8-2009

Differential Expression of Genes During Diapause in the Flesh Fly, *Sarcophaga crassipalpis*.

Puja Karki
East Tennessee State University

Differential Expression of Genes During Diapause in the Flesh Fly, *Sarcophaga crassipalpis*

A thesis

presented to

the faculty of the Department of Biological Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Biological Sciences

---

by

Puja Karki

August 2009

Dr. Karl H. Joplin, Chair

Dr. Hugh A. Miller III

Dr. Dhirendra Kumar

Key words: Suppression Subtractive Hybridization, Diapause, *Sarcophaga crassipalpis*

ABSTRACT

Differential Expression of Genes During Diapause

in the Flesh Fly, *Sarcophaga crassipalpis*

by

Puja Karki

The objective of this study was to identify genes that are differentially regulated during diapause when compared with nondiapausing pupae in *Sarcophaga crassipalpis*. The results of a Suppression Subtractive Hybridization procedure was used to indentify genes that are differentially regulated in both diapause and nondiapausing states while suppressing genes that are common to both states. Randomly picked colonies from both subtractive libraries were isolated and the inserts sequenced. The sequences were analyzed using the bioinformatics tools NCBI, BlastX, Clustal W, etc. Out of 384 clones, 59 genes were found to be upregulated during diapause and 37 genes were found to be upregulated during a nondiapause pupal stage, no genes were found to be expressed commonly in both the diapause and nondiapause constructed libraries.

DEDICATION

To my husband - Lok Raj Pokhrel, my daughter -Angel Pokhrel, and my parents - Jagan Nath Karki and Sabitri Karki.

ACKNOWLEDGMENTS

My first and foremost gratitude goes to my thesis advisor, Dr. Karl H. Joplin, Associate Professor, Department of Biological Sciences, East Tennessee State University for his scholastic guidance throughout the pursuit of this study and valuable feedbacks that helped improved this manuscript.

I would also like to sincerely thank my thesis committee members, Dr. Hugh A. Miller III and Dr. Dhirendra Kumar, for their valuable feedbacks and guidance during this study.

I would like to thank Department of Biological Sciences, East Tennessee State University, for providing me with Graduate Teaching Assistantship throughout my study period. I would also like to acknowledge the help provided by East Tennessee State University Graduate Council Research Grant, and HHMI grant that helped in making this study possible.

My sincere gratitude to my revered parents for bringing me thus far, and thanks also to my brothers, Sarjan Karki and Sagar Karki, for supporting me all the way.

Above all, I must thank my husband Lok Raj Pokhrel for bearing with me when I was with computer for hours and hours preparing this thesis, and encouraging me to take higher goals in life. I owe an immeasurable debt and deep affection to my beloved daughter Angel Pokhrel.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Many organisms have evolved to survive seasonally recurring unfavorable environmental conditions by going through a genetically programmed state; in insects this state is called diapause (Denlinger 1985). In particular, animals enter diapause as a response to environmental cue that predicts future environmental extremes. Once an animal enters diapause, development does not continue even when growth friendly conditions occur. Diapause is a dormant stage, which on one hand significantly influences the ability of an animal to resist extremes of environmental conditions, while additionally acting as a center around which growth, development, and reproduction are timed (Tauber et al. 1986). Interestingly, in mammals, diapause is an inducible, but yet reversible, program that can halt the propagation of early embryonic cells that have an inherent capacity to multiply at a rapid rate (Hondo et al. 2004).

Diapause in Insects

Many species of insects (including *S. crassipalpis*) use diapause as a strategy to overwinter. Diapause is specific in the sense that it occurs at a specific stage and during a specific season in an insect's life history. Insects undergoing diapause have an arrest in development along with a suppression of metabolic activity (Denlinger et al. 2005). In fact, diapause is not purely the cessation of growth and development, rather a distinct physiological state of a specific set of traits that are induced by the diapausing program. Such traits have been shown to be expressed in response to environmental stimuli that the animal perceives prior to diapause periods, and the expression of these traits is mediated via specific-endocrine changes (Tauber et al. 1986).

Diapause is a dynamic process that occurs in embryonic, larval, pupal, and adult life stages of many insects. It is characterized by a syndrome of changes such as consumption of

energy reserves, changes in pattern of oxygen consumption, ability to withstand environmental stress, and distinctive hormonal changes in gene expression (Denlinger et al. 2005). Before entering diapause, insects generally exhibit a prediapause developmental phase during which they acquire larger body size and develop enough waterproof cuticle and lipid reserves along with coloration to blend with the changing environments. Moreover, those insects that enter winter diapause incorporate adjustments in their metabolic activities that bolster cold hardiness (Denlinger et al. 2005).

Figure 1 shows the relationship between the photosensitive stage and the stage of diapause in three insect examples. The yellow box represents the stage where the circadian-gated photosensitive period is received, whereas the black box represents the diapausing stage. *S. crassipalpis* and *Sarcophaga argyrostoma* enter diapause in the pupal stage, but *Bombyx mori* enters diapause in the egg stage of its life cycle.

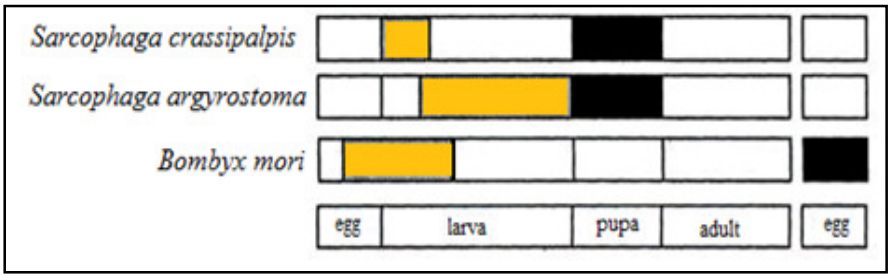


Figure 1 Photoperiodic Induction of Diapause for Different Species of Insects. Yellow box represents the photosensitive period and black box represents the diapause period. (Modified from Denlinger 1985; Nijhout 1998)

Insects that undergo diapause are totally dependent on the energy reserves that they have sequestered during earlier active phases of their life cycle (Flannagan et al. 1998). To extend their reserves, metabolism is typically suppressed to 10% of normal (nondiapause) levels. The polyols and other cryoprotective agents that are synthesized during diapause help them survive by reducing injury at low temperatures (Flannagan et al. 1998). This unique syndrome of events during diapause suggests that it is perhaps more appropriate to view diapause as an alternative developmental pathway and not just the halt in development

(Denlinger et al. 2005). In *S. crassipalpis*, the development of diapause-programmed (DP) and nondiapause-programmed (NDP) larvae is similar until the pupal stage. In approximately 2 weeks, NDP pupae go through metamorphosis, while DP pupae can remain in diapause for up to 2 years (Joplin, personal communication). It is assumed that insects make the decision to enter diapause during the preparatory phase and initiate necessary changes to accumulate additional energy reserves, generate waterproofing agents, and program specific behavioral patterns needed to find protective sites suitable for survival during overwintering (Denlinger et al. 2005).

Diapause in *Sarcophaga crassipalpis*

S. crassipalpis belongs to the order Diptera and family Sarcophagidae (Gr. *sarco-* corpse, *phage-* eating). It is multivoltine, i.e., more than one generation is produced a year. Thus, in order to have a specific diapause stage it needs an environmental signal to induce this alternative developmental stage.

The facultative pupal diapause in *S. crassipalpis* is induced by a circadian-gated, light /dark cycle (12:12 LD) (Denlinger 1972, Giebultowicz and Denlinger 1986) in which the dark phase must be at least 12 hours of uninterrupted darkness during a 24-hour period. Even short exposure of light during the scotophase is enough to prevent diapause initiation. But, this LD cycle must be received during a 4-day window of development including the last 2 days of embryonic development and the first 2 days of larval life (Denlinger 1972, Giebultowicz and Denlinger 1986). However, the observable initiation of the diapause programming occurs much later during the phanerocephalic pupal stage, approximately 3 days after puparium formation at 20°C (Denlinger 1981) and at least a week after the receipt of the LD signal. Interestingly, termination of diapause is mostly temperature dependent but can be broken by organic solvents (Denlinger 1972). Termination is accompanied by a rapid

decline in expression of the diapause upregulated genes and, conversely, an elevation in expression of many genes that are downregulated during diapause (Denlinger 2002).

The life cycle of this fly is shown in Figure 2. After mating, if the female fly receives 12 hours of light and 12 hours of dark during her photosensitive period, i.e., last 2 days of embryo development and first 2 days of larval development, i.e., L2 and L3, there will be the initiation of diapause program during the pupal stage that typically lasts for around 6 months, but can last for 2 years if low temperatures are maintained.

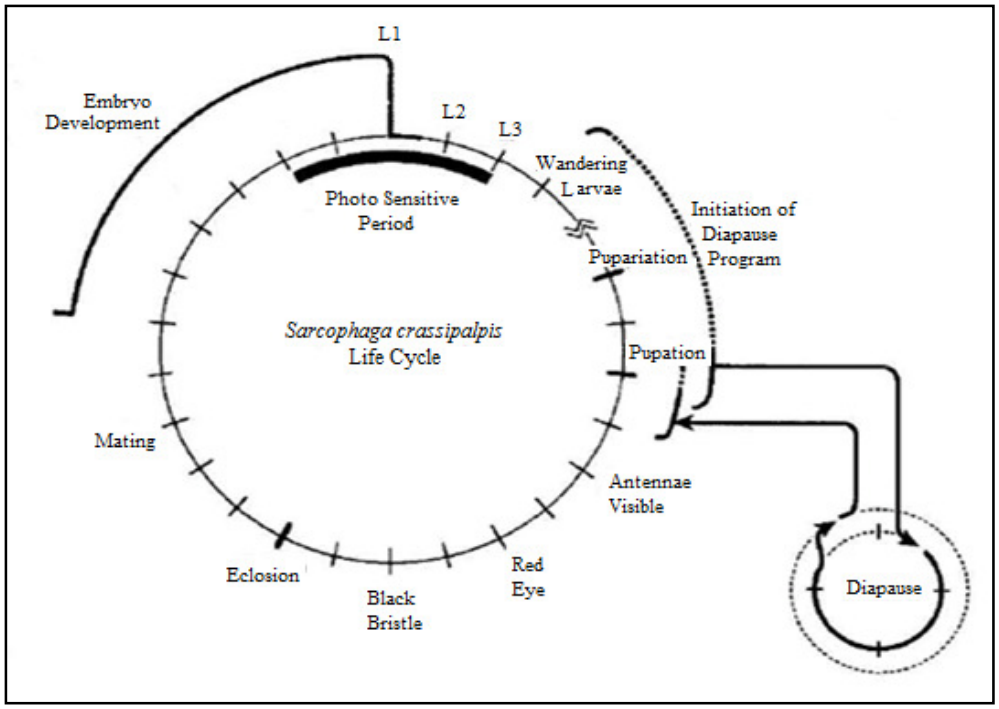


Figure 2 Life Cycle of *Sarcophaga crassipalpis* (Unpublished diagram, Karl Joplin)

Although animals are able to perceive day length and day number during the photosensitive phase, the mechanism underlying it is poorly known**.** The reception of light and the photoperiod clock might be regulated by the same elements that control the circadian clock because both mechanisms rely on measuring day length (Goto and Denlinger 2002b). It has been demonstrated that the underlying molecular mechanism involved in circadian clock consists of a set of genes that are controlled by feedback mechanisms where their protein

products, which form heterodimers, enter the cell nuclei at specific times during the day/night cycle and inhibit gene transcription (Goto and Denlinger 2002b). Several clock genes, *period, clock, cycle, doubletime, timeless,* and *cryptochrome,* have been identified and their roles in photosensitive period have been documented to be crucial, and more importantly they are presumed to preside over the programming of diapause (Denlinger 2002, Goto and Denlinger 2002b). When the flies are reared under short day or long day, in *S. crassipalpis*, clock genes like *period, cycle, clock,* and *cytochrome* are found to show little differences in expression patterns. However, the expression of *timeless* is greatly suppressed in long days as compared to short days. *Timeless* thus seems to be a potential player in the mechanism distinguishing short and long days (Goto and Denlinger 2002b).

Prothoracicotropic hormone (PTTH), the hormone from the brain that stimulates the prothoracic gland to produce ecdysone, is presumed to be shut down until the end of the diapause. Depending on the species, some metamorphic hormones such as juvenile hormone (JH) may or may not have the vital role in molting (Denlinger et al. 2005).

The production of ecdysone, juvenile hormone, and cyclic nucleotide activity shows distinct changes during the diapause developmental option (Denlinger 1981). During diapause, cycles of JH activity appear to be correlated with infradian cycles of O_2 consumption. This suggests the possible role of JH for the regulation of the length of diapause (Denlinger 1981). An examination of the cellular ultra structure of the ring gland, hormonal cells involved in the production and release of ecdysone and juvenile hormone verified a unique cellular organization during diapause (Joplin et al. 1993). The metabolic activity of flies going for pupal diapause drops to about 10% of the lowest activity as observed in nondiapausing pupae (Denlinger 1981). There is also a histological difference in tissue from diapausing pupae (Joplin et al. 1993).

It is this halt in nondiapause developmental processes and the induction of a syndrome of developmental and physiological processes that characterizes this strong facultative diapause process. This implies that there should be developmentally regulated gene expression patterns that are inducing these alternative processes. The identification of these genes is the basis of my research.

Diapause represents a shutdown in the expression of many genes as well as the upregulation of certain genes (Joplin et al. 1990; Flannagan et al. 1998). One important change in gene expression noted during diapause in *S. crassipalpis* is the downregulation of the cell cycle regulator, proliferating cell nuclear antigen (PCNA), an event that is involved in the G0/G1 cell cycle arrest (Tammariello and Denlinger 1998). Diapausing individuals demonstrate striking differences in gene expression when compared with nondiapausing ones in that many genes are silenced but a few are highly upregulated during diapause. Classes of diapause upregulated genes can be distinguished based on their expression patterns. Some are upregulated throughout diapause, and others are expressed only in early diapause, late diapause, or intermittently throughout diapause (Denlinger 2002). Two-dimensional gel electrophoresis of proteins, subtractive hybridization, and differential display of mRNA suggest that approximately 4% of the genes in flesh fly pupae are diapause upregulated and the expression patterns fall into several distinct categories (Denlinger 2002) such as genes unaffected by diapause, genes downregulated throughout diapause, genes upregulated throughout diapause, early diapause genes, late diapause genes, and those expressed irregularly throughout diapause. Genes unaffected by diapause in the flesh fly include those encoding ecdysone receptor (EcR), heat shock 70 cognate protein (HSC70), and 28S ribosomal protein (Denlinger 2002).

Denlinger (1991) has reported that diapausing insects are more resistant to stressful conditions than nondiapausing insects**.** Such resistance is thought to be linked with the

upregulation of many large and small heat shock proteins including HSP70 (Rinehart et al. 2000; Hayward et al. 2005) and Sc HSP23 (Denlinger et al. 1991; Yocum et al.1991, 1998; Kimura et al. 1998; Reinhart et al. 2000), while HSP90 was found to be downregulated during diapause in *S. crassipalpis* (Rinehart and Denlinger 2000). Diapausing insects are also found to exhibit higher resistance to free radicals associated with higher expression levels of superoxide and catalase (Morris et al.1996 and references therein).

