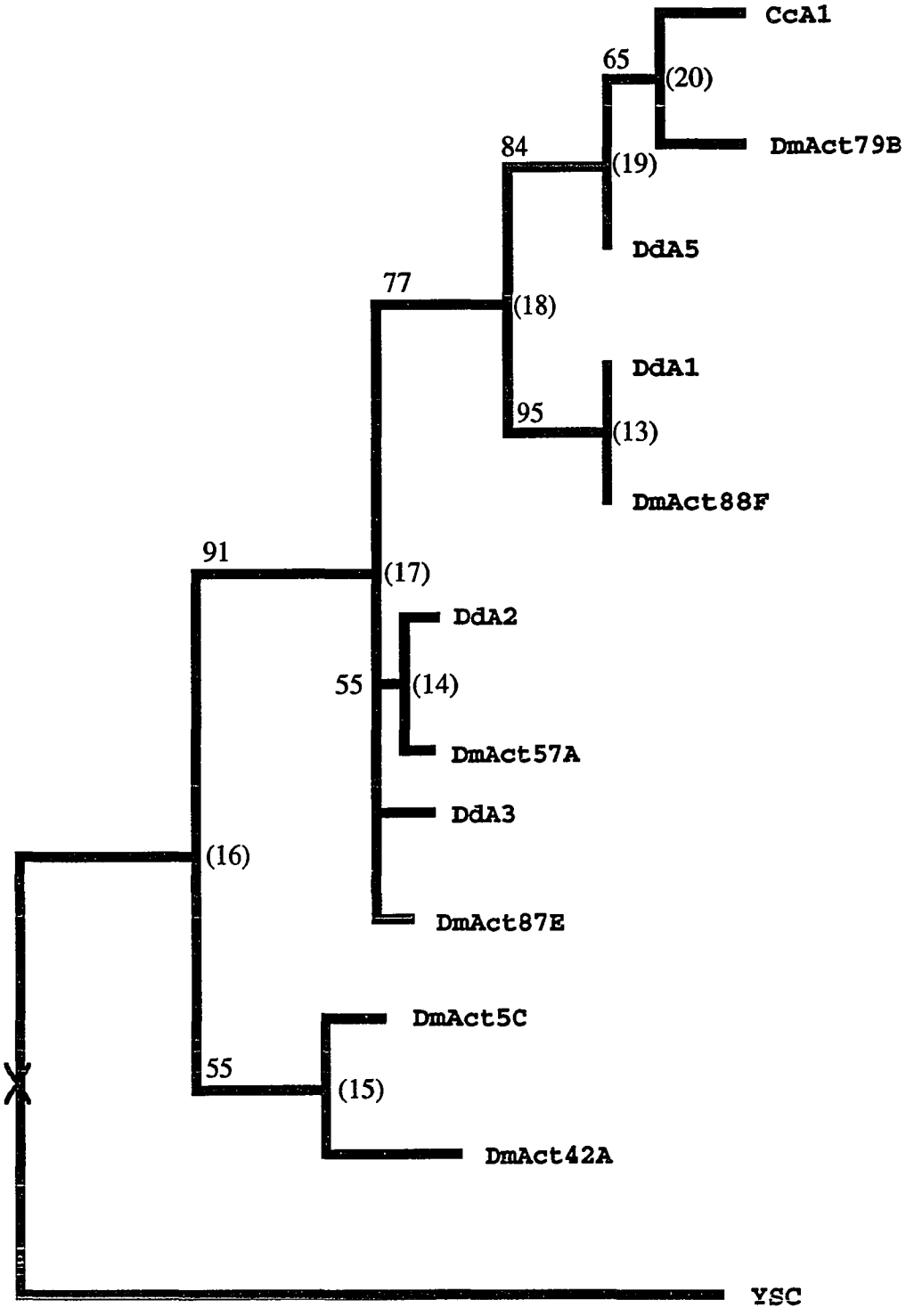
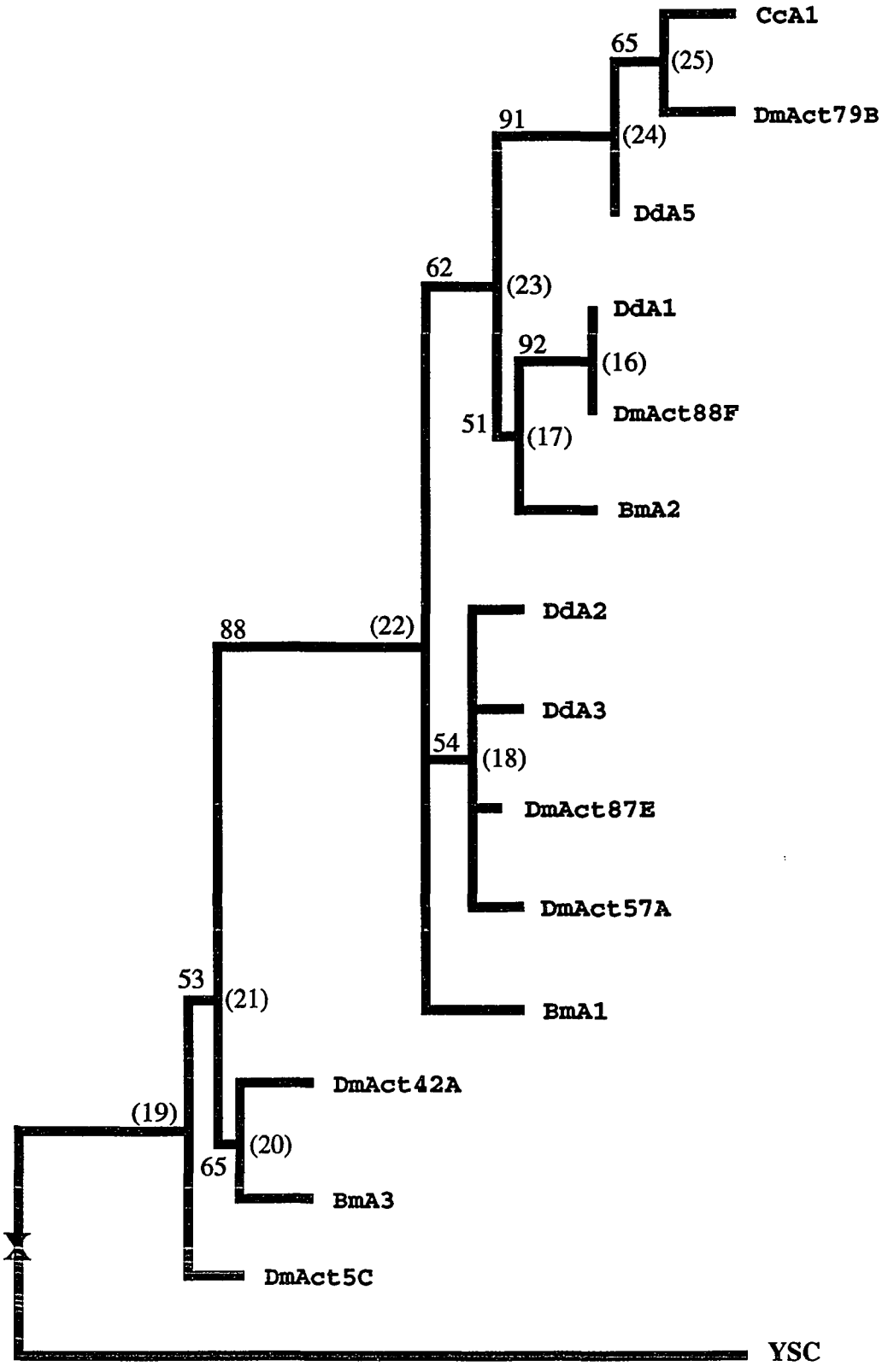

CcA1
65
(20)
DmAct79B
84
(19)
DdA5
77
(18)
DdA1
95
(13)
DmAct88F
91
(17)
DdA2
55
(14)
DmAct57A
DdA3
(16)
DmAct87E
DmAct5C
55
(15)
DmAct42A
YSC

Figure 3.2.15: Phylogenetic Tree of 15 Actin Protein Sequences Using the "PAUP" Method.

This tree was constructed by the "PAUP" method using amino acid sequences. The naked numbers are percentages of 100 bootstrap replicates where the same internal branch is recovered. The numbers in () are changes required for each branch. The tree was rooted by the yeast actin of *Saccharomyces cerevisiea.* Cc: *Ceratitis capitata*, A1: actin A1 gene; Dd: *Bactrocera dorsalis*; Dm: *Drosophila melanogaster*, Act79B: actin Act79B gene; Bm: Bombyx mori; YSC: yeast *Saccharomyces cerevisiea.*


CcA1
65
(25)
DmAct79B
91
(24)
DdA5
62
(23)
DdA1
92
(16)
DmAct88F
51
(17)
BmA2
DdA2
88
(22)
DdA3
54
(18)
DmAct87E
DmAct57A
53
(21)
BmA1
DmAct42A
(19)
(20)
65
BmA3
DmAct5C
YSC

As we know the intron positions for each gene in this phylogeny, one question is whether the tree relationships agree with relationships based on intron positions. All the adult muscle actin genes from the three Dipteran species share a common intron at the 307th codon. These genes are placed in the same group with one of the *Bombyx* genes (BmA2). This *Bombyx* gene does not have any introns, but it is grouped with two of the Dipteran genes DdA1 and act88F, seperate from the other three Dipteran genes. It is possible that the intron at the 307th codon was present in the ancestral gene of this clade, and somehow it was lost in the BmA2 gene. Alternatively, the ancestral gene of this clade possibly did not carry the intron, but the intron was gained later in the Dipteran lineage.

A duplication event appeared to have occured before the separation of the three Dipteran species. This event resulted in two copies of the adult muscle actin gene, specifically the paralogous genes act79B and act88F gene in *D. melanogaster* and the DdA1 and DdA5 genes in *B. dorsalis*. The original copy of this gene seems to have already gained the intron. After the duplication, the genes diverged, ultimately leading to orthologous groupings such as DdA1 and act88F. It is also possible that the relationship between the BmA2 gene and the two *Dipteran* genes DdA1 and act88F resulted from convergent evolution. They show close relationships in the protein sequences, possibly because of similar functions. Because only one adult musle like actin gene (CcA1) has been isolated from the *C. capitata*, it will be important to see if there is another adult muscle actin genes from this species. This may help distinguish these possibilities.

In the larval-and-adult actin clade, two Dipteran actin genes (DdA2 and act57A) share a common intron at the 13th codon, while the other three genes do not have any introns. Because the four Dipteran genes (act57A, DdA2, act87E and DdA3) are extremely conserved, showing 98.8% to 99.5% similarity in protein sequences (See Table 3.2.2), these genes might have also resulted from a duplication event before the

separation of these two species. The intron may have been either gained in one copy of the genes or lost in another copy of the genes, generating the orthologous genes DdA2 and act57A after the separation of these two species. In the three cytoplasmic actins, the BmA3 has an intron at the 116th codon, act5C has an intron in the 5' UT region, and act42A does not have any introns. The simplest explanation here is that the introns were not present in the ancestral gene but were gained independently in two of the genes. Overall, this would also be consistent with the model proposed by Dibb and Newman (1989) that the introns found in actin multigene families were gained rather than lost during eukaryotic evolution.

The two phylogenetic trees discussed above were constructed by using maximum parsimony method based on amino acid sequences. Because the actin proteins are highly conserved and the parsimony method requires the use of informative sites in calculation, the number of sites which can be used is thus limited. In this case, the overall differences as exhibited by distance values may preserve more information. To make a comparison with the parsimony trees, a distance matrix method has been used to construct phylogenetic trees based on data from DNA sequences.

In doing this, the rate of synonymous substitutions and nonsynonymous substitutions are first estimated by using a computer program "LWL, 85" of Li, Wu and Luo (1985). The results are summarized in Table 3.2.7. As expected, the synonymous substitution rate is much higher (with a mean 1.201) than the nonsynonymous substitution rate (with a mean 0.023), consistent with the notion previously discussed that there is a high level of silent substitutions and selective constraints at the protein level. In addition, except the Act42 gene, lower synonymous substitution rates are seen when comparison is made among *Drosophila* actin genes as opposed to between these actin genes and the Tephritid actin genes. However, for the nonsynonymous substitution rate there is essentially no difference for these same comparisons.

Table 3.2.7. Synonymous and Nonsynonymous Pairwise Comparisons.

Syn\Asyn	CcA1	DdA1	DdA2	DdA3	DdA5	79B	88F	87E	57A	5C	42A
CcA1	--	0.032	0.031	0.025	0.008	0.017	0.028	0.022	0.028	0.035	0.038
DdA1	0.892	--	0.027	0.025	0.021	0.027	0.005	0.018	0.021	0.034	0.039
DdA2	1.053	1.116	--	0.008	0.025	0.028	0.021	0.009	0.005	0.035	0.035
DdA3	0.878	0.873	0.508	--	0.022	0.028	0.021	0.008	0.008	0.030	0.029
DdA5	1.069	1.133	1.295	0.988	--	0.013	0.017	0.017	0.022	0.036	0.038
Act79B	1.455	2.002	1.488	1.805	2.029	--	0.020	0.020	0.024	0.032	0.030
Act88F	1.169	1.589	1.385	1.246	1.707	0.589	--	0.017	0.019	0 031	0.028
Act87E	1.149	1.464	1.056	1.166	1.651	0.705	0.631	--	0.007	0.024	0.025
Act57A	0.988	1.484	0.710	0.903	1.427	0.678	0.625	0.519	--	0.029	0.029
Act5C	1.103	1.495	1.015	1.119	1.728	0.630	0.637	0.481	0.441	--	0.015
Act42A	1.647	3.264	1.484	1.936	2.257	0.965	1.071	1.241	1.061	1.076	--

One tree is constructed based on nonsynonymous distance using Saito and Nei's (1987) "neighbor joining (NJ) method" from the "Phylip" computer program (Figure 3.2.16). The NJ method is one of the distance matrix methods. The distance matrix methods assumes that there is some kind of functional relationships among the overall distance value and the evolutionary distance is related to the times of divergence.

This tree includes 12 actin gene sequences, 11 from the three Dipteran species and 1 from yeast. The input matrix is the pairwise comparison of the nonsynonymous substitution rates. The yeast actin gene is used as an outside group. With respect to the major clades identified, this tree is essentially identical to the tree generated using maximum parsimony method (Figure 3.2.14). Small changes are seen within the clade including Act87E, DdA3, DdA2 and Act57A genes. However, nonsynonymous distances are indeed a reflection of amino acid sequences. For a protein such as actin, however, with apparent high levels of selective constraints, it is not easy to weight the selection value. Selection may also be different for different members of actin genes or actin genes from different species. There is the also possibility of convergent evolution.

In order to obtain a tree which represents a true phylogenetic relationship synonymous substitutions may be more appropriate, providing that the synonymous substitutions are selectively neutral. A tree has been constructed based on synonymous distances using the Fitch and Margoliash method (1967) from the "Phylip" computer program (Figure 3.2.17). The Fitch-Margoliash method, like the NJ method, is also one of the distance matrix methods. The input matrix is a pairwise comparison of synonymous subsitution rates (Table 3.2.7). This tree includes the 11 actin genes from the three Dipteran species. Unlike the trees discussed above, this tree clusters all the *D. melanogaster* genes together. A tree based on total DNA sequences using the "PAUP" parsimony method also generates a similar tree (not shown).

Figure 3.2.16: Phylogenetic Tree of 12 Actin Genes Using the NJ Method.

This tree was constructed by the neighbor-joining method. Genetic distances were determined from nonsynonymous substitution rates of coding region sequences. Horizontal distances are proportional to evolutionary distances. Scale is shown on the left side. The tree was rooted by the yeast actin of *Saccharomyces cerevisiea*. Cc: *Ceratitis capitata*, A1: actin A1 gene; Dd: *Bactrocera dorsalis*; Dm: *Drosophila melanogaster*, Act79B: actin Act79B gene; YSC: yeast *Saccharomyces cerevisiea*.

