ABSTRACT

Title of Document: EVOLUTION OF PAIR-RULE GENES

Yong Lu, Doctor of Philosophy, 2015

Directed By: Dr. Leslie Pick, Professor
Department of Entomology
University of Maryland, College Park

All insects have a segmented body. The genes controlling segment development have been well characterized in the fruit fly, *Drosophila melanogaster*. These genes were divided into three categories: gap genes specify several continuous segments over a broad region of the embryo; Pair-Rule Genes (PRG) are responsible for segment formation and are the first set of genes to be expressed in repetitive patterns in the embryo; Segment polarity genes define anterior and posterior polarities within each segment.To understand how PRGs evolve, I took a comparative approach in this thesis. First, I compared the function of the *Drosophila* PRG *ftz-f1* to that of its mammalian orthologs by expressing them all in *Drosophila* embryos. I found that the molecular function of this family of nuclear receptors has been highly conserved during evolution.

Next, I set out to establish new insect model systems to study PRG function. While, some PRGs have been studied in other insects, most of these studies focused on holometabolous insects. My work focused on the sister group to the holometabolous

insects, the Hemipteroid Assemblage. I participated in the genome annotation of a hemiptera insect, *Oncopeltus fasciatus*. I annotated nuclear receptor super family, Hox and PRGs in *Oncopeltus*. I further studied the expression and function of four PRGs in *Oncopeltus*. Using *in situ* hybridization and RNAi, I found that, *Of-ftz* and *Of-hairy* do not have segmentation function, while *Of-ftz-f1* has function in oogenesis and segmentation. *Of-runt* was found to induce cell death in oocytes, but its function in segmentation needs further analysis. Using the knowledge and expertise I gained from *Oncopeltus,* I successfully set up *in situ* hybridization, antibody staining and parental RNAi in an invasive hemipteran insect pest, the Brown Marmorated Stink Bug (BMSB) *Halyomorpha halys.* These studies show that the expression and function of PRGs varies extensively in diverse insects, despite the overall conservation of a segmented body plan.

EVOLUTION OF PAIR-RULE GENES

By

Yong Lu

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Dr. Leslie Pick, Chair
Dr. Alexandra Bely
Dr. Cerruti Hooks
Dr. Zhongchi Liu
Dr. Steve Mount

Dedication

To Qian, Daniel, Ian

Acknowledgements

First, I would like to thank my advisor, Leslie Pick, for her great support and outstanding mentorship. Her great passion and enthusiasm for science inspires me. It is my great pleasure and fortune to have an opportunity to work with her.

Second, I would like to thanks other members of the Pick lab, who provided great helps to me and make our lab an enjoyable place to stay: Amanda Field, Alison Heffer, Ray Anderson, Jie Xiang, Georgeta Crivat, Zhaoqing Jin, Caroline Li, Hua Zhang.

Lastly, I would like to thank my family, my parents and in-laws for their great support.

Table of Contents

Chapter 1: Introduction

Section 1.1: A cascade of genes controls **Drosophila** *development*

The genes controlling embryonic development have been well characterized in the fruit fly, *Drosophila melanogaster*. Decades of study culminating in a few massive genetic screens led to the identification of a cascade of genes that sequentially specify the basic body plan of the fly (Jürgens et al., 1984; Lewis, 1978; Nüsslein-Volhard et al., 1985; Nüsslein-Volhard and Wieschaus, 1980; Nüsslein-Volhard et al., 1984). In those screens, genes were divided into three categories based on mutant phenotypes: gap genes specify several continuous segments over a broad region of the embryo; mutations in pair-rule genes resulted in loss of every other segment; segment polarity genes affect each segment (Nüsslein-Volhard and Wieschaus, 1980). The Pair-Rule Genes (PRGs) are responsible for segment formation and are the first set of genes to be expressed in repetitive patterns in the embryo. Finally, the homeotic, *Hox* genes, confer unique identities to specific segments (Figure 1.1).

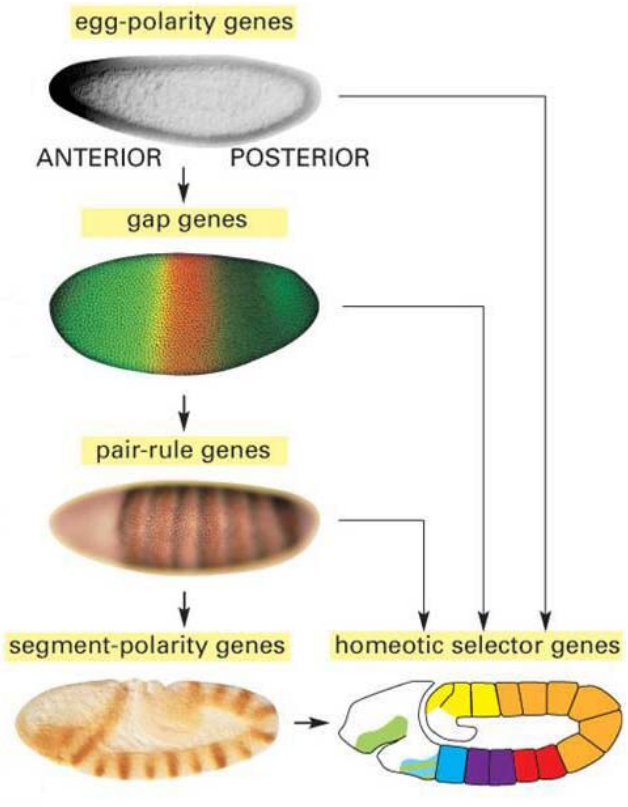

Figure 1.1 The regulatory hierarchy of egg-polarity, gap, segmentation, and homeotic selector genes. A cascade of regulatory genes divides the embryo into segments which have unique identifies. The genes were classified into categories on the basis of mutant phenotype. Genes are expressed in pre-patterns in the regions of the embryo they specify. Figure from Molecular Biology of the Cell. 4th edition, Alberts B, Johnson A, Lewis J, et al. 2002.

Among these genes, it is the PRGs that generate periodic, striped expression patterns from non-periodic signals. Their functions are thus critical to the establishment of the basic segmented patter of *Drosophila*. The fact that all insects share a segmented body plan suggests the function of PRGs is highly conserved throughout insects. However, data from our lab and others have suggested unexpected variation in the PRG network. In this chapter, I summarize a body of literature on PRGs in *Drosophila* and other insects. The expression patterns and functions of these genes from these studies are summarized in Figure 1.2.

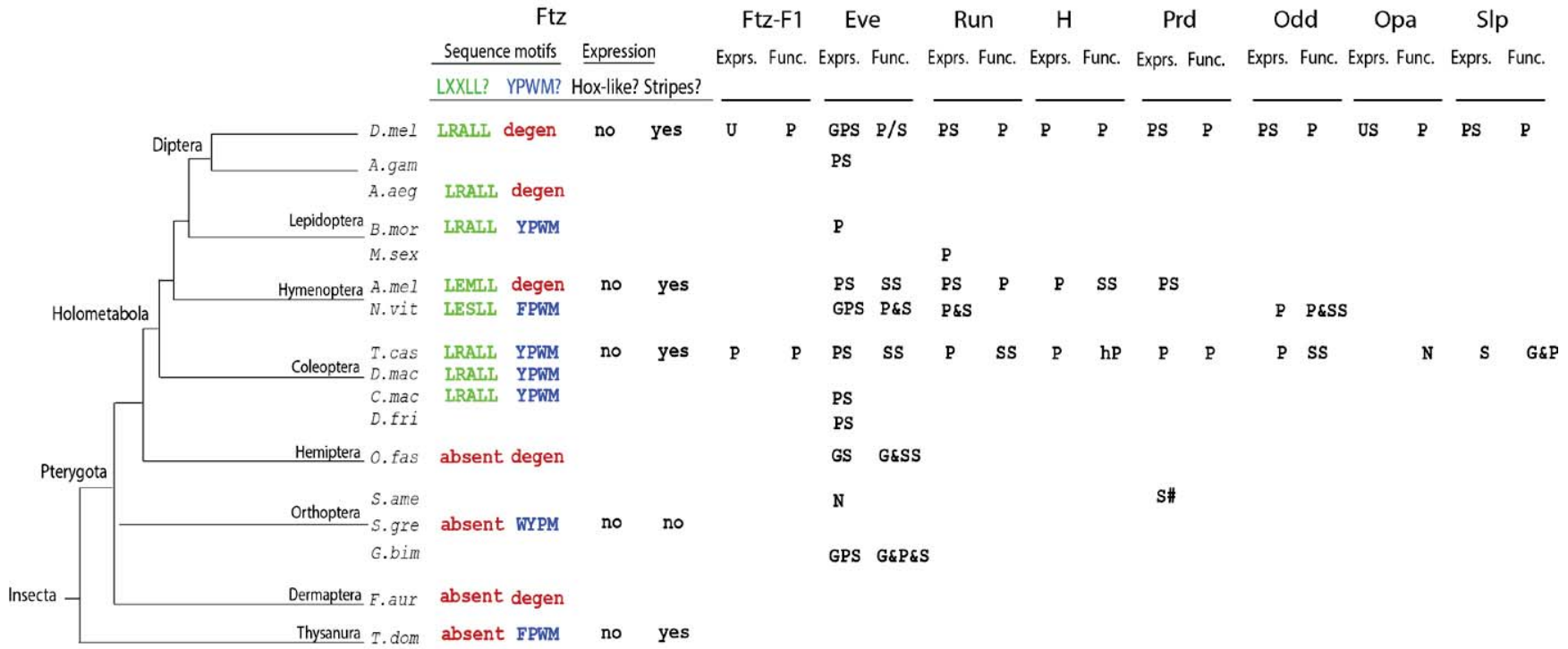

Figure 1.1 Summary of Pair-rule gene expression and function in insects. Information on Ftz is modified from Heffer et al. (Heffer et al., 2010). Left, Cladogram of major arthropod taxa is shown with divergence timeline below. The presence of cofactor interaction motifs (LXXLL motif, green; YPWM motif, blue; absent, red) and observed expression patterns (stripes; Hox-like) are indicated. Information for orthologs of the eight Drosophila PRGs, as indicated. Abbreviations: Exprs, expression, Func, function. G, gap-like; P, pair-rule; S; segmental, U, ubiquitous; N, Non-pair-rule, hP , head pair-rule; SS, severe segmental phenotype, in which most of segments are missing. If there is more than one letter, the first letter indicates an early expression pattern, the second, a later expression pattern. For an example, PS means the expression pattern is pair-rule like in early development stages and switches to segmental late. Under function, letters separated by "&" means the phenotypes of these letters represent are seen in the same embryo in different segments. Letters separated by / indicates that phenotypes are observed with different alleles. # refers to pby1 and pby2. Most of the expression patterns are mRNA expression patterns. References are listed in the text in order to save space.

Section 1.2: Pair-rule genes in Drosophila melanogaster *are required for formation of body segments*

As mentioned above, the PRGs are a class of genes that was first identified in *Drosophila* on the basis of mutant phenotype. Mutations in PRGs result in lethality accompanied by loss of alternate body segments (Jürgens et al., 1984; Nüsslein-Volhard and Wieschaus, 1980; Sander et al., 1980). In the original screens, a total of eight PRGs were identified. These genes are: *even-skipped (eve), hairy (h), runt (run), fushi-tarazu (ftz), odd-skipped (odd), paired (prd), odd-paired (opa),* and *sloppy-paired* (*slp*) (Nüsslein-Volhard and Wieschaus, 1980; Nüsslein-Volhard et al., 1984; Sander et al., 1980; Wakimoto and Kaufman, 1981). Subsequent screens for maternal effects of zygotic lethal alleles revealed one additional pair-rule gene, *ftz-f1* (Chou and Perrimon, 1996; Yu et al., 1997). All PRGs encode DNA-binding transcription factors. Below I describe each pair-rule gene, how it was discovered, its mutant phenotype, expression pattern, transcription factor activity and evolutionary conservation.

fushi tarazu

Drosophila fushi tarazu (ftz) is located in the *Drosophila Hox* complex (HOM-C) between the homeotic genes *Antennapedia (Antp)* and *Sex combs reduced (Scr)* (Lewis et al., 1980; Lindsley and Zimm, 1992). Although *Drosophila ftz (Dm-ftz)* is a homeobox-containing gene, it has lost its homeotic function in *Drosophila* where it acts as a pair-rule segmentation gene to direct the development of even-numbered parasegments (Hafen et al., 1984; Kuroiwa et al., 1984; Wakimoto et al., 1984). *Dm-ftz* is expressed in seven stripes in the primordia of these parasegmental regions