Four diapause upregulated clones, seven diapause downregulated clones, eight clones expressed equally in both diapause and nondiapause have been found from brain mRNA of diapausing pupae of *S. crassipalpis* (Flannagan et al. 1998). Sequencing has identified homology to stress response genes, cell cycle control genes and DNA repair genes from the same (Flannagan et al. 1998).

Joplin and Denlinger (1989) and Li et al. (2007) reported 37 diapause upregulated proteins and 43 diapause downregulated proteins. HSP70 were found to be most abundant during diapause.

Figure 3 shows the expression patterns of different genes and different stages in S. crassipalpis. *ecr and hsc70* genes are not influenced by diapause, *pcna and hsp70* genes are diapause downregulated, *pScD41* is early diapause gene, *usp* is late diapause gene, and *po* gene is expressed intermittently during diapause .

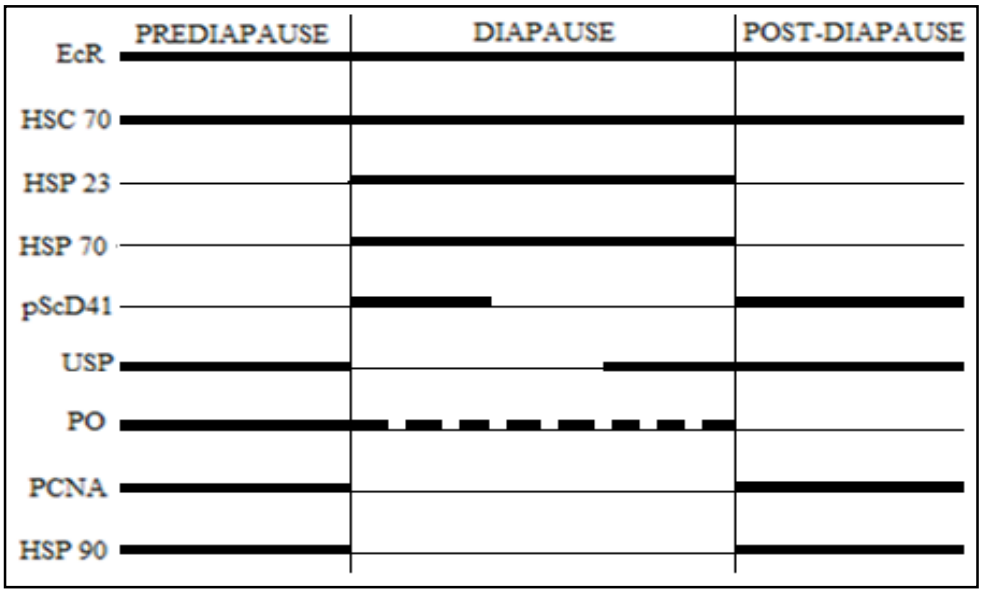


Figure 3 Gene Expression Patterns Noted in *Sarcophaga crassipalpis* in Relation to Pupal Diapause. Proteins encoded by genes are indicated on the left of the diagram. (Modified from Denlinger 2002)

Genes like sarcocystatin A (Scys-A) and sarcocystatin B (Scys-B), which have been cloned from *S. crassipalpis*, are developmentally regulated, though their expression patterns are not similar. Scys-A transcripts are present in both diapause and nondiapause destined third instar wandering larvae followed by downregulation throughout pupal diapause. By contrast, Scys-B transcripts are only weakly expressed during the third larval instar, but are highly upregulated in early diapause. The upregulation of Scys-B in early diapause suggests a possible role for this proteinase inhibitor in halting development (Goto and Denlinger 2002a).

Diapause and its regulation have greatly contributed to the understanding of insect phenology and its seasonal patterns. An in-depth understanding of its molecular function offers new tools that may prove valuable for developing biomarkers in order to monitor the ontogeny of pest species to enhance the agronomic yield and eliminate the vector-borne diseases that kill millions of people today. Work on diapause also has the potential to offer interesting insights on the molecular basis for aging. Diapause is, in fact, a dramatic example of extending the life span. Future diapause research offers a rich potential for contributing not

only to the aforementioned understanding, but also to other related and timely topics in the agricultural and biomedical sciences.

Suppression Subtractive Hybridization

The inheritable information in a gene (DNA sequence) is transcribed into a functional gene product (RNA /protein) during gene expression. The basis for cell differentiation, morphogenesis, and adaptability of all organisms is due to gene regulation that controls cell over structure and function that is in response to environmental conditions. Hence the identification, cloning of genes and categorization of differential expression of genes will grant significant insights into the molecular determinants of several life processes.

Suppression Subtractive Hybridization (SSH) is a powerful technology used for the amplification of PCR-based cDNA fragments that differ between a control and experimental transcriptome. This technique constructs a subtracted cDNA library by normalization and subtraction of mRNA (Munir et al. 2004). The normalization step normalizes the rare and abundant mRNA levels and subtracts the mRNA present in both the samples. Without knowing the identity of the genes, SSH isolates differentially expressed genes. In a number of experimental systems, for example, liver regeneration (Groenink and Aad 1996), embryo development (Simpson et al. 1999), and malignant melanoma (Hipfel et al. 2000), SSH had been used for the investigation of the differentially expressed genes (Gadgil et al. 2002).

As compared to the classical library without subtraction, the subtracted libraries should have number of specific sequences. According to the theory, similar sequences between the key tissue and the reference tissue should be subtracted by SSH (Sanchez et al. 2007).

Two populations of mRNA can be compared and the clones of the genes that are expressed only in one of the population can be obtained by SSH. mRNA populations are converted into cDNA by Reverse Transcription. cDNA from the state in question consisting

of the same and differentially expressed transcripts are considered as tester and the control state as driver. Both the cDNA from tester and driver are hybridized. The resulting hybridized sequences are then eliminated and the unhybridized cDNA found only in tester are the genes that are expressed differentially.

CHAPTER 2

MATERIALS AND METHODS

Rearing of *Sarcophaga crassipalpis*

Flesh flies, *S. crassipalpis*, have been reared in the laboratory for over 200 generations, and were established from a long standing colony at Ohio State University. Nondiapausing pupae were reared throughout their life cycles under long-day conditions, (light: dark 15:9h) at 25° C. Pupal diapause was induced by the exposure of adult females to short day conditions (light: dark 12:12h) at a temperature of 25° C during the photosensitive period until larviposition. After larviposition, the larvae were transferred to a condition, i.e., 12:12 light: dark and maintained at 20° C until the flies entered pupal diapause (Denlinger et al. 1972).

Suppression Subtractive Hybridization

PCR-Select cDNA Subtraction- Molecular Event

Genes that are upregulated in diapause are considered as tester and genes that are upregulated in nondiapause condition in the pupae of *S. crassipalpis* are considered as driver. Poly A+ mRNA of both samples was converted into cDNAs that were then digested with Rsa I, a four-base-cutting restriction enzyme yielding blunt ends. The tester cDNA was divided into two halves and were ligated with a different cDNA adaptor. Because there is not a phosphate group at the end of adaptor, only one strand for each adaptor attaches to the 5' ends of the cDNA.

Excess amount of nondiapause upregulated genes (driver) were added to both the halves of the tester separately; heat denatured and reannealed producing four different types of molecules **a**, **b**, **c** and **d** from each halves of tester (Figure 4). Type **a** molecules consist of amount of differentially expressed sequences as the concentration of high and low abundance

sequences is normalized. Type **c** molecules consist of cDNA that are not differentially expressed.

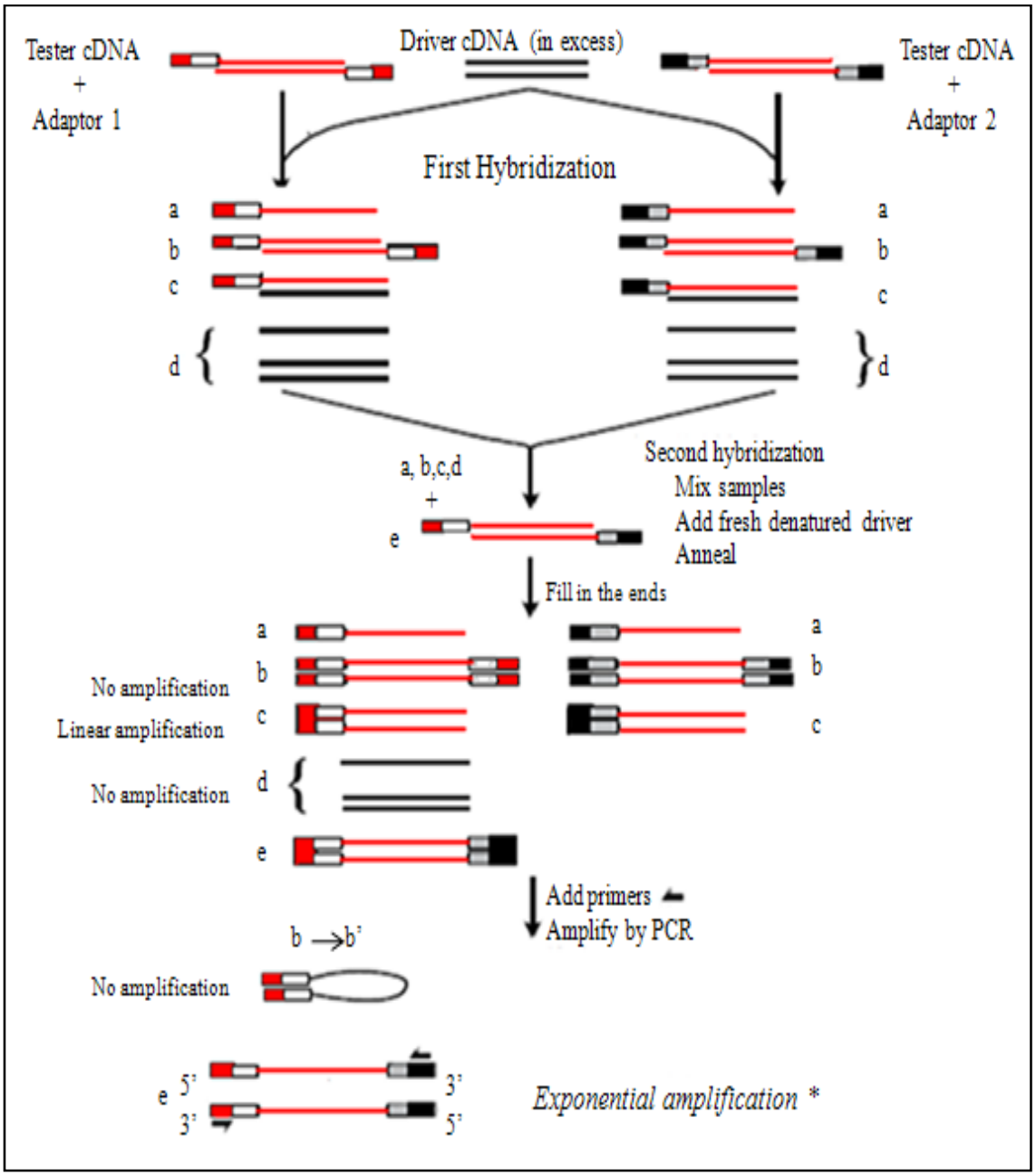


Figure 4 Flow Chart of Suppression Subtractive Hybridization

In the second hybridization step the samples obtained from first hybridization were mixed without denaturing them. Only the ss-tester cDNA will re-associate and form a new type of molecule-**e**. This new hybrid has ds-tester molecules with different ends

corresponding to the adaptor 1 and 2R sequences. DNA polymerase was added. Type e molecules will have different annealing sites for the nested primers on 5' and 3' ends.

Polymerase Chain Reaction was carried out to amplify the desired sequences. Type **a**, **b**, **c** and **d** cannot be amplified exponentially. Only type **e** molecule was amplified exponentially with two different adapters. Due to the lack of primer annealing site, type **a** and **d** molecules cannot be amplified. Type **b** molecule forms a pan like structure due to the suppression PCR, hence preventing amplification, whereas type **c** molecule has only one site for primer annealing so is amplified linearly.

Another PCR was done for the enrichment of the **e** molecules and elimination of background PCR products by using the nested primers. A cloning vector, pGEMT, was used to insert cDNAs.

The cDNA clones were then shotgun ligated into pGEMT- Easy cloning vector (Promega, Madison, WI). It was then transformed into DH5α cells and LB agar plates with ampicillin and oxacillin used for plating the recovered product. Colonies were picked from the LB plates, sequenced (Sanger), cleaned, and the cloning vector was removed. Annotations of the sequence were done by comparing it with the known sequences using BLAST. E value of $1E^{-4}$ or less was used to suggest a positive match between the cloned sequence and the available known sequence.

Bioinformatics Tools

The nucleic acid sequence obtained from the Ecoarray Institute was analyzed. Sequence similarity search were done by comparing it with the DNA public database found in GenBank, maintained by National Center for Biotechnology Information (NCBI). Database was compared with the results obtained from Basic Local Alignment Search Tools - BLAST. Clustal W available at European Molecular Biology Laboratory (EMBL) was used for the multiple sequence alignment for obtaining the consensus sequence. Likewise, Blastn

and BlastX searches in GenBank (http://www.ncbi.nlm.nih.gov/) were used to determine putative sequence identities. The closest match for each sequence was then listed in a table with the Accession number, Organism, E value, and Maximum Identity. The results with low percentage of maximum identity and high E value were considered as not significant similarity. These were listed as unknown. The consensus sequence was also used to carry out similarity search with protein database, i.e., Swissprot protein sequence using BlastX (Search protein database using a translated nucleotide query).

CHAPTER 3

RESULTS

As a result of this project, two SSH libraries were produced. The diapause upregulated library (DRL) used nondiapause mRNA as the driver to eliminate the nondiapause specific sequences from the library. The diapause downregulated library (NDRL) used diapause mRNA to eliminate diapause specific sequences. Analysis of the SSH Library has identified both diapause upregulated and nondiapause upregulated genes (diapause downregulated). 392 clones were randomly selected from the two libraries; from which 379 clones were used for the analyses of nucleotide sequence and were compared to the known genes with similar identity using public sequence databases like NCBI, ExPasy, GO, and other databases.

Table 1 shows the results of these sequences, with the number of clones and their identities from the DRL. None of these sequences was identified from the NDRL isolates. These results are supportive evidence that the subtractive procedure has produced sequences that are specific to each developmental state. Number of isolates represents the number of randomly selected clones that were identified as the Gene. The number of isoforms is the number of sequence that could be identified by the Clustal W analysis.

Table 2 lists the number of clones that were isolated and identified from the NDRL. As in the DRL isolates, none of these sequences was identified from the DRL from the sample of the diapause subtractive library.

Rest of the result section follows the examination of each group of genes from both libraries. Starting with the diapause upregulated genes, each of the types of isolates was examined based on the percentage of isolates.