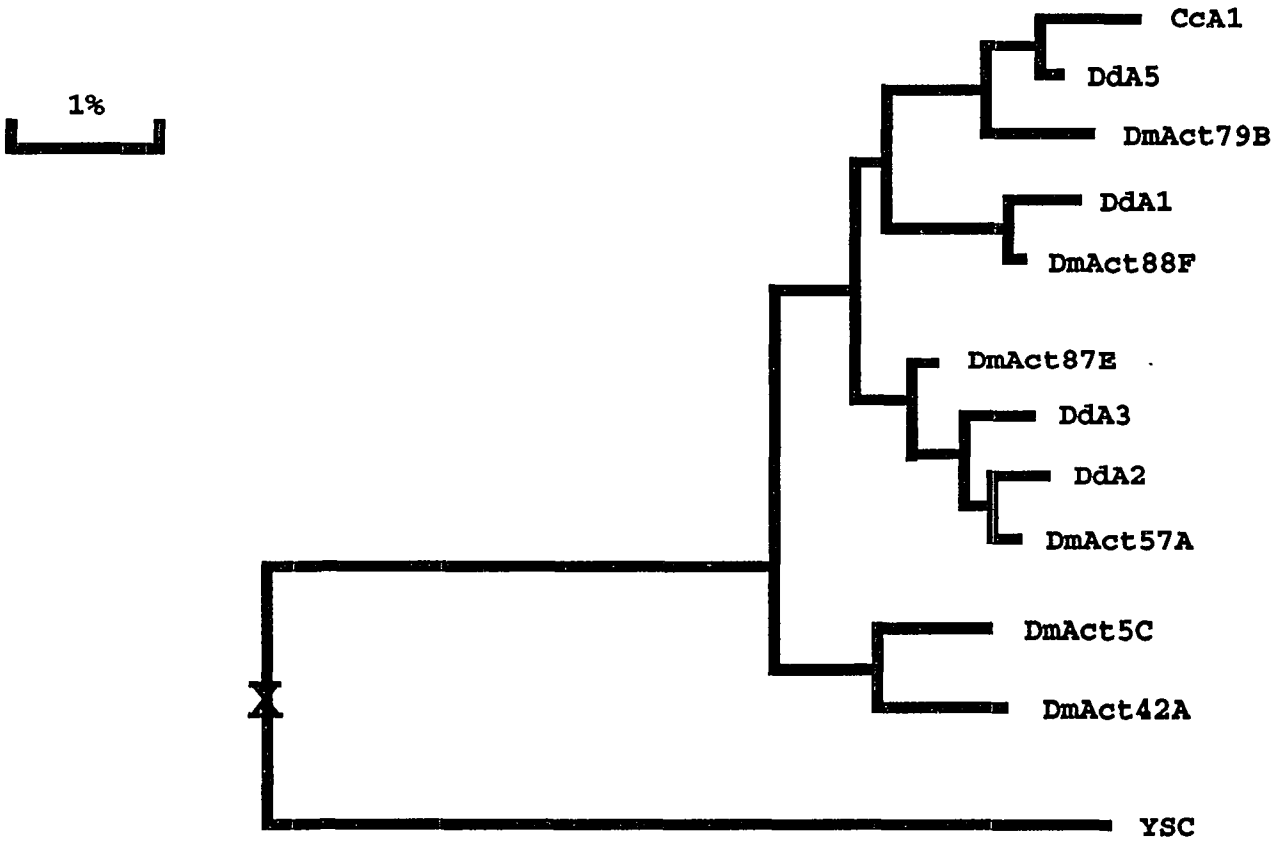
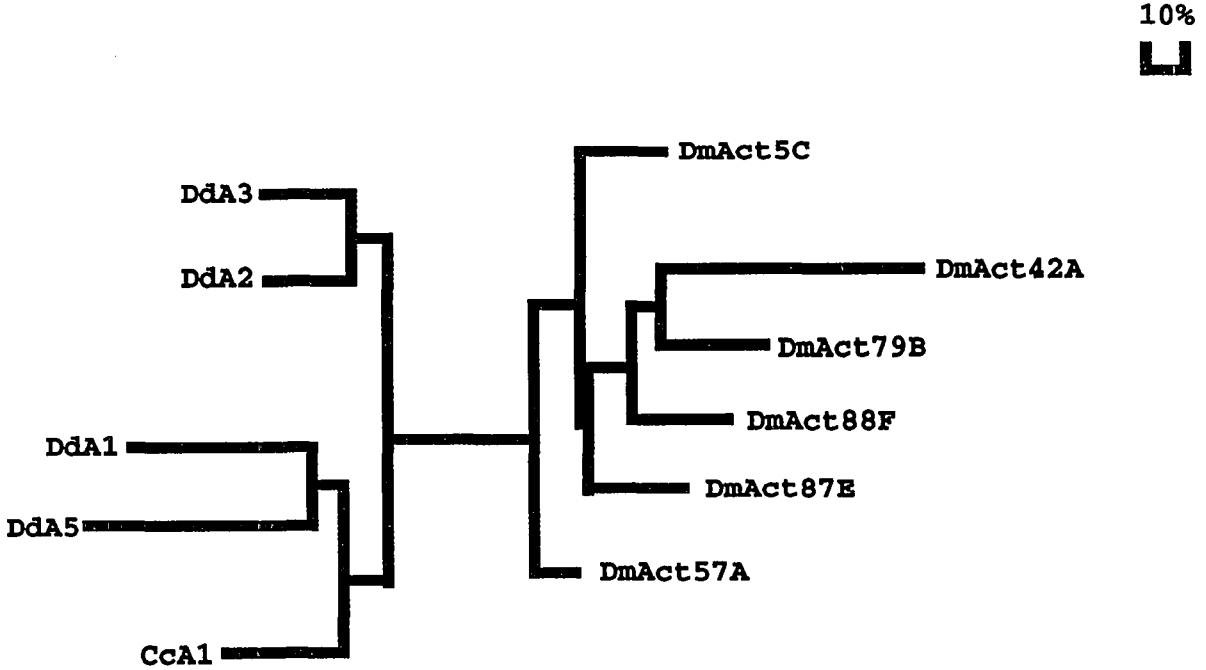

1%
CcA1
DdA5
DmAct79B
DdA1
DmAct88F
DmAct87E
DdA3
DdA2
DmAct57A
DmAct5C
DmAct42A
YSC

Figure 3.2.17: Phylogenetic Tree of 11 Actin Genes Using the Fitch-Margoliash Method.

This tree was constructed by the Fitch-Margoliash method. Genetic distances were determined from synonymous substitution rates of coding region sequences. Horizontal distances are proportional to evolutionary distances. Scale is shown on right side. This tree is unrooted. Cc: *Ceratitis capitata*, A1: actin A1 gene; Dd: *Bactrocera dorsalis*; Dm: *Drosophila melanogaster*, Act79B: actin Act79B gene.


10%
DmAct5C
DdA3
DmAct42A
DdA2
DmAct79B
DmAct88F
DdA1
DmAct87E
DdA5
DmAct57A
CcA1

The complete disagreement between this tree and the trees constructed based on protein sequences or nonsynonymous distances require an explanation. If the synonymous tree is more representative of the phylogenetic relationship, it is hard to explain the positions found for these actin genes. For example, the tree based on synonymous changes does not separate the two cytoplasmic actin genes (Act5C and act42) from the muscle specific actin genes, even within the *Drosophila* species. As discussed, trees based on protein sequences or nonsynonymous distance fit the intron position data very well. This can be taken as strong evidence for accepting the trees based on protein sequences or nonsynonymous distance over the tree based on synonymous distance.

3.2.3.2 Test for concerted evolution

"Concerted evolution" or "molecular drive" has been used to described the situation in which higher homology is found between multigene family members within a species as opposed to homologous genes between species (Hood et al., 1975; Dover, 1982; Dover and Tautz, 1986). The homogenization phenomenon found among the *Drosophila* actin genes as exhibited by the cluster in the synonymous distance tree (Figure 3.2.17) may be an example of such concerted evolution. Clearly, the non-independence of gene evolution in a multigene family, reflected by concerted evolution, has effects on the inference of phylogenetic relationships.

Several mechanisms have been shown to be responsible for concerted evolution such as unequal crossing over, gene conversion, transposition, and RNA-mediated exchanges (Hood et al., 1975; Dover, 1982; Dover and Tautz, 1986). All of them involve nonreciprocal exchanges. To detect whether nonreciprocal exchanges occurred between the actin genes from the three Dipteran species, a computer program "VTDIST" is used. This program, developed by Sawyer (1989), can be used as a statistical test for detecting gene conversion. Gene conversion is a process of nonreciprocal recombination

between two genes in which one genes is converted completely or partially by the other gene (Lewin, 1990). It has been suggested to be responsible for the homogenization in the γ-globin genes (Jefferys, 1979, Scott et al., 1984) and many other genes (Dover, 1982).

The results from the "VTDIST" are summarized in Table 3.2.8. In general, the program calculates the probability of random distributions of synonymous substitution. A test for 7 strains at GND locus in E.coli is used here as a positive control. The p value is zero for this set of genes and it indicates a high incidence of gene conversion. For the genes from the Dipteran species, three sets of test have been calculated, including within *D. melanogaster* actin genes, within *B. dorsalis* actin genes and all eleven genes from the three Dipteran species. All three tests have p value greater than 0.05. These results indicate an agreement of random distributions of synonymous substitutions and little chance of gene conversion. No evidence is found for gene conversion being responsible for the concerted evolution observed here. In addition, this kind of homogenization is not seen at the protein level, or in nonsynonymous changes. Apparently something else has an affect on the synonymous changes.

Table 3.2.8. Statistic for SSCF (Using VTDIST Program).

Genes/Species	Observed score	P value	SD above mean	SD of score
GND/E.coli	5,213	0.0000	7.21	203.00
Actin genes				
within D.m.	5,745	0.4675	0.02	224.44
within B.d.	1,598	0.3851	0.21	96.69
all 11 genes	36,620	0.7979	-0.85	867.82

*SSCF: the sum of the lengths of the condensed fragments, for all sets of silent polymorphic sites in a fragment for all pairs of sequences.

*P value: the proportion of permuted data sets (obtained from 10,000 random permutations of the orders of the s silent polymorphic sites) that have SSCF scores that are greater than or equal to the original score.

3.2.3.3 Test for codon bias

Codon bias is a phenomenon of nonrandom usage of synonymous codons (Grantham et al., 1980). In general it is species specific. Closely related species tend to show similar patterns of codon preference (Grantham et al., 1980). The choice of synonymous codons for a particular gene has been found to be related to the abundance of tRNA species in the cell. The correlation is especially strong for highly expressed genes (Post et al., 1979; Ikemura, 1981;1982). Because of codon bias, synonymous mutations may not be selectively neutral. The genes with higher codon bias experience high selective constraints by tRNA availability and other factors, and therefore may have lower rates of synonymous substitutions (Sharp and Li, 1986).

Codon bias has already been suggested from results in Table 3.2.5 where a lower ratio of transitions and transversions is seen among the genes within *D. melanogaster* species. It has been found in many studies that the ratio of transitions to transversions is usually around 2.0 between closely related species. A lower ratio indicates a high incidence of transversion substitutions. This high rate of transversion substitutions may be maintained by codon bias which limits substitutions to certain types of changes such as G<=>C or A<=>T. In the case of the *Drosophila* actin genes, the high GC content suggests that the G<=>C transversion substitutions may be favored. To estimate the codon bias in all the eleven genes from the three Dipteran species, the "scaled χ^2" method is used. This method was first introduced by Shields et al. (1988) when they examined the codon bias of 91 genes from *D. melanogaster.* In their studies they found a positive correlation between the scaled χ^2 and the GC content at synonymous sites. This correlation is not found with GC content at nonsynonymous sites and in intron regions. When they compared the genes between *D. melanogaster* and closely related sibling species, the genes with a higher scaled χ^2 showed reduced synonymous substitutions compared to the genes with a lower scaled χ^2.

The results of the test for the eleven actin genes is shown in Table 3.2.9. This table also shows the GC content of synonymous sites, nonsynonymous sites as well as the total GC content. Among *D. melanogaster* actin genes, five have very high scaled χ^2 value (0.813-0.966) and one (act42A) has a low scaled χ^2 (0.364). Among Tephritid actin genes, one (DdA2) has a very high scaled χ^2 (0.846) and other four have low scale χ^2 (0.267-0.508). The positive correlation between scaled χ^2 and the synonymous GC content is roughly there within *D. melanogaster* as compared to the similar GC content seen in nonsynonymous sites. This correlation is not obvious among the Tephritid actin genes.

Table 3.2.9. Codon Bias and GC Content Comparisons.

Actin gene	Scale (X^2)	$(G+C)_S$ %	$(G+C)_A$ %	$(G+C)_T$ %
Act57A	0.966	0.735	0.546	0.580
Act79B	0.920	0.814	0.562	0.610
Act88F	0.880	0.798	0.555	0.600
Act5C	0.830	0.759	0.551	0.590
Act87E	0.813	0.745	0.543	0.580
Act42A	0.364	0.647	0.529	0.550
DdA2	0.846	0.558	0.498	0.510
DdA3	0.508	0.535	0.493	0.500
DdA1	0.338	0.554	0.477	0.490
DdA5	0.267	0.506	0.503	0.500
CcA1	0.396	0.536	0.520	0.520

*Scale: deviation per codon

*$(G+C)_S$ %: GC content at synonymous sites

*$(G+C)_A$ %: GC content at nonsynonymous sites

*$(G+C)_T$ %: Total GC content

The scale χ^2 makes an indication of the extent of codon bias. The *Drosophila* actin genes tend to have much higher codon bias than the Tephritid actin genes, except for DdA2 and act42A. It is also possible to test if there are differences in codon preferences among the actin genes in *D. melanogaster* and the Tephritid species. For this purpose, "Monte Carlo 2xN contingency tests" were used. This test checks whether the codon usage in any pair of the 11 actin genes is statistically the same. Pairwise calculations of codon usage are summarized in Table 3.2.10. At least 10,000 trials were done for each pair. The numbers in the table are the probability (p) value for the pairwise comparisons. At the 0.05 significance level, a p value greater than 0.05 means there is no significant difference in codon usage between the pair while a p value lower than 0.05 means there is significant difference in the codon usage between the pair.

The table clearly documents that the five *Drosophila* genes which have a high scaled χ^2 are significantly different from the genes of the Tephritid species. The p values are not significant among the five *Drosophila* actin genes or among the four Tephritid genes which have a low scaled χ^2. The act42A gene is an exception in *Drosophila* while the DdA2 gene is an exception in Tephritids. The DdA2 gene has a much higher scale χ^2 than other Tephritid actin genes. It shows significant differences with all the actin genes compared here, except for DdA3. Although the DdA2 gene has high codon bias, it appears that this gene has a different pattern of codon preference from the *Drosophila* actin genes. This is also reflected in the low GC content for this gene. Act42 has much lower scaled χ^2 than any of the other *Drosophila* actin genes. It shows significance compared to some of the genes (DdA1, DdA2, DdA3, act57A) but not with others. From the synonymous substitutions, the 42A gene shows about twice the rate (1.08 on average) when compared with the other five *Drosophila* actin genes as opposed to the rate observed between the five genes (0.59 on average), even between the act5C gene which is also a cytoplasmic actin gene. This rate is about the same as the rate

observed between the Tephritid genes (0.98 on average). The rate is high between the five *Drosophila* actin genes and the Tephritid actin genes (1.37 on average). It appears that the five *Drosophila* actin genes have experienced a reduced synonymous substitution rate, possibly because of the selective constaints related to the codon bias. The codon bias may be responsible for the homogenization phenomenon observed among the *Drosophila* actin genes.