(Carroll et al., 1988a; Hafen et al., 1984). *Dm*-Ftz is required for expression of alternate *engrailed (en)* stripes, directly regulating *en* by binding to an intronic enhancer along with its partner Ftz-F1 (see below). *Dm*-Ftz binds to similar DNA sequences as other Antp-class homeodomain proteins and functions as a transcription activator (Pick, 1990; Florence 1997). *Dm*-Ftz has been shown to directly regulate the expression of *ftz, en*, *drm* and likely regulates the expression of 50-100 genes in total in early embryos, all in conjunction with Ftz-F1 (Bowler and Field – manuscript in preparation; Florence et al., 1997; Han et al., 1998; Hou et al., 2009; Pick et al., 1990; Yu et al., 1997).

ftz-f1

Ftz-F1 (Ftz-Factor 1) was first isolated as a transcription factor that binds to the zebra element, a cis-regulatory element of *ftz,* potentially activating *ftz* gene expression (Ueda et al., 1990). However, the pair-rule function of *ftz-f1* was revealed in a genetic screen for maternal effects of zygotic lethal alleles (Yu et al., 1997) and simultaneously in a screen for maternal-effect genes (Guichet et al., 1997). Ftz-F1 encodes an orphan nuclear receptor and is the founding member of the NR5A nuclear receptor proteins that includes mammalian SF-1 and LRH-1 (Pick et al., 2006). Embryos derived from *ftz-f1* germline clones (Yu et al., 1997) or from mothers expressing a maternal allele of *ftz-f1* (Guichet et al., 1997) display pair-rule phenotypes indistinguishable from *ftz*, deletion of the even-numbered parasegments. Similar to *opa*, *ftz-f1* is expressed ubiquitously in embryos, despite the fact that loss of function mutations result in pair-rule defects. As discussed in more detail below, this phenotype is explained by the obligate interaction of Ftz-F1 with Ftz. These

proteins function as partners to coordinately bind DNA and regulate transcription (Guichet et al., 1997; Schwartz et al., 2001; Yu et al., 1997; Yu et al., 1999; Yussa et al., 2001a).

even-skipped

even-skipped (*eve*) was first isolated by Nüsslein-Volhard and co-workers in the genetic screen described above (Nüsslein-Volhard and Wieschaus, 1980). As its name indicates, *eve* mutations resulted in a typical pair-rule phenotype with embryos missing portions of the even-numbered segments, corresponding to loss of odd-numbered parasegments. Note that parasegments correspond to segmental-width regions of the embryo that are offset by a half-segment unit from the body segments that form later during development. The parasegments appear to be directly specified by the PRGs and are likely the first step in the establishment of metameric units in the embryo (Lawrence, 1981). It was later found that the mutations associated with the *eve* pair-rule phenotype are hypomorphic, while null mutations result in embryos with a "lawn of denticles" phenotype, similar to *en* mutants, lacking overt signs of segmentation (Macdonald et al., 1986; Nüsslein-Volhard et al., 1985; Nüsslein-Volhard and Wieschaus, 1980; Nüsslein-Volhard et al., 1984). At the early syncitial blastoderm stage, *eve* mRNA and protein were found to be expressed in a broad band spanning the central region of the embryo (Frasch et al., 1987; Macdonald et al., 1986). In keeping with the pair-rule phenotype, *eve* is then expressed in seven stripes at the blastoderm stage, in the primordia of regions missing in *eve* pair-rule mutants. At slightly later stages, an additional seven *eve* stripes arise de novo between the seven primary stripes, resulting in a total of 14 stripes in the presumptive posterior

compartments of each segment (Macdonald et al., 1986). *eve* encodes a transcription factor containing a divergent homeodomain that binds to DNA and represses the expression of *ftz, Ultrabithroax (Ubx) and wingless (wg)* (Carroll and Scott, 1986; Ingham et al., 1988; Martinez-Arias and White, 1988)

runt

runt was first identified in a screen for X-linked lethals (Lifschytz and Falk, 1968) but its pair-rule function was recognized by Nüsslein-Volhard and Wieschaus (Nüsslein-Volhard and Wieschaus, 1980). At the syncitial blastoderm stage, *runt* is expressed in a broad central region and during cellularization *runt* is expressed in seven pair-rule stripes, which later split into 14 stripes (Klingler and Gergen, 1993). *runt* encodes a unique transcription factor with a DNA binding domain named the Runt domain (Gergen and Butler, 1988), different from the homeodomain, zinc finger or other previously characterized DNA binding domains (Kagoshima et al., 1993; Pepling and Gergen, 1995). *Drosophila* Runt is the founding member of the Runx (Runt-related transcription factor) transcription factor family. In *Drosophila,* there are three homologues of *runt,* named l*ozenge (lz), runxA* and *runxB (Bao and Friedrich, 2008). Dm-runt* functions in segmentation (Nüsslein-Volhard and Wieschaus, 1980), neurogenesis (Dormand and Brand, 1998; Duffy et al., 1991) and sex determination (Cline, 1986; Duffy and Gergen, 1991). In *Drosophila, lz* plays roles in eye and antenna development, hematopoiesis, and fertility (Batterham et al., 1996; Gupta and Rodrigues, 1995). The function of RunxA and RunxB have not been studied (St Pierre et al., 2014). In mamals, there are three Runt homologues *RUNX1, RUNX2* and *RUNX3.* RUNX1 plays a similar role in hematopoiesis as its *Drosophila*

homolog Lz; RUNX2 is involved in skeletal development and RUNX3 is considered to be a major tumor suppressor in many tumor types, while both RUNX1 and RUNX2 were found to have oncogenic potenial (Ito, 2008; Levanon and Groner, 2008). All Runt domain family members have a VWRPY motif at the C terminus of the protein. The VWRPY motif interacts with Groucho and plays a very important role in repression of some target genes, although Runt appears to repress *en* expression independently of VWRPY (Aronson et al., 1997).

hairy

The *hairy (h)* mutation was first found by Dr. O.L. Mohr in the early 1900's (Lindsley and Zimm, 1992). The name *hairy* was given to these mutants because of their phenotype: *hairy* flies possess hair on the scutellum in a region lacking hair in wild type flies. The role of *hairy* in *Drosophila* development was first discovered by Nüsslein-Volhard and co-workers who initially named this mutation *barrel*. Later *barrel* and *hairy* were found to be the same gene (Nüsslein-Volhard et al., 1984). *hairy* was cloned by Holmgren (Holmgren, 1984). *hairy* encodes a basic helix-loop-helix (bHLH) transcription factor and binds to DNA (Holmgren, 1984). Hairy has a WRPW motif, which interacts with the co-repressor Groucho, functioning as a repressor (Jimenez et al., 1997), While most of Hairy's repressive effects depend on WRPW, its repression of Scute was found to be independent of the WRPW motif and Groucho (Dawson et al., 1995). *hairy* is expressed in stripes in *Drosophila* embryos (Carroll et al., 1988b; Ingham et al., 1985) and mutants are missing the posterior part of odd-numbered segments (Jürgens et al., 1984). In addition to playing an important role in *Drosophila* segmentation, *hairy* is also required for bristle patterning

(Rushlow et al., 1989) and has been found to be a quantitative trait locus for *Drosophila* sternopleural bristle number (Robin et al., 2002). In *Drosophila,* a few *hairy* related genes have been found including *Hey (Hairy/E(spl)-related with YRPW motif)* and *E(spl) (Enhancer of Split)*, both of which are transcription repressors that play important roles in neurogenesis (Fisher and Caudy, 1998).

paired

paired (prd) was discovered by Nüsslein-Volhard et al. and Sander et al. around the same time (Nüsslein-Volhard and Wieschaus, 1980; Sander et al., 1980). *prd* mutants show deletions of the posterior part of the odd-numbered segments and the anterior part of even-numbered segments. *prd* is expressed in seven primary stripes at the syncitial blastoderm stage; each primary stripe later splits into two stripes, giving rise to 14 stripes at the cellular blastoderm stage (Kilchherr et al., 1986). *prd* encodes a transcription factor with two DNA binding domains: a Paired-Domain (PD) and Prd-type Homeodomain (HD). Both the PD and HD are required to activate expression of target genes such as *en, wg* and *eve* (Gutjahr et al., 1993b; Ingham and Hidalgo, 1993; Lan et al., 1998).

In *Drosophila*, there are two homolgs of *paired*, *gooseberry (gsb)* and *gooseberry-neuro (gsbn)* (Baumgartner et al., 1987; Gutjahr et al., 1993). All of them belong to the PAX group III family (Noll, 1993). *gsb* was first discovered as a segment polarity gene (Nüsslein-Volhard and Wieschaus, 1980). Both *gsb* and *gsbn* are expressed in the central nervous system and *gsb* regulates the expression of *gsbn,* which in turn is involved in neural specification (Gutjahr et al., 1993; Gutjahr et al., 1994; He and Noll, 2013).

odd-skipped

odd-skipped (odd) was first identified in the genetic screen described above based on its mutant phenotype - posterior denticle rows of the odd-numbered segments are deleted (Nüsslein-Volhard and Wieschaus, 1980). *odd* is expressed in seven primary stripes at the cellular blastoderm stage; at gastrulation, those seven primary stripes narrow and at the same time eight new stripes arise (Coulter et al., 1990; Coulter and Wieschaus, 1988). *Odd* encodes a transcription factor with a zinc finger DNA-binding domain (Coulter et al., 1990; Coulter and Wieschaus, 1988). It negatively regulates *en,* limiting the expression of *en* within even-numbered parasegments. In *Drosophila*, there are two genes related to *odd*, sister of *odd (sob*) and *brother of odd with entrails limited (bowl)*, which play roles in embryonic hindgut development (Iwaki et al., 2001).

odd-paired

odd-paired was first identified by Jurgens et al. on the basis of phenotype (Jürgens et al., 1984). Although *opa* mutants display a typical pair-rule phenotype with deletion of odd-numbered parasegments, *opa* is not expressed in a pair-rule pattern (Benedyk et al., 1994). Rather, from cellularization to gastrulation, *opa* is ubiquitouslly expressed in a broad region of the central of the embryos; as germ band extension begins, its expression changes to 14 weak stripes with low backgroud level of expression throughout the embryo (Benedyk et al., 1994). *opa* encodes a zinc finger transcription factor that is thought to regulate the expression level of *wg* (Benedyk et al., 1994).

sloppy paired

sloppy paired (slp) was first identified by Nüsslein-Volhard et al. on the basis of its pair-rule mutant phenotype – deletion of odd-numbered abdominal segments and the mesothorax segment (Nüsslein-Volhard et al., 1984). When Grossnikilaus et al. cloned the *slp* locus, they found that *slp* is composed of two genes, *slpl* and *slp2* (Grossniklaus et al., 1994). *slp1* is first expressed in a broad region (gap-like) in the anterior of the embryo at the syncitial blastoderm stage; towards the end of the syncitial blastoderm stage the broad expression pattern narrows into a circumferential ring; during cellular blastoderm, seven primary stripes appear and seven more secondary stripes are added between the primary stripes. Expression of *slp2* starts later than *slp1,* and follows a similar pattern to *slp1* except that it is not expressed in a gap-like pattern, and the circumferential ring is one to two cells narrower than that of *slp1* (Grossniklaus et al., 1992). s*lp1* is required for head formation, while *slp2* is redundant to *slp1* for segmentation but plays no role in head formation (Cadigan et al., 1994a). *slpl* and *slp2* encode related transcription factors with forkhead DNA binding domains. Both Slp1 and Slp2 regulate expression of the segmentation polarity genes *en and wg* (Cadigan et al., 1994b).

Interactions between pair-rule genes

A large body of work has documented regulatory interactions between the *Drosophila* PRGs. These studies suggest that in *Drosophila*, the PRGs are not created equal. Some, known as primariy PRGs, are activated by the maternal and gap genes and regulate other PRGs. Others are regulated by primary PRGs and are called secondary or tertiary PRGs (Ingham, 1988; Noll, 1993). Traditionally, *hairy, runt and*

eve were thought to be the primary PRGs; other PRGs were thought to be secondary or tertiary PRGs (Akam, 1989). This classification is likely oversimplified as, for example, the *ftz* pair-rule stripes were found to be activated by non-periodic cues (Yu and Pick, 1995); *prd* stripes are established by gap genes (Gutjahr et al., 1993a); and *runt,* classified as a primary PRG, is regulated by the so call secondary PRGs such as *ftz* and *prd (Klingler and Gergen, 1993)*.

In *Drosophila*, Hairy was found to negatively regulate *runt* and *ftz* (Ingham and Gergen, 1988). Runt was found to negatively regulate *h, eve and odd* (Ingham and Gergen, 1988; Jaynes and Fujioka, 2004). Eve negatively regulates *slp, prd, odd,*and *run* (Baumgartner and Noll, 1990; Coulter and Wieschaus, 1988; DiNardo and O'Farrell, 1987; Fujioka et al., 1995; Jaynes and Fujioka, 2004). Ftz activates *odd* (Nasiadka and Krause, 1999) but Odd negatively regulates *ftz, slp, prd* (Baumgartner and Noll, 1990; DiNardo and O'Farrell, 1987; Mullen and DiNardo, 1995). Slp negatively regulates *ftz, odd, eve* (Cadigan et al., 1994b; Jaynes and Fujioka, 2004). Prd positively regulates the expression of *eve*, but does not regulate other PRGs as the expression of *runt and eve* are unchanged in *prd* null mutants (Frasch and Levine, 1987; Hooper et al., 1989; Klingler and Gergen, 1993).