Table 1 Number of Isolates and Genes from DRL

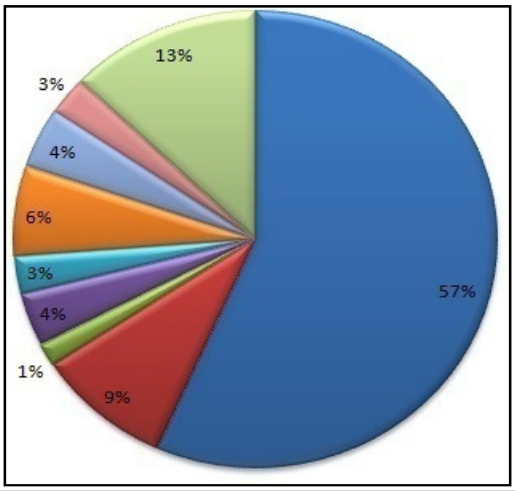
S.N.	Gene	DRL		# of isolates NDRL
		# of isolates	# of genes	
1.	25kDa small heat shock protein (*S. crassipalpis*)	49	4	0
2.	70kDa heat shock protein (*S. crassipalpis*)	35	10	0
3.	Not Identified	24	15	0
4.	23kDa heat shock protein (*S. crassipalpis*)	10	2	0
5.	CG12000-PA,isoform A (*D. melanogaster*)/ Proteasome beta type - 4 subunit	7	1	0
6.	Cognate 70kDa heat shock protein (*S. crassipalpis*)	6	1	0
7.	GA 19395-PA (*D. pseudoobscura*)/ Unknown	5	1	0
8.	Ribosomal protein L37a CG 9091-PA (*D. melanogaster*)	5	2	0
9.	CG 9836-PA (*D. melanogaster*)/ NifU-like N terminal domain	4	1	0
10.	GA 12626-PA (*D. pseudoobscura*)/ Unknown	4	1	0
11.	RE 38876P (*D. melanogaster*)/ Subfamily of SANT domain	4	1	0
12.	Small heat shock protein (*Culex pipiens pipiens*)/ 18 kDa small heat shock protein	4	1	0
13.	Similar to *D. melanogaster* RpS6 (*D. yakuba*)/ Ribosomal protein S6e	3	1	0
14.	Ribosomal protein L13a CG 1475-PB (*D. melanogaster*)	3	1	0
15.	Ribosomal protein P2 (*C. capitata*)/ 60s Acidic ribosomal protein	3	2	0
16.	Unnamed protein product (*D. melanogaster*)/ Phosphoenolpyruvate Carboxykinase (PEPCK)	2	1	0
17.	Beta tubulin at 60D CG3401-PA (*D. melanogaster*)	2	1	0
18.	CG 7998-PA (*D. melanogaster*)/ Malate dehydrogenases (MDH)	2	1	0
19.	CG4111-like protein(*D. miranda*)/ Ribosomal L29 protein	2	1	0
20.	Hypothetical protein PF14_0710 (*P. falciparum*)/ Unknown	2	1	0
21.	Singed CG32858-PC, isoform (*D. melanogaster*)/ Fascin-like domain	2	1	0
22.	GA 10081-PA (*D. pseudoobscura*)/ Insect cuticle protein,	1	1	0
23.	Similar to *D. melanogaster* CG7808 (*D. yakuba*)/ Ribosomal protein S8e	1	1	0
24.	Heat shock protein 70 (*S. crassipalpis*)	1	1	0
25.	Beta tubulin (*C. p.pipiens*)	1	1	0
26.	ATP synthase beta (*D. simulans*)	1	1	0
27.	Chicade (*D. simulans*)	1	1	0
28.	Cyclophilin 1 CG 9916-PA (*D. melanogaster*)	1	1	0
29.	Unknown, PE-PGRS family protein (*M. tuberculosis*F11)	1	1	0
30.	Heat shock protein 60 (*Liriomyza huidobrensis*)	1	1	0
	TOTAL	187	59	0

Table 2 Number of Isolates and Genes from NDRL

S.N.	Gene	# of isolates DRL	NDRL	
			# of isolates	# of genes
1.	Non Diapause Not Identified	0	85	9
2.	NFE2L2 (*Pan troglodytes*)/ bZIP transcription factor	0	15	1
3.	Putative SPT transcriptional factor family membrane (*Lepephtheirus salmonis*)/ Unknown	0	15	1
4.	Hypothetical protein (*T. domestica*)/ Unknown	0	12	2
5.	Variable membrane protein precursor (*Trichomonas vaginalis*)	0	12	1
6.	Serine protease inhibitor 1 (*G. m morsitans*)	0	10	1
7.	GA15651-PA (*D. pseudoobscura*)/ Unknown	0	6	1
8.	GA21196-PA (*D. pseudoobscura*)/ Unknown	0	5	2
9.	Unnamed protein product (*M. fascicularis*)/ Unknown	0	4	1
10.	Storage protein binding protein (*S. peregrine*)/ Hemocyanin	0	4	1
11.	Enterobacteria phage (*Coliphage phix* 174)/ Microvirus A* protein	0	4	1
12.	Hypothetical protein SPAC1B2.03C (*Schizosaccharomyces prombe*)/ GNS1/SUR4 family	0	3	1
13.	Transferring precursor (*S. peregrine*)	0	2	2
14.	Hypothetical protein (*P. tetraurelia*)/ Unknown	0	2	1
15.	NADH dehydrogenase subunit 2 (*Steinernema carpocapsae*)/ Unknown	0	2	1
16.	Predicted hypothetical protein isoform 1 (*Nasonia vitripennis*)/ Histone 2A	0	1	1
17.	Similar to Mediator of RNA polymerase II transcription subunit 129 (*Rattus norvegicus*)/ Transcription mediator subunit Med12	0	1	1
18.	Casein kinase lalpha (*D. melanogaster*)	0	1	1
19.	GA14517-PA (*D. pseudoobscura*)/ ATP synthase subunit C	0	1	1
20.	Pollen coat oleosin –glycine rich protein (*Olimarabidopsis pumila*)	0	1	1
21.	CG 8092-PA, isoform A (*D. melanogaster*)/ Unknown	0	1	1
22.	Hypothetical protein AaeL_AAEL003275 (*Aedes aegypti*)/ Zinc-finger associated domain	0	1	1
23.	Hypothetical protein (*Mus musculus*)/ Unknown	0	1	1
24.	RNA helicase/ DEAD-box helicase	0	1	1
25.	Cyclin box fold, Cyclin K (*Gallus gallus*)	0	1	1
26.	Hypothetical protein DDBDRAFT (*Dictyostelium discoideum*)/ DEAD-box helicase	0	1	1
	TOTAL	0	192	37

Diapause Upregulated Genes

Fifty-seven percent of the sequences are heat shock protein transcripts. Nine percent of the isolates were identified as subunits of the ribosomal protein complex. This class of genes has not been previously reported as diapause upregulated transcripts. One percent of beta tubulin was present. The rest of the sequences were found to be present as unique sequences. Thirteen percent of isolates were unidentified sequences (Figure 5).



- **Heat shock proteins (57%)** → 25kDa small heat shock protein (*Sarcophaga crassipalpis*)
  70kDa heat shock protein (*Sarcophaga crassipalpis*)
  23kDa heat shock protein (*Sarcophaga crassipalpis*)
  Cognate 70kDa heat shock protein (*Sarcophaga crassipalpis*)
  Small heat shock protein (*Culex pipiens pipiens*)
  Heat shock protein 60 (*Liriomyza huidobrensis*)
  Heat shock protein 70(*Sarcophaga crassipalpis*)
- **Ribosomal protein (9%)** → Ribosomal protein L37a CG 9091-PA (*Drosophila melanogaster*)
  Ribosomal protein L13a CG 1475-PB (*Drosophila melanogaster*)
  Ribosomal protein P2 (*Ceratitis capitata*)
  Similar to *Drosophila melanogaster* RpS6 (*Drosophila yakuba*)
  Similar to *Drosophila melanogaster* CG7808 (*Drosophila yakuba*)
  CG4111-like protein (*Drosophila miranda*)
- **Beta tubulin (1%)** → Beta tubulin at 60D CG3401-PA (*Drosophila melanogaster*)
  Beta tubulin (*Culex pipiens pipiens*)
- **Other number of hits 7 (4%)** → CG12000-PA, isoform A (*Drosophila melanogaster*)
- **Other number of hits 5 (3%)** → GA 19395-PA (*Drosophila pseudoobscura*)
- **Other number of hits 4 (6%)** → GA 12626-PA (*Drosophila pseudoobscura*)
  CG 9836-PA (*Drosophila melanogaster*)
  RE 38876P (*Drosophila melanogaster*)
- **Other number of hits 2 (4%)** → CG 7998-PA (*Drosophila melanogaster*)
  Hypothetical protein PF14_0710 (*Plasmodium falciparum*)
  Unnamed protein product (*Drosophila melanogaster*)
  Singed CG32858-PC, isoform (*Drosophila melanogaster*)
- **Other number of hits 1 (3%)** → GA 10081-PA (*Drosophila pseudoobscura*)
  ATP synthase beta (*Drosophila simulans*)
  Chicade (*Drosophila simulans*)
  Cyclophilin 1 CG 9916-PA (*Drosophila melanogaster*)
  PE-PGRS family protein (*Mycobacterium tuberculosis* F11)]
- **No hits (13%)**

Figure 5 Type of Genes (shown as % of each type) Expressed During Diapause

Different databases like All GenBank* and BlastX were used for the search of similarity of differentially expressed sequences. Conserved domain search was used to find the main domain and function of each of these sequences.

Seven different types of heat shock proteins were found with different number of isolates (Table 3).

Table 3 Heat Shock Proteins

S.N.	Gene	No. of Isolates
1	25kDa small heat shock protein (*S. crassipalpis*)	49
2	70kDa heat shock protein (*S. crassipalpis*)	35
3	23kDa heat shock protein (*S. crassipalpis*)	10
4	Cognate 70kDa heat shock protein (*S. crassipalpis*)	6
5	Small heat shock protein (*C. p. pipiens*)/18 kDa small hsp	4
6	Heat shock protein 70 (*S. crassipalpis*)	1
7	Heat shock protein 60 (*Liriomyza huidobrensis*)	1

The isolates were grouped according to the types of genes as identified by GO function. During diapause, seven different types of heat shock sequence (25 kDa small heat shock protein, 23kDa heat shock protein, 70 kDa heat shock protein, cognate 70 kDa heat shock protein, small heat shock protein, heat shock protein 70, and heat shock protein 60) were found to be upregulated out of which five genes (25 kDa small heat shock protein, 23kDa heat shock protein, 70 kDa heat shock protein, cognate 70 kDa heat shock protein, and heat shock protein 70) were found previously as *S. crassipalpis* genes. Most of these have been identified as diapause upregulated transcripts except for an isolate of the small heat shock protein from *C. p. pipiens*.

1. 25kDa small heat shock protein (*S. crassipalpis*)

Twenty-six percent (49) of the isolates were identified as hsp 25 related genes. These were a direct match for the *S. crassipalpis* sequences that were previously identified as diapause upregulated (Rinehart et al. 2007).

25kDa small heat shock proteins showed 100% similarity with *S. crassipalpis* clone 7H7 25 kDa small heat shock protein mRNA with accession number EF103577.1 and E value 0. BlastX result also gave maximum identity of 100% with E value 2e-38 with the accession number gb|ABL06941.1| and description of 25 kDa small heat shock protein (*S. crassipalpis).*

This analysis verifies that 25kDa hsp were isolated from *S. crassipalpis.* A total number of 49 isolates of 25 kDa small heat shock protein were sequenced, and Clustal W analysis identified four different types of sequence. Out of four different isoforms, 25 of them were grouped as similar sequences as isoform A, 11 formed isoform B, another 11 formed isoform C, and 2 sequences formed isoform D.

Dendogram result showed the closest relationship of 25kDa hsp- isoform B identified by this study to the 25kDa hsp of accession number EF103577 (Figure 6).

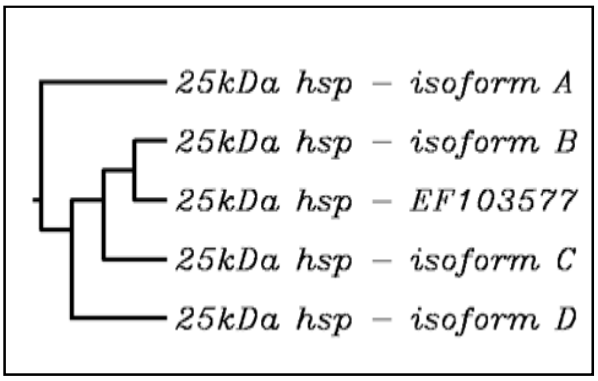


Figure 6 Dendrogram Showing Interrelationships Between 25kDa hsp Genes

2. 70kDa heat shock protein (*S. crassipalpis*)

Nineteen percent (35) of isolates were identified as *S. crassipalpis* hsp 70 genes by GenBank analysis (Denlinger 1991; Yocum et al. 1991, 1998; Kimura et al.1998; Reinhart et al. 2000; Hayward et al. 2005).

The accession number was AF107338.2 and E value 0 with 97% identity. BlastX result on the other hand gave 97% identity with 70 kDa heat shock protein ScHSP70 (*S. crassipalpis*) with accession number gb|AAD17995.2| and an E value of 1e-41.

Out of the 35 isolates identified, Clustal W identified11 sequences as isoform A, 6 sequences as isoform B, 5 as isoform C, 4 as isoform D, 2 as isoform E, and 2 as isoform F. Five sequences were identified as isoforms G, H, I, and J for a total of 11 different HSP 70 isoforms. Figure 7 shows the interrelationships between the isoforms of 70kDa hsp found by this study with accession number AF107338.

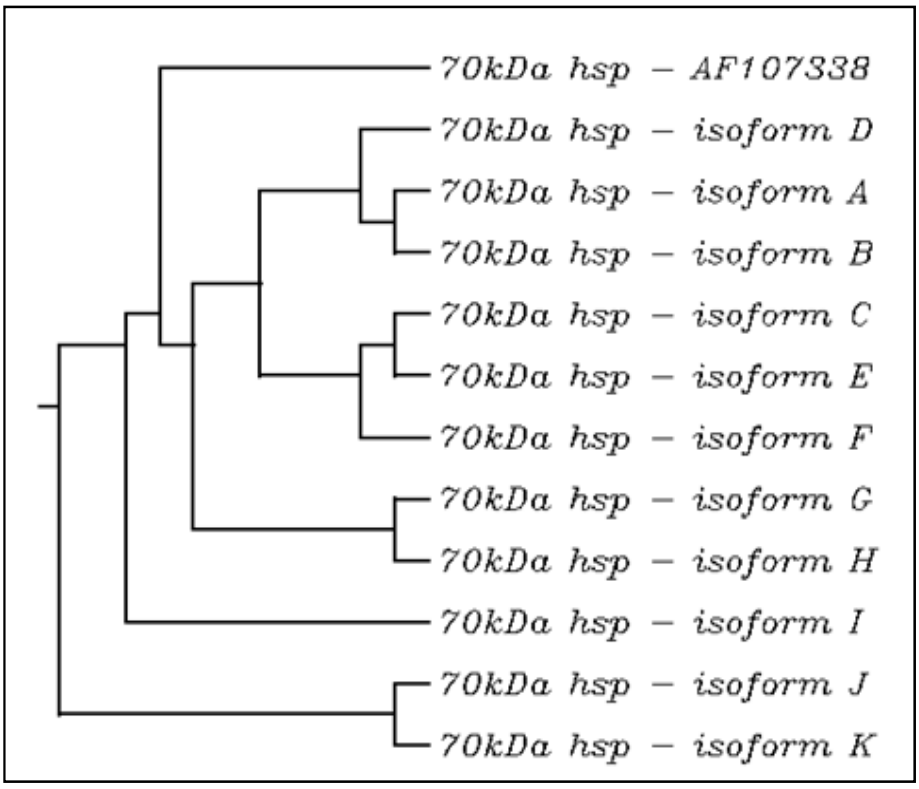


Figure 7 Dendrogram Showing Interrelationships Between 70kDa hsp Genes

3. 23kDa heat shock protein (*S. crassipalpis*)

Five percent (10) of the isolates were identified by BlastX as previously identified *S. crassipalpis* upregulated HSP 23 kDa (Denlinger 1991; Yocum et al. 1991, 1998; Kimura et al.1998; Reinhart et al. 2000; Hayward et al. 2005).