Table 3.2.10. Monte Carlo 2xN Contingency Table Test

p value	CcA1	DdA1	DdA2	DdA3	DdA5	79B	88F	87E	57A	5C	42A
CcA1	--										
DdA1	0.9950										
DdA2	0.0047	0.0007									
DdA3	0.9423	0.9546	0.7922								
DdA5	1.0000	1.0000	0.0010	0.8412							
Act79B	0.0001	0.0000	0.0000	0.0000	0.0000						
Act88F	0.0008	0.0000	0.0000	0.0000	0.0000	0.8546					
Act87E	0.0068	0.0000	0.0000	0.0000	0.0004	0.9481	0.9849				
Act57A	0.0001	0.0000	0.0000	0.0000	0.0000	0.6859	0.6299	0.9968			
Act5C	0.0010	0.0000	0.0000	0.0000	0.0000	0.8900	0.9836	0.9992	0.9525		
Act42A	0.4694	0.0030	0.0000	0.0007	0.3317	0.3403	0.0900	0.3348	0.0085	0.3445	--

3.3 ISOLATION OF GENE SPECIFIC PROBES

Gene specific probes were isolated by using the nested deletion method, and the regions were confirmed by sequencing. In the case of CcA1, the cDNA subclone pmedC1 was used as gene specific probe. Because the subclone (pmed5) for CcA2 gene only contains the 5' portion of the coding region, the 5' untranslated and flanking region was used as a gene specific probe. The subclone pmed5(5') is about 1.3kb in length, containing the sequences up to about 150bp before the start codon (see Figure 3.2.2).

The gene specific probes of the four genes from *B. dorsalis* were constructed using the 3' untranslated regions. The probe for DdA1 gene is named pDdA1(3'). This deletion subclone was constructed from an *Eco*RI-*Hin*dIII fragment in pDdA1.2. It is about 310bp in length, containing sequences from 80bp after the stop codon to the end of pDdA1.2 subclone at *Eco*RI site (see Figure 3.2.3). For the DdA2 gene, the subclone is named pDdA2(3'). This deletion subclone was constructed from an *Eco*RI-*Hin*dIII fragment in pDdA2. It is about 570bp in length, containing the sequences from 100bp after the stop codon to a *Hin*dIII site (see Figure 3.2.4). For the DdA3 gene, the subclone is named pDdA3(3'). It was constructed from an *Eco*RI-*Bgl*II fragment in pDdA3. It is about 530bp in length, containing the sequences from 60bp after the stop codon to a *Bgl*II site (see Figure 3.2.5). Two subclones have been constructed for DdA5 gene. The pDdA5(3')S deletion subclone was constructed from a *Pst*I-*Xba*I fragment in pDdA5. It is about 248bp in length, containing the sequences from 8bp after the stop codon to a *Xba*I site. The pDdA5(3')L was constructed by ligating adjacent *Xba*I-*Sst*I fragment in pDdA5 to the pDdA5(3')S subclone. It is about 740bp in length, containing the sequences in pDdA5(3')S subclone plus the *Xba*I-*Sst*I fragment (see Figure 3.2.6).

The specificity for each probe was tested by both computer analysis using MacVector as well as Southern Blotting. No apparent homology is seen between

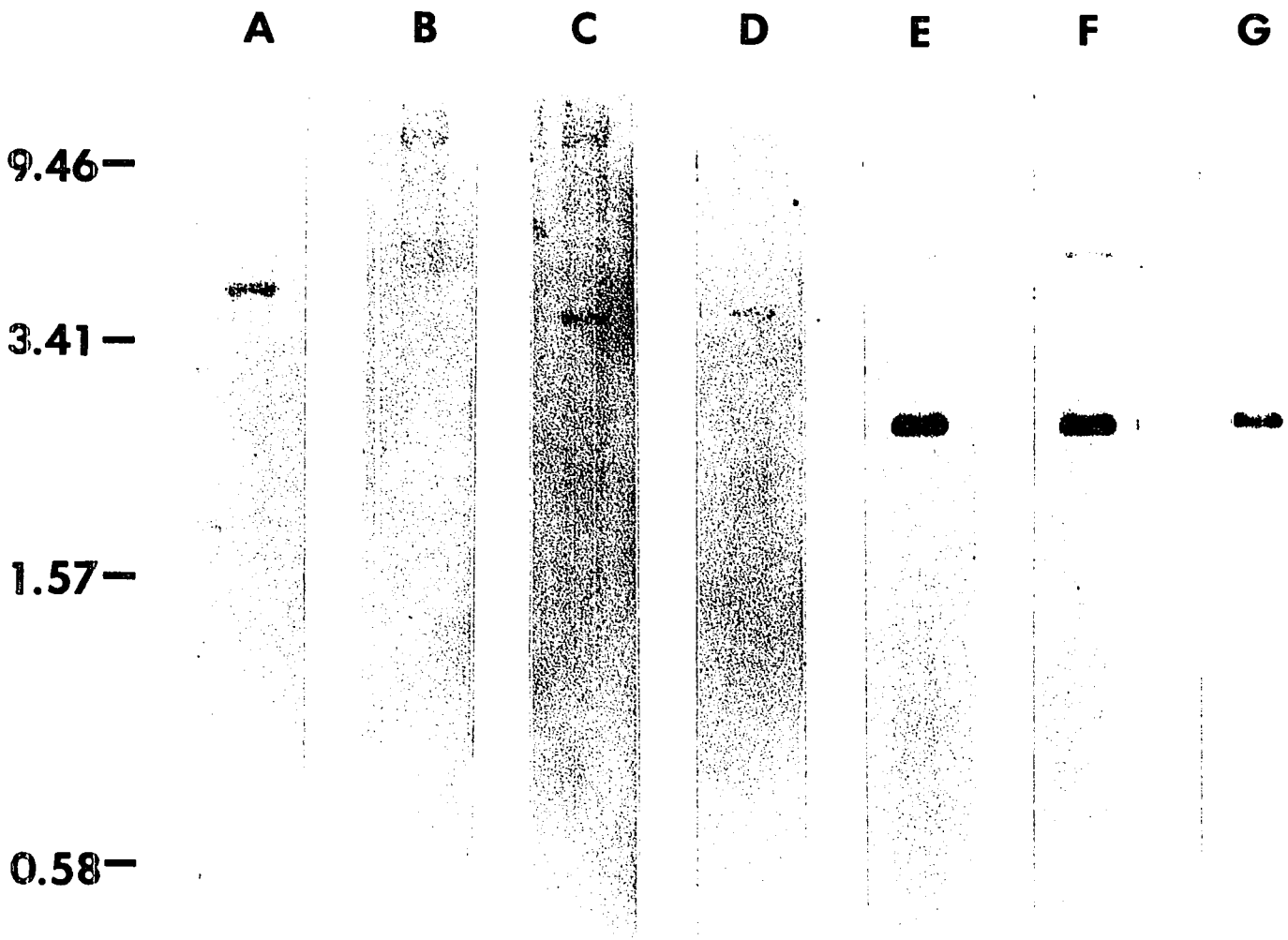
different probes within a species by the computer analysis. Southern blots were carried out under stringent conditions (50% formamide, 60°C washes). The results are shown in Figure 3.3.1. Each probe was hybridized to *Eco*RI digests of genomic DNA of either *C. capitata* (Lane A probed with pmed5(5') or *B. dorsalis* (Lane B through lane F probed with pDdA1(3'), pDdA2(3'), pDdA3(3') pDdA5(3')S and pDdA5(3')L respectively, lane G probed with pBR322 plasmid DNA). The specificity for pmedC1 has been sureyed in previous studies (He and Haymer, 1992). A single band of hybridization, which correspond to its own representitive gene of each probe, is seen for most probes except for the two DdA5 probes. In addition to the fragment representing the DdA5 gene, another hybridization band of higher molecular weight is seen. This band appears to correspond to the third band of the seven bands observed in the genomic *Eco*RI digests when a probe containing coding region was used (He and Haymer, 1991). At this point, it is not certain what the nature of this additional fragment is. It could be another actin gene which has homology in 3' untranslated region with the DdA5 gene. Or it could be a duplication of the DdA5 gene in the genome.

In addition, *B. dorsalis* species seems to have problems of bacterial infections potentially carrying plasmids. Recent extractions of the genomic DNA of *B. dorsalis* from different methods all showed the contamination with these plasmid homologous sequences. This was demonstrated by using only the plasmid DNA (pBR322) as a probe to a *B. dorsalis Eco*RI digest of genomic DNA which produced a single band of hybridization about 2.7kb in length, a size typical of pUC vectors (Lane G). This band is also seen in the two DdA5 probes where the recent extraction of the genomic DNA was used (Lane E and F). Even when only the insert fragment has been used to make the probe, some part of the polylinker sequences from the vector are still attached to the insert. Because of the lack of restriction enzyme sites in these deletion subclones as well

as the small size of the subclones, it is very difficult to isolate pure insert fragments lacking any vector sequences.

Figure 3.3.1: Genomic Southern for Testing Gene Specificity of Gene Specific Probes.

Genomic DNA samples (3ug) from *C. capitata* (Lane A) or *B. dorsalis* (Lane B to Lane G) were digested with the *Eco*RI restriction enzyme and fractionated on a 0.7% agarose electrophoresis gel. The DNA was transfered to a nylon filter by Southern blotting. Hybridization and detection was performed using the nonradioactive system with gene specific probes. The probes include: Lane A, pmed(5'); LaneB, pDdA1(3'); Lane C, pDdA2(3'); Lane D, pDdA3(3'); Lane E, pDdA5(3')S; LaneF, pDdA5(3')L; Lane G, pBR322. The size markers are indicated at the left side, using both *Eco*RI and *Eco*RI-*Hin*dIII digests of lambda DNA as standards.


A
B
C
D
E
F
G
9.46
3.41
1.57
0.58

3.4 DIFFERENTIAL PATTERNS OF EXPRESSION

Many studies have indicated that each member of the actin gene family shows a specific pattern of expression during development. Similar results from actin genes in Tephritid species can be used to make comparisons of expression patterns between actin genes with higher homology in their amino acid sequences. This can also be used to study correlation between protein structure and function. It may also provide evidence for identifying orthologous genes between species.

In order to detect the expression pattern in a temporal and spatial manner, RNA samples have been isolated from different stages as well as different body parts. In previous studies carried out in *C. capitata* pmed21 was used as a probe (Haymer et al., 1990). Although these studies indicated a typical muscle specific expression pattern, it was not clear that if the expression observed here came only from the CcA1 gene in pmed21 because the coding region was also included in the probe. The 3' or 5' gene specific probes should give a more precise expression pattern of each gene. The use of 3' untranslated region as a gene specific probe to detect differential patterns of expression of individual genes has been used in *D. melanogaster* (Fyrberg et al., 1983).

Total RNA samples from *C. capitata* and *B. dorsalis* were extracted from various temporal stages during development and from specific tissues. The samples from *C. capitata* were extracted in previous work (Haymer et al., 1990; He MS thesis). Material from early embryos refers to egg collections less than three hours old, late embryos are approximately 24 hours old. The late larval stage is 9 days after egg laying (AEL). Early pupal material is 10 days AEL, midpupal corresponds to 14 days AEL and late pupal stage is 20 days AEL. Whole adults and adult tissues (head, thorax, abdomen and leg) were obtained from material collected within 24 hours after adult emergence. For *B. dorsalis*, RNA samples were extracted in this work, including early embryos (within 3 hours), late embryos (about 24 hours), third instar larvae (12 day AEL), early pupae (13

day AEL), mid pupae (18 day AEL), late pupae (24 day AEL) and adults (one day after emergence and one week after emergence). RNAs were also extracted from different body parts of newly emerged adult flies including head, thorax, leg and abdomen. RNA slot blots were prepared, each loaded with about 3ug of RNA. The blots were then hybridized to ^{32}P labelled gene specific probes. The results are shown in Figure 3.4.1.

As can be seen, all the genes examined here exhibit differential patterns of expression in both temporal and spatial manners. Five of the genes (CcA1, DdA1, DdA2, DdA3 and DdA5) show muscle specific patterns of expression, and one (CcA2) shows a cytoskeletal pattern of expression. The expression of muscle specific actin genes is found to be in abundance corresponding with two musculature preparation phases, larval musculature and adult musculature. The larval musculature development starts from the mid-late embryonic stage and continues through out the whole larval stage. Adult musculature development starts from the late pupal stage and continues into the adult stage.

The five muscle specific actin genes show higher expression in developmental stages as well as tissues where considerable muscle differentiation is known to be occurring (Crossley 1978). For the CcA1 gene, abundant transcripts are detected only during the late pupal and newly emerged adult stages of development. This is also true for the DdA1 and DdA3 genes. For the DdA5 gene, abundant transcripts are also detected during late larval stage in addition to the late pupal and adult stages. For the DdA2 gene, abundant transcripts are detect during the late larval and late pupal stages. Only weak expression is detected in the adult stage. In terms of tissue specificity, abundant expression of the CcA1 gene is detectable only in samples derived from thoracic and leg tissues. For the DdA1 gene, abundant expression is detectable only in thorax. The remaining three genes (DdA2, DdA3, DdA5) show expression in all of the four tissues but with differing intensities. The expression of DdA2 is strong in head,

abdomen and leg. The expression of DdA3 is strong in leg and thorax. The expression of DdA5 is strong in leg, head and thorax.

The expression patterns of the four muscle specific actins genes from *D. melanogaster* have been reviewed in Chapter 1. The act88F gene is correlated with adult musculature and only generates the myofibrils for the construction of indirect flight muscles in the thorax (Mahaffey et al., 1985; Hiromi & Hotta, 1985). This gene has 100% similarity in protein sequence with the DdA1 gene. The patterns of expression of these two genes are very similar. Both of them show tissue specific expression only in thorax. This is also consistent with the relationship exhibited by the phylogenetic tree. The act79B gene shows a similar pattern of transcripts as that of act88F. Both of them show expression at about the same time during development in the thorax (Sanchez et al., 1983; Fyrberg et al., 1983). However, the expression of act79B gene expression is more broad and is most abundant in such tissues as leg muscles, direct flight muscles in thorax and the muscles which support the abdomen and head, and the male-specific muscles as well (Ball et al., 1987; Courchesne-Smith and Tobin, 1989). This gene shows a high degree of homology at the protein level with both CcA1 and DdA5. The expression pattern of CcA1 gene is very similar to that of 79B gene, However, the expression pattern of DdA5 gene is not so clear. In addition, the 3' gene specific probe of this also detect an extra fragment on the genomic DNA sothern blot (see Figure 3-3-1). Therefore, it is not clear at this moment whether the transcripts detected here are exclusively all from the DdA5 gene. The transcripts of both act57A and act87E accumulate in larval intersegmental muscle but at different levels and in different patterns (Fyrberg et al., 1983; Tobin et al., 1990). These are parts of the larval musculature which later on form the origin of the adult abdominal muscles. These two genes are most closely related to DdA2 and DdA3 genes in amino acid seequences respectively. The patterns of expression are also similar between these genes.

The expression of CcA2 represents a cytoplasmic actin gene. Abundant transcripts are detected in both early and late embryos. Detectable expression is also seen in early and mid pupal stages. These are stages corresponding to cytoplasmic actin accumulation. No expression is detectable in any of tissues from newly emerged adult samples. This gene is most closely related to the two cytoskeletal actin genes act5C and act42A from *D. melanogaster* based on the partial protein sequence. The expression patterns are also very similar between these three genes. Both act5C and act42A appear to peak during early-to-mid embryogenesis and early pupal development. They show no correlation with muscle differentiation. In addition, they have no expression in muscle rich tissues, but appear in undifferentiated tissues such as ovary and undifferentiated cell lines where cell division is occurring.

The patterns of expression detected for each gene are consistent with the phylogenetic tree based on the protein sequences (Figure 3.2.14 and Figure 3.2.15). It confirms that the DdA1, DdA5 and CcA1 genes, similar to the act79B and act88F genes, encode adult muscle actins. The DdA2 and DdA3 genes, similar to the act57A and act87E genes, encode larval-and-muscle actins. It also demonstrates a correlation between protein sequence and protein function of these genes. Figure 3.4.2 summarizes relationships of the 12 Dipteran actin genes, including phylogenetic comparisons, intron position data and the expression pattern of each gene.

The patterns of expression for CcA1, as detected by the gene specific probe pmedC1, is indeed more specific than those detected by pmed21 where coding region sequence is also included. When pmed21 was used as probe, intense hybridization was seen at the late third instar larval and late pupal stages as well as hybridization at the late embryonic and adult stages, also in tissues like leg, thorax and head (Haymer et al., 1990). This demonstrates that the 3' or 5' gene specific probe can be very useful in providing information of individual genes.