In addition to this*, runt* and *slp1* have been found to have gap-like functions (Tsai and Gergen, 1994). *runt* is expressed in a broad domain in the central region of the embryo (Klingler and Gergen, 1993) and regulates the express of two gap genes, *Kruppel* and *hunchback (Tsai and Gergen, 1994). Slp1* has gap-like functiona in the head where it is expressed in a gap-like pattern. This early expression of *slp1* in the head is not regulated by any other PRG (Grossniklaus et al., 1994). Overall, the

interactions among PRGs and interactions between PRG and gap genes are complicated. Some interactions are partially redundant, for an example, both Runt and Slp repress *eve,* setting up the anterior border of the odd-numbered *eve*-stripes. This complexity and redundancy indicate the importance of correct expression patterns of PRGs in *Drosophila* development.

Section 1.3*: *Evolution of PRGs in insects

In order to understand how the PRG network changed during evolution, these genes have been examined in other insects. Based on how their germ band deveolps, insects have been classified into two categories, sequentially segmenting and long germ band (Peel et al., 2005). In sequentially segmenting insects, the anterior segments are specified in blastoderm and posterior segments are formed by sequential addition from a cellularized growth zone after gastrulation. In long germ insects, all their segments are patterned in the blastoderm. Sequential segmentation is believed to be more basal (Anderson, 1973). Below, I summarize what is known about the sequence, expression and function of orthologs of the *Drosophila* PRGs in other insect species, including sequentially segmenting species and others with a long germ mode of development similar to *Drosophila*.

fushi tarazu (ftz)* and *fushi tarazu- factor 1 (ftz-f1)

As mentioned above, Ftz and Ftz-F1 function as obligate partners in *Drosophila*. Accordingly, the functional evolution of these two genes is presented in one section here. Years of works from many labs, especially the Pick lab demonstrated extensive variation in the expression and function of these genes, providing a broad view about the evolution of this partner pair.

ftz is thought to have arisen from a duplication of an *Antp*-like ancestral homeotic gene and it is expressed in a *Hox*-like pattern in distant arthropods (Figure 1.3 C), such as the mite *Archegozetes longisetosu*s (Telford, 2000), the water flea *Daphnia pulex* (Papillon and Telford, 2007), the centipede*, Lithobius atkinsoni*. (Hughes and Kaufman, 2002) and a spider *Cupiennius* (Damen et al., 2005). *ftz* is expressed in pair-rule stripes in *Drosophila melanogaster* (Figure 1.3 B), *D. hydei* (Jost et al., 1995), *Tribolium* (Brown et al., 1994a), *Apis mellifer*a (Dearden et al., 2006) and *Thermobia domestica* (Hughes et al., 2004). For *Drosophila*, these stripes are localized to the primoridia of regions missing in *ftz* mutants and striped expression is necessary for proper segmentation, expression of alternate *en* stripes, and viability.

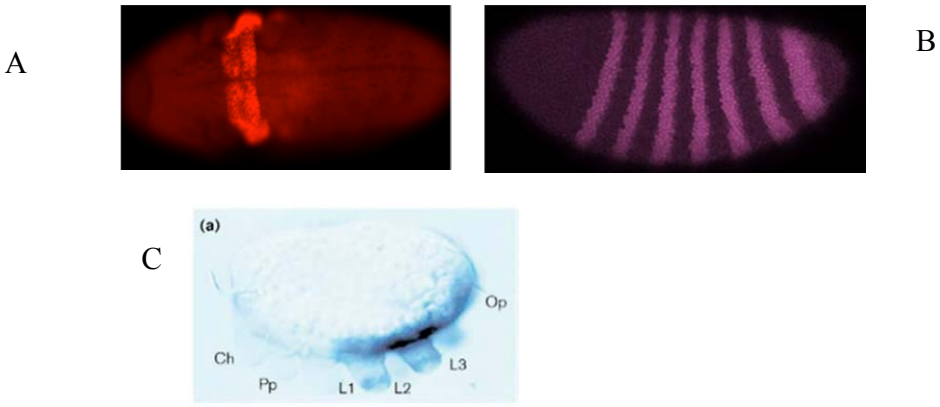

Figure 1.3 The expression of *ftz* switched from *Hox*-like to stripes. (A) *Drosophila Scr* expressed in a Hox-like pattern, showing how a typical *Hox* gene expression pattern in *Drosophila*. (B) *Drosophila ftz* expressed in seven stripes in the primordia of the parasegments it promotes. (C) *Archegozetes longisetosus ftz* is expressed in a *Hox*-like pattern, thought to reflect the ancestral state. Panel C from (Telford, 2000).

Studies in our lab showed that *Dm*-Ftz changed its function during evolution due to changes in its protein sequence and its expression pattern (Lohr and Pick, 2005; Lohr et al., 2001). *Dm*-Ftz acquired an LXXLL motif which mediates interaction with Ftz-

F1 and is required for its segmentation function; it lost its ancestral YPWM motif which mediates interaction with the Hox co-factor Extradenticle (Exd) (Lohr and Pick, 2005; Lohr et al., 2001).

As discussed above, *ftz* likely arose from an *Antp*-like ancestor, and it changed its expression pattern from *Hox*-like to a seven stripe pattern, seen in *Drosophila*. Therefore, its function in segmentation may be explained solely by its changed expression pattern. If this were the case, *Dm*-Ftz would retain the ability to function as a homeotic/Hox protein. To test this, Lohr et al. ectopically expressed *Dm-ftz* in imaginal discs to see whether it would mimic the homeotic phenotype seen when classic *Hox* genes are expressed in this way. To do this, they generated *UAS-Dm-ftz* transgenic flies and crossed them with *dll-GAL4*, which drives expression of *UAS-Dm-ftz* in the imaginal discs (Lohr and Pick, 2005). When Antp was expressed in the same way, it caused antenna-to-leg transformation, which serves as a indicator of its homeotic function. Using this approach, Lohr et al. found that *Dm-ftz* lost its homeotic potential (Figure 1.4 B), while distant *ftz* genes such as *Tribolium castaneum ftz (Tc-ftz)* and *Schistocerca gregaria ftz (Sg-ftz*) caused antenna-to-leg transformations in *Drosophila* (Figure1.4 D and E) (Lohr and Pick, 2005). Since the YPWM motif had been found to play an important role in the interaction between Hox proteins and Exd (Johnson et al., 1995), as expected, when a YPWM motif was added to *Dm*-Ftz, it increased the homeotic potential of *Dm*-Ftz (Lohr and Pick, 2005). Surprisingly, when the YPWM motif was changed to AAAA in *Tc*-Ftz, the mutated *Tc*-Ftz still had homeotic potential (Figure 1.4 H) (Lohr and Pick, 2005). This suggests that *Tc*-Ftz has another domain necessary for homeotic function. Since

without the YPWM motif, *Dm*-Ftz does not have homeotic potential, it is possible that *Dm*-Ftz lost the second motif which confers *Tc*-Ftz homeotic potential. Alternatively, *Tc*-Ftz independently acquired this motif or *Dm*-Ftz acquired another motif which inhibits homeotic function.

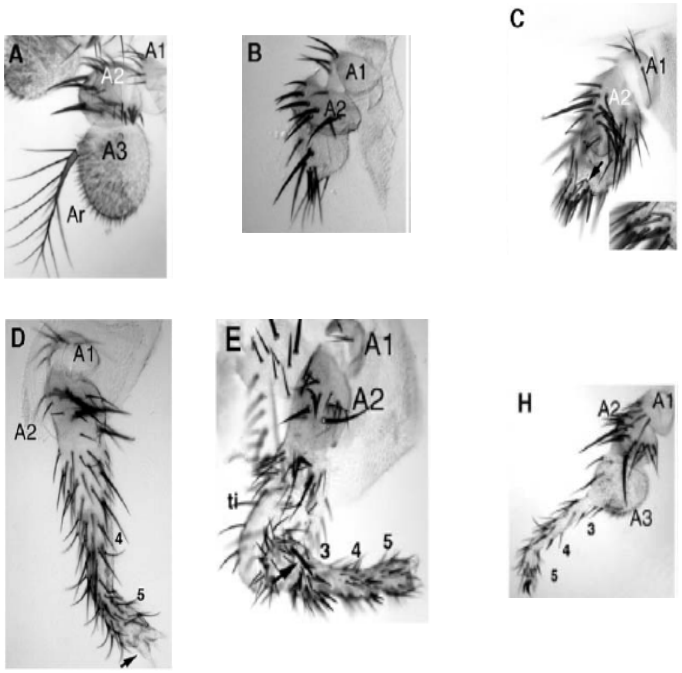

Figure.1.4 Addition of YPWM is enough to confer homeotic function to *Dm*-FTZ but is not necessary for *Tc*-Ftz homeotic activity. (A)Control. Arista (Ar) expressing *lacZ*. All antennal segments (A1–A3) developed normally. (B) Expression of *Dm*-Ftz caused the deletion of the arista and the truncation of the A3 segment.(C) *Dm*-FtzLRAAA caused a similar phenotype. (D) *Dm*-FtzYPWM caused weak antenna-to-leg transformations. (E) Expression of Dm-FtzLRAAA/YPWM caused a strong antenna-to-leg transformation. (H) Expression of Tc-FtzAAAA, in which the YPWM was changed to AAAA.

As mentioned above, Ftz depends on the interaction with Ftz-F1 for its function in segmentation in *Drosophila*. Ftz-F1 belongs to the nuclear receptor (NR) superfamily, and has been classified as an orphan receptor (reviewed in Pick, 2005). Like all NRs, Ftz-F1 has a conserved DNA-binding domain (DBD) and a ligand-binding domain (LBD). At the C-terminal end of the LBD, there is highly conserved activating factor (AF-2) motif, which had been shown to interact with LXXLL motifs in other nuclear

receptor co-factors. Dm-Ftz-F1 is ubiquitously expressed in the *Drosophila* embryo during development (Figure 1.5 d). *Dm-ftz-f1* mutants lack all *ftz*-dependent segments and are missing the *Dm*-Ftz-dependent *en* stripes, resulting in an embryonic phenotype identical to that of *ftz* mutants (Figure. 1.5 b and 1.5 c). *Dm*-Ftz and *Dm*-Ftz-F1 were found to interact with each other both *in vitro* and *in vivo* (Florence et al., 1997; Guichet et al., 1997; Yu et al., 1997; Yu et al., 1999; Yussa et al., 2001b).

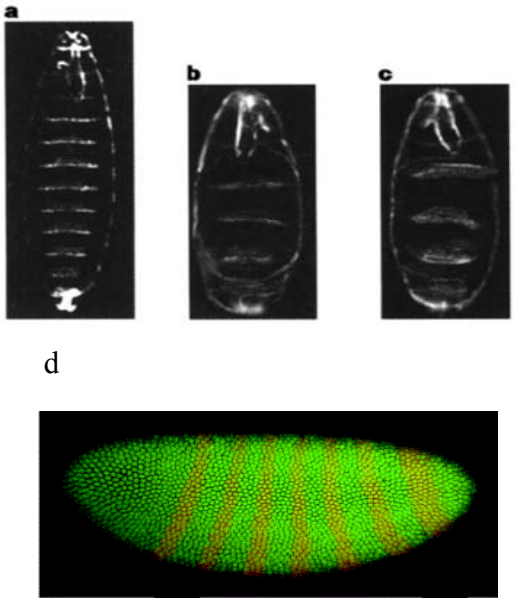

Figure 1.5. Mutations in *ftz* and *ftz-f1* cause identical pair-rule mutant phenotypes (a) Wild-type cuticle, (b) *ftz* mutants are missing every other parasegment, (c) *ftz-f1* mutants are identical to *ftz* mutants, (d) Expression of *ftz-f*1 (green) overlapping *ftz* (red; overlap yellow). From Yu et. al, 1997; Yussa et. al, 2001.

The function of Ftz has been studied in two holometabolous insects in addition to *Drosophila*. In the honeybee *Apis mellifera, Am-ftz* was expressed in a pair-rule pattern. Knockdown of *Am-ftz* with embryonic RNAi (eRNAi) affected the formation of anterior segments but the thoracic and abdominal segments were unaffected (Wilson and Dearden, 2012). In the sequentially segmenting beetle *Triboliu castanem*, *ftz* is expressed in pair-rule stripes, and its stripes partially overlap with the primary stripes of *Tc-eve* (Brown et al., 1994a). Surprisingly, Ftz does not have any

segmentation function, as embryos with a deletion of part of the homeotic complex that includes *Tc-ftz* did not display segmentation defects (Brown et al., 1994b; Stuart et al., 1991). Similarly, parental RNAi (pRNAi) of *Tc-ftz* did not cause defects in segmentation (Choe et al., 2006). Note that when *Tc-ftz* was expressed in *Drosophila* embryos, it showed both segmentation and homeotic activity. Further, *Tc*-Ftz has an LXXLL motif and interacted in vitro with Ftz-F1 (Lohr and Pick, 2005). Therefore, its apparent lack of function in *Tribolium* segmentation is unexpected.