BlastX results of ScHSP23 accession number gb|AAC63387.1| , gave maximum identity of 77%, and an E value of 5e-22.

Out of 10 isolates found by SSH, two different types of sequence were found to be similar by Clustal W analysis. Eight of them were identified as isoform A and 2 sequences as

isoform B. Figure 8 shows the distance relation between the isoforms of 23kDa hsp that has been identified by this study with the 23kDa hsp of accession number SCU96099.

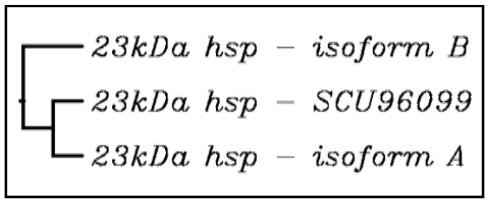


Figure 8 Dendrogram Showing Interrelationships Between 23kDa hsp Genes

4. Cognate 70kDa heat shock protein (*S. crassipalpis*)

Three percent (6) of the cognate 70 kDa heat shock sequence were found to be present and identified as previously upregulated *S. crassipalpis* sequences (Rinhart et al. 2000).

GenBank analysis showed 100% identity with *S. crassipalpis* cognate 70 kDa heat shock protein scHSC70 (HSC70) gene with an accession number of AF107339.1, and E value 0.

Clustal W identified all six isolates of cognate 70kDa heat shock proteins to be same.

5. Small heat shock proteins (*C. p.pipiens)/* 18 kDa small heat shock protein

Two percent (4) of the small heat shock sequence from *C. p. pipiens* were found to be present and identified as previously upregulated *S. crassipalpis* sequences.

GenBank analysis identified these sequences as *S. crassipalpis* clone 9D11 18 kDa small heat shock protein mRNA with E value 0 , and accession number EF103578.1, with 99% identity .

Clustal W showed all four sequences were similar.

6. Heat shock protein 70 (*S. crassipalpis)*

The sequence of a single isolate was identified as another heat shock protein 70 (*S. crassipalpis).*

S. crassipalpis 70 kDa heat shock protein ScHSP70 (HSP70) mRNA is the description shown by GenBank analysis with an accession number AF107338.2, E value of 8e-170 and identity 91%.

7. Heat shock protein 60 (*Liriomyza huidobrensis)*

One isolate sequence was identified as heat shock protein 60 (*Liriomyza huidobrensis*).

Heat shock protein 60 (*Liriomyza huidobrensis*- pea leafminer) consists of GroEL-like type I chaperonin. Chaperonins are involved in productive folding of proteins and promotes folding by using energy derived from ATP hydrolysis (Kusmierczyk and Martin 2001).

BlastX result showed identity with heat shock protein 60 (*Liriomyza huidobrensis*) with an accession number gb|AAW30392.2|, and E value 2e-40 with 92% similarity. This represents the first identity of a mitochondrial hsp upregulated in diapause.

Ribosomal proteins

Six different types of ribosomal proteins were identified by this study, and Table 4 lists the ribosomal proteins according to the number of isolates.

Table 4 Ribosomal Proteins

S.N.	Gene	No of Isolates
1	Ribosomal protein L37a CG 9091-PA(*D. melanogaster*)	5
2	Ribosomal protein L13a CG 1475-PB(*D. melanogaster*)	3
3	Ribosomal protein P2(*Ceratitis capitata*)/ 60s Acidic ribosomal protein	3
4	Similar to *D. melanogaster* RpS6 (*D. yakuba*)/ Ribosomal protein S6e	3
5	CG4111-like protein (*D. miranda*)/ Ribosomal L29 protein	2
6	Similar to *D. melanogaster* CG7808(*D. yakuba*)/ Ribosomal protein S8e	1

1. Ribosomal protein L37a CG 9091-PA (*D. melanogaster)*

Three percent (5) of isolates were identified as the ribosomal protein subunit L37a CG9091- PA in *D. melanogaster.* This gene family includes ribosomal protein L37 from eukaryotes and archaebacteria, and contains many conserved cysteines and histidines suggesting that this protein may bind to zinc.

BlastX showed 83% similarity with ribosomal protein L37a CG9091-PA (*D. melanogaster*), accession number ref|NP_573005.1|, with an E value of 3e-07.

Clustal W analysis showed two different types with 3 of them forming isoform A and 2 were isoform B.

2. Ribosomal protein L13a CG 1475-PB (*D. melanogaster*)

Two percent (3) of the isolates were identified as ribosomal protein L13a CG 1475-PB (*D. melanogaster*). Protein L13, a large ribosomal subunit protein, is one of the five proteins required for an early folding intermediate of 23S rRNA in the assembly of the large subunit upregulated in a variety of human gastrointestinal cancers. BlastX results gave 86% identity with ribosomal protein L13A (*D. melanogaster*), accession number ref|NP_649560.1| with an E value 2e-35.

All three sequences were found to be similar by Clustal W analysis.

3. Ribosomal protein P2 *(Ceratitis capitata)*/60s Acidic ribosomal protein

Two percent (3) of the isolates were identified as 60s acidic ribosomal ribosomal protein P2 (*Ceratitis capitata*), accession number = Y09056.1, E value = 7e-72 from GenBank analysis and 87% identity with ribosomal protein P2 (*Ceratitis capitata*) (accession number = emb|CAA70259.1|, E value = 9e-22) from BlastX result. Another identity found by BlastX result was 60S acidic ribosomal protein P2 (*Aedes aegypti*) with 82% identity and accession number ref|XP_001649060.1| and E value 1e-19.

Out of the three sequences, two different types of isoform were found (2 similar sequences as isoform A and 1 sequence as isoform B).

4. Similar to *D. melanogaster* RpS6 (*D. yakuba*)/ Ribosomal protein S6e

Two percent (3) of the isolates were identified similar to *D. melanogaster* RpS6 (*D. yakuba*), ribosomal protein L13a CG 1475-PB (*D. melanogaster*) and ribosomal protein P2 (*Ceratitis capitata*).

The main domain is similar to *D. melanogaster* RpS6 (*D. yakuba*). BlastX gave accession number gb|AAR10071.1|, E value of 1e-21 and maximum identity of 91%.

All three isolates obtained via Clustal W analysis were the same.

5. CG4111-like protein (*D. miranda*)/ Ribosomal L29 protein

One percent (2) of isolates was similar to CG4111-like protein (*D. Miranda*). BlastX gave 67% identity, accession number gb|AAS87309.1|, and an E value of 1e-07.

Ribosomal L29 protein is a protein of the large ribosomal subunit. In mammals, it is believed to play a role in cell adhesion and modulation of blood coagulation. It has also been shown to inhibit apoptosis in cancer cells.

Two similar sequences were found by Clustal W analysis giving single isoform.

6. Similar to *D. melanogaster* CG7808 (*D. yakuba*)/ Ribosomal protein S8e

One isolate was identified as *D. melanogaster* CG7808.

Ribosomal protein S8e is the main domain found. BlastX accession number ref|NP_651740.1| gave 91% identity with the ribosomal protein S8 CG7808-PC (*D. melanogaster*) with an E value of 7e-17.

Table 5 lists the number of beta tubulin identified according to the number of isolates found.

Table 5 Beta Tubulin

S.N.	Gene	No of Isolates
1	Beta tubulin at 60D CG3401-PA(*D. melanogaster*)	2
2	Beta tubulin (*C. p.pipiens*)	1

1. Beta tubulin at 60D CG3401-PA (*D. melanogaster*)

Two isolates were similar to Beta tubulin at GK22203 Dwil (*D. willistoni*). Genbank gave the accession number XM_002066109.1, E value of 3e-91, and maximum identity of 84%. Both sequences found by SSH were similar to each other.

2. Beta tubulin (*C. p. pipiens*)

One isolate was found to be similar to Beta tubulin (*Lucilia cericata*). GenBank analysis gave accession number EF056211.1, E value of 0, with 92% identity.

BlastX gave 97% identity with Beta tubulin (*C. p. pipiens*) with 2e-131 E value and accession number gb|ABG29421.1|.

Fifteen other different types of genes were identified as diapause upregulated genes that are shown in Table 6 according to the number of isolates found.

Table 6 Other Isolates from DRL

S.N.	Gene	No of Isolates
1	Not identified	24
2	CG12000-PA,isoform A (*D. melanogaster*)/ Proteasome beta type - 4 subunit	7
3	GA 19395-PA (*D. pseudoobscura*)/ Unknown	5
4	GA 12626-PA (*D. pseudoobscura*)/ Unknown	4
5	RE 38876P (*D. melanogaster*)/ Subfamily of SANT domain	4
6	CG 9836-PA (*D. melanogaster*)/ NifU-like N terminal domain	4
7	Hypothetical protein PF14_0710 (*Plasmodium falciparum*)	2
8	Unnamed protein product (*D. melanogaster*)/ Phosphoenolpyruvate Carboxykinase (PEPCK)	2

Table 6 Continued

9	Singed CG32858-PC, isoform (*D. melanogaster*)/ Fascin-like domain	2
10	CG 7998-PA (*D. melanogaster*)/ Malate dehydrogenases (MDH)	2
11	ATP synthase beta (*D. simulans*)	1
12	Chicade (*D. simulans*)	1
13	Cyclophilin 1 CG 9916-PA (*D. melanogaster*)	1
14	PE-PGRS family protein (*M. tuberculosis* F11)	1
15	GA 10081-PA (*D. pseudoobscura*)/ Insect cuticle protein	1

1. Isolates Not Identified

Thirteen percent (24) of isolates represented unidentified sequences. Additional searches with FASTA and ExPasy failed to give any identified sequences. Six sequences were very short (< 100 nucleotide). Fifteen different types of isoforms were found by Clustal W (including the 6 short sequences). 5 sequences were similar- Unidentified Diapause gene (UDG) A , 4 sequences were similar- UDG B, 2 sequences UDG C, 2 sequences UDG D, and eleven additional sequences were UDG- E, F, G, H, I, J, K, L, M, N and O.

2. CG12000-PA,isoform A (*D. melanogaster*)/Proteasome beta type-4 subunit

Four percent (7) isolates were identified as CG 12000-PA, isoform A found in *D. melanogaster,* a proteasome beta type-4 subunit. The 20S proteasome, multisubunit proteolytic complex, is the central enzyme of nonlysosomal protein degradation in both the cytosol and nucleus. BlastX results gave no significant similarity. Seven isolates were found by SSH and all of them were similar to each other.

3. GA 19395-PA (*D. pseudoobscura*)/Unknown

Three percent (5) isolates were identified as GA 19395 (*D. pseudoobscura*) by BlastX accession number of ref|XP_001355885.1|, E value 3e-88 and gave 80% identity. It has domain of unknown function (DUF1000). No significant similarity was found by GenBank analysis. Clustal W analysis found all the sequences similar.

4. GA 12626-PA (*D. pseudoobscura*)/ Unknown

Two percent (4) of the isolates were identified as GA 12626- PA (*D. pseudoobscura*). BlastX accession number of ref|XP_001955986.1|, E value 2e-41 and showed 86% similarity with GF24822 (*D. ananassae*). Clustal W analysis found all the sequences similar.

5. RE 38876P (*D. melanogaster*)/ SANT domain

Two percent (4) isolates were identified as RE 38876P (*D. melanogaster*) with 51% of maximum identity, accession number gb|AAL48974.1| and E value of 7e-17. The main domain found was a subfamily of the SANT domain. Clustal W analysis found all the sequences similar.

6. CG 9836-PA (*D. melanogaster*)/ NifU-like N terminal domain

Two percent (4) isolates were identified as CG 9836-PA (*D. melanogaster*).

The main domain is NifU-like N terminal domain. This domain is found in several bacterial species. These proteins appear to be scaffold proteins for iron-sulfur clusters. No significant similarity was found by GenBank analysis, whereas BlastX showed 97% identity with CG9836-PA (*D. melanogaster*) (accession number of ref|NP_649840.1|, E value of 3e-45).

Clustal W analysis found all the sequences similar.

7. Hypothetical protein PF14_0710 (*Plasmodium falciparum*)

One percent (2) of isolates was identified as hypothetical protein PF14_0710 (*Plasmodium falciparum*).

No conserved domains as well as no significant similarity by both GenBank analysis and BlastX result have been identified.

Two sequences were found both giving a single isoform.

8. Unnamed protein product (*D. melanogaster*)/ Phosphoenolpyruvate carboxykinase

One percent (2) of isolates was identified as an unnamed protein product (*D. melanogaster*).

Phosphoenolpyruvate carboxykinase (PEPCK), a critical gluconeogenic enzyme, catalyses the first committed step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis. BlastX result gave 70% identity to unnamed protein product (*D. melanogaster*) with E value 1e-44 and accession number emb|CAA68463.1|.

Both the isolates obtained by SSH were similar.

9. Singed CG32858-PC, isoform (*D. melanogaster*)/ Fascin-like domain

One percent (2) of isolates was identified as Singed CG32858-PC, isoform (*D. melanogaster*).

The main domain was Fascin-like domain. GenBank analysis gave 100% identity to *Agaricus bisporus* partial mRNA for putative inorganic phosphate transporter (ipt gene), clone pm127 with E value of 8e-05 (accession number = AJ534339.1). BlastX result gave 87% identity to singed CG32858-PC, isoform C (*D. melanogaster*) (accession number = ref|NP_511076.2|, E value = 4e-47).

Two sequences were found by SSH and both the sequence showed similarity.

10. CG 7998-PA (*D. melanogaster*)/ Malate dehydrogenases

One percent (2) of isolates was identified as CG 7998-PA (*D. melanogaster*).

Malate dehydrogenases (MDH) are one of the key enzymes in the citric acid cycle. GenBank analysis gave 86% identity with *Pichia stipitis* CBS 6054 malate dehydrogenase partial mRNA (accession number = XM_001384913.1, E value = 1e-09). BlastX result gave 88% identity with CG7998-PA (*D. melanogaster*) (accession number = 2e-36, E value = 2e-36).

Two sequences were found by SSH and both the sequence showed similarity.

11. ATP synthase beta (*D. simulans*)

One percent (1) of isolates was identified as ATP synthase beta (*D. simulans*).

The domain is F1 ATP synthase beta subunit, nucleotide-binding domain. It uses a proton gradient to drive ATP synthesis and hydrolyzes ATP to build the proton gradient. Maximum identity of 86%, E value 3e-104 with an accession number EU207929.1 was found with an identity to *Armigeres subalbatus* ASAP ID ACN-0188192 ATP synthase mRNA sequence by GenBank analysis. Maximum identity of 100%, E value 7e-62 with an accession number of ref|XP_002105781.1| was found giving an identity of ATPsyn-beta (*D. simulans*) by BlastX result.

Only one sequence was found by SSH.

12. Chicade (*D. simulans*)

One percent (1) of isolates was identified as chicade (*D. simulans*).