Figure 3.4.1: Differential Patterns of Expression of Actin Genes Detected by the Gene Specific Probes Using RNA Slot Blots.

RNA samples were extracted from various developmental stages as well as four different body parts. Material from early embryos refers to egg collections less than three hours old, late embryos are approximately 24 hours old. The late larval stage is 9 days after egg laying (AEL). Early pupal material is 10 days AEL, midpupal corresponds to 14 days AEL and late pupal stage is 20 days AEL. Whole adults and adult tissues (head, thorax, abdomen and leg) were obtained from material collected within 24 hours after adult emergence. 3ug of each RNA sample was subjected to electrophoresis on a formadehyde-agarose gel and transfered to a nitrocellulose filter. The gene specific probes were labeled with ^{32}P. The gene specific probe for each case is shown at the bottom.

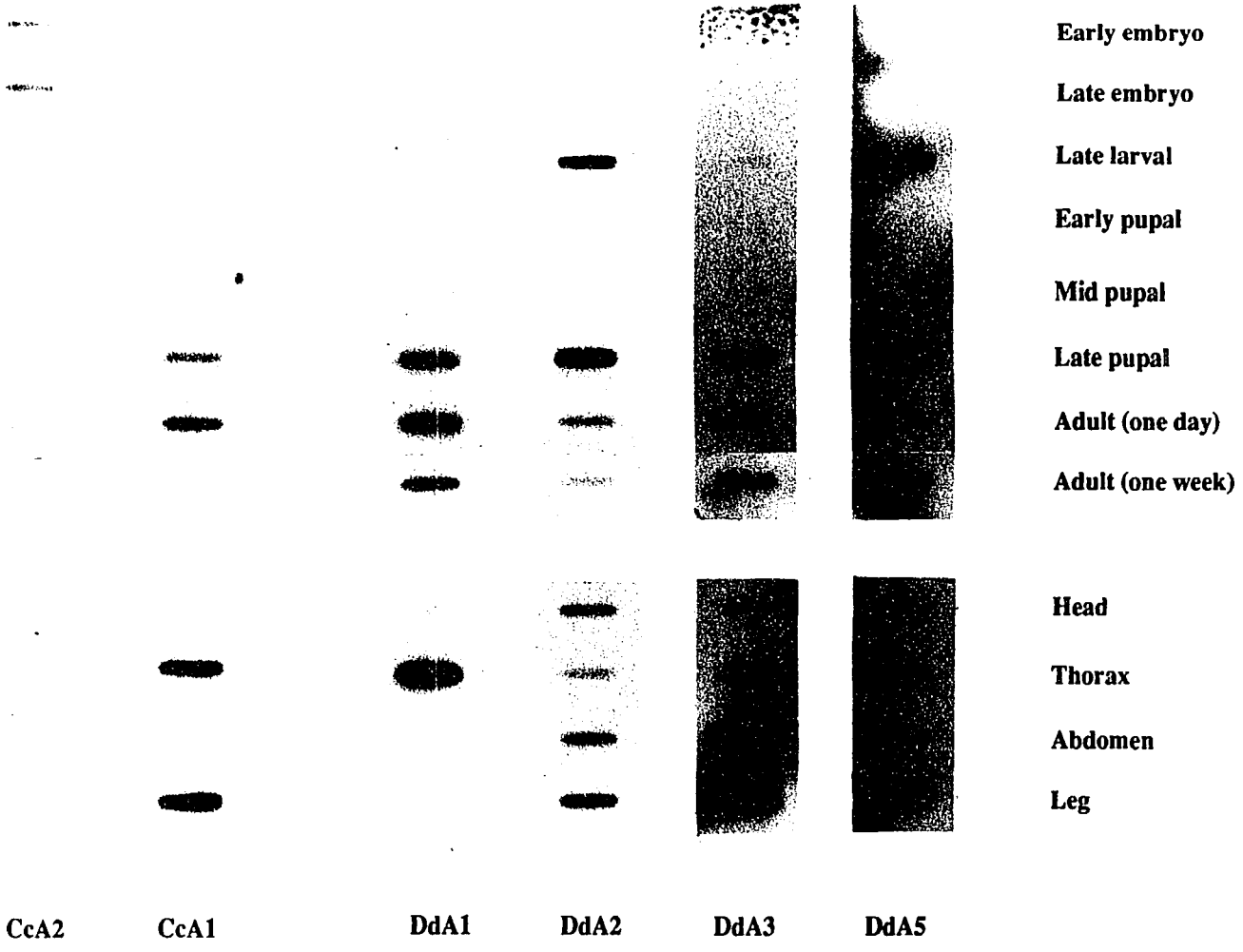
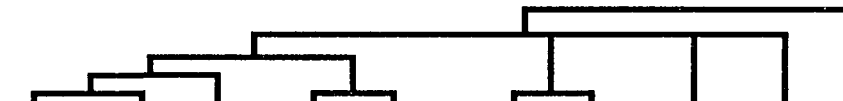

Early embryo
Late embryo
Late larval
Early pupal
Mid pupal
Late pupal
Adult (one day)
Adult (one week)
Head
Thorax
Abdomen
Leg
CcA2
CcA1
DdA1
DdA2
DdA3
DdA5

Figure 3.4.2: Summary of the Relationships of the 12 Dipteran Actin Genes.

The phylogenetic relationship are drawn based on the phylogenetic trees from amino acid sequences or nonsynonymous distances. The actin types are based on the studies of actin genes of *D. melanogaster*. Differential patterns of epression are summarized from this studies for the Tephritid actin genes and studies by Fyrberg et al. (1983) for *Drosophila* actin genes. A blank space indicates no detectable expression. Where expression was detected, a single plus sign represents weak expression while three pluses represent abundant expression. Material from early embryos refers to egg collections less than three hours old for Tephritid species, 0-2 hours for *D. melanogaster*, late embryos are approximately 24 hours old for Tephritid species, 18-24 hours for *D. melanogaster*. 1st instar larva was used for *D. melanogaster*, while the 3rd instar larva was used for Tephritid species.. Early pupal material is 10 days AEL, midpupal corresponds to 14 days AEL and late pupal stage is 20 days AEL for Tephritid species while it is about 6 days, 7 days and 8 days for *D. melanogaster*, respectively. Whole adults and adult tissues (head, thorax, abdomen and leg) were obtained from material collected within 24 hours after adult emergence for Tephritid species, while for *D. melanogster*, the whole adult was collected from 5 days after emergence and the tissues were collected from late pupae.



Genes	Act79B	CcA1	DdA5	DdA1	Act88F	DdA2	Act57A	DdA3	Act87E	Act5C	Act42A	CcA2
Types	Adult muscle actin					Larval-and-adult muscle actin				Cytoplasmic actin		
Intron	307	307	307	307	307	13	13	None	None	5'UTR	None	?
Expression												
Early embryo										++	++	++
Late embryo							++			+		++
Larval			++			++	++					
Early pupal										++	++	+
Mid pupal	++				++				++	++	++	+
Late pupal	+++	++	+++	+++	+++	+++	++	++	++			
Adult	+	++	++	+++	+	+	+	++	++	++	++	+
Head			+			++	++	+	++			
Thorax	+++	+++	++	+++	++	+		++				
Abdomen			+			++	++	+	++			
Leg	+++	+++	+++			++	+	+++	++			

3.5 LOCALIZATION OF ACTIN GENES BY IN SITU HYBRIDIZATION

3.5.1 Karyotype Studies in *B. dorsalis*

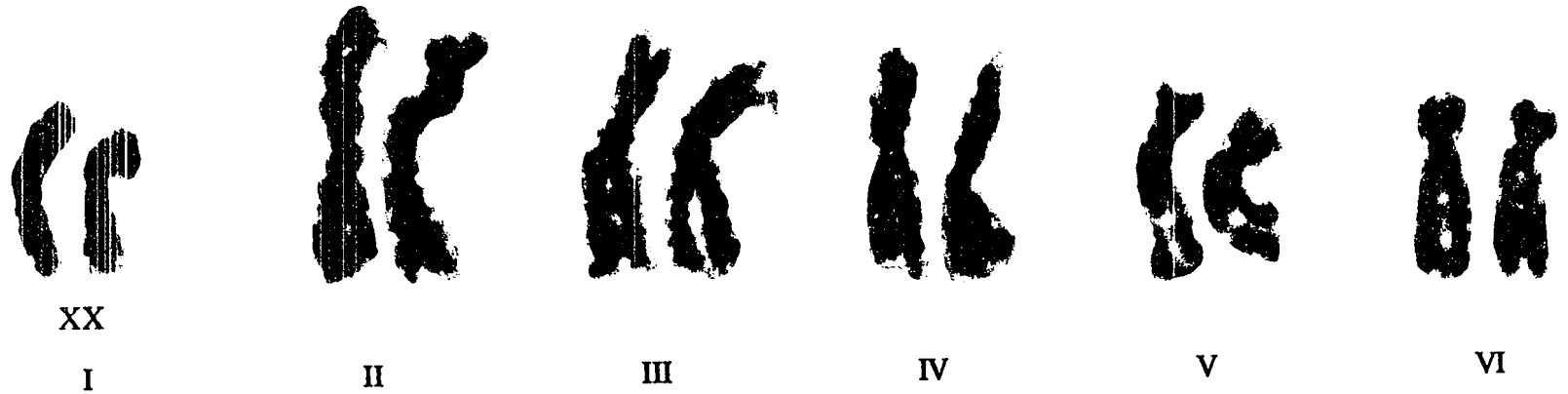
The karyotype of *C. capitata* has been very well characterized (Radu et al., 1975). However, there have been no studies reported yet in *B. dorsalis*. Before doing in situ hybridization, a karyotype study of this species was carried out. Brains dissected from third instar larval were used to collect mitotic chromosomes. Chromosomes were stained and visualized under microscope. The karyotype as chromosome pairs is shown in Figure 3.5.1 and the discription for each pair of chromosomes is summarized in Table 3.5.1. The chromosomes are numbered based on their relative sizes. The arm ratios are also listed in the table. Like all the Tephritid species studied so far, there are six pairs of chromosomes in *B. dorsalis*, including one pair of sex chromosomes. The Y is a small dot chromosome. Most chromosomes are submetacentric with arm ratios between 1.59 and 2.12, except for chromosome VI which is a metacentric chromosome with arm ratio of 1.10. In *C. capitata* the X chromosome is the largest chromosome and named as chromosome I (Radu et al., 1975), while in *B. dorsalis* the X chromosome is the smallest chromosome but still named as chromosome I.

Table 3.5.1. Karyotype Characterization of Mitotic Chromosomes in *B. dorsalis*.

Chromosome	relative size (%)	arm ratio	types of chromosome
Chromosome I (X)	9.27	1.83	submatecentric
Chromosome II	22.78	1.59	submatecentric
Chromosome III	20.62	2.12	submatecentric
Chromosome IV	17.19	1.10	matecentric
Chromosome V	15.59	1.79	submatecentric
Chromosome VI	14.55	2.04	submatecentic

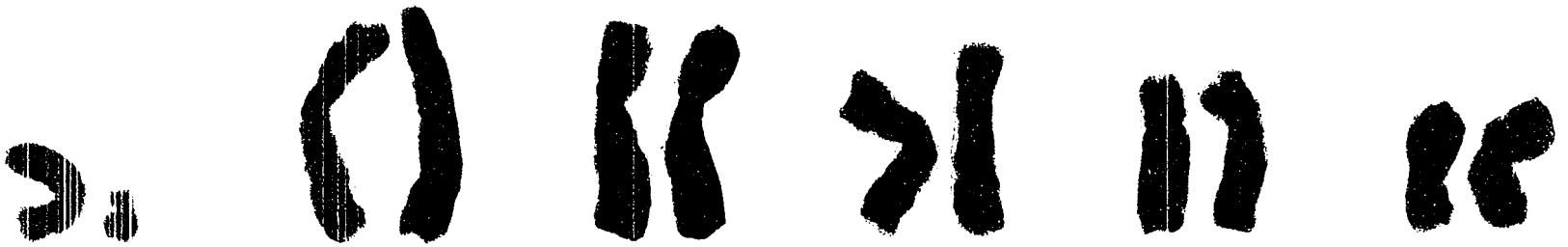
Figure 3.5.1: Mitotic Chromosomes and Karyotypes of *B. dorsalis*.

The upper portion shows the chromosome karyotype from a female fly (XX). The bottom portion shows the chromosome karyotype from a male fly (XY).



XX

I II III IV V VI



XY

I II III IV V VI

3.5.2 In Situ Hybridization

Actin gene organization in a genome is highly variable. The genes can be dispersed throughout the chromosomes as the case in *Drosophila* (Tobin et al., 1980; Fryberg et al., 1980), chicken (Cleveland et al., 1981), and mammals (Czosnek et al., 1983; Minty et al., 1983; Gunning et al., 1984b; Robert et al., 1985); or they can be in clusters as the case in sea urchin (Scheller et al., 1981) and the nematode *C. elegans* (Files et al., 1983; Krause & Hirsh, 1986). The distribution of different actin genes in these two Tephritid species can be used to address questions related to evolutionary history. In addition, the localization of the different actin genes can also be used as chromosomal markers, especially for *B. dorsalis* species.

To assign each actin gene to chromosomes, in situ hybridizations were carried out on mitotic chromosomes using the gene specific probes. Examples of hybridization results are shown in Figure 3.5.2 and 3.5.3. Figure 3.5.2 is the in situ hybridization on mitotic chromosomes from *B. dorsalis* probed with gene specific probes for the DdA1, DdA2, DdA3 and DdA5 genes while Figure 3.5.3 is the in situ hybridization on mitotic chromosomes from *C. capitata* probed with gene specific probes for the CcA1 and CcA2 genes. The arrow marked dark spots represent the sites of hybridization. The hybidization spots are shown on both homologous chromosomes in all the cases. The chromosome with positive hybridization for each gene specific probe is identified based on the relative sizes as well as the chromosome type if it can be recognized. The results are summarized in Table 3.5.2. In *B. dorsalis*, the DdA1 gene is located on chromosome IV, a metacentric chromosome. The DdA2 gene is located on the short arm of chromosome VI, a submetacentric chromosome. The DdA3 gene is difficult to identified, it could be on chromosome IV as the DdA1 gene, or it could be on the long arm of chromosome V, a submetacentric chromosome. The DdA5 is located on the long arm of chromosome III, a submetacentric chromosome. In *C. capitata*, the CcA1 gene is

located on the tip of chromosome III, a subtelocentric chromosome while the CcA2 gene is located close to the centromere on the short arm of chromosome IV. The distributions of these actin genes indicate that the actin genes in these two Tephritid species are dispersed throughout the genome rather than in a cluster.

Table 3.5.2. Localization of Actin Genes in *B. dorsalis* and in *C. capitata.*

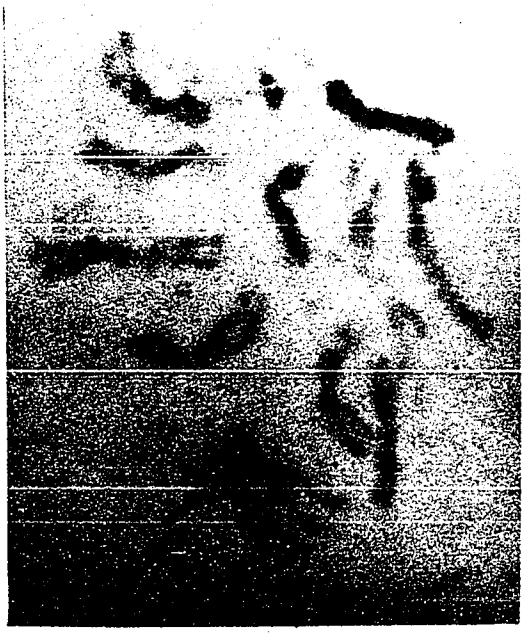
Actin genes	Location
DdA1	Chromosome IV
DdA2	Chromosome VI, short arm
DdA3	Chromosome IV, or Chromosome V long arm
DdA5	Chromosome III, long arm
CcA1	Chromosome III, on the tip
CcA2	Chromosome IV, short arm close to centromere

The locations of the six actin genes in *D. melanogaster* have been studied (see Loucas and Kafatos, 1986). The act79B gene is located on the left arm of chromosome 3; both act42A and act57A are located on the right arm of chromosome 2; both act 87E and act88F are located on the right arm of chromosome 3; act5C is located on the X chromosome. Mapping of the actin genes in six distantly *Drosophila* species including *D. melanogaster* indicated that six loci existed, and were similarly dispersed on four chromosomal elements. However due to the lack of studies on the relationships between the chromosomes from the three Dipteran species, it is impossible to make correlations on the localization of the actin genes in these species.