In *Drosophila*, in addition to its role in segmentation, Ftz is also expressed and functions in the central nervous system (CNS) (Carroll and Scott, 1985; Doe et al., 1988). Its CNS expression has been found in all the arthropods examined, which includes myriapods (Damen, 2002; Hughes and Kaufman, 2002; Janssen and Damen, 2006), crustaceans (Heffer et al., 2010; Mouchel-Vielh et al., 2002), and insects (Brown et al., 1994a; Dawes et al., 1994; Hughes et al., 2004). Thus, despite change in protein sequence and expression pattern during embryonic development, Ftz expression in the CNS seems to be maintained in all arthropods. It was proposed that *ftz* is maintained in the arthropod genome because of an indispensable and conserved function in the CNS (Heffer and Pick, 2013).

Ftz-F1 is required for segmentation of *Drosophila* (Guichet et al., 1997; Yu et al., 1997). Its role is as important as Ftz if not more, especially if we consider the observation that the segmentation defects in *ftz* mutants can be rescued by a partial Ftz protein without the DNA binding domain (Copeland et al., 1996; Fitzpatrick et al., 1992). The function of Ftz-F1 orthologs in other insects has only been studied in *Tribolium* (Heffer et al., 2013). Heffer et al. found that *Tc-ftz-f1* is expressed in one

single stripe at blastoderm stage, and in a pair-rule pattern at the germband extension stage (Heffer et al., 2013). Using eRNAi, Heffer et al. found that *Tc-ftz-f1* dsRNA injected embryos lost even-numbered abdominal segments and when the concentration of dsRNA was increased, embryos failed to hatch with no cuticle formed. These findings indicate that in *Tribolium, Tc-ftz-f1* functions as a pair-rule gene and is necessary for cuticle development (Heffer et al., 2013). Interestingly, *Tc-ftz-f1* is expressed in stripes in *Tribolium* that overlap with the *Tc-ftz* stripes (Heffer et al., 2013; Heffer, 2012). This differs from the ubiquitous Ftz-F1 expression pattern seen in *Drosophila*. The authors suggested that the dependence of *Dm*-Ftz-F1 on Ftz released constraints on *ftz-f1* expression, allowing it to be expressed throughout the embryo without causing activation of target gene expression outside of the stripe domains in which *ftz* is expressed. Thus, for *ftz*, both the expression pattern and protein sequence vary in different extant arthropod, while the expression pattern of *ftz-f1* has changed at least once during insect radiations.

even-skipped

Orthologs of *eve* have been studied in a number of holometabolous insects. In a long germ band insect, the mosquito *Anopheles gambiae* (Diptera, Culicidae), expression of *eve* mRNA was found to be similar to that seen in *Drosophila* (Goltsev et al., 2004). No functional analysis was performed in this study. In another long germ band insect, the honeybee *Apis mellifera* (Hymenoptera: Apidae), Wilson et al. detected *Am-eve* expression in the ovarioles of the queen ovary, maturing oocytes and the posterior nurse cells (Wilson and Dearden, 2012). During embronic development, *Am-eve* mRNA was detected throughout the embryo in early embryos (stage 1-4), and

later in a broad domain similar to that seen in *Drosophila* (Wilson and Dearden, 2012*)*. In early stage 5, *Am-eve* is expressed in stripes with dual segment periodicity that later split into segmental stripes at stage 5 and stage 6. RNAi produced larva with fused central, posterior and terminal segments or larva without segments (Wilson and Dearden, 2012).

eve has also been studied in several sequentially segmenting holometabolous insects. In the silk worm *Bombyx mori* (Lepidoptera), a sequentially segmenting insect, Nakao et al. found that *Bm-eve* is expressed in a broad domain in the central region of the embryo (Nakao, 2010). Eight pair-rule stripes then emerge from this broad domain in an anterior to posterior order. According to the authors, all the eight pair-rule stripes appear before "marked elongation occurs." No functional data are available for *Bm-eve*. In the mothmidge *Clogmia albipunctata* (Diptera, Psychodidae), another sequentially segmenting insect, *Ca-eve* was found to be expressed in seven stripes but the secondary *eve* stripes seen in *Drosophlia* were not observed (Rohr et al., 1999). Again, no functional analysis was performed. However, functional studies were carried out for one sequentially segmenting holometabolous insect, the jewel wasp *Nasonia vitripennis* (Hymenoptera, Pteromalidae). *Nv-eve* was found to be expressed first in a broad domain of the embryo, then a stripe appeared at the posterior region (stripe 6) and the broad expression domain later split into three stripes (stripe 1, 2 and 3); at the start of cellularization, a faint stripe (stripe 4/5) appeared between stripe 3 and stripe 6. Then from anterior to posterior, each stripe of stripe 1 to stripe 5 split into two stripes; stripe 6 then gave rise to six segmental stripes (Rosenberg et al., 2014). Using morpholinos to target *Nv-eve*, Rosenberg et al.

found that knocking down *Nv-eve* caused partial pair-rule phenotypes for anterior segments and truncation of A5-A10 segments (Rosenberg et al., 2014), which represents a segmental phenotype.

In the sequentially segmenting red flour beetle *Tribolium castaneum (*Coleoptera, Tenebrionidae), *Tc-eve* was expressed in primary stripes and secondary stripes similar to that seen in *Drosophila,* except that the secondary stripes split off from the primary stripes (Brown et al., 1997; Patel et al., 1994). *Tc-eve* RNAi produced asegmental embryos that contained labrum, antennae, and telson but no gnathal or trunk segments (Choe et al., 2006). Eve expression was examined in two other beetle species, the sequentially segmenting beetle, *Dermestes frischi* (Coleoptera, Dermestidae) and a long germ beetle, *Callosobruchus maculatus* (Coleoptera, Chrysomelidae). *eve* orthologs were found to be expressed in stripes in both species in similar patterns to those see in *Tribolium* but functional studies were not carried out (Patel et al., 1994).

eve expression and function have also been examined in hemiptera (sister group to the holometabolous insects) and in more basally branching species. In the sequentially segmenting insect *Oncopeltus fasciatus* (Hemiptera, Lygaeidae), *Of-eve* was found to be expressed in a broad region of the early blastoderm stage embryo, spanning the posterior two thirds of the blastoderm. At late blastoderm stage, the broad expression pattern of *Of-eve* changed into six stripes (Liu and Kaufman, 2005). Since *Of-en* is also expressed in six stripes pattern at late blastoderm (Liu and Kaufman, 2004a), it appears that *Of-eve* is expressed in each segment in blastoderm stage embryos. In germband extension stage embryos, *Of-eve* was also expressed in a segmental pattern in contrast to the pair-rule expression in alternate segment

primordia seen in *Drosophila* (Liu and Kaufman, 2005). Knockdown of *Of-eve* with parental RNAi resulted in embryos with large deletions of almost the entire body (Liu and Kaufman, 2005). In mildly or weakly affected embryos, almost all had defective abdomens, with normal head segments but posterior thoracic segments were affected in some embryos (Liu and Kaufman, 2005). Further study indicated that *Of-eve* affects the expression of two gap genes, *hunchback* and *Krüppel*. Based on the RNAi phenotypes and the fact that *Of-eve* affects expression of gap genes, Liu et al. (Liu and Kaufman, 2005) classified *Of-eve* as a gap gene in *Oncopeltus*. The broad *Of-eve* expression pattern observed in early blastoderm is consistent with a gap gene function.

In the more basally branching sequentially segmenting cricket *Gryllus bimaculatus* (Orthoptera, Gryllidae), *Gb-eve* was found to expressed in a pair-rule pattern from the maxillary to the T2 segment, and secondary stripes arose by splitting of the primary stripes. A mandibular stripe formed in a segmental manner, while it is not clear if the T3 and A1 stripes were expressed in pair-rule or segmental patterns (Mito et al., 2007; Mito et al., 2006). In segments A4, A5, A8, and A9 , *Gb-eve* arose in a pair-rule to segmental pattern (one primary stripe split into two stripes), while in segments A2, A3, A6, A7, and A10, *Gb-eve* arose in a segmental pattern (Mito et al., 2007). *Gb-eve* embryonic RNAi resulted in embryos with defects in anterior segmentation, showing reduction of thoracic legs and fusion of labial and T1 segments, fusion of T2 and T3 segments, and some showing a deletion from T1 to T3 (Mito et al., 2007). Mito et al. interpreted these phenotype as pair-rule like and gap like (Mito et al., 2007). The *Gb-eve* RNAi embryos also had abdominal segmentation defects. Some embryos lost one, two or three abdominal segments; some embryos had

irregular abdominal segment borders; some embryos had defective, short posterior regions (Mito et al., 2007). These experiments also showed that *Gb-eve* regulates the gap genes *hunchback* and *Krüppel.* The authors suggested that the *Gb-eve* functions partially as a pair-rule gene in posterior segmentation. In one additional Orthoptera insect, the grasshopper *Schistocerca americana,* Eve protein was not expressed in stripy pattern; its expression was only detected in the CNS and growth zone (Patel et al., 1992). The absence of segmental expression suggests that *Sg-eve* does not function in that process at all in *Schistocerca*.

runt

In the long germ band insect *Apis mellifera, Am-runt* was found to be expressed in nurse cells and the oocytes. In early embryos, weak expression of *Am-runt* was detected broadly in the abdominal region; at stage 6, pair-rule stripes appeared in the thoracic and abdominal regions and those stripes then split to generate secondary stripes expressed in a segmental pattern (Wilson and Dearden, 2012). *Am-runt* RNAi larvae showed defects in segmentation. Segments appeared to be more widely spaced than in controls with segmentally-organized trachea less densely spaced (Wilson and Dearden, 2012). The authors suggested that this may indicate loss of alternating segments or an expansion of remaining segments.

In *Nasonia,* a sequentially segmenting insect, *Nv-runt* was found to be expressed in six pair-rule stripes before cellularization and two more posterior pair-rule stripes are added during gastrulation. At full germ band extension, *Nv-runt*'s expression changes into single-segment stripes (Rosenberg et al., 2014). Functional analysis has not been reported. In a sequentially segmenting insect *Manduca sexta* (Lepidoptera),

Ms-runt was expressed in eight pair-rule stripes at the blastoderm stage before germ band elongation starts (Kraft and Jackle, 1994). Finally, in *Tribolium,* a sequentially segmenting insect, *Tc-runt* was reported to be expressed in primary pair-rule stripes and no secondary segmental stripes were detected (Brown and Denell, 1996). *Tc-runt* pRNAi created almost completely asegmental larvae; only labium, antennae and mandibles were found in the embryonic cuticles (Choe et al., 2006). I did not find any study of *runt* in more basally branching insects that have been published to date.

hairy

The expression and funciton of *h* have been examined in several holometabolous insects in addition to *Drosophila.* In a long germ band insect, the honeybee *Apis mellifera* (Hymenoptera: Apidae), *hairy (Am-h)* is expressed in mid-stage oocytes with a weak expression in the nurse cells, and during embryonic development, first as a broad thoracic stripe, and later in eight stripes (Wilson and Dearden, 2012). Given the fact that *en* is expressed in 16 stripes in *Apis* at the late germ band extension stage (Fleig, 1990), the eight-stripe expression pattern suggests that *Am-h* is expressed in a pair-rule pattern. *Am-h* dsRNA injected embryos developed into larva with fused thoracic and anterior abdominal segments, and some larva had all segments fused (Wilson and Dearden, 2012). The most severly affected *Am-h* RNAi larva lost all *en (e30)* stripes and mildly affected larva had disorganized *en* stripes (Wilson and Dearden, 2012). This suggests that *Am-h* has a role in segmentation but one that is different from that of its *Drosophila* ortholog.

hairy orthologs have been studied in two sequentially segmenting holometabolous insects, both *Coleoptera*: the red flour beetle (*Tribolium castaneum*) and the confused

flour beetle (*Tribolium confusum*). In *Tribolium castaneum, Tc-h* is expressed in pair-rule stripes (Aranda et al., 2008; Sommer and Tautz, 1993). However, pRNAi knock down affected head development but the posterior thoracic and abdominal segments developed normally (Choe et al., 2006). Aranda et al. compared the expressionof *hairy* between *T. castaneum* and *T. confusum* (Aranda et al., 2008). They found that in both species *hairy* is expressed in stripes, but the stripes in *T. confusum* were more distinct and more persistent. Using pRNAi, Aranda et al. confirmed the results from Choe et al. (Aranda et al., 2008; Choe et al., 2006), in *T. castaneum,* showing that *Tc-h* does not have a function in the formation of the abdominal segments. In *T. castaneum*, using *Tc-gsb* as a segmental marker, Aranda et al. found that *Tc-h* RNAi mainly affected mandibular and labial segments. This suggested to the authors that *Tc-h* could have pair-rule function at the blastoderm stage, when *Tc-h* appears to be expressed in a stripy pattern.

paired

prd expression and function have been studied in two holometabolous insects, in addition to *Drosophila.* In the long germ band insect *Apis mellifera, Am-prd* was found to be expressed in a primary pair-rule pattern and every primary stripe split into two stripes (Osborne and Dearden, 2005). Different from *Drosophila,* in *Apis* the anterior primary stripes split before posterior primary stripes appear (Dearden et al., 2006). Functional data are not available.