Profilin may link the cytoskeleton with major signaling pathways by interacting with components of the phosphatidylinositol cycle and Ras pathway. Ninety percent of identity was found with *D. yakuba* GE18963 (Dyak\GE18963), mRNA (accession number = XM_002089115.1, E value = 1e-33) from GenBank analysis and from BlastX result, 100% identity with chickadee CG9553-PA, isoform A (*D. melanogaster*) (accession number = ref|NP_477016.1|, E value = 2e-15).

Only one sequence was found by SSH.

13. Cyclophilin 1 CG 9916-PA (*D. melanogaster*)

One percent (1) of isolates was identified as Cyclophilin 1 CG 9916-PA (*D. melanogaster*).

Cyclophilin_ABH_like: Cyclophilin A, B, and H-like cyclophilin-type peptidylprolyl cis- trans isomerase (PPIase) domain. These enzymes have been implicated in protein folding processes which depend on catalytic/chaperone-like activities. GenBank analysis gave 77%

identity with *Vicia faba* vcCyP mRNA (E value = 2e-29, accession number = AB012947.1) and BlastX result gave 85% identity to Cyclophilin 1 CG9916-PA (*D. melanogaster*) (accession number = ref|NP_523366.2|, E value = 3e-82).

Only one sequence was found by SSH.

14. PE-PGRS family protein (*M. tuberculosis* F11)

One percent (1) of isolates was identified as PE-PGRS family protein (*M. tuberculosis* F11).

No conserved domains have been identified for this query sequence. BlastX result gave no significant similarity.

Only one sequence was obtained by SSH.

15. GA 10081-PA (*D. pseudoobscura*)/ Insect cuticle protein

One percent (1) of isolates was identified as GA 10081-PA (*D. pseudoobscura*).

Insect cuticle protein is the main domain of GA 10081-PA (*D. pseudoobscura*). The extensive conservation of this region might indicate its function in binding chitin. GenBank analysis showed no significant similarity. From BlastX results, 70% identity was found with GA10081 (*D. pseudoobscura pseudoobscura*) with an E value of 5e-19 having an accession number ref|XP_001360597.1|.

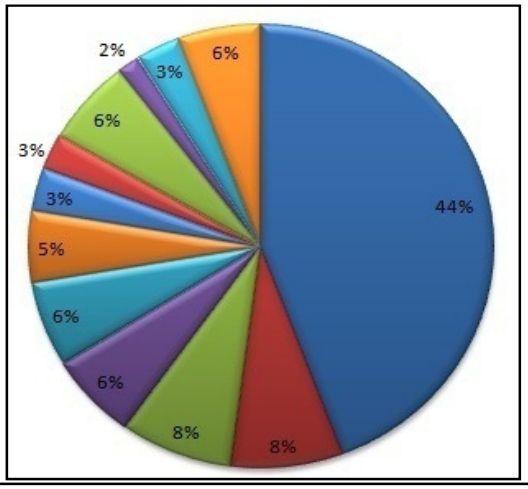
Only one sequence was obtained by SSH.

Nondiapause Upregulated Genes

As with the DRL results, none of the gene sequences isolated from the NDRL were found in the DRL isolates. Thus, there does not appear to be any overlap between both the sequences. Except for the unidentified sequences, there was not the large concentration of sequences in one group, as with the HSP and ribosomal subunit sequences in the DRL (Tables 2 and 3). Twenty-five sequences were found to be identifiable, while 85 sequences were not. Of the 85 unidentified sequences, 52 were identified as a single sequence by Clustal

W analysis. This is one of the largest groups of nondiapause upregulated genes to be identified from the nondiapause papal stage of *S. crassipalpis*.

Figure 9 shows different percentage of nondiapause upregulated genes identified by this study in the form of pie chart.



- Non diapause No hit
- NFE2L2 [*Pan troglodytes*]
- Putative SPT transcriptional factor family membrane [*Lepephtheirus salmonis*]
- Hypothetical protein [*Thermobia domestica*]
- Variable membrane protein precursor [*Trichomonas vaginalis*]
- Serine protease inhibitor 1 [*Glossina morsitans morsitans*]
- GA15651-PA [*Drosophila pseudoobscura*]
- GA21196-PA [*Drosophila pseudoobscura*]
- Number of hit 4
  - Unnamed protein product [*Macaca fascicularis*]
  - Storage protein binding protein [*Sarcophaga peregrine*]
  - Enterobacteria phage [*Coliphage phix* 174]
- Number of hit 3
  - Hypothetical protein SPAC1B2.03C [*Schizosaccharomyces prombe*]
- Number of hit 2
  - Transferring precursor [*Sarcophaga peregrina*]
  - Hypothetical protein [*Paramecium tetraurelia*]
  - NADH dehydrogenase subunit 2 [*Steinernema carpocapsae*]
- Number of hit 1
  - Predicted hypothetical protein isoform 1 [*Nasonia vitripennis*]
  - PREDICTED: similar to Mediator of RNA polymerase II transcription subunit 129 [*Rattus norvegicus*]
  - Casein kinase lalpha [*Drosophila melanogaster*]
  - GA14517-PA [*Drosophila pseudoobscura*]
  - Pollen coat oleosin –glycine rich protein [*Olimarabidopsis pumila*]
  - CG 8092-PA, isoform A [*Drosophila melanogaster*]
  - Hypothetical protein AaeL_AAEL003275 [*Aedes aegypti*]
  - Hypothetical protein [*Mus musculus*]
  - RNA helicase
  - Cyclin K [*Gallus gallus*]
  - Hypothetical protein DDBDRAFT [*Dictyostelium discoideum*]

Figure 9 The Percentage of Number of Isolates of Different Genes During Nondiapause

1. Unidentified isolates

From the total unidentified isolates for nondiapause, 9 different types of sequence were identified. 52 sequences were identical to each other giving isoform A, 20 sequences were identified as isoform B, 3 of them as isoform C, 2 sequences as isoform D, 2 as isoform E, similarly 2 sequences as isoform F, 2 similar sequences giving isoform G, and 2 sequences were unique (isoform H and I) giving a total of 9 different isoforms.

2. NFE2L2 (*Pan troglodytes*)/ bZIP transcription factor

The main domain of NFE2L2 (*Pan troglodytes*) is bZIP transcription factor.

GenBank analysis gave maximum identity of 99% with *Pan troglodytes* NFE2L2 (NFE2L2) gene with E value 0 and accession number DQ977335.1. On the other hand, BlastX result gave maximum identity of 100% to NFE2L2 (*Pan troglodytes*) and E value 3e-94 (accession number = gb|ABM91944.1|).

Fifteen sequences of NFE2L2 (*Pan troglodytes*) were identified by SSH. All 15 sequences showed similarity by Clustal W analysis.

3. Putative SPT transcription factor family member (*Lepeophtheirus salmonis*)

No conserved domains have been identified for this query sequence.

BlastX result showed 32% identity with putative SPT transcription factor family member (*Lepeophtheirus salmonis*) with E value of 2e-31 and accession number gb|ABU41130.1|.

A total of 15 sequences were obtained, and all were shown similar by Clustal W analysis.

4. Hypothetical protein (*Thermobia domestica*)/ Unknown

No conserved domains have been identified for this query sequence.

From the GenBank analysis, 96% identity was found with *Chrysomya putoria* mitochondrion, complete genome with Evalue 0 and accession number AF352790.1. BlastX

result gave 75% identity with hypothetical protein (*Thermobia domestica*) with E value 1e-07 and accession number emb|CAM36311.1|.

Out of 12 sequences identified by SSH, 2 different types of isoforms were found. Eight of them were identified as isoform A and 4 of them were identified as isoform B.

5. Variable membrane protein precursor (*Trichomonas vaginalis* G3)

RNA recognition motif and is probably diagnostic of an RNA binding protein. The motif also appears in a few single stranded DNA binding proteins.

No significant similarity was found by both GenBank analysis and BlastX result.

Twelve isolates of variable membrane protein precursor were identified, and all of them showed similarity by Clustal W analysis.

6. Serine protease inhibitor 1 (*Glossina morsitans morsitans*)

SERine Proteinase INhibitors (serpins). Serpins are of medical interest because mutants have been associated with blood clotting disorders, emphysema, cirrhosis, and dementia.

GenBank analysis showed no significant similarity but BlastX result showed 35% identity with serine protease inhibitor 1 (*Glossina morsitans morsitans*) with E value 9e-40 and accession number gb|ABC25072.1|.

All 10 isolates obtained by SSH were similar to each other.

7. GA15651-PA (*D. pseudoobscura*)/ Unknown

No conserved domains have been identified for this query sequence.

No significant similarity found by GenBank analysis. Seventy-five percent identity was found with GA24726 (*D. pseudoobscura pseudoobscura*) with E value 3e-04 (accession number = ref|XP_002138362.1|) by BlastX result.

Six similar sequences were present for GA15651-PA giving one isoform.

8. GA21196-PA (*D. pseudoobscura*)/ Unknown

No conserved domains have been identified for this query sequence.

BlastX result showed E value 4e-32 with maximum identity of 66% with GA21196 (*D. pseudoobscura pseudoobscura*) (accession number = ref|XP_001352687.1|).

Out of five isolates, two different types of isoform were found (4 of them were similar, while 1 was different) by Clustal W analysis.

9. Unnamed protein product (*Macaca fascicularis*)/ Unknown

No conserved domains have been identified for this query sequence.

GenBank analysis showed 100% identity to *Homo sapiens* haplotype HV1 mitochondrion, complete genome with E value of 0 and accession number of FJ210914.1. BlastX result showed 82% identity to unnamed protein product (*Macaca fascicularis*) with accession number dbj|BAE87620.1| and E value 5e-04.

Four isolates were identified, and all of them showed similarity by Clustal W analysis.

10. Storage protein-binding protein (*S. peregrina*)/ Hemocyanin

Domain of storage protein-binding protein is Hemocyanin. This family includes arthropod hemocyanins and insect larval storage proteins.

Ninety-three percent identity was found with *S. peregrina* mRNA for storage protein-binding protein, complete cds with E value=0 and D29741.1 accession numer by GenBank analysis. Eighty-four percent identities was seen with storage protein-binding protein (*S. Peregrina*) with E value of 8e-78 and accession number dbj|BAA06161.1| by BlastX result.

Four sequences were identified, and all of them were similar.

11. *Enterobacteria phage* phiX174 (*Coliphage phi*X174)/ Microvirus A* protein

Microvirus A* protein is the main domain. This family contains several microvirus A* proteins. The A* protein binds to double stranded DNA and prevents their hydrolysis by nucleases.

GenBank analysis gave 99% identity with *Enterobacteria phage* phiX174 isolate DEL4, complete genome with E value 2e-137 and accession number EF380032.1. BlastX result gave 97% identity to protein A (*Enterobacteria phage* phiX174). Its E value is 2e-46 and accession number gb|ABN49798.1|.

Four isolates were identified, and all of them were similar.

12. Hypothetical protein SPAC1B2.03c (*Schizosaccharomyces pombe*)/GNS1/SUR4 family

The main domain is GNS1/SUR4 family. Members of this family are involved in long chain fatty acid elongation systems that produce the 26-carbon precursors for ceramide and sphingolipid synthesis.

No significant similarity was found by GenBank analysis and BlastX result.

Three similar sequences were identified by Clustal W analysis.

13. Transferrin precursor (*S. peregrina*)

Transferrin is the main domain found.

GenBank analysis gave 94% identity to *S. peregrina* mRNA for transferrin, complete cds with E value 0 and accession number D28940.1. BlastX result, on the other hand, gave 97% identity with Transferrin precursor and E value 2e-90 and accession number sp|Q26643|TRF_SARPE.

Out of two isolates obtained by SSH, both are found to be different.

14. Hypothetical protein (*Paramecium tetraurelia*)/ Unknown

No conserved domains have been identified for this query sequence.

One hundred percent of maximum identity was seen with *Danio rerio* pnrc2 mRNA by GenBank analysis having an E value 3e-04 (accession number = AM422120.2). No significant similarity found by BlastX result.

Two similar sequences were identified by Clustal W analysis.

15. NADH dehydrogenase subunit 2 (*Steinernema carpocapsae*)

No conserved domains have been identified for this query sequence.

No significant similarity found by both GenBank analysis and BlastX result

Two similar sequences were identified by Clustal W analysis.

16. Hypothetical protein isoform 1 (*Nasonia vitripennis*)/ Histone 2A

The domain is Histone 2A, which is a subunit of the nucleosome. The H2A subunit performs essential roles in maintaining structural integrity of the nucleosome, chromatin condensation, and binding of specific chromatin-associated proteins.

GenBank analysis showed 87% identity to *D. virilis* GJ15308 (Dvir\GJ15308), mRNA with E value 2e-76 and accession number XM_002058913.1. BlastX result gave 100% identity to PREDICTED: hypothetical protein isoform 1 (*Nasonia vitripennis*) (E value = 3e-41, accession number = ref|XP_001608164.1|).

Only one sequence was identified by SSH.

17. Similar to Mediator of RNA polymerase II transcription subunit 129 (*Rattus norvegicus*)

The main domain is transcription mediator subunit Med12 which is responsible for suppression of transcription.

GenBank analysis and BlastX result showed no significant similarity.

Only one sequence was identified by SSH.

18. Casein kinase Ialpha CG2028-PB, isoform B (*D. melanogaster*)

Serine/Threonine protein kinases, catalytic domain are the main domain found.

GenBank analysis gave no significant similarity whereas Blast result showed 93% identity with Casein kinase Ialpha (*D. melanogaster*) with E value 1e-16 and accession number ref|NP_511140.1|.

Only one sequence was identified by SSH.

19. GA14517-PA (*D. pseudoobscura*)/ ATP synthase subunit C

ATP synthase subunit C is the main domain.

91% identity was found with *D. pseudoobscura pseudoobscura* GA14517 (Dpse\GA14517), mRNA (E value= 5e-137, accession number= XM_001357527.2) and BlastX result gave 98% identity with GA14517 (*D. pseudoobscura pseudoobscura*)(E value = 8e-36, accession number = ref|XP_001357564.1|).

Only one sequence was obtained by SSH.

20. Pollen coat oleosin-glycine rich protein (*Olimarabidopsis pumila*)

Oleosin is the main domain.

No significant similarity was found by both GenBank analysis and BlastX result.

Only one sequence was obtained by SSH.

21. CG8092-PA, isoform A (*D. melanogaster*)/ Unknown

No conserved domains have been identified for this query sequence.

No significant similarity was found by GenBank analysis but BlastX result gave 81% identity with CG8092 CG8092-PA, isoform A (*D. melanogaster*) and E value 2e-67 (accession number = ref|NP_611019.2|).

Only one sequence was obtained by SSH.

22. Hypothetical protein AaeL_AAEL003275 (*Aedes aegypti*)/ Zinc-finger associated domain

Zinc-finger associated domain (zf-AD). The zf-AD domain is thought to be involved in mediating dimer formation but does not bind to DNA.

GenBank analysis showed no significant similarity. Thirty-one percent identity was found by BlastX result with hypothetical protein AaeL_AAEL003275 (*Aedes aegypti*) with very low E value (0.002) and accession number ref|XP_001656653.1|.

Only one sequence was identified by SSH.

23. PREDICTED: hypothetical protein (*Mus musculus*)

No conserved domains have been identified for this query sequence.

No significant similarity seen by either GenBank analysis or BlastX result.

Only one sequence was obtained by SSH.

24. RNA helicase

DEAD-box helicases is the main domain. They are diverse family of proteins needed in a variety of cellular processes including splicing, ribosome biogenesis and RNA degradation.