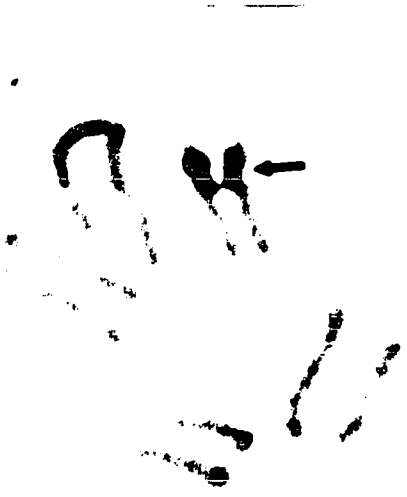
Figure 3.5.2: *In Situ* Hybridization on Mitotic Chromosomes from *B. dorsalis* Using Gene Specific Probes.

Mitotic chromosomes from *B. dorsalis* were immobilized on slides. Hybridization and detection was performed using the nonradioactive system with gene specific probes. Arrow pointed dark spots indicates the sites of hybridization. The probes include pDdA1(3') for DdA1; pDdA2(3') for DdA2; pDdA3(3') for DdA3; pDdA5(3')S or pDdA5(3')L for DdA5.

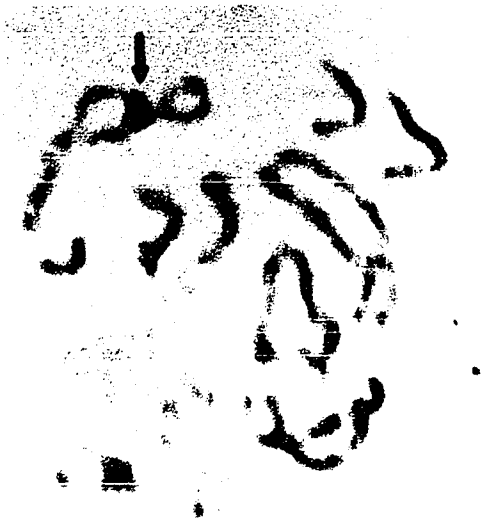
DdA1



DdA2



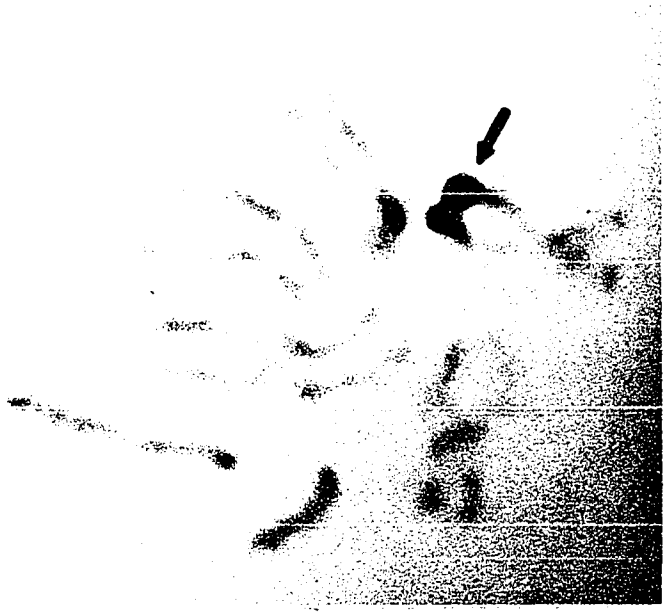
DdA3



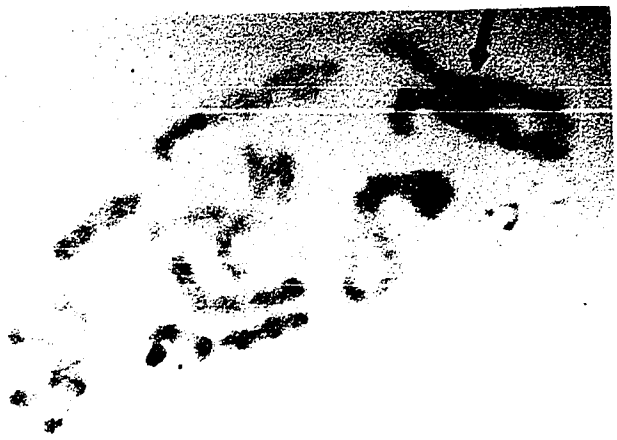
DdA5

Figure 3.5.3: *In Situ* Hybridization on Mitotic Chromosomes from *C. capitata* Using Gene Specific Probes.

Mitotic chromosomes from *C. capitata* were immobilized on slides. Hybridization and detection was performed using the nonradioactive system with gene specific probes. Arrow pointed dark spots indicates the sites of hybridization. The probes include pmdC1 for CcA1; pmed(5') for CcA2.



CcA1



CcA2

CHAPTER 4
CONCLUSIONS

Actin proteins perform critical functions in the cells of eukaryotic organisms. The genes encoding these proteins appear to have evolved conservatively, especially in the coding regions. Despite this, the existence of actin as a multigene family has allowed the divergence of certain sequences, especially in noncoding regions such as 5' and 3' untranslated regions. Intron positions among the different members of actin gene family have also been shown to be highly variable. Actin genes also display precisely regulated patterns of expression which differ among the members of the family, showing temporal and spatial specificity. The factors responsible for the specific expression of different members may be found in the 5' and 3' flanking and untranslated regions. In fact, studies in many organisms have indicated that these regions show specificity within species. Previous work has shown that there are multiple actin genes in both genomes of the Oriental fruit fly, *Bactrocera dorsalis* and the Mediterranean fruit fly, *Ceratitis capitata* (Haymer et al. 1990; He and Haymer, 1991). The evidence presented here shows that a total of six actin genes have been cloned from two Tephritid fruit fly species, four from the genome of the Oriental fruit fly *Bactrocera dorsalis* and two from the genome of the Mediterranean fruit fly *Ceratitis capitata.*

The DNA sequence data presented here has revealed several characteristics of these actin genes. To begin with, perhaps not surprisingly, the coding regions of these genes are highly conserved at the protein level when compared to actin genes from *Drosophila melanogaster*, exibiting 95.0% to 100% similarity. At the nucleotide sequence level, greater diversity is observed (9.0% to 22.0% diversity) for these same comparisons. This apparent discrepancy between these two types of comparison can, of course, be explained by a high level of synonymous changes seen in terms of the DNA

sequence evolution. However within these synonymous changes, some very different patterns emerge. For example, in the Tephritid species there is a clear preference for synonymous substitutions of C with T. Also, in the Tephritids, codons with T in the third position are, in most cases, used more frequently than in the *Drosophila* genes. This is consistent with the fact that the *Drosophila* actin genes have a higher GC content than the Tephritid actin genes.

In addition to the coding sequence homology, three of the Tephritid genes (DdA1, DdA5 and CcA1) have an intron in exactly the same position as act79B and act88F, two muscle specific actin genes of *Drosophila.* Another Tephritid gene (DdA2) has an intron in exactly the same position as act57A, another muscle specific actin gene of *Drosophila.* These particular intron positions have not been seen previously in other species. This suggests a common ancestral relationship for these genes, based on the sharing of intron positions. It has been previously suggested that the act79B and act88F genes may have arisen from a duplication event (Sanchez et al. 1983). The two *B. dorsalis* actin genes DdA1 and DdA5 may also reflect the same duplication event. If *C. capitata* also has a second actin gene with this intron positioning, it would be clear that the duplication event occurred before the divergence of these three species.

Beyond the coding region, however, there is little evidence for conservation of sequences, even in 5' and 3' flanking regions that are transcribed but not translated. The partial cDNA (pmedC1) sequenced here corresponds to the 3' untranslated region of the CcA1 actin gene. The sequence of the cDNA matches precisely with that of the genomic sequence, except for a CG dinucleotide which is reversed in the cDNA sequence. Previous work has shown that the CcA1 gene can be uniquely identified by this cDNA subclone (He, 1990, Master Degree Thesis; He and Haymer, 1992). This is consistent with what has been seen in actin genes from other species, e.g. that even while the coding regions of these genes are highly conserved, flanking regions (including regions

transcribed but not translated) are gene specific (Fyrberg et al., 1983; Minty et al., 1983). In other systems, such as *D. melanogaster,* these long 3' untranslated regions have provided unique sequence markers for monitoring the expression of individual actin genes during development (Fyrberg et al., 1983). Based on this, gene specific sequences have also been isolated from the other four complete Tephritid actin genes characterized here. For the CcA2 gene, 5' flanking region sequences have been isolated. These gene specific sequences, either from 3' untranslated regions or 5' untranslated regions, also represent unique sequences in the genome.

Using these gene specific probes, patterns of expression of individual genes have been documented in temporal and tissue specific terms. The pattern of expression exhibited by each gene specific probe during development suggests that five of the actin genes identified here encode muscle actins, and one encodes a cytoplasmic actin. Of the muscle actin genes, two encode larval-and-adult muscle actins while three encode adult muscle actins. The gene specific probes have also been used to localize the individual genes to chromosomes by in situ hybridization. The results suggest that the actin genes in these two Tephritid species are dispersed throughout the genome. This kind of distribution has been seen in many other species such as *Drosophila* (Tobin et al., 1980; Fryberg et al., 1980), chicken (Cleveland et al., 1981), and mammals (Czosnek et al., 1983; Minty et al., 1983; Gunning et al., 1984b; Robert et al., 1985).

In order to infer phylogenetic relationships between the actin genes characterized here and actin genes from other species, especially the *Drosophila* genes, phylogenetic trees have been constructed. Orthologous gene relationships revealed by these phylogenetic trees are consistent with the relationships suggested by comparisons of RNA expression patterns as well as by positioning of introns. The results indicate that CcA1 and DdA5 are orthologs of act79B; DdA1 is an ortholog of act88F; DdA2 is an

ortholog of act57A; DdA3 is an ortholog of act87E. The CcA2 analysis is incomplete because of the lack of complete coding region sequences.

The phenomenon of concerted evolution is evident among the actin genes from *D. melanogaster*. No evidence of gene conversion has been found on the basis of VTDIST analysis, although extreme codon bias has been found in most of the *Drosophila* actin genes. The same bias is not obvious in most of the Tephritid actin genes. In addition, there are substantial differences in codon preference between *D. melanogaster* and the two Tephritid species. Furthermore, the *Drosophila* actin genes with high codon bias appear to have experienced a reduced synonymous substitution rate. It has been suggested that because of the codon bias, synonymous mutations may not be selectively neutral. The concerted evolution observed among the *Drosophila* actin genes may reflect, at least in part, codon bias and differences in codon preferences among different species.

Thus despite the similarities between genes in these species, there are major differences between the Tephritid and *Drosophila* genomes, such as in codon usage and G-C content (He unpublished; Berry, 1985). The time of divergence between the Tephritidae and the Drosophilidae has been estimated at between 80 million years to 120 million years (Beverly and Wilson, 1984; Kwiatowski et al., 1992). The differences between these genomes may help to explain why, for example, *Drosophila* based gene transfer technologies do not function in Tephritid species (O'Brochta and Handler 1988; Haymer 1990). This also suggests that what is known about the molecular makeup of certain well characterized insect species such as *Drosophila* cannot be taken as a paradigm for all insects, and that individual species must be understood and appreciated in light of their own biology.

BIBLIOGRAPHY

Ankenbauer, T., Kleinschmidt, J. A., Walsh, M. J., Weiner, O. H. and Franke, W. W. (1989). Identification of a widespread nuclear actin binding protein. Nature 342, 822-25.

Bains, W., Ponte, P., Blau, H. and Kedes, L. (1984). Cardiac actin is the major actin gene product in skeletal muscle cell differentiation in vitro. Mol. Cell Biol. 4, 1449-53.

Baird, W. V. and Meagher, R. B. (1987). A complex gene superfamily encodes actin in *Petunia*. EMBO J. 6, 3223-31.

Ball, E., Karlik, C. C., Beall, C. J., Saville, D. L., Sparrow, J. C., Bullard, B., and Fyrberg, E. A. (1987). Arthrin, a myofibrillar protein of insect flight muscle, is an actin-ubiquitin conjugate. Cell 51, 221-28.

Beach, R. L. and Jeffery, W. R. (1990). Temporal and spatial expression of a cytoskeletal actin gene in the Ascidian *Stryela clava*. Dev. Genet. 11, 2-14.

Benton, W. D. and Davis, R. W. (1977). Screeing λgt recombinant clones by hybridization to single plaques in situ. Science 196, 180.

Bernatsky, R., and Tanksley, S. D. (1986). Genetics of actin-related sequences in tomato. Theor. Appl. Genet. 72, 314-21.

Berry, S. (1985). Insect nucleic acids. In "Comprehensive Insect Physiology, Biochemistry and Pharmacology", Kerkut, G and Gilbert, L. (eds), pp. 219-54, vol 10. Pergamon press, NY.

Bess, H. A. (1953). Status of *Ceratitis capitata* in Hawaii following the introduction of *Dacus dorsalis* and its parasites. Proc. Haw. Ent. Soc. vol 15, 221-34.

Beverly, S. and Wilson, A. (1984). Molecular evolution and the higher Diptera II. A time scale for fly evolution. J. Mol. Evol. 21, 1-13.

Bhattacharya, D., Stickel, S. K., and Sogin, M. L. (1991). Molecular phylogenetic analysis of actin genic regions from *Achlya bisexualis* (Oomycota) and *Costaria costata* (Chromophyta). J. Mol. Evol. 33, 525-36.

Blin, N. and Stafford, D. W. (1976). Isolation of high-molecular-weight DNA. Nucleic Acids Res. 3, 2303.

Boxer, L. M., Miwa, T., Gustafson, T. A., and Kedes, L. H. (1989a). Identification and characterization of a factor that binds to two human sarcomeric actin promoters. J. Biol. Chem. 264, 1284-1292.

Boxer, L. M., Prywes, R., Roeder, R. G., and Kedes, L. H. (1989b). The sarcomeric actin CArG binding factor is indistinguishable from the c-fos serum response factor. Mol. Cell. Biol. 9, 515-522.

Brawley, S. and Quatrano, R. (1979). Sulfation of fucoidin in *Fucus* embryos. Dev. Biol. 73, 193-205.

Breitbart, R. E., Nguyen, H. T., Medford, R. M., Destree, A. T., Mahdavi, V. and Nadal-Ginard, B. (1985). Intricate combinatorial patterns of exon splicing generate multiple regulated troponin T isoforms from a single gene. Cell 41, 67-82.

Buckingham, M. E. (1985). Actin and myosin multigene families: Their expression during the formation of skeletal muscle. Essays Biochem. 20, 77-109.

Buckingham, M. E. and Minty, A. (1983). Contractile protein genes. In " Eukaryotic genes: Their Structure, Activity and Regulation." Machean, N., Gregory, S. and Flavell, R. (eds.) Butterworths, London.

Bullock, B. P., Nisson, P. E. and Crain, W. R. (1988). The timing of expression of four actin genes and RNA polymerase III-transcribed repeated sequence is correct in hybrid embryos of the sea urchin species *Strongylocentrotus purpuratus* and *Lytechinus pictus*. Devl. Biol. 130, 335-47.

Burn, T. C., Vigoreaux, F.O. and Tobin, S. L. (1989). Alternative 5C actin transcripts are localized in different patterns during *Drosophila* embryogenesis. Developmental Biology 131, 345-55.