In the sequentially segmenting insect *Tribolium castaneum, Tc-prd* was expressed in a primary pair-rule pattern and the primary stripes split to create secondary stripes (Choe and Brown, 2007). Similar to the *Am-prd, Tc-prd* anterior primary stripes split

before posterior primary stripes appear (Choe and Brown, 2007). Knockdown of *Tc-prd* with pRNAi revealed pair-rule function for this gene, with loss of odd-numbered segments and alternate *en* stripe expression, similar to what is seen in *Drosophila* (Choe and Brown, 2007).

Prd has only been examined in one non-holometabolous insect, the sequentially segmenting insect *Schistocerca americana.* Here, two *paired* homologs were found (Davis et al., 2001). These were named *pairberry1(pby1)* and *pairberry2 (pby2)*. Both *pby1* and pby2 were found to be expressed in segmental stripes (Davis et al., 2001). Phylogenetic analysis indicates that both *pby1* and *pby2* belong to PAX group III, and *pby1* and *pby2* are more closely related to each other than to *Drosophila prd, gsb* or *gsbn* (Davis et al., 2001). The authors suggesetd that *pby1* and *pby2* arose through a duplication event independent of that leading to the *Drosophila* paralogs .

odd-skipped

As for several other *Drosophila* PRG orthologs, *odd* has been studied in only two other insects, both holometabolous. In a sequentially segmenting insect *Nasonia, Nv-odd* was found to be expressed in double-segment stripes (Rosenberg et al., 2014). The authors stated that the third *Nv-odd* pair-rule stripe arises from the second pair-rule stripe; pair-rule stripes 4 to 6 are all generated by an advancing "wave" (Rosenberg et al., 2014). Knocking down *Nv-Odd* with morpholinos generated embryos missing most posterior segments (A5-A10) or embryos missing T2, A1, A3 and A5 segments (Rosenberg et al., 2014), which indicates that *Nv-odd* acts as a pair-rule gene in the central region of the embryo. In the sequentially segmenting insect *Tribolium, Tc-odd* is expressed in a pair-rule pattern (Choe et al., 2006). *Tc-odd*

RNAi embryos are severely affected, with most embryos being truncated and asegmental (Choe et al., 2006). In a separate study, Sarrazin et al. found that over time, expression of *Tc-odd* in the growth zone changes from high to low levels, and back to high levels during production of new primary stripes. Using transgenic *Tribolium* expressing nuclear-localized GFP to monitor cell movements with live imaging, they showed that this oscillating expression is due to changes in expression level. They concluded that there is a segmentation clock in the growth zone of *Tribolium* (Sarrazin et al., 2012).

odd paired

Opa is a zinc finger transcription factor. Since zinc finger transcription factors comprise the largest group of transcription factors, it is hard to find the ortholog of *opa* in other insects. This may be the reason why we have very limited data on its expression or function in other insects. So far it has only been studied in *Tribolium castaneum* (Choe et al., 2006). Choe et al. used pRNAi to knock down *Tc-opa* and reported that larvae shown no segmentation defects (Choe et al., 2006). No expression data were shown.

sloppy-paired

slp has been studied in one insect outside *Drosophila***.** In the sequentially segmenting insect *Tribolium, Tc-slp* was found to be expressed in one stripe and then two stripes at the blastoderm stage; during germ band elongation, pairs of *Tc-slp* stripes are generated in the anterior region of the growth zone, with the anterior stripe usually narrower and weaker than the posterior stripe (Choe and Brown, 2007). The authors suggested that the difference in intensity within each pair of stripes may

indicate they have different roles in segmentation. By double-staining with *en, Tc-slp* stripes were confirmed to be expressed in a single segment pattern (Choe and Brown, 2007). In *Tc-pRNAi* embryos, all the gnathal segments are affected; in the thoracic and abdominal regions, T1, T3 and A4 or A5 segments were deleted (Choe and Brown, 2007). The authors concluded that *Tc-slp* acts like a head gap gene, and a pair-rule gene in the thoracic and abdominal segments (Choe and Brown, 2007). It is interesting that in *Drosophila, slp* mutations affect odd-numbered segments (Grossniklaus et al., 1992), while in *Tribolium, slp* RNAi affects even-numbered segments (Choe and Brown, 2007; Choe et al., 2006)

Section 1.4: Other genes that have pair-rule function in insects

In *Oncopeltus, Ecdysone-induced protein 75A (E75a)* was found to be expressed in a pair-rule pattern at the blastoderm stage. During germ band elongation, *Of-E75A* abdominal segmental stripes appeared by splitting from primary stripes, which mimics the expression pattern of several *Drosophila* PRGs (Erezyilmaz et al., 2009). Interestingly, no stripes were detected after the *en* stripe is formed in A6. This suggested that *Of-E75A* may not be expressed in segments A7 to A10. In *Of-E75A* pRNAi treated animals, fusion between labial and T1, T2 and T3, T3 and A1 were seen; segmentation defects were observed between A3 and A4, A5 and A6, A7 and A8 (Erezyilmaz et al., 2009). These observations indicate that *Of-E75A* primarily affects odd-numbered parasegments. The authors did find rare defects in even-numbered parasegments (ps12 and ps14), and more frequent defects in ps6 (Erezyilmaz et al., 2009). In contrast to this pair-rule function in *Oncopeltus*,

Drosophila E75A is required for regulation of ecdysteroid biosynthesis, and null mutations do not affect segmentation (Bialecki et al., 2002).

In the silk worm *Bombyx mori,* Liu et al. found that knockdown of *Bm-groucho* resulted in larvae with very typical pair-rule phenotypes (Liu, 2012). The larvae miss mandibles, labium, one thoracic segment and half of the abdominal segments. The identity of missing segments in thoracic and abdominal region could not be unequivocally determined. However, the author assumed they are T2, A1, A3, A5, A7 and A9, which would be suggestive of pair-rule function. No expression data are available for *Bm-groucho.* In *Drosophila,* the pair-rule proteins Hairy, Runt, Eve and Slp depend on Groucho*,* which functions as a corepressor (Andrioli et al., 2004; Aronson et al., 1997; Jimenez et al., 1997; Kobayashi et al., 2001). Some *groucho* alleles do cause segmentation defects but not a typical pair-rule defect, since Groucho interacts with multiple PRG repressors (Chen and Courey, 2000; Paroush et al., 1994). Interestingly, a mutant in a histone deacetylase Rpd3, another corepressor, causes pair-rule defects in *Drosophila* (Mannervik and Levine, 1999). The authors proposed that Rpd3 functions as a corepressor with Eve (Mannervik and Levine, 1999). It is possible that the phenotype in *Bm-groucho* knockdown animals may be caused by a similar mechanism.

In a genetic screen for *Tribolium* mutants with larval cuticle patterning defects, Maderspacher et al. isolated four mutants with segmentation defects (Maderspacher et al., 1998). Two of them, *Scratchy* and *itchy* have clear pair-rule defects. In *Scratchy* mutants, a point mutation in the homeodomain of Paired was detected, and in *Itchy,* a

mutation was found in the forkhead domain of Slp (Choe and Brown, 2007). It is likely that those mutations are responsible for the phenotypes of *Scratchy* and *Itchy*, but the possibility that other genes may be involved is not ruled out.

Section 1.5: Conclusions

Except for *ftz* and *eve,* the evolution of other PRGs has not been studied systematically. Some of the pair-rule orthologs were only studied in one or two species, such as *opa, slp* and *odd*. Most of the others were only studied in Holometabolous insects. In *T. castaneum,* all the *Drosophila* PRG orthologs have been studied. *Tc-h, Tc-ftz, and Tc-opa* do not have segmentation function, and interaction between other PRGs differs dramatically from that in *Drosophila* (Choe and Brown, 2007; Choe et al., 2006). Apparently, the function of *Drosophila* PRGs are not as conserved as segmentation polarity genes such as *engrailed,* and *Hox* genes, which confer segment identity. However, studies in more basally branching insects are quite limited.

Not only do the function of PRGs vary during evolution, the mechanism underlying similar phenotypes caused by orthologous genes can be different too. For example, in *Drosophila, eve* hypermorphic mutations cause pair-rule phenotypes and *eve* null mutation produce asegmental embryos covered by a continuous lawn of ventral hairs (Nüsslein-Volhard et al., 1985; Nüsslein-Volhard and Wieschaus, 1980). The null mutant phenotype was explained as a result of the deletion of both the odd-numbered parasegments and portions of the even-numbered parasegments (Macdonald et al., 1986). In *Oncopeltus* and *Tribolium, eve* RNAi generates asegmental embryos in both species. According to Liu et al., *Of-eve* acts as a gap

gene and causes segmental defects (Liu and Kaufman, 2005). *Of-eve* is not expressed in a pair-rule pattern but in a segmental pattern. The gap-gene properties of *Of-eve* were explained by its early gap-like expression pattern (the same pattern is observed in early *Drosophila* and *Tribolium* embryos), and disruption of expression of other gap genes (Liu and Kaufman, 2005). In *Tribolium, Tc-eve* is first expressed in a pair-rule pattern, and later each pair-rule stripe generates a secondary stripe. *Tc-eve* was considered to be a primary pair-rule gene. *Tc-eve* does not regulate other gap genes and the asegmental phenotype observed in *Tc-eve* RNAi was explained by *Tc*-eve's role in activating *Tc-run* and *Tc-odd* (Choe et al., 2006).

In *Tribolium,* a few *Drosophila* pair-rule gene orthologs do not have segmentation function. Instead of assuming *Tribolium* does not need as many as PRGs as *Drosophila* does, it is tempting to think that their roles in segmentation are overtaken by some of their homologues. For example, gene *Tc-opa*, which encodes a zinc-finger transcription factor, does not play a role in segmentation of *Tribolium.* It is possible that another zinc-finger transcription factor replaced the segmentation function of *Tc-opa*. Substituting one gene with its homolog may cause less disruption to a regulatory network than replacing it with an unrelated transcription factor. Gene homologues usually bind to very similar DNA sequences, and their requirements for co-factors may also be very similar. Thus, once a gene acquires a new expression pattern, making it is possible to interact in the segmentation process, it can easily take over the function of its homolog. In *Drosophila,* homologues *slp1* and *slp2, en* and *inv*, and *prd, gsb* and *gsbn* all have functions in segmentation and most of these are redundant to each other. This supports the idea homologues can substitute each other's roles

without disrupting an existing regulatory network. The discovery of *E75A* as a pair-rule gene in *Oncopeltus* is in agreement with this hypothesis. E75A, a homolog of Ftz-F1, belongs to the nuclear receptor superfamily. Its role in *Oncopeltus* can be viewed as a substitution/redundancy to the function of Ftz-F1 in the pair-rule network. RNAi in *Tribolium* with dsRNA targeting the most conserved domains of some of the PRGs that do not show segmentation function would test this hypothesis.

There are limited PRG expression data in two Lepidoptera insects, *Bombyx mori* and *Manduca sexta*. In *B mori,* eight *Bm-eve* stripes appeared before "marked elongation occurs" (Nakao, 2010). In *M. sexta,* a similar phenomenon was observed with *Ms-runt (Kraft and Jackle, 1994).* If these two genes have pair-rule function in these two species, then their expression patterns suggest a long germ band development mode, since all their pair-rule stripes have been established before germ band elongation. In these two species, the way their embryos develop is unique. For example, in *M. sexta,* at the cellular blastoderm stage, a large round germ anlage develops, then two head lobes form at the anterior part. The ventral furrow then forms and extends from anterior to posterior and, at the same time, the germ anlage constricts laterally. When the ventral furrow reaches the posterior end, segmentation starts from anterior to posterior. This germ band elongation is driven by cell movements rather than cell proliferation (Kraft and Jackle, 1994). This mode of embryonic development is hard to classify into either long or sequentially segmenting modes, leaving their classification controversial (Davis and Patel, 2002; Nakao, 2010; Peel, 2008; Xu et al., 1997). It was suggested that *B.mori* might represent an

intermediate state in the transition from sequentially segmenting to long germ band development (Nakao, 2010). More PRG expression and functional data could help us better understand the embryonic development in these species.