GenBank analysis showed 82% identity to *D. mojavensis* GI18279 (Dmoj\GI18279), mRNA with E value 4e-83 and accession number XM_002004086.1. BlastX result showed 97% identity to RNA helicase (E value= 6e-65, accession number = gb|AAA28603.1|).

Only one sequence was obtained by SSH.

25. Cyclin K (*Gallus gallus*)

Cyclin box fold is the main domain. This domain is the protein binding domain functioning in cell-cycle and transcription control.

No significant similarity was found by GenBank analysis, and BlastX result showed 54% identity to cyclin K (*Gallus gallus*) with very low E value (0.001) and accession number ref|NP_001026380.1|.

Only one sequence was obtained by SSH.

26. Hypothetical protein DDBDRAFT_0206406 (*Dictyostelium discoideum* AX4)/DEAD-box helicases

The main domain is DEAD-box helicases: a diverse family of proteins involved in ATP-dependent RNA unwinding, needed in a variety of cellular processes including splicing, ribosome biogenesis and RNA degradation.

GenBank analysis gave 100% identity to *Belgica antarctica* clone Ba-U40 CG32816-like mRNA with an accession number DQ507301.1 and E value 3e-04. BlastX result gave no significant similarity.

Only one sequence was obtained by SSH.

This list has been combined with all previously listed differentially and non differentially regulated genes isolated from studies of *S. crassipalpis,* with references to the relevant papers .The sequences from this study will be deposited in GenBank for an accession number assignment.

Table 7 lists all the genes that are found to be diapause upregulated in *S. crassipalpis*. Thirty-four different types of genes that are upregulated during diapause in *S. crassipalpis* have already been identified. This study identified 59 genes as diapause upregulated genes.

Table 7 List of Diapause Upregulated Genes in *S. crassipalpis*

S.N.	Name of the identified sequence or GenBank accession number	Organism	Author
1	ScHsp23	*S. crassipalpis*	Denlinger 1991; Yocum et al. 1991, 1998; Kimura et al.1998; Reinhart et al. 2000; Hayward et al. 2005
2	23kDa heat shock protein, isoform A	*S. crassipalpis*	sscDNA (2009)
3	23kDa heat shock protein, isoform B	*S. crassipalpis*	sscDNA (2009)
4	HSP70	*S. crassipalpis*	Denlinger 1991; Yocum et al. 1991, 1998; Kimura et al.1998; Reinhart et al. 2000; Hayward et al. 2005
5	70kDa heat shock protein, isoform A	*S. crassipalpis*	sscDNA (2009)
6	70kDa heat shock protein, isoform B	*S. crassipalpis*	sscDNA (2009)
7	70kDa heat shock protein, isoform C	*S. crassipalpis*	sscDNA (2009)
8	70kDa heat shock protein, isoform D	*S. crassipalpis*	sscDNA (2009)
9	70kDa heat shock protein, isoform E	*S. crassipalpis*	sscDNA (2009)
10	70kDa heat shock protein, isoform F	*S. crassipalpis*	sscDNA (2009)

Table 7 Continued

11	70kDa heat shock protein, isoform G	*S. crassipalpis*	sscDNA (2009)
12	70kDa heat shock protein, isoform H	*S. crassipalpis*	sscDNA (2009)
13	70kDa heat shock protein, isoform I	*S. crassipalpis*	sscDNA (2009)
14	70kDa heat shock protein, isoform J	*S. crassipalpis*	sscDNA (2009)
15	Hsc70	*S. crassipalpis*	Rinehart, Yocum and Denlinger 2000
16	Heat shock protein 70	*S. crassipalpis*	sscDNA (2009)
17	Cognate 70 kDa heat shock protein scHSC70 (HSC70) gene	*S. crassipalpis*	Rinehart , Yocum and Denlinger 2000
18	Cognate 70kDa heat shock protein	*S. crassipalpis*	sscDNA (2009)
19	25kDa small heat shock protein	*S. crassipalpis*	Rinehart et al. 2007
20	25kDa small heat shock protein, isoform A	*S. crassipalpis*	sscDNA (2009)
21	25kDa small heat shock protein, isoform B	*S. crassipalpis*	sscDNA (2009)
22	25kDa small heat shock protein, isoform C	*S. crassipalpis*	sscDNA (2009)
23	25kDa small heat shock protein, isoform D	*S. crassipalpis*	sscDNA (2009)
24	Small heat shock protein	(*C. p.pipiens*)	sscDNA (2009)
25	Heat shock protein 60	(*L. huidobrensis*)	sscDNA (2009)
26	60S acidic ribosomal protein PO	*S. crassipalpis*	Craig and Denlinger 2000
27	60s acidic ribosomal protein/Similar to D. melanogaster RpS6	(*D. yakuba*)	sscDNA (2009)
28	Ribosomal protein L37a CG 9091-PA, isoform A	(*D. melanogaster*)	sscDNA (2009)
29	Ribosomal protein L37a CG 9091-PA, isoform B	(*D. melanogaster*)	sscDNA (2009)
30	Ribosomal protein L13a CG 1475-PB	(*D. melanogaster*)	sscDNA (2009)
31	Ribosomal protein S6e/Ribosomal protein P2, isoform A	*(C. capitata*)	sscDNA (2009)
32	Ribosomal protein S6e/Ribosomal protein P2, isoform B	*(C. capitata*)	sscDNA (2009)
33	Ribosomal L29 protein/ CG4111-like protein	(*D. miranda*)	sscDNA (2009)
34	Ribosomal protein S8e/ Similar to D. melanogaster CG7808	(*D. yakuba*)	sscDNA (2009)
35	Beta tubulin at 60D CG3401-PA	(*D. melanogaster*)	sscDNA (2009)
36	Beta tubulin	(*C. p.pipiens*)	sscDNA (2009)
37	Ultraspiracle (USP)	*S. crassipalpis*	Henrich and Brown 1995
38	pScD41	*S. crassipalpis*	Flannagan et al. 1998
39	Clock	*S. crassipalpis*	Denlinger 2002; Goto and Denlinger 2002b
40	Doubletime	*S. crassipalpis*	Denlinger 2002; Goto and Denlinger 2002b

Table 7 Continued

41	Timeless	*S. crassipalpis*	Denlinger 2002; Goto and Denlinger 2002b
42	Cycle	*S. crassipalpis*	Denlinger 2002; Goto and Denlinger 2002b
43	Cryptochrome	*S. crassipalpis*	Denlinger 2002; Goto and Denlinger 2002b
44	Period	*S. crassipalpis*	Denlinger 2002; Goto and Denlinger 2002b
45	Sarcocystatin B	*S. crassipalpis*	Goto and Denlinger 2002a
46	13560880	(*D. melanogaster*)	Li et al. 2007
47	1037174	*(Elmeria maxima)*	Li et al. 2007
48	13560880	(*D. melanogaster*)	Li et al. 2007
49	5570	*(Aplysia californica)*	Li et al. 2007
50	1037174	*(Elmeria maxima)*	Li et al. 2007
51	2058737	*S. crassipalpis*	Li et al. 2007
52	2058737	*S. crassipalpis*	Li et al. 2007
53	2058737	*S. crassipalpis*	Li et al. 2007
54	2058737	*S. crassipalpis*	Li et al. 2007
55	2058737	*S. crassipalpis*	Li et al. 2007
56	2058737	*S. crassipalpis*	Li et al. 2007
57	66548185	*(Apis mellifera)*	Li et al. 2007
58	67083335	*(Ixodes scapularis)*	Li et al. 2007
59	125090	*(Ovis aries)*	Li et al. 2007
60	Unknown		Li et al. 2007
61	Unknown		Li et al. 2007
62	Unknown		Li et al. 2007
63	Unknown		Li et al. 2007
64	Unknown		Li et al. 2007
65	Unknown, Hypothetical protein PF14_0710	(*Plasmodium falciparum*)	sscDNA (2009)
66	Unknown,GA 12626-PA	(*D. pseudoobscura*)	sscDNA (2009)
67	Unknown, PE-PGRS family protein	(*M. tuberculosis* F11)	sscDNA (2009)
68	Proteasome beta type-4 subunit,CG12000-PA,isoform A	(*D. melanogaster*)	sscDNA (2009)
69	Domain of Unknown Function ,GA 19395-PA	(*D. pseudoobscura*)	sscDNA (2009)
70	Subfamily of SANT domain , RE 38876P	(*D. melanogaster*)	sscDNA (2009)
71	NifU-like N terminal domain, CG 9836-PA	(*D. melanogaster*)	sscDNA (2009)

Table 7 Continued

72	Unnamed protein product	(*D. melanogaster*)	sscDNA (2009)
73	Fascin-like domain/Singed CG32858-PC, isoform	(*D. melanogaster*)	sscDNA (2009)
74	Malate dehydrogenases (MDH), CG 7998-PA	(*D. melanogaster*)	sscDNA (2009)
75	ATP synthase beta	(*D. simulans*)	sscDNA (2009)
76	Chicade	(*D. simulans*)	sscDNA (2009)
77	Cyclophilin 1 CG 9916-PA	(*D. melanogaster*)	sscDNA (2009)
78	Insect cuticle protein, GA 10081-PA	(*D. pseudoobscura*)	sscDNA (2009)
79	Diapause No Hit, isoform A	*S. crassipalpis*	sscDNA (2009)
80	Diapause No Hit, isoform B	*S. crassipalpis*	sscDNA (2009)
81	Diapause No Hit, isoform C	*S. crassipalpis*	sscDNA (2009)
82	Diapause No Hit, isoform D	*S. crassipalpis*	sscDNA (2009)
83	Diapause No Hit, isoform E	*S. crassipalpis*	sscDNA (2009)
84	Diapause No Hit, isoform F	*S. crassipalpis*	sscDNA (2009)
85	Diapause No Hit, isoform G	*S. crassipalpis*	sscDNA (2009)
86	Diapause No Hit, isoform H	*S. crassipalpis*	sscDNA (2009)
87	Diapause No Hit, isoform I	*S. crassipalpis*	sscDNA (2009)
88	Diapause No Hit, isoform J	*S. crassipalpis*	sscDNA (2009)
89	Diapause No Hit, isoform K	*S. crassipalpis*	sscDNA (2009)
90	Diapause No Hit, isoform L	*S. crassipalpis*	sscDNA (2009)
91	Diapause No Hit, isoform M	*S. crassipalpis*	sscDNA (2009)
92	Diapause No Hit, isoform N	*S. crassipalpis*	sscDNA (2009)
93	Diapause No Hit, isoform O	*S. crassipalpis*	sscDNA (2009)

The genes that are found to be diapause downregulated (nondiapause upregulated) are enlisted in Table 8. Ten genes have already been reported to be diapause downregulated genes from *S. crassipalpis,* and this study added 37 more genes to be diapause upregulated genes.

Table 8 List of Nondiapause Upregulated Genes in *S. crassipalpis*

S.N.	Name of the identified sequence or GenBank accession number	Organism	Author
1	PCNA	*S. crassipalpis*	Tammariello and Denlinger 1998
2	HSP 90	*S. crassipalpis*	Rinehart and Denlinger 2000; Denlinzger 2002
3	Sarcocystatin A	*S. crassipalpis*	Goto and Denlinger 2002a
4	1591246	*(Methanocaldococcus jannaschii)*	Li et al. 2007
5	160843	*(Schistocerca gregaria)*	Li et al. 2007
6	3757564	*(D. melanogaster)*	Li et al. 2007

Table 8 Continued

7	74310551	*(Epidermophyton floccosum)*	Li et al. 2007
8	4001821	*(Ambystoma tigrinum)*	Li et al. 2007
9	50552978	*(Yarrowia lipolytica)*	Li et al. 2007
10	39721833	*(Onion yellows phytoplasma)*	Li et al. 2007
11	bZIP transcription factor, NFE2L2	(*Pan troglodytes*)	sscDNA (2009)
12	Unknown,Putative SPT transcriptional factor family membrane	(*Lepephtheirus salmonis*)	sscDNA (2009)
13	Unknown, Hypothetical protein , isoform A	(*Thermobia domestica*)	sscDNA (2009)
14	Unknown, Hypothetical protein, isoform B	(*Thermobia domestica*)	sscDNA (2009)
15	RNA binding protein,/Variable membrane protein precursor	(*Trichomonas vaginalis*)	sscDNA (2009)
16	Serine protease inhibitor 1	(*Glossina morsitans morsitans*)	sscDNA (2009)
17	Unknown , GA15651-PA	(*D. pseudoobscura*)	sscDNA (2009)
18	Unknown , GA21196-PA , isoform A	(*D. pseudoobscura*)	sscDNA (2009)
19	Unknown , GA21196-PA , isoform B	(*D. pseudoobscura*)	sscDNA (2009)
20	Unknown , Unnamed protein product	(*Macaca fascicularis*)	sscDNA (2009)
21	Hemocyanin, ig-like domain/ Storage protein binding protein	(*S. peregrine*)	sscDNA (2009)
22	Microvirus A* protein/Enterobacteria phage	(*Coliphage phix* 174)	sscDNA (2009)
23	GNS1/SUR4 family/Hypothetical protein SPAC1B2.03C	(*Schizosaccharomyces prombe*)	sscDNA (2009)
24	Transferring precursor , isoform A	(*S. peregrine*)	sscDNA (2009)
25	Transferring precursor, isoform B	(*S. peregrine*)	sscDNA (2009)
26	Unknown , Hypothetical protein	(*Paramecium tetraurelia*)	sscDNA (2009)
27	Unknown , NADH dehydrogenase subunit 2	(*Steinernema carpocapsae*)	sscDNA (2009)
28	Histone 2A/ Predicted hypothetical protein isoform 1	(*Nasonia vitripennis*)	sscDNA (2009)
29	Transcription mediator subunit Med12/ Similar to Mediator of RNA polymerase II transcription subunit 129	(*Rattus norvegicus*)	sscDNA (2009)
30	Casein kinase lalpha	(*D. melanogaster*)	sscDNA (2009)
31	ATP synthase subunit C/GA14517-PA	(*D. pseudoobscura*)	sscDNA (2009)

Table 8 Continued

32	Pollen coat oleosin –glycine rich protein	(*Olimarabidopsis pumila*)	sscDNA (2009)
33	Unknown , CG 8092-PA, isoform A	(*D. melanogaster*)	sscDNA (2009)
34	Zinc-finger associated domain/Hypothetical protein AaeL_AAEL003275	(*Aedes aegypti*)	sscDNA (2009)
35	Unknown , Hypothetical protein	(*Mus musculus*)	sscDNA (2009)
36	RNA helicase / DEAD-box helicase ,		sscDNA (2009)
37	Cyclin K	(*Gallus gallus*)	sscDNA (2009)
38	DEAD-box helicase ,Hypothetical protein DDBDRAFT	(*Dictyostelium discoideum*)	sscDNA (2009)
39	Non Diapause No Hit, isoform A	*S. crassipalpis*	sscDNA (2009)
40	Non Diapause No Hit, isoform B	*S. crassipalpis*	sscDNA (2009)
41	Non Diapause No Hit, isoform C	*S. crassipalpis*	sscDNA (2009)
42	Non Diapause No Hit, isoform D	*S. crassipalpis*	sscDNA (2009)
43	Non Diapause No Hit, isoform E	*S. crassipalpis*	sscDNA (2009)
44	Non Diapause No Hit, isoform F	*S. crassipalpis*	sscDNA (2009)
45	Non Diapause No Hit, isoform G	*S. crassipalpis*	sscDNA (2009)
46	Non Diapause No Hit, isoform H	*S. crassipalpis*	sscDNA (2009)
47	Non Diapause No Hit, isoform I	*S. crassipalpis*	sscDNA (2009)

Table 9 lists all the genes that have an unknown regulation. Seventeen genes were reported from *S. crassipalpis* to have unknown regulation.