Campisi, J., Grey, H. E., Pardee, A. B., Dean, M. and Sonenshein, G. E. (1984). Cell-cycle control of c-myc but not c-ras expression is lost following chemical transformation. Cell 36, 241.

Campos, A., Bernard, P., Fauconnier, A., Landa, A., Gomez, E., Hernandez, R., Willms, K., and Laclette, J. P. (1990). Clonign and sequencing of two actin genes from *Taenia solium.* (Cestoda). Mol. Biochem. Parasitol. 40, 87-94.

Capco, D. G., Wan, K. M. and Penman, S. (1982). The nuclear matrix: Three dimensional architecture and protein composition. Cell 29, 847.

Carroll, S. L., Bergsma, D. J. and Schwartz, R. J. (1986). Structure and complete nucleotide sequence of the chicken a-smooth muscle (aortic) actin gene. J. Biol. Chem. 261, 8965-76.

Chang, K. S., Rothblum, K. N., and Schwartz, R. J. (1985). The complete sequence of the chichen α-cardiac actin genes: a highly conserved vertebrate gene. Nucleic Acids Res. 13, 1223-37.

Clark, T. G. and Merriam, R. W. (1977). Diffusible and bound actin in nuclei of Xanopus laevis oocytes. Cell 12, 883-91.

Cleveland, D. W., Hghes, S. M., Stubblefield, E., Kirsehner, M. W. and Varmus, H. E. (1981). Multiple α and β tubulin genes represent unlinked and dispersed gene families. J. Biol. Chem. 256, 3130-34.

Cleveland, D. W., Lopata, M. A., MacDonald, R. J., Cowan, N. J., Rutter, W. J. and Kirschner, M. W. (1980). Number and evolutionary conservation of α- and β-tubulin and cytoplasmic β- and γ-actin genes using specific cloned cDNA probes. Cell 20, 95-105.

Collins, F. S. and Weissman, S. M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. Proc. Natl. Acad. Sci. USA 81, 6812-16.

Courchesne-smith, C. L. and Tobin, S. L. (1989). Tissue-specific expression of the 79B actin gene during *Drosophila* development. Dev. Biol. 133, 313-21.

Cox, K. H., Angerer, L. M., Lee, J. J., Davidson, E. H., and Angerer, R. C. (1986). Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. J. Mol. Biol. 188, 159-172.

Crain, W. R., Boshar, M. F., Cooper, A. D., Durica, D. S., Nagy, A., and Steffen, D. (1987). The sequence of a sea urchin muscle actin gene suggests a gene conversion with a cytoskeletal actin gene. J. Mol. Evol. 25, 37-45.

Cross, G. S., wilson, C., Erba, H. P., and Woodland, H. R. (1988). Cytoskeletal actin gene families of *Xenopus borealis* and *Xenopus laevis*. J. Mol. Evol. 27, 17-28.

Crossley, A. C. (1978). The morphology and development of the *Drosophila* muscular system. In "The Genetics and Biology of *Drosophila*, Vol. 2b.", M. Ashburner and T. R. F. Wright (eds.), pp. 499-560. Academic Press, NY.

Cupples, G. C. and Pearlman, R. E. (1986). Isolation and characterization of the actin gene from *Tetrahymena thermophila*. Proc. Natl. Acad. Sci. USA 83, 5160-64.

Czosnek, H., Nudel, U., Mayer, Y., Barker, P. E., Pravtcheva, D. D., Ruddle, F. H. and Yaffe, D. (1983). The Genes coding for the cardiac muscle actin, the skeletal muscle actin and the cytoplasmic β-actin are located on three different mouse chromosomes. EMBO J. 2, 1977-79.

Davidson, E. H., Thomas, T. L., Scheller, R. H. and Britten, R. J. (1982). The *Sea Urchin* actin genes and a speculation on the evolutionary significance of small

gene families. In " Genome Evolution" Dove, G. A. and Flavell, R. A. (eds.) pp. 177-92. NY: Academic Press.

DePonti-Zilli, L., Seiler- Tuyns, A. and Paterson, B. M. (1988). A 40-base-pair sequence in the 3' end of the β-actin gene regulates β-actin mRNA transcription during myogenesis. Proc. Natl. Acad. Sci. USA 85, 1389-93.

DesGroseillers, L., Auslair, D., and Wickham, L. (1990). Nucleotide sequence of an actin cDNA gene from *Aplysia califonica*. Nucleic Acids Res. 18, 3654.

Dibb, N. J. and Newman, A. J. (1989). Evidence that introns arose at proto-slice sites. EMBO J. 8, 2015-21.

Dover, G. A. (1982). Molecular drive: A cohesive mode of species evolution. Nature 299, 111-17.

Dover, G. A. (1987). DNA turnover and the molecular clock. J. Mol. Evol. 26, 47-58.

Dover, G. A. and Tautz, D. (1986). Conservation and divergence in multigene families: alternatives to selection and drift. Phil. Trans. R. Soc. Lond. B. 312, 275-89.

Drouin, G., and Dover, G. A. (1987). A plant processed pseudogene. Nature 328, 557-58.

Drouin, G., and Dover, G. A. (1990). Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. J. Mol. Evol. 31, 132-50.

Drummond, D. R., Peckham, M., Sparrow, J. C., and White, D. C. S. (1990). Alteration in crossbridge kinetics caused by mutations in actin. Nature 348, 440-42.

Drummond, D. R., Hennessey, E. S. and Sparrow, J. C. (1991a). Stability of mutant actins. Biochem. J. 274, 301-30.

Drummond, D. R., Hennessey, E. S. and Sparrow, J. C. (1991b). Characterization of missense mutations in the Act88F gene of *Drosophila melanogaster*. Mol. Gen. Genet. 226, 70-80.

Dudler, R. (1990). The single-copy actin gene os *Phytophthora megasperma* encodes a protein consideratly diverged from any other known actin. Plant Mol. Biol. 14, 415-22.

Durica, D. S., Schloss, J. A., and Crain, W. R. (1980). Organization of actin gene sequence in the *Sea urchin*: molecular cloning of an intron-containing DNA sequence coding for a cytoplasmic actin. Proc. Natl. Acad. Sci. USA 77, 5683-87.

Efstradiatis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. and Proudfoot, N. J. (1980). The structure and evolution of the human b-gobin gene family. Cell 21, 653-68.

Egelman, E. H. (1985). J. Muscle Res. & Cell Motil. 6, 129-151.

Egly, J. M., Miyamoto, V. and Chambon, P. (1984). Is actin a transcription factor for RNA polymerase B? EMBO J. 3, 2363-71.

Einat, P., Shani, M., and Yaffe, D. (1990). The amount of the endogenous and exogenous skeletal muscle actin mRNA in the heart of transgenic mice is affected by the genotype of the cardiac actin gene. Differentiation 44, 36-41.

Elder, P. K., Schmidt, L. J., Ono, T. and Getz, M. J. (1984). Specific stimulation of actin gene transcription by epidermal growth factor and cycloheximide. Proc. Natl. Acad. Sci. USA 81, 7476.

Ellison, J., Buxbaum, J. and Hood, L. (1981). The nucleotide sequence of a human Immunoglobulin Cg4 gene. DNA 1(1), 11-18.

Engel, J. N., Gunning, P. W. and Kedes, L. (1981). Isolation and characterization of human actin genes. Proc. Natl. Acad. Sci. U.S.A. 78, 4674-78.

Engel, J. N., Gunning, P. W. and Kedes, L. (1982). Human cytoplasmic actin proteins are encoded by a multigene family. Mol. Cell. Biol. 2, 674-84.

Erba, H. P., Gunning, P. and Kedes, L. (1986). Nucleotide sequence of the human γ-cytoskeletal actin mRNA: anomalous evolution of vertebrate non-muscle actin genes. Nucleic Acids Res. 14, 5275-94.

Farmer, S. R., Wan, K. M., Ben-Ze'ev, A. and Penman, S. (1983). Regulation of actin mRNA levels and translation responds to changes in cell configuration. Mol. Cell. Biol. 3, 182.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6.

Feinberg, A. P. and Vogelstein, B. (1984). Addendum "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". Anal. Biochem. 137, 266.

Fidel, S., Doonan, J. H., and Morris, N. R. (1988). *Aspergillus nidulans* contains a single actin gene which has unique intron locations and encodes an actin. Gene 70, 283-93.

Files, J. G., Car r S. and Hirsh, D. (1983). Actin gene family of *Caenorhabolitis elegans*. J. Mol. Biol. 164, 355-76.

Fitch, W. M. and Margoliash, E. (1967). Construction of phylogenetic trees. A method based on mutation distances as estimated from cytochrome c sequences is of general applicability. Science 155, 279-84.

Fontes, M., Coulon, J., Delgrossi, M. H. and Thouveny, Y. (1983). Muscle dedifferentiation and contractile protein synthesis during post-traumatic

regeneration by *Owenia fusiformis* (Polychaete Annelid). Cell Differ. 13, 267-82.

Fornwald, J. A., Junico, G., Peng, I., and Ordahl, C. P. (1982). The complete nucleotide sequence of the chicken α-actin gene and its evolutionary relationship to the actin gene family. Nucleic Acids Res. 10, 3861-76.

Foster, G. G., Whitten, M. J., Konovalov, C., Arnold, J. T. A., Maffi, G. (1981). Autosomal genetic maps of the Australian sheep blowfly, *Lucilia cuprina dorsalis* R-D (Diptera: Calliphoridae), and possible correlations with the linkage maps of *Musca domestica* L and *Drosophila melanogaster* Mg. Genet Res. 37, 55-69.

Fowler, V. and Pollard, H. (1982). Chromaffin granule membrane-F-actin interactions are calcium sensitive. Nature 295, 336-39.

Frieden, C. and Patane, K. (1988). Mechanism for nucleotide exchang in monomeric actin. Biochemistry 27, 3812-2820.

Fritsch, E. F., Lawn, R. M. and Maniatis, T. (1980). Molecular cloning and characterization of the human beta-like globin gene cluster. Cell 19: 959.

Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. 170, 827-842.

Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S. and Davidson, N. (1981). The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. Cell 24, 107-16.

Fyrberg, E. A. and Donady, J. J. (1979). Actin heterogeneity in primary embryonic culture cells from *Drosophila melanogaster.* Dev. Biol. 68, 487-502.

Fyrberg, E. A., Kindle, K. L. Davidson, N. and Sodja, A. (1980). The actin genes of *Drosophila:* a dispersed multigene family.. Cell 19, 365-378.

Fyrberg, E. A., Mahaffey, J. W., bond, B. J. and Davidson, N. (1983). Transcripts of the six *Drosophila* actin genes accumulate in a stage- and tissue-specific manner. Cell 33, 115-23.

Gallwitz, D. and Sures, I. (1980). Structure of a split yeast gene: Complete nucleotide sequence of the actin gene in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. 77, 2546-50.

Garcia, R., Paz-Aliaga, B., Ernest S. G., and Crain, W. R. Jr (1984). Three sea urchin actin genes show different patterns of expression: muscle specific, embryo specific, and constitutive. Mol. Cell. Biol. 4, 840-45.

Garner, I., Sassoon, D., Vandekerckhove, J., Alonso, S. and Buckingham, M. E. (1989). A developmental study of the abnormal expression of a-cardiac and a-skeletal actins in the striated muscle of a mutant mouse. Dev. Biol. 134, 236-45.

Garrels, J. I. and Gibson, W. (1976). Identification and characterization of multiple forms of actin. Cell 9, 793-805.

Gasperi, G., Malacrida, A., Tosetti, M., and Milani, R. (1986). Enzyme variability: a tool for investigating the genome organization and the population strcture of *Ceratitis capitata*. In: Cavalloro, R. (ed.), "Fruit Flies of Economic Inportance 84". A. A. Balkema, Rotterdam, pp. 153-61.

Gershman, L. C., Selden, L. A., and Estes, J. E. (1986). High affinity binding of divalent cation to actin monomer is much stronger than previouly reported. Biochem. Biophys. Res. Commun. 135, 607-614.

Geyer, P. K. and Fyrberg, E. A. (1986). 5'-Flanking sequence required for regulated expression of a muscle-specific *Drosophila melanogaster* actin gene. Mol. Cell. Biol. 6, 3388-96.

Gilbert, W. (1978). News and views. Nature 271, 501.

Goldstein, L., Rubin, R. and Ko, C. (1977). The present of actin in nuclei: a critical appraisal. Cell 12, 601-08.

Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pave, A. (1980). Codon ctalog usage and the genome hypothesis. Nucleic Acids Res. 8, r49-r62.

Greenberg, M. E. and Ziff, E. B. (1984). Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. Nature 311, 433.

Gunning, P., Mohun, T., Ng, S. -Y., Ponte, P. and Kedes. L. (1984). Evolution of the human sarcomeric-actin genes: evidence for units of selection within the 3' untranslated regions of the mRNAs. J. Mol. Evol. 20, 202-214.

Gunning, P., Ponte, P., Blau, H. and Kedes, L. (1983). α-skeletal and α-cardiac actin genes are coexpressed in adult human skeletal muscle and heart. Mol. Cell. Biol. 3, 1985-95.

Gunning, P., Ponte, P., Kedes, L., Eddy, R. and Shows, T. (1984). Chromosomal location of the co-expressed human skeletal and cardiac actin genes. Proc. Acid. Sci. USA 81, 1813-17.

Gunning, P., Ponte, P., Okayama, H., Engel, S., Blau, H. and Kedes, L. (1983). Isolation and characterization of full-length cDNA clones for human a-, b-, and g-actin m RNA: skeletal but not cytoplasmic actins have an amino-terminal cysteine that is subsequently removed. Mol. Cell. Biol. 3, 787-95.

Gustafson, R. A. and Kedes, L. (1989). Identification of multiple proteins which interact with function regions of the human cardiac α-actin promoter. Mol. Cell. Biol. 9, 3269-83.

Gustafson, T. A., Miwa, T., Boxer, L., and Kedes, L. (1988). Interaction of nuclear proteins with muscle-specific regulatory sequences of the human cardiac α-actin promoter. Mol. Cell. Biol. 7, 4100-4119.

Hallauer, P.L., Hastings, K.E.M., Baldwin, A.S., Pearson-White, S., Merrifield, P.A. and Emerson, C. P., Jr. (1987). Closely related a-tropomyosis mRNA in Quail fibrolasts and skeletal muscle cells. J. Biol. Chem. 262, 3590-96.

Hamada, H., Petrino, M. G., and Kakunaga, T. (1982). Molecular structure and evolutionary origin of human cardiac muscle actin gene. Proc. Natl. Acad. Sci. USA 79, 5901-05.

Hanauer, A., Levin, M., Heillig, R., Daegelen, D., Kahn, A., and Mandel, J. L. (1983). Isolation and characterization of cDNA clones for human skeletal muscle α actin. Nucleic Acids Res. 11, 3503-16.

Hartl, D. L. and Clark, A. G. (1989). "Principles of Population Genetics", 2nd Ed. Sinauer Associates, Sunderland, MA.

Haymer, D. S. (1990). Work towards a transformation system for the Mediterranean fruit fly, *Ceratitis capitata.* Insect Mol. Genet. Newslet 5, 2-3.

Haymer, D. S., Anleitner, J., He, M., Thanaphum, S., Saul, S., Ivy, J., Houtchen, K., and Arcangeli, L. (1990). Actin genes in the Mediterranean fruit fly, *Ceratitis capitata.* Genetics 125, 155-60.

Hayward, L. J. and Schwartz, R.J. (1986). Sequential expression of chicken actin genes during myogenesis. J. Cell Biol. 102, 1485-93.

He, M. and Haymer, D. S. (1991). Isolation and characterization of actin homologous sequences from *Dacus dorsalis* (Hendel) (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 84, 601-607.

He, M. and Haymer, D. S. (1992). A muscle-specific actin gene from the Mediterranean fruit fly, *Ceratitis capitata.* Insect Mol. Biol. 1, 15-24.

Héilig, R., Perrin, F., Gannon, F., Mandel, J. L. and Chambon, P. (1980). The Ovalbumin gene family: structure of the X gene and evolution of duplicated split genes. Cell 20, 625-37.

Hightower, R. C. and Meagher, R. B. (1986). The molecular evolution of actin. Genetics 114, 315-32.