In the studies of *Drosophila* PRGs and their evolution in other insects, there are many examples which indicate that the function of a gene cannot be automatically inferred from its expression pattern. For example, in *Tribolium, Tc-ftz* is expressed in stripes, but it does not have any function in segmentation. In *Drosophila, ftz-f1* and *rdp3* are expressed ubiquitously, by interact with other co-factors, and thus regulate genes in only a subset of the cells in which they are expressed. These studies highlight the need for functional analysis, as opposed to only analysis of spatio-temporal gene expression patterns, in additional species.

In this thesis, I address some of the issues raised here. From the above review, we noticed that most studies of segmentation focused on holometabolous insects. Data about PRGs in non-holometabolous are sparse and indispensable for a better understanding of PRG evolution.

I studied the ligand dependency of mammalian Ftz-F1 homologues. Structural evidence suggested that SF-1 and human LRH-1 bind regulatory ligands, but mouse LRH-1 and *Drosophila* FTZ-F1 are active in the absence of ligand. I found that *Dm*-Ftz-F1 and mLRH-1, though not to binding ligand, or mSF-1 and hLRH-1, predicted to bind ligand, each efficiently rescued the defects of *Drosophila ftz-f1* mutants. This result indicates that the activation of NR5A family members is not initiated by ligand binding.

In order to facilitate the study of non-holometabolous insects, I participated in the genome annotation of a Hemiptera insect, *Oncopeltus fasciatus.* I annotated nuclear receptor super family genes, a few *Hox* genes and PRGs in *Oncopeltus*. I further studied the function and expression pattern of four PRGs in *Oncopeltus.* Using *in situ* hybridization and RNAi, I found that Of-*ftz* and Of-*hairy* do not have segmentation function, while *Of-ftz-f1* functions in oogenesis and segmentation. *Of-runt* was found to induce cell death in oocytes. Using the knowledge and expertise I gained from *Oncopeltus,* I successfully set up *in situ* hybridization, antibody staining and pRNAi in an invasive Hemiptera insect pest, the brown marmorated stink bug (BMSB) *Halyomorpha halys.*

Chapter 2: Functional conservation of *Drosophila* FTZ-F1 and its mammalian homologs suggests ligand-independent activation of NR5A family members

[Published Lu et al., Development Genes and Evolution, 2013]

Section 2.1 Abstract

Drosophila Ftz-F1 is an orphan nuclear receptor required for segmentation and metamorphosis. Its mammalian orthologs, SF-1 and LRH-1, function in sexual development and homeostasis, and have been implicated in stem cell pluripotency maintenance and tumorigenesis. These NR5A family members bind DNA as monomers and strongly activate transcription. However, controversy exists as to whether their activity is regulated by ligand-binding. Structural evidence suggested that SF-1 and human LRH-1 bind regulatory ligands, but mouse LRH-1 and *Drosophila* FTZ-F1 are active in the absence of ligand. We found that Dm-Ftz-F1 and mLRH-1, thought not to bind ligand, or mSF-1 and hLRH-1, predicted to bind ligand, each efficiently rescued the defects of *Drosophila ftz-f1* mutants. Further, each correctly activated expression of a Dm-Ftz-F1 target gene in *Drosophila* embryos. The functional equivalence of *ftz-f1* orthologs in these sensitive *in vivo* assays argues against specific activating ligands for NR5A family members.

Section 2.2 Introduction

Diverse nuclear receptors (NRs) play important roles in development, differentiation, reproduction and homeostasis. The activity of some of these

transcription factors is regulated by small molecule ligands that act as molecular switches to control transcriptional activity (McKenna et al., 2009). For a number of NRs, no natural ligand has been identified. These so-called 'orphan' NRs share with ligand-activated NRs a typical domain structure, including a variable N-terminal domain, a DNA-binding domain, a hinge region and a ligand-binding domain (Benoit et al., 2006). For ligand-regulated NRs, the binding of ligand induces a conformational change that exposes the AF-2 domain at the C-terminus of the LBD to allow for interaction with LXXLL motifs in NR coactivators and thus, activation of transcription. Although orphan receptors harbor AF-2 domains, it is not certain how their activity is modulated. Protein–protein interaction, post-translational modifications such as phosphorylation, acetylation and sumoylation affect the activity of many NRs and could potentially substitute for ligand-mediated activation.

Drosophila Ftz-F1 is the founding member of the NR5A family of orphan receptors (Pick et al., 2006). Ftz-F1 proteins bind DNA as monomers and appear to be constitutive activators of transcription in a range of cell types (Pick et al., 2006). In *Drosophila*, two isoforms of Ftz-F1 have been identified: *αftz-f1* is maternally expressed and is required in the early embryo to establish the basic segmented body plan of the fly (Guichet et al., 1997; Yu et al., 1997). *βftz-f1* is expressed during larval molting and is required for elaboration of the cuticle pattern in larvae (Ruaud et al., 2010). It further serves as a competence factor for metamorphosis in part by regulating the breakdown of fat for fueling this process (Bond et al., 2011; Broadus et al., 1999). *Drosophila αftz-f1* mutant embryos (derived from germline clones or from females homozygous for maternal-specific alleles; referred to here collectively as *ftz-*

fl mutants) display a pair-rule segmentation phenotype in which alternate, even-numbered parasegments are missing (Guichet et al., 1997; Yu et al., 1997), and see Figure 2.2. This phenotype is identical to that of *Drosophila ftz*, a homeobox-containing gene, which is expressed in stripes in the primordia of the alternate parasegments that are missing in *ftz* and *ftz-fl* mutants. Thus, although Ftz-F1 is present in all somatic cells of the embryo, its activity is limited to cells in which Ftz is present (Ftz+ cells). In these Ftz+ cells, Ftz and Ftz-F1 interact to form a stable complex, immunoprecipitable from wild-type *Drosophila* embryos (Yu et al., 1997). Ftz and Ftz-F1 bind cooperatively to heterodimeric DNA target sites *in vivo* where they activate the transcription of at least 10 genes involved in segmentation (manuscript in preparation), while also positively auto-regulating the expression of *ftz* in stripes (Yu et al., 1997). The interaction between Ftz and Ftz-F1 is dependent upon an LRALL sequence (LXXLL motif) in Ftz (Schwartz et al., 2001). This NR coactivator-like LXXLL motif in Ftz led our lab and others to propose that the binding of Ftz to Ftz-F1 obviates ligand binding, with the protein–protein interaction between Ftz and Ftz-F1 serving as the molecular switch that activates the transcriptional potential of Ftz-F1 (Schwartz et al., 2001; Suzuki et al., 2002; Yussa et al., 2001a). This explains why Ftz-F1, although present and constitutively nuclear throughout the *Drosophila* embryo, only activates target genes in cells that co-express Ftz. This is much like the situation noted first for mammalian SF-1: protein–protein interaction with the homeodomain protein Ptx1 activates its transcriptional potential (Tremblay et al., 1999a). Similarly, protein–protein interaction seems to be sufficient to regulate the biological activity of Dm-Ftz-F1.

In mammals, two NR5A family members, SF-1 and LRH-1, play important roles in development, differentiation, tumorigenesis and embryonic stem cell pluripotency (Fernandez-Marcos et al., 2011). The potential clinical importance of these NR5A family members bolstered interest in finding small molecule regulators of their activity. However, searches for endogenous NR5A family ligands have led to controversial findings. Several groups published crystal structures of mammalian NR5A LBDs. These studies revealed the presence of phospholipids in the binding pockets of human and mouse SF-1 (hSF-1, mSF-1) and human LRH-1 (Krylova et al., 2005; Li et al., 2005; Ortlund et al., 2005). In contrast, the LBD of mouse LRH-1 (mLRH-1) was found to adopt an active conformation in the absence of bound ligand (Sablin et al., 2003). Recently, the LBD of Dm-FTZ-F1 was shown to be in an active conformation without ligand binding (Yoo et al., 2011b). For all of these studies, proteins were produced in bacteria, raising the possibility that the small molecules identified were fortuitous ligands. However, mutations in the ligand-binding pocket decreased transcriptional activity in cell based assays. Further, removal of phospholipid from bound receptors decreased their ability to activate transcription of a number of target genes in cell culture experiments, and "humanization" of mLRH-1 rendered its transcriptional activity ligand-dependent (Forman, 2005; Krylova et al., 2005; Wang et al., 2005). More recently, sphingosine was identified as a candidate repressory ligand for SF-1 (Urs et al., 2007). Together, these findings, while revealing the importance of occupancy of the LBD pocket, do not distinguish between activating ligands and stabilizing compounds that are necessary as constitutive cofactors for protein folding (Forman 2005). Phosphatidylcholine and

phosphatidylethanolamine bind SF-1 with very high affinity (~100 nM), are abundant in mammalian membranes, and are present in the nuclear matrix (Forman 2005). Thus, it is possible that there is sufficient phospholipid in the nucleus under most physiological conditions to constitutively bind and stabilize NR structure. In sum, despite extensive analysis, there is still controversy as to whether ligand binding serves as a switch to activate NR5A family proteins.

Here, we compared the ability of mammalian orthologs of *Drosophila* Ftz-F1 to functionally substitute for Ftz-F1 *in vivo*. Our expectation was that if orthologs differed in their requirements for activation by ligand they would differ in their abilities to activate target gene expression. We found that mSF-1 and hLRH-1, which binds ligand in vitro, as well as mLRH-1, which does not bind ligand, each rescue *ftz-f1* mutants with no obvious difference in efficiency. These results, combined with analysis of the crystal structure of the Ftz-F1 LBD, suggest that specific activating ligands are not required for the function of NR5A family members *in vivo*. Rather, activity of all family members was limited by Ftz in *Drosophila* embryos, suggesting control by protein-protein interaction.

Section 2.3 Material and methods

Plasmid construction and transgenic flies

For expression in *Drosophila*, cDNAs were inserted into the P-element vector pUAS-T and transgenic flies were generated by P-element mediated transformation. To express Dm-Ftz-F1, a BamHI/EcoRI fragment from pGEX5X1-αftz-f1 was isolated and inserted along with a PCR generated fragment encoding a 3xFLAG

epitope, to which EcoRI/BamHI sites had been added, into the EcoRI site of pUAS-T to produce Dm-Ftz-F1 protein with an N-terminal FLAG tag. Reading frame and directionality were verified by sequencing. To express mammalian NR5A proteins, cDNAs encoding full-length NR5A family members with N-terminal HA tags were inserted into pUAS-T as follows: mouse Lrh-1 was isolated from pCI-Neo-HA-mLRH using EcoRI and inserted into the EcoRI site of pUAS-T to generate UAS-mLrh-1; mouse Sf-1 was isolated from pCI-Neo-HA-mSf using EcoRI and NotI, and inserted into the EcoRI/NotI sites of pUAS-T to generate UAS-mSf-1; human LRH-1 was isolated from pCI-Neo-HA-hLRH using EcoRI and NotI, and inserted into the EcoRI/NotI sites of pUAS-T to generate UAS-hLRH-1. The full-length sequence of each insert in pUAS-T was verified. Transgenic *Drosophila* were generated by Rainbow Transgenic Flies, Inc. (Newbury Park, CA, USA). Multiple independent lines were established for each transgene. At least three independent lines were tested for each and all gave similar results.

Functional assays in Drosophila

To test the ability of mammalian NR5A family members to functionally substitute for Dm-Ftz-F1 *in vivo*, they were expressed in embryos derived from females homozygous for a maternal-specific *ftz-f1* allele (Guichet et al., 1997). Transgenes were expressed ubiquitously in these *ftz-f1* mutant embryos with the UAS/GAL4 system using an NGT40 driver that directs expression ubiquitously in blastoderm embryos, mimicking the endogenous expression pattern of Dm-*ftz-f1*. The experiments were carried out as follows: NGT40; ftz-f1 ^{209}/ftz-f1 ^{209} virgin females were crossed to males carrying UAS transgenes: UAS-Dm-*ftz-f1* (positive control),

UAS-lacZ (negative control), UAS-mSf-1, UAS-mLrh-1 or UAS-hLRH-1. The ability of transgenes to rescue the *ftz-f1* mutants was assessed by examining larval cuticles using standard methods. Except for the UAS-lacZ negative control, three independent lines were tested for each transgenic construct. For each individual line tested, at least 100 cuticles were counted. Statistical analysis was carried out using SAS software. Immunofluoresence staining with an anti-Engrailed antibody was carried using standard methods. The anti-Engrailed antibody (4D9) was obtained from the Developmental Studies Hybridoma Bank (University of Iowa).