Table 9 List of Genes with Unknown Regulation in *S. crassipalpis*

S.N.	Gene	Author
1	Cytochrome oxidase subunit I	Chen et al. 2008
2	Neuropeptide-like precursor 4	Li , Rinehart and Denlinger 2008
3	Extracellular signal-regulated kinase	Fujiwara and Denlinger 2007
4	p38 mitogen-activated protein kinase	Fujiwara and Denlinger 2007
5	TCP-1 containing chaperonin	Rinehart et al. 2006
6	18 kDa small heat shock protein	Rinehart et al. 2006
7	Ribosomal protein L32	Goto and Denlinger 2002
8	Fat body protein 2	Goto and Denlinger 2002
9	Regucalcin	Goto and Denlinger 2002
10	Cytochrome b	Song et al. 2005

Table 9 Continued

11	Ribosomal protein S9	Michaud and Denlinger 2003
12	Putative ATP synthase-beta subunit	Hayward, Rinehart and Denlinger 2003
13	Sarcotoxin IIa	Rinehart, Diakoff and Denlinger 2002
14	Ecdysone receptor	Rinehart et al. 2001
15	Arylphorin	Yeager-Hall and Tammariello 2000
16	NADH dehydrogenase subunit 5 (ND5)	Zehner et al. 2003
17	Small heat shock protein	Rinehart et al. 2007

CHAPTER 4

DISCUSSION

This study has identified 59 different diapause up-regulated genes and 37 different non-diapause upregulated genes by the process of suppression subtractive library construction and randomly isolated and sequenced clones. None of the diapause-specific sequences found in the DRL overlapped with the sequences isolated from the NDRL. These results support the premise that the technique of SSH library construction can be used to isolate differentially regulated genes because the ones isolated were found to be expressed in either one or the other of the libraries, but not in both.

The results of these randomly isolated clones, many are gene sequences reported for the first time, along with previous published reports, continue to indicate that the diapause programming is complex and are being regulated at many levels of control. Prior studies include extensive and complex differential gene expression to accompany results in brain protein expression (Joplin et al. 1990; Li et al. 2008). The finding, that we have identified 30 different sequence types out of a cDNA library of unknown complexity, suggests that there are many more gene sequences that still need to be identified. This supports previous studies that have estimated between 500-1000 genes that are differentially regulated in diapause (Joplin et al. 1990; Flanagan et al.1998; Joplin un-published research).

Although the inducible HSP genes have previously been found to be expressed constitutively during diapause, our results, along with the report on brain protein identification (Li et al. 2008), further confirm that there is a large expression of multiple copies of the diapause expressed HSP (dHSP). Out of seven different types of upregulated dHSP, five of them (25kDa small heat shock protein , 70kDa heat shock protein, 23kDa heat shock protein, cognate 70kDa heat shock protein, heat shock protein 70) have already been identified as being in *S. crassipalpis* (ScHSP), and been shown to be expressed during

diapause (Rinehart et al. 2000, Denlinger 2002). In addition to the previously identified ScHSP, one isolate shows similarity to a small heat shock protein found in *C. p. pipiens*, and one isolate shows similarity with heat shock protein 60 from *Liriomyza huidobrensis* (*Lh*hsp 60). This HSP appears to be similar to the HSP found in mitochondria. Thus, these results verify and extend the findings that HSP, which are normally induced in stress conditions, are found to be constitutively expressed during the pupal diapause in this species and further strengthens this common feature of pupal diapause (Denlinger et al. 2001), but not adult diapause (Goto and Kimura 2004; Robich et al. 2007).

Dendrograms constructed from Clustal W show that 25kDa hsp with accession number EF103577 has close relation to isoform B and distance to isoform A (Figure 6). 70kDa hsp showed close relation to the gene that is already reported (AF107338) with many of the isoforms that this study identified (Figure 7). Likewise, the already identified 23 kDa hsp gave close relation with isoform A (Figure 8).

A surprising finding is that, even though aspects of the SSH cDNA library construction technique is supposed to normalize transcript clones to some degree, a number of isolates represent multiple different sequences with the same identity. This indicates that there are multiple copies and high expression levels of all types of the ScdHSP mRNA during diapause. These results suggest that the expression level of all ScdHSP is associated with the increased stress resistance observed in diapausing pupae and thus are involved in diapause syndrome genes (Fig. 10) (Denlinger et al. 2001).

Interestingly, the major large HSP 90 is diapause down-regulated (Rinehart and Denlinger 2000; Denlinger 2002) and does not appear in our libraries. The HSP90 is expressed during normal cellular functions and is involved in binding steroid receptors, such as the ecdysone receptor (Joplin and Denlinger 1990). Although the ecdysone receptor is not differentially expressed during diapause, the HSP 90 maybe required to maintain the

ecdysone receptor in a control system for metamorphic development. Although HSP 90 is upregulated during stress, its function during stress response is open to question.

Most of the rest of the diapause isolates are unique or only represented by a few genes, and do not show the variation in sequence found in the HSP. These variations in number can be used to estimate the number of separate isolates needed for further sequencing and can be used to determine the limit of the number that we have to sequence to get a representative sampling of the number of sequences that represent the diapause upregulated mRNA population. If 4.8% (9/187) of the DRL are unique sequences, then app. 4000 clones will need to be sequenced to get a good sampling of the number of unique clones. Unfortunately, of these at least 2000 clones will be the HSP and ribosomal subunit sequences.

Nine diapause upregulated genes represent multiple types of ribosomal protein subunits in *Sarcophaga* pupal diapause. A single ribosomal subunit gene has been previously reported in diapause (Craig and Denlinger 2000), but the 9 additional genes in this study represent a significant increase in this class of genes during pupal diapause. These may act similar to the prokaryotic "sigma" factors, controlling translation of a subset of the diapause-specific mRNA (Gruber and Gross, 2003). Sigma factors are found in prokaryotic systems that allow groups of mRNAs to be expressed. *E coli* has at least 8 sigma factors that express subsets of genes in response to normal conditions for housekeeping genes, nitrogen limitation conditions, nutrient starvation, hest shock or stress among other conditions. If the subunits found here have similar functions, such a finding would be the first report of this method for regulating gene expression during diapause. In the adult diapause of *C. pipiens*, 8 ribosomal subunits genes were up-regulated (Robich et al. 2007). This may indicate that the developmental state of diapause may have this expression pattern in common.

Other types of genes that are diapause upregulated are those involved with cellular structural genes. These include previously unreported two beta tubulin genes, the fascilin-like

actin-binding protein, and perhaps the insect cuticle protein (GA 10081-PA). These could be involved in maintaining cellular structure or cuticle structure in the absence of restructuring during metamorphosis.

There are a few genes that are apparently associated with cell cycle control [cyclophilin 1 (GC 9916-PA)] that could be associated with the down regulation of the previously identified PCNA (Tammariello and Denlinger 1998) and the cyclin K (this study). Since one of the characteristics of diapause is the morphological stasis of the organism, the down regulation of some of the cell cycle control genes along with the genes that are upregulated may be involved in maintaining this aspect of the diapause syndrome. These would be part of the diapause control functions that are involved in the maintenance of the diapause developmental state. These genes could be responsive to the genes that are hypothesized to be expressed after the reception of the environmental signals that trigger the diapause developmental state (Figure 10).

Down regulation of metabolism is a major indication of the diapause programming. The up-regulation of malate dehydrogenase, ATP synthase beta, and perhaps the proteasome beta subunit may reflect control of this portion of the controlling factors of the syndrome. Malate dehydrogenase was also up-regulated in the adult diapause of *C. pipiens* (Robich et al. 2007). Only a few genes in this category were isolated in this study which may indicate that there are more genes in the library that have not yet been sampled. Likewise, many of the unidentified sequences could simply indicate that only a partial sequence was available from the subtractive process, the cloning procedure, or the sequencing reaction. Further work remains to be done on each of these gene fragments.

Diapause Down-Regulated Genes

In contrast to the genes up-regulated in diapause, the genes up-regulated in the non-diapause pupae (diapause downregulated) do not show the clustering of sequences identified

in the upregulated library. With few exceptions, most of the sequences were only from low numbers of isolates. There were 20 isolates of an unknown gene, but only 3 sequences were represented by two isoforms (Table 2). None of the stress related genes or ribosomal subunits were isolated in the nondiapause library.

The largest group appears to involve 6 transcription factors (NFE2L2, SPT, RNA pol II, AaeL, RNA helicase and DEAD-box helicase). Since the nondiapause state is more active than the diapause state, these transcription factors should represent the control of a larger set of genes involved in the increased metabolic and developmental activity of the metamorphic developmental pupal state.

Diapause Gene Control Networks

There are few studies that have examined the genes that are upregulated in the nondiapause stage. One of the closest studies is the microarray analysis of genes expressed in all life stages in *Drosophila* (Arbeitman et al. 2002). Although there are differences between the methodology of these two studies, particularly since the control stage is different, a comparison of the results deserve a more complete analysis that is beyond the scope of this thesis. Genes from this study that are expressed during the pupal stage in *Drosophila* could be a starting point for comparison with nondiapause upregulated genes.

The genes identified in this study during the pupal stage of *Drosophila* were separated into the following 10 GO classes of genes: Endopeptidase, Enzyme, Structural, Cell cycle, Chaperone, Cytoskeleton, Signaling, Transcription factor, Metabolism, and Cell adhesion. The genes isolated from this study show differential expression during diapause particularly the ScdHSP and the diapause ribosomal subunits during diapause and the upregulation of transcription factors during the nondiapause stage. The identification of the unknown genes as well as a more through sampling of the library will begin to fill in many of these other classes of gene.

A second study that has applicability to this present study is the use of SSH library construction to examine the genes differentially regulated during the late summer-early fall adult diapause of the northern house mosquito, *Culex pipiens* (Robich et al. 2007). The physiological state of the pupal diapause of *Sarcophaga* and the adult diapause of *Culex* are very different, but there are still some interesting comparisons. There was no upregulation of HSP genes in contrast to the extreme upregulation of all the HSP in the *Sarcophaga* pupal diapause. However, a similar situation occurs in the adult diapause of *Drosophila triauriaria* in which HSP are not upregulated in adult diapause (Goto and Kimura 2004). Thus, this may reflect the difference between pupal and adult diapause programming.

One aspect of similarity between these two studies is the upregulation of a number of ribosomal protein subunits. Seven were found upregulated in adult diapause in *Culex* and 9 were found to be upregulated in the *Sarcophaga* pupal diapause. More study is needed to determine the functions of this set of genes during the diapause state. One aspect that prevents a more comprehensive comparison is the limited number of clones that have been identified in both studies. Only 95 clones were examined in the *Culex* study while the *Sarcophaga* study examined almost 400 clones. A more through sampling of both libraries will be required to assemble a more comprehensive set of genes for comparison.

Despite the need to sequence a more extensive set of isolates, the results of this study have allowed us to propose a Hypothetical Gene Control Network for the establishment and control of pupal diapause (Figure 7) in *S. crassipalpis*. Diapause inducing photosensitive stage is the stage in the life cycle of this fly when diapause can be induced i.e., the last two days of embryonic life and first two days of larval life. During the diapausing inducing photosensitive stage if the fly receives the short day length signal, i.e., 12L: 12D, then there will be the initiation of diapause which takes place in pupal stage. In this diapause inducing photosensitive stage, there must be the differential regulation of genes in embryonic/ larval

stage for the initiation of diapause in pupal stage. A genetic study of the pupal diapause in a similar fly, *Sarcophaga bullata*, suggests that the induction is associated with a single gene, or a set of a few tightly linked genes, possibly connected with the photoperiodic clock control genes (Han and Denlinger, 2009).

Since diapause takes place at the pupal stage, there is an up-regulation of diapause program genes. A number of genes would be predicted to be differentially regulated at this stage and our project verifies it.

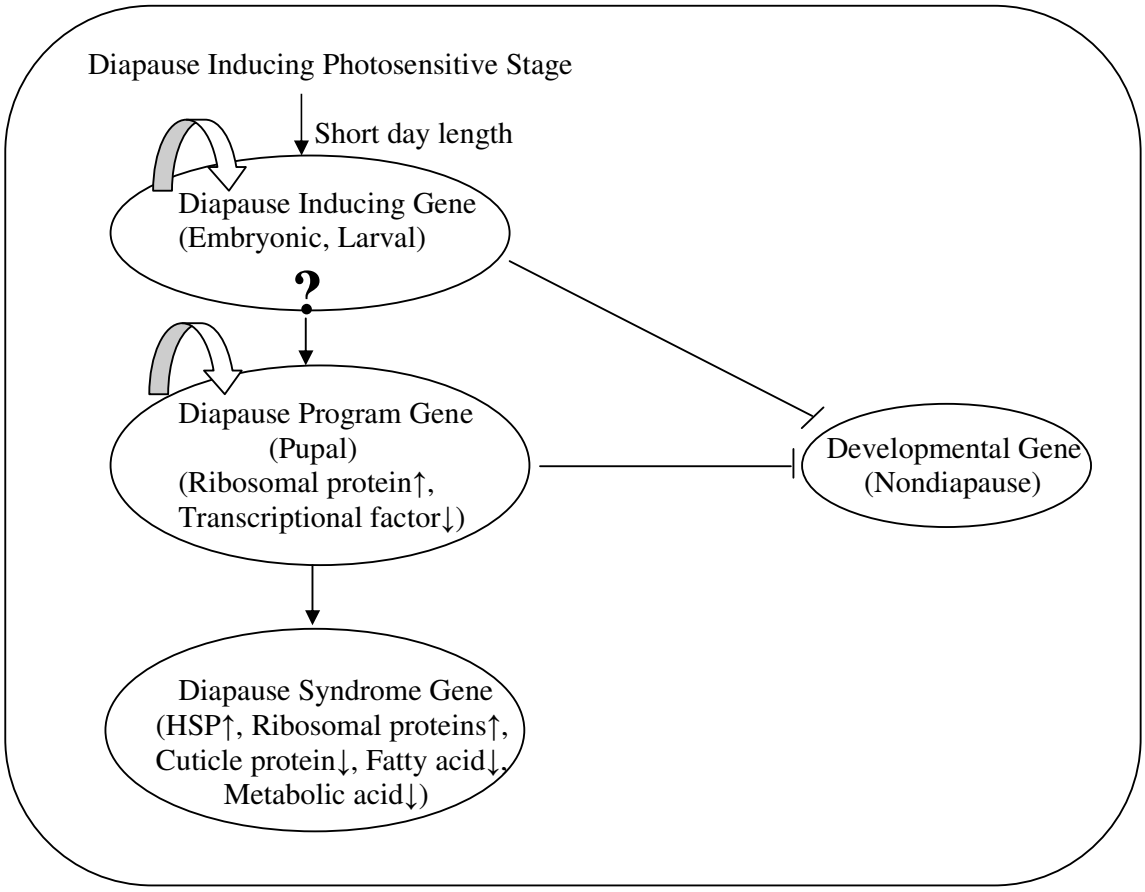


Figure 10 Hypothetical Gene Control Network for the Establishment and Control of Diapause

Only few genes are reported in GenBank to be *S. crassipalpis* specific genes and their possible functions. Different heat shock proteins may have different functions. Role of hsp23 and hsp70 could be the taking care of the enzymes that play a role in metabolism during

diapause, also these dHSP may regulate diapause by arresting cell cycle hence ceasing development in insects during diapause. The molecular function of Ribosomal protein L37a CG 9091-PA is to provide structural constituent of ribosome. The expression of ribosomal protein subunits in diapause may be as diapause program gene as its biological function is translation. Its function during diapause is unknown. *Chicadae* may play an important role in the cytoskeleton with function in major signaling pathways by interacting with components of the phosphatidylinositol cycle and *Ras* pathway.