Hightower, R. C. and Meagher, R. B. (1985). Divergence and differential expression of soybean actin genes. EMBO J. 4, 1-8.

Hiromi, Y. and Hotta, Y. (1985). Actin gene mutation in *Drosophila*: Heat shock activation in the indirect flight muscles. EMBO J. 4, 1681-87.

Hiromi, Y., Okamoto, Y., Gehring, W. and Hotta, Y. (1986). Germline transformation with *Drosophila* mutant actin genes induces constitutive expression of heat shock genes. Cell 44, 293-301.

Hirono, M., Endoh, H., Okada, N., Numata, O., and Watanabe, Y. (1987). Clonign and sequencing of the *Tetrahymena* actin gene and identification of its gene product. J. Mol. Biol. 194, 181-92.

Holmes, K. C., Popp, D., Gebhard, W., and Kabsch, W. (1990). Atomic model of the actin filament. Nature 347, 45-49.

Holmes, D. S. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114, 193.

Hood, L., Campbell, J. H., and Elgin, S. C. R. (1975). The organization, expression and evolution of antibody genes and other multigene families. Annu. Rev. Genet. 9, 305-53.

Horovitch, S. J., Storti, R. V., Rich, A., and Pardue, M. L. (1979). Multiple actins in *Drosophila melanogaster*. J. Cell. Biol. 82, 86-92.

Hsu, C.-Y. and Frankel, F. R. (1987). Effect of estrogen on the expression of mRNAs of different actin isoforms in immature rat uterus. J. Biol. Chem. 262, 9594-9600.

Hu, M. C.-T., Sharp, S. B., and Davidson, N. (1986). The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein coding region. Mol. Cell. Biol.6, 15-25.

Iatrou, K., Tsitilou, S. G., Goldsmith, M. R. and Kafatos, F. C. (1980). Molecular analysis of the Gr^B mutation in *Bombyx mori* through the use of a chorion cDNA library. Cell 20, 659-69.

Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli*: Transer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optional for the *E. coli* translationa system. J. Mol. Biol. 151, 389-409.

Ikemura, T. (1982). Correlation between the abundance of yeast transer RNAs and the occurrence of the respective codons in its protein genes: Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. 158, 573-97.

Jefferys, A. (1979). DNA sequence variants in $^G\gamma$-, $^A\gamma$-, δ- and β-globin genes of man. Cell 18, 1-10.

Johannes, F.-J. and Gallwitz, D. (1991). Site-directed mutagenesis of the yeast actin gene: a test for actin function in *vivo*. EMBO J. 10(12), 3951-58.

Jones, C W. and Kafatos, F. C. (1980). Coordinately expressed members of two chorion multigene families are clustered, alternating and divergently orientated. Nature 284, 635-38.

Jowett, T. (1986). Preparation of nucleic acids. In "Drosophila, a practical approach", Roberts, D. B. (ed.), p276. IRL Press, Oxford, Washington, DC.

Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F., and Holmes, K. C. (1990). Atomic structure of the actin:DNase I complex. Nature 347, 37-44.

Kaine, B. P. and Spear, B. B. (1982). Nucleotide sequence of a macro-nuclear gene for actin in *Oxytricha fallax*. Nature 295, 430-32.

Kamada, S. and Kakunaga, T. (1989). The nucleotide sequence of a human smooth muscle alpha-actin (aortic type) cDNA. Nucleic Acids Res. 17, 1767.

Karlik, C. C., Mahaffey, J. W., Couto, M. D. and Fyrberg, E. A. (1984). Organization of contractile protein genes within the 88F subdivision of the *D. melanogaster* third chromosome. Cell 37, 469-81.

Karlik, C. C., Saville, D. L., and Fyrberg, E. A. (1987). Two missense alleles of the *Drosophila melanogaster* Act88F actin gene are strongly antimorphic but only weakly induce synthesis of heat shock proteins. Mol. Cell. Biol. 7, 3084-91.

Kelly, K., Cochran, B. H., Stiles, C. D. and Leder, P. (1983). Cell-specific regulation of the c-myc gene by lymphocyte mitogens and platelet-derived growth factor. Cell 35, 603.

Khrestchatisky, M., Djabali, M., Thouveny, Y. and Fontes, M. (1988). Expression of muscle actin genes in early differentiation stages of tail regeneration of the urodele amphibian *Pleurodeles waltlii*. Dell Differ. and Dev. 25, 203-12.

Kim, K., Gooze, L., Petersen, C., Gut, J., and Nelson, R. G. (1992). Isolation, sequence and molecular karyotype analysis of the actin gene of *Cryptosporidium parvum*. Mol. Biochem. Parasitol. 50, 105-14.

Kindle, K. N. and Firtel, R. A. (1978). Identification and analysis or *Dietyosteluim* actin genes, a family of moderately repeated genes. Cell 15, 763-78.

Kocher, O. and Gabbiani, G. (1987). Analysis of a-smooth muscle actin mRNA expression in rat aortic smooth-muscle cells using a specific cDNA probe. Differentiation 34, 201-13.

Koeller, D. M., Casey, J. L., Hentze, M. W., Gerhardt, E. M., Chan, L-N. L., Klausner, R. D., and Harford, J. B. (1989). A cytosolic protein binds to

structural elements within the iron regulatory region of the transferin receptor mRNA. Proc. Natl. Acad. Sci. USA 86, 3574-78.

Konsolaki, M., Komitopoulou, K., Tolias, P., King, D., Swimmer, C., and Kafatos, F. (1990). The chorion genes of the medfly, *Ceratitis capitata*, I: Structural and regulatory conservation of the s36 gene relative to two Drosophilar species. Nucl. Acids Res. 18, 1731-37.

Korn, E. D. (1978). Biochemistry of actomysin-dependent cell motility (A Review). Proc. Natl. Acad. Sci. USA 75, 588-99.

Kost, T. A., Theodorakis, N., and Hughes, S. H. (1983). The nucleotide sequence of the chick cytoplasmic β-actin gene. Nucleic Acids Res. 11, 8287-8301.

Krause, M. and Hirsh, D. (1986). The actin genes in *Caenorhabditis elegans*. In " Cell and Molecular Biology of the Cytoskeleton." Shay, J. W. (ed.) pp.151-78. NY: Plenum Press.

Krause, M., Wild, M., Rosenzweig, B., and Hirsh, D. (1989). Wild-type and mutant actin genes in *Caenorhabditis elegans*. J. Mol. Biol. 208, 381-92.

Kwiatowski, J., Skarecky, D., and Ayala, F. J. (1992). Structure and sequence of the *Cu, Zn Sod* gene in the Mediterranean fruit fly, Ceratitis capitata: Intron insertion/deletion and evolution of the gene. Mol. Phylogenet. Evol. 1, in press.

LeMeur, M., Glanville, N., Mandel, J. L., Gerlinger, P., Palmiter, R. and Chambon, P. (1981). The ovalbumin gene family: Hormonal control of X and Y gene transcription and mRNA accumulation. Cell 23, 561-71.

Levin, B (1990). "Genes IV". Oxford University Press, Oxford.

Li, W.-H, Wu, C.-I, and Luo, C.-C. (1985). A new method for estimating synonymous and nonsynoymous rates of nucleotide substitution considering the relative likelyhood of nucleotide and codon changes. Mol. Biol. Evol. 25, 330-42.

Loukas, M. and Kafatos, F. C. (1986). The actin CI in the genus *Drosophila:* establishment of chromosomal homologies among distantly related species by *in situ* hybridization Chromosoma. 94, 297-308.

Luscher, B., Stauber, C., Schindler, R., Schumperli, D. (1985). Faithful cell-cycle regulation of a recombinant mouse histone H_4 gene is controlled by sequences in the 3' terminal part of the gene. Proc. Natl. Acad. Sci. USA 82, 4389-93.

Maeda, N. and Smithies, O. (1986). The evolution of multigene families: Human haptoglobin genes. Annu. Rev. Genet. 20, 81-108.

Mahaffey, J. W., Coutu, M. D., Fyrberg, E. A. (1985). The flightless *Drosophila mutant* raised has two distinct genetic lesions affecting accumulation of myofibrillar proteins in flight muscles. Cell 40, 101-10.

Malacrida, A., Gasperi, G., Biscaldi, G. F., Milani, R. (1984). Functional significance of gene clusters in the housefly, *Musca domestica* and in other Diptera. Atti. 2' Congr. Soc. Ital. E.

Maness, P. E. and Walsh, R. C. (1982). Dihydrocytochalasin B disorganizes actin cytoarchitecture and inhibits initiation of DNA synthesis in 3T3 cells. Cell 30, 253.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular Cloning. Cold Spring Harbor. NY: Cold Spring Harbor Laboratory.

Markert, C. L. (1987). Isozymes and the regulatory structure of the genome. In "Current Topics in Biological and Medical Research', Markert, C. L. (ed.), vol 14, Molecular and Cellular Biology, Alan R. Liss, NY.

McCairns, E., Fahey, D., Muscat, G. E. O., Murray, M. and Rowe, P. B. (1984). Changes in levels of actin and tubulin mRNAs upon lectin activation of lymphocytes. Mol. Cell. Biol. 4, 1754.

McElroy, D., Rothenberg, M., Reece, K. S., and Wu, R. (1990a). Characterization of the rice (*Oryza sativa*) actin gene family. Plant Mol. Biol. 15, 257-68.

McElroy, D., Rothenberg, M., and Wu, R. (1990b). Structural characterization of a rice actin gene. Plant Mol. Biol. 14, 163-71.

McElroy, D., Zhang, W., Cao, J., and Wu, R. (1990c). Isolation of an efficient actin promoter for use in rice transformation. Plant Cell 2, 163-71.

McHugh, K. M. and Lessard, J. L. (1988). The development expression of the rat α-vascular and γ-enteric smooth muscle isoactins: Isolation and characterization of α rat g-enteric actin cDNA. Mol. Cell. Biol. 8, 5224-31.

McKenna, N., Meigs, J. B. and Wang, Y. L. (1985). Identical distribution of fluorescently labeled brain and muscle actins in living cardiac fibroblasts and myocytes. J. Cell. Biol. 100, 292.

McKeown, M. and Firtel, R. A. (1981). Evidence for subfamilies of actin genes in *Dictyostelium* as determined by comparisons of 3' end sequences. J. Mol. Biol. 151(4), 593-606.

Mckeown, M., Taylor, W. C., Kindle, K., Firtel, R. A., Bender, W. and Davidson, N. (1978). Multiple, heterogeneous actin genes in *Dictyostelium.* Cell 15, 789-800.

McKnight, T. D. (1983). "DNA sequences at the five prime ends of plants genes." Ph.D. Thesis, Univ. of Geogia, Athens, GA.

McLean, B. G., Eubanks, S., and Meagher, R. B. (1990a). Tissue specific expression of divergent actins in soybean root. Plant Cell J. 2, 335-44.

McLean, M. D., Baird, W. V., Gerats, A. G. M., and Meagher, R. B. (1988). Determination of copy number and linkage relationships among five actin gene subfamilies in *Petunia hybrida.* Plant Mol. Biol. 11, 663-72.

Mclean, M. D., Gerats, A. G. M., Baird, W. V., and Meagher, R. B. (1990b). Six actin gene subfamilies map to five chromosomes of *Petunia hybrida*. J. Heredity 81, 341-46.

Meagher, R. B. (1991). Divergence and differential expression of actin gene families in higher plants. Int. Rev. Cytol. 125, 139-163.

Minty, A. J., Alonso, S., Caravatti, M. and Buckingham, M. E. (1982). A fetal skeletal muscle actin mRNA in the mouse and its identity with cardiac actin mRNA. Cell 30, 185-92.

Minty, A. and Kedes, L. (1986). Upstream regions of the human Cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. Mol. Cell. Biol. 6, 2125-36.

Minty, A. J., Alonso, S., Guenet, J. L. and Buckingham, M. E. (1983). Number and organization of actin related sequences in the mouse genomes. J. Mol. Biol. 167, 77-101.

Minty, A., Blau, H. and Kedes, L. (1986). Two-level regulation of Cardiac actin gene transcription: muscle-specific modulating factors can accumulate before gene activation. Molecular and Cellular Biol. 6, 2137-48.

Miwa, T., Manabe, Y., Kurokawa, K., Kamada, S., Kanda, N., Bruns, G., Ueyama, H., and Kakunaga, T. (1991). Structure, chromosome location, and expression of the human smooth muscle (enteric type) γ-actin gene: evolution of six human actin genes. Mol. Cell. Biol. 11(6), 3296-3306.

Mogami, K. and Hotta, Y. (1981). Isolation of *Drosophila* flightless mutants which affect myofibrillar proteins of indirect flight muscle. Mol. Gen. Genet. 183, 409-17.

Mohun, T. J., Brennan, S., Dathan, N., Fairman, S., and Gurcon, J. G. (1984). Cell type-specific activation of actin genes in the early amphibian embryo. Nature 311, 716-21.

Mohun, T. J., Garrett, N., Gurdon, J. B. (1986). Upstream sequences required for tissue-specific activation of the cardiac actin gene in *Xenopus laevis* embryos. EMBO J. 5, 3185-93.

Mohun, T. J., Taylor, M. V., Garrett, N., and Gurdon, J. B.(1989). The CArG promoter sequence is necessary for muscle-specific transcription of the cardiac actin gene in *Xenopus* embryos. EMBO J. 8, 1153-61.

Mollenhauer, H. and Morre, D. (1976). Cytochalasin but not colchicine inhibits migration of secretory vesicles in root tips of maize. Protoplasma 87, 39-48.

Mounier, N., Gaillard, J., and Prudhomme, J. C. (1987). Nucleotide sequence of the coding regions of two actin genes in *Bombyx mori*. Nucleic Acids Res. 15, 2781.

Mounier, N., Gouy, M., Mouchiroud, D., and Prudhomme, C. (1992). Insect muscle actins differ distinctly from invertebrate and vertebrate cytoplasmic actins. J. Mol. Evol. 34, 406-15.

Mounier, N. and Prudhomme, J.-C. (1986). Isolation of actin genes in *Bombyx mori*: the coding sequence of cytoplasmic actin gene expressed in the silk gland is interrupted by a single intron in an unusual position. Biochimie 88, 1053-61.

Nairn, C. J., Winesett, L., and Ferl, R. J. (1988). Nucleotide sequence of an actin gene from *Arobidopsis thaliana*. Gene 65, 247-57.

Nakajima-Iijima, S., Hamada, H., Reddy, P., and Kakunaga, T. (1985). Molecular structure of the human cytoplasmic β-actin gene: interspecies homology of sequences in the introns. Proc. Natl. Acad. Sci. USA 82, 6133-37.

Nellen, W. and Gallwitz, D. (1982). Actin genes and actin messenger RNA in *Acanthamoeba castellani.* Nucleotide sequence of the split actin gene. J. Mol. Biol. 159, 1-18.

Ng, R. and Abelson, J (1980) Isolation and sequence of the gene for actin in *Sacchromyces Cervisiae.* Pro. Natl. Acad. Sci. U.S.A. 77, 3912-16.

Nisson, P. E., Dike, L. E. and Crain, W. R. (1989). Three *strongylocentrotus purpuratus* actin genes show correct cell-specific expression in hybrid embryos of *S. purpuratus* and *Lytechinus pictus.* Development 105, 407-13.

Novick, P. and Botstein, D. (1985). Phenotypic analysis of temperature-sensitive yeast actin mutants. Cell 40, 405-16.

Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z., and Yaffe, D. (1983). The nucleotide sequence of the rat cytoplasmic β-actin gene. Nucleic Acids Res. 11, 1759-71.

O'Brochta, D. and Handler, A. (1988). Mobility of P elements in Drosophilids and nondrosophilids. Proc. Natl. Acad. Sci. USA 85, 6052-56.

Ochman, H., Gerber, A. S., and Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. Genetics 120, 621-23.

Oefner, C. and Suck, D. (1986). Crystallographic refinement and structure of DNase I at 2A resolution. J. Mol. Biol.192, 605-632.