Section 2.4 Results and discussion

Mouse and human Ftz-F1 orthologs correctly regulate target gene expression in Drosophila

Drosophila Ftz-F1, with its partner Ftz, directly regulates expression of the engrailed gene (en) in alternate parasegments via direct binding to an *en* enhancer (Florence et al., 1997). Thus, in *ftz-f1* mutants, as in *ftz* mutants, alternating En stripes are missing (Figure. 2.11 A). As shown previously (Yussa et al., 2001a), expression of Dm-Ftz-F1 at the blastoderm stage in *ftz-f1* mutant embryos rescued expression of alternate en stripes (Figure 2.1 B). We reasoned that if NR5A proteins from mouse and human have differential requirements for ligand binding, different effects on target gene expression would be seen in *Drosophila* embryos. For example, if an activating ligand were constitutively present in *Drosophila* embryo, ligand-responsive hLRH-1 would be transcriptionally competent throughout the embryo, and could potentially interact with ubiquitously expressed NR coactivators to activate en expression outside of its endogenous striped domain. If *Drosophila* embryos lack an

activating ligand, ligand-dependent NRs would remain inactive and only the ligand-independent family members would rescue en expression. However, we found that mSF-1 (Figure. 2.1 C), mLRH-1 (Figure. 2.1 D) and hLRH-1 (Figure 2.1 E) each correctly activated En expression in alternate stripes. In addition, no ectopic En expression was observed in any of these experiments, suggesting that proteins were not ectopically activated outside of the Ftz expression domain.

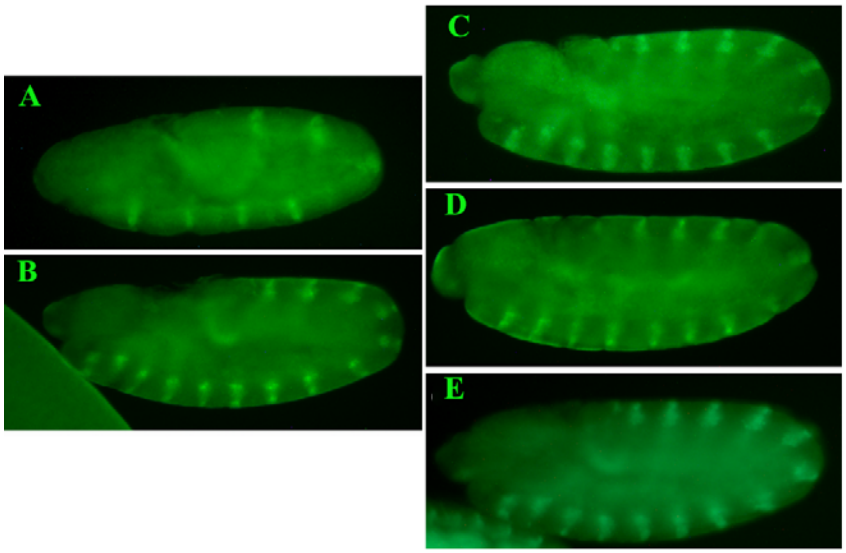

Figure 2.1 Mammalian NR5A family members correctly regulate a Ftz-F1 target gene *in vivo*. The *UAS/GAL4* system was used to express NR5A family members ubiquitously in *Drosophila* embryos. Embryos were analyzed for expression of the *Dm*-Ftz-F1 target gene *engrailed*, revealed by staining with an anti-Engrailed (En) antibody. (A) *UAS-lacZ* served as a negative control, revealing loss of alternate En stripes in *ftz-f1* mutants. (B) *Dm*-Ftz-F1, (C) mSF-1, (D) mLRH-1, (E) hLRH-1. Each mammalian NR5A family member tested rescued expression of alternate En stripes. Embryos are oriented anterior (left), dorsal (top).

Mouse and human Ftz-F1 orthologs rescue Drosophila ftz-f1 mutants

As shown above, mouse and human Ftz-F1 orthologs rescued the expression of En in *ftz-f1* mutant embryos. To test whether they could substitute for Dm-Ftz-F1 in

other aspects of segmentation, transgenes were expressed as described above and cuticle preparations were examined. In *Drosophila*, loss of maternal *ftz-f1* results in embryonic lethality due to defects in segmentation (Guichet et al., 1997; Yu et al., 1997). Embryos fail to hatch and display a typical pair-rule phenotype–missing alternate parasegments (Figure 2.2 A). Ubiquitous expression of UAS-Dm-*ftz-f1* at the blastoderm stage in *ftz-f1* mutants rescued these cuticle defects (Yussa et al., 2001a). Using a Flag-tagged Dm-Ftz-F1 generated similar results, with complete rescue of cuticle defects in most embryos (Figure 2.2 B and Table 2.1). Similarly, expression of mSf-1 (Figure 2.2 C), mLrh-1 (Figure 2.2 D) or hLRH-1 (Figure 2.2 E) rescued the cuticle defects in the *ftz-f1* mutants. No defects were observed in the Ftz-F1 independent portions of the cuticle, suggesting that none of the transgenes were active outside of the Ftz+ cells, although they were expressed throughout the embryo.

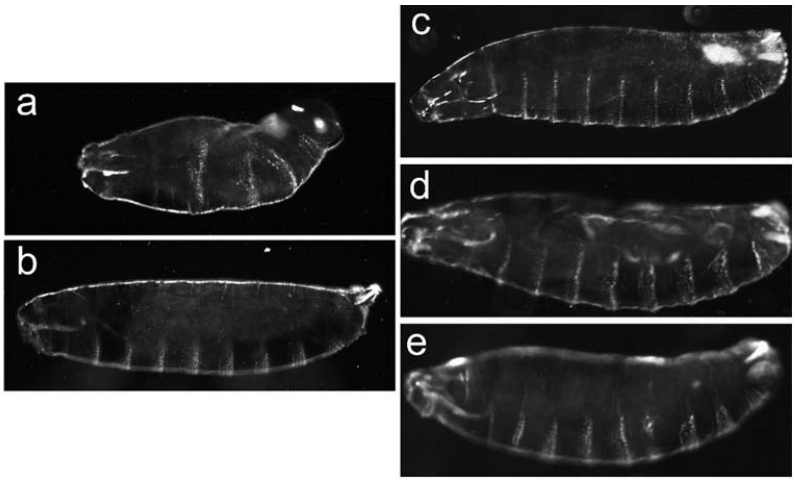

Figure 2.2 Mammalian NR5A family members rescue *Dm-ftz*-f1 mutants. The *UAS/GAL4* system was used to test the ability of NR5A family members to rescue *ftz-f1* mutant phenotypes in *Drosophila* embryos. Embryos were analyzed for rescue of defects in larval cuticle preparations. (A) UAS-lacZ served as a negative control, revealing the typical *ftz*-f1 pair-rule defect, with alternate parasegments missing. (B) Dm-Ftz-F1, (C) mSF-1, (D) mLRH-1, and (E) hLRH-1. Each mammalian NR5A family members tested rescued pair-rule segmentation defects of ftz-f1 mutants

Table 2.1. Rescue efficiency of NR5A family member transgenes.

Transgenic line	Rescued (n=number of embryos)	Not rescued	Percent Rescued (%)
UAS-lacZ	27	316	7.9
UAS-Dm-ftz-f1	310	9	97.2
UAS-m*Sf*-1	260	60	81.3
UAS-m*Lrh*-1	318	63	83.5
UAS-h*LRH*-1	268	82	76.6

To test the relative effectiveness of different orthologs, transgenes were expressed in *ftz-f1* mutant embryos and larval cuticles were scored as "rescued:" presence of three thoracic and eight abdominal denticle belts, or "non-rescued:" absence of all *ftz/ftz-f1*-dependent denticle belts. Absence of only a few segments, defined as partial rescue in previous analyses, was very rare in the experiments reported here. Three independent transformant lines were tested for each transgene and the data were pooled (Table 1). In the absence of functional Ftz-F1 (*UAS-lacZ*), 7.9 % of embryos display wild-type cuticles in these experiments because of leakage of the *ftz-f1* [209]

allele (Florence et al. 1997; Guichet et al. 1997). Ubiquitously expressed Dm-*ftz-f1* rescued ~97 % of the *ftz-f1* mutants. For mammalian NR5A family members, rescue efficiency was slightly lower: mSf-1 rescued ~81 % of the mutants, mLrh-1 rescued ~84 % of the mutants, and hLRH rescued ~77 % of the mutants. A Chi-square experimental-wise comparison between the rescue efficiency of human and mouse orthologs was carried out using SAS. The calculated chi value was 5.688, with a p value of 0.058. In sum, although the rescue efficiency of mammalian NR5A family members is slightly lower than that of *Drosophila ftz-f1*, the mammalian NR5A transgenes all effectively rescue *ftz-f1* mutants, with no significant difference ($p > 0.05$) between the ligand-bound and ligand-independent orthologs.

Section 2.5 Discussion

After their initial discovery, many orphan nuclear receptors have been 'adopted' as ligands have been identified (reviewed in Forman 2005; Benoit et al. 2006). In some cases, ligands were identified by virtue of crystal structures that revealed ligands residing in pockets of various sizes in the LBDs of purified proteins. In other cases, these fortuitous ligands, bound to the receptor after expression in bacteria, were either not endogenous ligands or were shown to function as structural rather than activating ligands (reviewed in Forman 2005). Studies to date on the ligand status for NR5A family members do not distinguish between structural and activating ligands for the phospholipid bound family members (see "Introduction" section). Yoo et al. studied the crystal structure of the *Dm*-Ftz-F1 LBD (Yoo et al., 2011b). A surprising finding was that the LBD of *Dm*-FTZ-F1 was in an active conformation in the presence of the *Dm*-Ftz LRALL motif, without any ligand. Further, the ligand-

binding pocket of *Dm*-Ftz-F1 was filled with helix 6 of its own LBD (Yoo et al., 2011). These data suggest that *Dm*-Ftz-F1 functions in a ligand-independent fashion, similar to mLRH-1.

Previous parsimony analysis of mammalian NR5A proteins suggested that the ancestral NR5A family member was ligand activated and that loss of ligand dependence occurred in rodents (Krylova et al. 2005). The results presented here add an outgroup to this analysis and extend the data available for reconstruction of the origins of ligand binding for the NR5A family. Together, they suggest that if the ancestral NR5A family member was ligand activated, at least two independent losses of ligand dependence occurred, one within insect lineages and one in rodent lineages. Alternatively, if ancestral NR5A was ligand independent, one gain of ligand dependence occurred in lineages leading to mammals and one subsequent loss occurred in rodents. Thus, with the additional data on *Dm*-Ftz-F1, strict parsimony analysis cannot distinguish between a ligand dependent or independent ancestral state for NRA5 family members, as two independent events must have occurred, either two independent losses or one gain and one loss of ligand binding.

To further analyze potential functional differences between the ligand-bound and ligand-independent NR5A family members, we used *Drosophila* as an *in vivo* model. These experiments tested whether mammalian orthologs were sufficiently similar to *Dm*-Ftz-F1 to be able to complement loss-of-function mutations. This very stringent assay—rescue of a whole animal mutant phenotype—requires proper regulation of multiple target genes to allow progression through development of the embryo to the fully differentiated larval stage. Further, this assay is highly sensitive in that it

requires that target genes be activated in the correct cells at the correct times, as either lack of activation or the opposite, ectopic activation, would cause changes in the patterns of segmentation that are readily observed in cuticle preparations (Figure 2.2). Despite these rigorous requirements, we found that human LRH-1, mouse Lrh-1 and Sf-1 could each fully rescue *Dm-ftz-f1* mutants. This suggests that the molecular function of these proteins has been highly conserved since the divergence of mammalian and insect lineages over 500 MY ago.

What does this tell us about the status of ligand-dependent activation for those family members found to bind phospholipids? As discussed above, *Dm*-Ftz-F1 appears to activate target gene expression in the absence of a ligand, with its activity in the embryo limited to the cells that co-express the homeodomain protein Ftz. However, if other family members require a specific activating ligand and a general cofactor, that ligand would have to be present in the *Drosophila* embryo at the correct time, in the correct cells and in sufficient quantity during development to allow for function of the receptor. If such a ligand is not present in *Drosophila*, no activity would be expected for ligand dependent NRs. If ligand were expressed only in discrete regions of the embryo, ligand-dependent NRs would be active only in those domains. At the other extreme, if a ligand were present ubiquitously, ligand-dependent NRs would be transcriptionally competent throughout the embryo. Several NR coactivators are ubiquitously expressed in *Drosophila* embryos, including dCBP, the mammal ortholog of which interacts with SF-1 (Monte et al., 1998; Waltzer and Bienz, 1999), Taiman, whose mosquito orthologs interacts with βFtz-F1 (Bai et al., 2000; Zhu et al., 2006), Cryptocephal (Gauthier et al., 2012) and others. Thus, it is

reasonable to expect that ubiquitously ligand-activated NR5A family members would activate transcription of engrailed and/or other target genes ectopically in this case. Given these constraints, a minimal expectation is that some difference in function would be seen between ligand-dependent and ligand-independent family members.