Likewise, and perhaps just as important for diapause, is the down regulation of genes that are expressed during the nondiapause pupal developmental stage. It is quite possible that these genes need to be downregulated in order for the diapause controlling genes to function. As shown in Figure 11, the expression of some of these genes may shut down the expression of the diapause upregulated thereby preventing their expression and preventing the diapause programming to run. The upregulation of the cell cycle genes, hsp90, the various transcription factors and the metabolic enzyme genes would be expected in an actively developing and metabolic state.

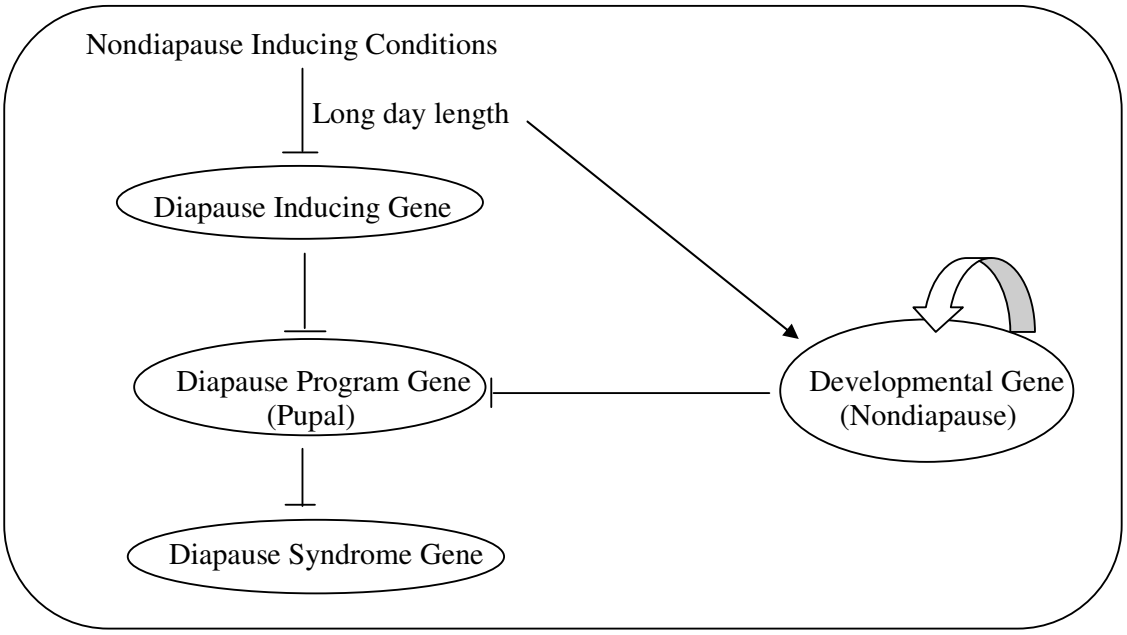


Figure.11 Hypothetical Gene Control Network for the Establishment and Control of the Nondiapause Pupal Developmental State.

This study is an initial partial survey of genes that are differentially regulated during the pupal diapause developmental state in *S. crassipalpis*. Of the 96 differentially genes, many could be assigned to specific classes of protein that could have a regulatory function in controlling this state. Many of these genes are identified for the first time to be differentially regulated in this state. Much more work is needed to more fully sample the clones in these libraries to create a more complete picture of the genetic network between the diapause and nondiapause pupal developmental states of *S. crassipalpis*.

REFERENCES

Arbeitman MN, EEM Furlong, F Imam, E Johnson, BH Null, BS Baker, MA Krasnow, MP Scott, RW Davis and KP White. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. Science 297:2270–2275.

Craig TL and DL Denlinger. 2000. Sequence and transcription patterns of 60S ribosomal protein PO, a diapause-regulated AP endonuclease in the flesh fly, *Sarcophaga crassipalpis*. Gene 255:381-88.

Denlinger DL. 1972. Induction and termination of pupal diapause in *Sarcophaga* (dipteral: Sarcophagidae). Biol bull. Wood's Hole 142:11-24.

Denlinger DL. 1981. Hormonal and metabolic aspects of pupal diapause in diptera. Entomologia Generalis 7:245-259.

Denlinger DL. 1985. The hormonal control of diapause. In: Comprehensive Insect Physiology. Biochemistry and Pharmacology. Ed. GA Kerkut and LI Gilbert. Pergamon Press, Oxford 8:353-412.

Denlinger DL. 1986. Dormancy in tropical insects. Annu. Rev. Entomol. 31:239-264.

Denlinger DL, KH Joplin, CP Chen, and RE Lee. 1991. Cold shock and heat shock: In Insects at Low Temperature. Ed. R.E. Lee and D.L. Denlinger. Chapman and Hall, New York.

Denlinger DL, JP Rinehart and GD Yocum. 2001. Stress proteins: a role in insect diapause? 155-71.

Denlinger DL. 2002. Regulation of diapause. Annu. Rev. Entomol. 47:93-122.

Denlinger DL, GD Yocum and JP Rinehart. 2005. Hormonal control of diapause. Elsevier BV 615-650.

Flannagan RD, SP Tammariello, KH Joplin, RA Cikta-Ireland, GD Yocum and DL Denlinger. 1998. Diapauses –specific gene expression in pupae of flesh fly *Sarcophaga crassipalpis.* Proc. Natl Acad. Sci. USA 95:5616-5620.

Gadgil C, A Rink, C Beattie and WS Hu. 2002. A mathematical model for suppression subtractive hybridization. Comp Funct Genom 3:405-422.

Giebultowicz JM and DL Denlinger. 1986. Role of the brain and the ring gland in relation to pupal diapause in the flesh fly *Sarcophaga crassipalpis.* J. Insect Physiol. 32:161-166.

Goto SG and DL Denlinger. 2002a. Genes encoding two cystatins in the flesh fly *Sarcophaga crassipalpis* and their distinct expression patterns in relation to pupal diapause. Gene 29:121-127.

Goto SG and DL Denlinger. 2002. Short-day and long-day expression patterns of genes involved in the flesh fly clock mechanism: *period, timeless, cycle* and *cryptochrome*. J Insect Physiology 48:803-816.

Goto SG and MT Kimura. 2004. Heat-shock-responsive genes are not involved in the adult diapause of *Drosophila triauraria.* Gene 326:117-122.

Groenink M and CJ Aad. 1996. Isolation of delayed early genes associated with liver regeneration using the Clontech PCRSelect ^{TM} subtraction technique. CLONTECHniques XI:7–8.

Gruber TM, Gross CA. 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol. 57:441-66.

Han B and DL Denlinger. 2009. Mendelian inheritance of puapl diapause in the flesh fly, *Sarcophaga bullata.* J Heredity 100:251-255.

Hayward SAL, SC Oavlides, SP Tammariello, JP Rinehart and DL Denlinger. 2005. Temporal expression patterns of diapause-associated genes in the flesh fly pupae from the onset of diapause through post- diapause quiescence. Journal of Insect Physiology 51:631-640.

Henrich VC and NE Brown. 1995. Insect nuclear receptors: a developmental and comparative perspective. Insect Biochem.Mol. Biol. 25:881-97

Hipfel R, B Schittek, Y Bodingbauer and C Garbe. 2000. Specifically regulated genes in malignant melanoma tissues identified by subtractive hybridization. Br J Cancer 82: 1149–1157.

Hondo E and CL Stewart. 2004. Profiling gene expression in growth- arrested mouse embryos in diapause. Genome Biology 6:202.

Joplin KH and DL Denlinger. 1989. Cycles of protein synthesis during pupal diapause in the flesh fly, *Sarcophaga crassipalpis.* Arch. of Insect Biochemistry and Physiology 12: 111-122.

Joplin KH, GD Yocum, DL Denlinger. 1990. Diapause specific proteins expressed by the brain during pupal diapause of the flesh fly *Sarcophaga crassipalpis*. J. Insect Physiol. 36:775-783.

Joplin KH, DL Stetson, JG Diaz and DL Denlinger. 1993. Cellular differences in ring glands of flesh fly pupae as a consequence of diapauses programming. Tissue and Cell 25:245-257.

Kimura MT, KM Yoshida and SG Goto. 1998. Accumulation of Hsp70 mRNA under environmental stresses in diapausing and nondiapausing adults of *Drosophila triauraria*. J Insect Physiol. 44:1009-1015.

Kusmierczyk AR and J Martin. 2001. Chaperonins- keeping a lid on folding proteins. Elsevier Science B.V. FEBS Letters 505:343-347.

Li AQ, A Popova-Butler, DH Dean and DL Denlinger. 2007. Proteomics of the flesh fly brain reveals an abundance of upregulated heat shock proteins during pupal diapause. Journal of insect Physiology 53:385-391.

Morris J, H Tissenbaum and G Ruvkin. 1996. A phosphatidylinositol-3-OH kinase family member regulating longevity and diapause in *Caenorhabidits elegans*. Nature 382: 356-539.

Munir S, S Singh, K Kaur and V Kapur. 2004. Suppression subtractive hybridization coupled with microarray analysis to examine differential expression of genes in virus infected cells. Biol Proced. 6(1): 94-104.

Nijhout HF. 1998. Insect hormones. Princeton University Press, 163.

Rinehart JP and DL Denlinger. 2000. Heat shock protein 90 is downregulated during pupal diapause in the flesh fly, *Sarcophaga crassipalpis*, but remains responsive to thermal stress. Insect Molecular Biology 9:641-645.

Rinehart JP, GD Yocum, and DL Denlinger. 2000. Developmental upregulation of inducible hsp70 transcripts, but not the cognate form, during pupal diapause in the flesh fly, *Sarcophaga crassipalpis*. Insect Biochem. and Mol. Biol. 30:515-521.

Robich, RM, JP Rinehart, LJ Kitchen, and DL Denlinger. 2007. Diapause-specific gene expression in the northern house mosquito, *Culex pipiens* L., identified by suppressive subtractive hybridization. J. Insect Physiol. 53:235-245.

Sanchez S, S Hourdez and FH Lallier*. 2007. Identification of proteins involved in the functioning of *Riftia pachyptila* symbiosis by Subtractive Suppression Hybridization. BMC Genomics, 8:337 doi: 10.1186/1471-2164-8-337

Simpson KS, MH Adams, CY Behrendt-Adam, C Ben Baker and KJ McDowell. 1999. Identification and initial characterization of calcyclin and phospholipase a (2) in equine conceptuses. Mol. Reprod Dev 53:179–187.

Tauber MJ, CA Tauber and S Masaki. 1986. Seasonal adaptations of Insects. New York, Oxford University Press.

Tammariello SP and DL Denlinger. 1998. Go/G1 cell cycle arrest in the brain of *Sarcophaga crassipalpis* during pupal diapause and expression pattern of the cell cycle regulator, proliferating cell nuclear antigen. Insect Biochemistry and Molecular Biology 28:83-89.

Yocum GD, KH Joplin and DL Denlinger. 1991. Expression of heat shock proteins in response to both heat and cold shock in diapausing pharate larvae of the gypsy moth, Lymantria dispar. Arch. of Insect Biochem. and Physiol. 18:239-249.

Yocum GD, KH Joplin and DL Denlinger. 1998. Upregulation of a 23 kDa small heat shock protein transcript during pupal diapause in the flesh fly, *Sarcophaga crassipalpis*. Insect Biochem Mol Biol. 28:677-82.

Yoder JA and DL Denlinger, MW Dennis and PE Kolattukudy. 1992. Enhancement of diapausing flesh fly puparia with additional hydrocarbons and evidence for alkane biosynthesis by a decarbonylation mechanism. Insect Biochem Mol Biol 22:237-43.

APPENDIX

Abbreviations

amp - ampicillin

bp - base pairs

BLAST - Basic Local Alignment Tool

$CaCl_2 \bullet 2H_2O$ - calcium chloride

CDMC - The Canadian Drosophila Microarray Centre

cDNA - complementary deoxyribonucleotide acid

cGMP - cyclic guanosine mono-phosphate

DP - diapause-programmed

Desat - Stearoyl CoA Desaturase

DEPC - diethyl pyrocarbonate

dHSP - diapause expressed Heat Shock Protein

dNTP - deoxyribonucleoside triphosphate

dsRNA - double-stranded ribonucleic acid

DTT – dithiothreitol

DRL – diapause upregulated library

NDRL – non diapause upregulated library

Eb1 - EB1 microtubule binding

EcR - ecdysone receptor

EDTA - ethylenediaminetetraacetic acid

EtBR - ethidium bromide

EtOH - ethanol

GSP - gene specific primer

HSC 70 - heat shock 70 cognate protein

HSP - heat shock proteins

JH - juvenile hormone

KCl - potassium chloride

LB - luria-bertani

L1, L2, L3 - larval developmental stages

LD - light dark

MESK2 - Misexpression suppressor of KSR 2

M - molar

mg - milligram

mL - milliliter

mM - millimolar

mRNA - messenger ribonucleic acid

NaCl - sodium chloride

$NaHCO_3$ - sodium bicarbonate

NDP - non diapause-programmed

NCBI - National Center for Biotechnology Information

EMBL-EBI - European Bioinformatics Institute

nm - nanometer

O_2 - oxygen

PCR - polymerase chain reaction

PTTH - prothoracicotropic hormone

PCNA - proliferating cell nuclear antigen

RNA - ribonucleic acid

RT-PCR - reverse transcription polymerase chain reaction

Scys-A - sarcocystatin A

Scys-B - sarcocystatin B

SSH - Suppression Subtractive Hybridization

TR - Transcriptional Repressor

TAE - Tris-acetate EDTA buffer

TE - Tris-EDTA buffer

µg - microgram

µL- microliter

µM - micromolar

X-GAL - 5-Bromo-Chloro-3-Indolyl-β-D-Galactopyranoside

VITA

PUJA KARKI

Place of Birth:	Biratnagar, Nepal
Education:	M. S. Biology, East Tennessee State University, Johnson City, TN 2009 M. Sc. Botany, Degree Campus, Tribhuvan University, Nepal 2002 B. Sc. Biology, Mahendra Morang Adarsh Multiple Campus, Nepal 2000
Professional experience:	Graduate Teaching Assistant**,** East Tennessee State University, TN, USA 2007- 2009 Laboratory Technician, East Tennessee State University, TN, USA 05/2008-08/2008
Publications:	Karki, Puja. 1998. General survey of fodder plants of Janpath tole of Biratnagar, Nepal. Karki, Puja. 1999. General survey and study of non-timber forest plant products of Charkoshe, Nepal.
Scholarships and Grants:	Graduate Teaching Assistantship, East Tennessee State University, TN 01/2007- 05/ 2009. East Tennessee State University Graduate Council Research Grant 2008