Ohta, T. (1982). Some evolutionary problems of multigene families considered from the standpoint of population genetics. In " Molecular Evolution, Protein Polymorphism and the Neutral Theory" Kimura, M. (ed.), pp. 59-77. Tokyo: Japan Scientific Press.

Ohta, T. (1991). Multigene families and the evolution of complexity. J. Mol. Evol. 33, 34-41.

Okamoto, H., Hiromi, Y., Ishikawa, E., Yamada, T., Isoda, K., Maekawa, H., and Hotta, Y. (1986). Molecular characterization of mutant actin genes which induce heat-shock proteins in *Drosophila* flight muscles. EMBO J. 5, 589-96.

Oldroyd, H. (1964). "The Nature History of Flies". Weidendield and Nicolson, London. 324pp.

Ordahl, C. P. and Cooper, T. A. (1983). Strong homology in promoter and 3'-untranslated regions of chick and rat a-actin genes. Nature 303, 348-49.

Ordahl, C. P., Tilghman, S. M., Ovitt, C., Fornwald, J. and Largen, M. T. (1980). Structure and developmental expression of the chick α-actin gene. Nucleic. Acids Res. 8, 4988-5005.

Pearson, L. and Meagher, R. B. (1990). Diverse soybean actin transcripts contain a large intron in the 5' untranslated leader: Structural similarity to vertebrate muscle actin genes. Plant Mol. Biol. 14, 513-25.

Perutz, M. F. (1983). Species adaptation in a protein molecule. Mol. Biol. Evol. 1, 1-28.

Pirrotta, V. (1986). Cloning *Drosophila* genes. In "Drosophila, a practical approach". Robert, D. B. (ed.), pp. 88. Oxford: IRL Press.

Pollard, T. D. and Cooper, J. A. (1986). Actin and actin-binding proteins. A critical evaluation of mechanisms and functions. Ann. Rev. Biochem. 55, 587-1035.

Pollard, T. D. (1990). Actin. Curr. Opin. Cell. Biol. 2, 33-41.

Pollard, T., Selden, C. and Maupin, P. (1984). interaction of actin filaments with microtubules. J. Cell Biol. 99, 335-75.

Ponte, P., Gunning, P., Blau, H. and Kedes, L. (1983). Human actin genes are single copy for a-skeletal and a-cardiac actin but multicopy for b- and g-cytoskeletal genes: 3' untranslated regions are isotype specific but are conserved in evolution. Mol. Cell. Biol. 3, 1783-91.

Ponte, P., Ng, S. -Y., Engel, J., Gunning, P. and Kedes. L. (1984). Evolutionary conservation in the untranslated regions of actin mRNAs: DNA sequence of a human beta-actin cDNA. Nucleic Acids Res. 12, 1687-96.

Post, L. E., Strycharz, G. D., Nomura, H., Lewis, H., and Dennis, P. P. (1979). Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit β in *Escherichia coli.* Proc. Natl. Acad. Sci. USA 76, 1697-1701.

Prokopy, R. J. and Roitberg, B. D. (1984). Foraging behavior of true fruit flies. American Scientist 72, 41-49.

Radu, M., Rassler, Y., and Koltin, Y. (1975). The chromosomes of the Mediterranean fruit fly *Ceratitis capitata* (Weid.): Karyotype and chromosomal organization. Cytologia 40, 823-28.

Reece, K. S., McElroy, D., and Wu, R. (1990). Genomic nucleotide sequence of four rice (*Oryza sativa*) actin genes. Plant. Mol. Biol. 14, 621-24.

Robert, B., Barton, P., Minty, A. J., Daubas, P., Weydert, A., Bonhomme, F., Catalan, J. Chazottes, D., Guenet, J-L. and Buckingham, M. E. (1985). Investigation of genetic linkage between myosin and actin genes using an interspecific mouse back-cross. Nature 413, 181-83.

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J. L., LeMeur, M., Bregegegre, F., Gannon, F., LePennec, J. P., Chambon, P. and Kourilsky, P. (1979). The Ovalbumin gene region: common features in the organization of three genes expressed in chicken oviduct under hormonal control. Nature 279, 125-32.

Saborio, J. L., Segura, M., Flores, M., Garcia, R. and Palmer, E. (1979). Differential expression of gizzard actin genes during chick embryogenesis. J. Biol. Chem. 254, 11119-25.

Saito, N. and Nei, M. (1987). The Neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406-25.

Sakai, Y., Okamoto, H., Mogami, K., Yamada, T., and Hotta, Y. (1990). Actin with tumor-related mutation is antimorphic in *Drosophila* muscle: two distinct modes of myofibrillar disruption by antimorphic actins. J. Biochem. 107, 599-505.

Sanchez, F., Tobin, S. L., Rdest,V., Zulauf, E. and McCarthy, B. J. (1983). Two *Drosophila* actin genes in detail gene structure, protein structure and transcription during development. J. Mol. Biol. 163, 533-51.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463.

Sartorelli, V., Webster, K. A., and Kedes, L. (1990). Muscle-specific expression of the cardiac α-actin gene requires MyoD1, CArG-box binding factor, and Sp1. Genes Dev. 4, 1811-1822.

Sawyer, S. (1989). Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6, 526-38.

Scheer, U., Hinssen, H., Franke, W. W., and Jockusch, B. M. (1984). Microinjection of actin-binding proteins and actinantibodies demonstrates involvement of nuclear actin in transcription of lampbrush chromosomes. Cell 39, 111-22.

Scheller, R. H., McAllister, L. B., Crain, W. R., Durica, D. S., Posakony, J. W., Thomas, T., Brillen, R. J. and Davidson E. H. (1981). Organization and expression of multiple actin genes in the *Sea Urchin*. Mol. cell. Biol. 1, 609-28.

Schroeder, T. (1976). Actin in dividing cells: Evidence for its role in cleavage but not mitosis. In " Cell Motility " Goldman, A. R., Pollard, T. and Rosenbaum, J. (eds.) Cold Spring Harbor, New York. Cold Spring Harbor Laboratory, pp. 265-77.

Schuler, M. A., McOsker, P., and Keller, E. B. (1983). DNA sequences of two linked actin genes of sea urchin. Mol. Cell. Biol. 3, 448-56.

Schwartz, R. J. and Ruthblum, K. N. (1981). Gene switching in myogenesis: Differential expression of the chicken actin multigene family. Biochem. 20, 4122-29.

Scott, A. L., Dinman, J., Sussman, D. J., and Ward, S. (1989). Major sperm protein and actin genes in free-living and parasitic nematodes. Parasitology 98, 471-78.

Scott, A. L., Heath, P. Trusko, S., Boyer, S. H., Prass, W., Goodmn, M., Czelusniak, J., Chang, L.-Y. E., and Slightom, L. (1984). The sequence of the gorilla fetal globin genes: Evidence for multiple gene conversions in human evolution. Mol. biol. Evol. 1, 371-389.

Shah, D. M., Hightower, R. C., and Meagher, R. B. (1982). Complete nucleotide sequence of a soybean actin gene. Proc. Natl. Acad. Sci. USA 79, 1022-26.

Shah, D. M., Hightower, R. C., and Meagher, R. B. (1983). Genes encoding actin in higher plants: intron positions are highly conserved but the coding sequences are not. J. Mol. Appl. Gen. 2, 111-26.

Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., Katcoff, D., Carmon, Y., Reiter, J., Frischauf, A. M. and Yaffe, D. (1981). Skeletal muscle actin mRNA. Characterization of the 3' untranslated region. Nucleic Acids Res. 9, 579-89.

Sharp, P. M. and Li, W.-H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evo. 24, 28-38.

Shaw, G. and Kamen, R. (1988). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46, 659-67.

Shields, D. C., Sharp, P. M., Higgins, D. G., and Wright, F. (1988). "Silent" sites in Drosophila genes are not neutral: Evidence of selection among synonymous codons. Mol. Biol. Evol. 5, 704-16.

Shortle, D., Haber, J. and Botstein, D. (1982). Lethal disruption of the yeast actin gene by integrative DNA transformation. Science 217, 371-73.

Shortle, D., Novick, P., and Botstein, D. (1984). Construction and genetic characterization of temperature-sensitive alleles of the yeast actin gene. Proc. Natl. Acad. Sci. USA 81, 4889-93.

Soloman, L. R. and Rubenstein, P. A. (1985). Correct NH_2-terminal processing of cardiac muscle α-isoactin (class II) in a nonmuscle mouse cell. J. Biol. chem. 260, 7659-7664.

Soloman, L. R. and Rubenstein, P. A. (1987). Studies on the role of actin's N^T-methylhistidine using oligodeoxynucleotide-directed site-specific mutagenesis. J. Biol. chem. 262, 11382-11388.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503.

Sparrow, J., Reedy, M., Ball, E., Kyrtatas, V., Molloy, J., Durston, J., Hennessey, E., and White, D. (1991). Functional and ultrastructural effects of a missense mutation in the indirect flight muscle-specific actin gene of *Drosophila melanogaster*. J. Mol. Biol. 222, 963-82.

Storti, R. V., Horovitch, S. J., Scott, M. P., Rich, A., and Pardue, M., L. (1978). Myogenesis in primary cell cultures from *Drosophila melanogaster*: Protein synthesis and actin heterogeneity during development. Cell 13, 589-98.

Stranathan, M., Hastings, C., Trinh, H., and Zimmerman, J. I. (1989). Molecular evolution of two actin genes from carrot. Plant Mol. Biol. 13, 375-83.

Strauch, A.R. and Reeser, J. C. (1989). Sequential expression of smooth muscle and sarcomeric α-actin isoforms during BC3H1 cell differentiation. J. Biol. Chem. 264, 8345-55.

Suck, D., Oefner, C., and Kabsch, W. (1984). Three-dimensional structure of bovine pancreatic DNase I at 2.5A resolution. EMBO J. 3, 2423-2430.

Swofford, D. L. (1990). Phylogenetic Analysis Using Parsimony, Version 3.0, Computer Program. Illinois Natural History Survey, Champaign.

Swofford, D. L. and Olsen, G. J. (1990). Phylogeny reconstruction. In "Molecular systematics", Hillis, D. S. and Moritz, C. (eds.), pp 411-501. Sinauer Associates, Inc., Sunderland, Massachusetts.

Tellam, R. L., Morton, D. J., and Clarke, F. M. (1989). A common theme in the amino acid sequences of actin and many actin-binding proteins? TIBS 14, 131-133.

Tobin, S. L., Cook, P. J., and Burn, T. C. (1990). Transcripts of individual *Drosophila* actin genes are differentially distributed during embryogenesis. Dev. genet. 11, 15-26.

Tobin, S. L., Zulauf, E., Sanchez, F., Craig, E. A. and McCarthy, B. L. (1980). Multiple actin-related sequences in the *Drosophila melanogaster* genome. Cell 19, 121-31.

Tolias, P. P., Konsolaki, M., Komitopoulou, K., and Kafatos, F. C. (1990). The chorion genes of the medfly, *Ceratitis capitata.* II. Characterization of three novel cDNA clones obtained by differential screening of an ovarian library. Devel. Biol. 140, 105-12.

Toyota, N. and Shimada, Y. (1983). Isoform variants of troponin in skeletal and cardiac muscle cells cultured with and without nerves. Cell 33, 297-304.

Ueyama, H., Hamada, H., Battula, N., and Kakunaga, T. (1984). Structure of a human smooth muscle actin gene (aortic type) with a unique intron site. Mol. Cell. Biol. 4, 1073-78.

Vandekerckhove, J., Bugaisky, G. and Buckingham, M. (1986). Simultaneous expression of skeletal muscle and heart actin proteins in various striated muscle tissues and cells. A quantitative determination of the two actin isoforms, J. Biol. Chem, 261, 1838-43.

Vandekerckhove, J. and Weber, K. (1980). Vegetative *Dictyostelium* cells containing 17 actin genes express a single major actin. Nature 284, 475-77.

Vandekerckhove, J. and Weber, K. (1978a). At least six different actins are expressed in a higher mammal: an analysis based on the amino acid sequence of the amino-terminal tryptic peptide. J. Mol. Biol. 126, 783-802.

Vandekerckhove, J. and Weber, K. (1978b). Mammalian cytoplasmic actins are the products of at least two genes and differ in primary structure in at least 25 identified positions from skeletal muscle actins. Proc. Natl. Acad. Sci. USA 75, 1106-10.

Vandekerckhove, J. and Weber, K. (1979a). The amino acid sequence of actin from chicken skeletal muscle actin and chichen gizzard smooth muscle actin. FEBS Lett. 102, 219-22.

Vandekerckhove, J. and Weber, K. (1979b). The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle and rabbit slow skeletal muscle. Differentiation 14, 123-33.

Vandekerckhove, J. and Weber, K. (1981). Actin typing on total cellular extracts: A highly sensitive protein-chemical procedure able to distinguish different actins. Eur. J. Biochem. 113, 595-603.

Vandekerckhove, J. and Weber, K. (1984). Chordate muscle actins differ distinctly from invertebrate muscle actins. J. Mol. Biol. 179, 391-413.

Van Zwaluwenberg, R. H. (1947). Notes and exhibitions. Proc. Hawaii. Ent. Soc. 13, 8.

Vigoreaux, J. O. and Tobin, S. L. (1987). Stage-specific selection of alternative trascriptional initiation sites form the 5C actin gene of *Drosophila melanogaster.* Genes and Devel. 1, 1161-71.

Wahli, W., Dawid, I. B., Wyler, T., Weber, R., and Ryffel, G. U. (1980). Comparative analysis of the structural organization of two closely related vitellogenin genes in *X. laevis.* Cell 20, 107-117. Cell 20, 107-117.

Whalen, R. G., Butler-Browne, G. S. and Gros, F. (1976). Protein synthesis and actin heterogeneity in calf muscle cells culture. Proc. Natl. Acad. Sci. USA 73, 2018-22.

Whalen, R. G., Butler-Browne, G. S., Sell, S. and Gros, F. (1979). Transition in contractile protein isozymes during muscle cell differentiation. Biochimie 61, 625-32.

Winrow, M. A., and Sodja, A. (1991). Initial characterization of several actin genes in the parasitic nematode, *Ascaris suum.* Biochem. Biophy. Res. Comm. 178, 578-85.

Woodcock-mitchell et al., (1988). Smooth muscle actin is transiently expressed in embryonic rat cardiac and skeletal muscles. Differentiation 39,161-66.

Woodrum, D. T., Rich, S. A., and Pollard, T. D. (1975). Evidence for biased unidirectional polymerization of actin filaments using heavy meromyosin by an improved method. J. Cell. Biol. 67, 231-237.

Yaffe, D., Nudel, U., Mayer, Y. and Neuman, S. (1985). Highly conserved sequences in the 3' untranslated region of mRNAs coding for homologous proteins in distantly related species. Nucleic Acids Res. 13, 3723-37.

Yamawaki-Kataoka, Y., Miyata, T. and Honjo, T. (1981). The complete nucleotide sequence of mouse immunoglobulin g22 gene and evolution of heavy chain genes: further evidence for intervening sequence mediated gene transfer. Nucl. Acids Res. 9, 1365-81.

Zacharopoulou, A. (1987). Cytogenetic analysis of mitotic and salivery gland chromosomes in the Medfly *Ceratitis capitata.* Genome 29, 67-71.

Zafar, R. S. and Sodja, A. (1983). Homology between actin coding and its adjacent sequences in widely divergent species. Biochem. Biophys. Res. commun. 1, 67-73.

Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D., and Nudel, U. (1982). The nucleotide sequence of the rat skeletal muscle actin gene. Nature 298, 857-59.

Zimmer, E. A., Martin S. L., Beverly S. M., Kan, Y. W., and Wilson, A. C. (1980). Rapid duplication and loss of genes for the alpha-chains of hemoglobin. Proc. Natl. Acad. Sci. USA 77, 2158-62.

Zimmerle, C. T., Patane, K., and Frieden, C. (1987). Divalent cation binding to high- and low-affinity sites on G-actin. Biochemistry 26, 6545-6552.

Zulauf, E., Sanchez, F., Tobin, S. L., Rdest, U.and McCarthy, B. J. (1981). Developmental expression of a *Drosophila* actin gene encoding actin I. Nature 293, 556-8.