In contrast to this, we found mSF-1, mLRH-1 and hLRH-1 were all remarkably effective in recapitulating the activity of endogenous *Dm*-Ftz-F1 with respect to activation of one key target gene (Figure 2.1) and rescue of segmentation defects (Figure 2.2 and Table 2.1). There was little qualitative or quantitative difference in the activity of mammalian NR5A family members tested and no ectopic activity of any family member was observed. Rather than showing specificity for ligand, these family members appear to share with *Drosophila* Ftz-F1 the requirement for interaction with Ftz for function. This is rather surprising as the *ftz* gene is not present in mouse or human and these proteins have been diverging for millions of years.

One final potential interpretation of this Ftz-restricted activity of NR5A proteins is that an activating ligand is present in only the Ftz+ cells of the embryo. This scenario is very difficult to test, but we think it unlikely for the following reasons: At the time of Ftz/Ftz-F1 expression, cells have only recently formed from a syncytium and there would be little to restrict the movement of small molecules (potential ligands) to specific domains of the embryo. Zygotic transcription is also beginning around this time and the only gene that is expressed in only the Ftz+ cells is *ftz* itself. Thus, if an enzyme or other protein that synthesizes (or binds, releases or modulates) a ligand were responsible for presence of a ligand in only Ftz+ cells, the gene encoding this product would have to be regulated by Ftz. However, the evidence that

Ftz cannot activate specific gene expression without Ftz-F1 is compelling. Thus, it becomes a logical improbability that Ftz would activate transcription of a gene product required to generate a localized ligand that then activates Ftz-F1 transcriptional activity.

Taken together, these arguments, along with the finding that different NR5A family members function similarly—in an *in vivo* environment with full-length proteins—argue against a specific activating ligand for a subset of these proteins. Rather, it is more likely that a phospholipid(s) fortuitously and ubiquitously distributed in the *Drosophila* embryo functions as a stabilizing ligand for mSf-1 and hLRH-1. In keeping with this, Laudet and co-workers proposed that ancient NRs were orphan receptors and that the ability to bind ligand was acquired during evolution (Escriva et al., 2004). The NR5A family may represent a family in the proposed intermediate stage of evolution in terms of ligand binding: They have the ability to bind to some small molecules ("ligands") but the binding does not change the status of their activity.

Chapter 3: Expression and function of pair-rule genes in *Oncopeltus fasciatus*

Section 3.1 Oncopeltus fasciatus *is an emerging model system*

The studies described in this section give general intro to the section

Oncopeltus fasciatus (large milkweed bug) is a member of the diverse order Hemiptera (Superoder, Paraneoptera). Paraneoptera includes a group of hemimetabolous insects, phylogenetically positioned as the sister group of Holometabola (Kristensen, 1991). Both *Drosophila melanogaster* and *Tribolium castaneum*, which are the two best-studied insect model systems in term of molecular biology and developmental biology, belong to Holometabola. Studies in hemimetabolous insects can provide useful information, helping to elucidate ancestral states and the sequence of evolutionary events within insects.

3.1.1 FTZ evolution in insects

As discussed in Chapter 1, our data suggest that the LXXLL motif of Ftz was acquired at the stem of Holometablous insects while the presence of a YPWM motif varies (Figure 3.1). The stripy expression of *ftz* and the LXXLL motif has been observed in all holometabolous insects examined, suggesting that these traits were acquired and stably maintained in groups such as Coleoptera, Hymemptera and Diptera (Figure 3.1;)(Heffer et al., 2010). Ftz from basally branching insects such as Orthoptera, Dermaptera, Thysanura and Archeaognatha do not have LXXLL motifs and the stripy expression pattern has only been found in Thysanura (Heffer et al., 2010). *Oncopeltus fasciatus* belongs to the Hemipteroid Assemblage, which is the sister group of Holometabolous insects. Since the LXXLL motif of Ftz was acquired

at the stem of holometabolous insects, it is interesting to know the expression pattern and function of Ftz before acquiring LXXLL motif. Thus, in terms of both *ftz* evolution and insect phylogeny, *Oncopeltus* serves as intermediate group.

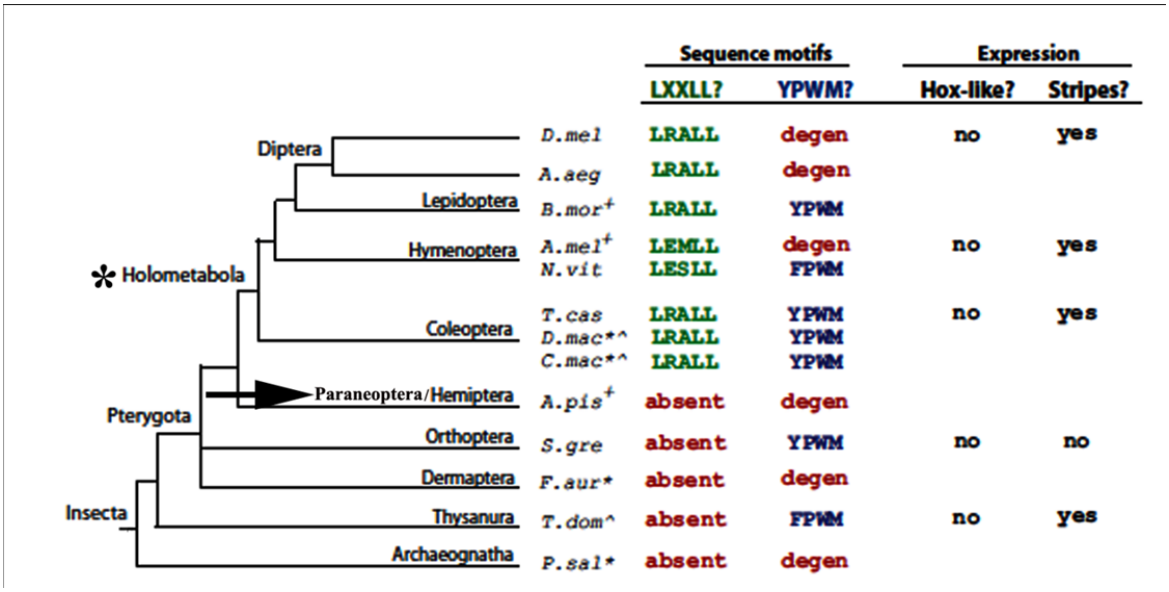

Figure3.1 FTZ evolution in insects. Cladogram of major insect taxa is shown. The Paraneoptera is indicated by an arrow, the Holometabola is indicated by a *. The presence of cofactor interaction motifs (LXXLL motif, green; YPWM motif, blue; absent, red) and observed expression patterns (stripes; *Hox*-like) are indicated. Modified from Heffer (Heffer et al., 2010)

3.1.2 Sequentially segmenting and long germ band modes of insect development.

All insects have a segmented body plan. Their metameric bodies have a head of six segments, a thorax of three and an abdomen of 8 to 11 segments (Chapman, 1998). Although their body plan is conserved, how these segments develop during embryogenesis varies. Based on how many segments have been specified by the end of the blastoderm stage, insects have been classified as short, intermediate or long germ band (Sander, 1976). For the long germ band insects, all the segments are specified simultaneously during the blastoderm stage (Figure 3.2 A). On the other

hand, short and intermediate germ band insects only pattern the anterior segments at the blastoderm stage (Figure 3.2 B). The posterior segments are specified during a phase of secondary growth and arise from the posterior region of the germ band, the growth zone (Figure 3.2 C). Posterior segments are added one-by-one in an anterior to posterior progression (Sander, 1976).

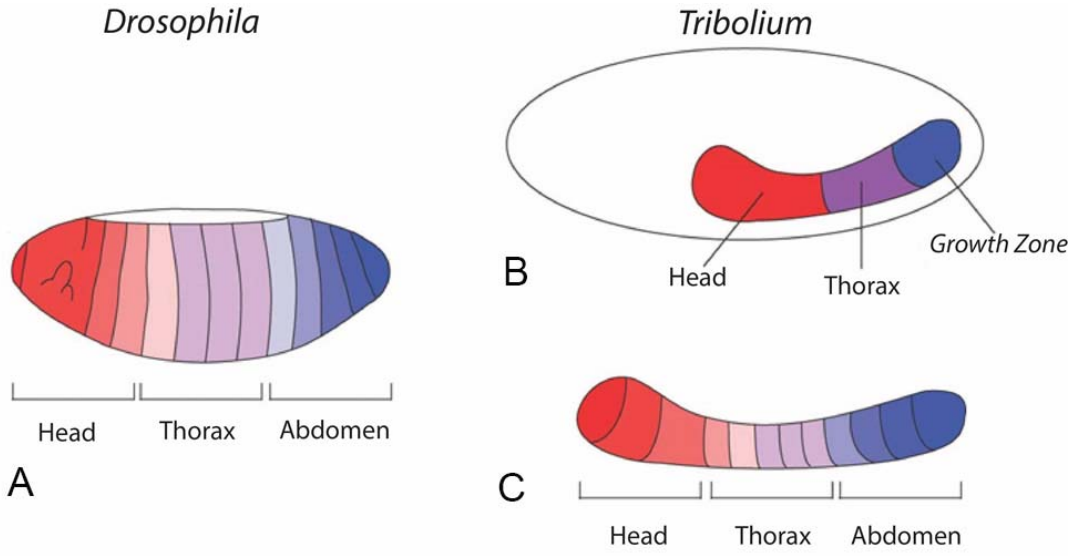

Figure 3.2 Long germ band and sequential segmentation. (A) In long germ band insect *Drosophila,* all segments are specified before the end of blastoderm stage. (B) In sequentially segmenting insects such as *Tribolium* (short germ band), only the head and thorax segments are specified at the end of blastoderm stage. (C) Posterior segments form from the growth zone.

Recently, this classification has been challenged (Davis and Patel, 2002; Peel et al., 2005). Peel at al. suggested the use of "sequentially segmenting" instead of "short/intermediate germ" for insects, in which the segments are added sequentially (Peel et al., 2005). Sequential segmentation is found among basally branching insect orders, while long germ band segmentation has only been found in the more derived holometabolous insects, such as *Drosophila*. Thus, sequential segmentation is thought to be ancestral (Anderson, 1973; Davis and Patel, 2002).

Oncopeltus is an sequentially segmenting (intermediate germ band) insect (Liu et al., 2004a). Its segmentation is thus expected to be different from the long germ band *Drosophila*, in which the segmentation regulation network had been intensively studied (Gilbert, 2010). For the purpose of comparison, studying the segmentation of *Oncopeltus* will allow us to better understand ancestral insect development and will provide information about the functional evolution of segmentation gene networks.

3.1.3 Oncopeltus fasciatus *as a research model*

Hemiptera is the fifth largest group of insects after Coleoptera, Diptera, Hymenoptera, and Lepidoptera (Forero, 2008). Many of them are important pest species to crops and gardens, such as aphids and brown marmorated stink bugs, and some are significant vectors of human diseases, such as lice and kissing bugs (*Panstrongylus megistus* and *Paratriatoma hirsuta*) (Bern et al., 2011). Studying their molecular developmental biology may lead to new methods to control these insect pests.

Compared with other Hemiptera, *Oncopeltus* also have many characteristics that are advantageous for development as a model system. They are easy to culture in the lab, and they only need water and sunflower seeds to survive and breed. Their life-cycle is relatively short: the eggs hatch within seven days, and the adults only live for about 30 days (Feir, 1974). Both adults and eggs are relatively large, making them easy to handle.

Kaufman's lab has shown that RNA *in situ* hybridization and RNAi work very well in *Oncopeltus* (Hughes and Kaufman, 2000). Liu et al. showed that injecting female *Oncopeltus* with dsRNA, referred to as parental RNAi (pRNAi), led to gene

knockdown effects in the eggs laid by those females (Liu and Kaufman, 2004a; Liu and Kaufman, 2005). The effectiveness and efficiency of parental RNAi in *Oncopeltus* makes it a good tool for knocking down gene expression for functional studies.

Using pRNAi, several labs have carried out studies in *Oncopeltus*. For example, Liu et al. showed that knockdown of *Krüppel* caused a gap phenotype similar to that seen in *Drosophila*, while *eve-skipped* RNAi caused deletion of almost the entire body, which suggests that *even-skipped* does not act as a PRG in *Oncopeltus* (Liu and Kaufman, 2004b, 2005). Chesebro et al. injected *Scr* dsRNA into *Oncopeltus* nymphs. This allowed them to study the function of *Scr* in post-embryonic stages (Chesebro et al., 2009). They found that pRNAi of *Scr* caused only minor changes in the labium, while injecting dsRNA into nymphs caused T1 to T2 leg transformation and ectopic wings on T1. These kinds of phenotypes indicate that the function of *Scr* is conserved in both holo- and hemi-metabolous insects.

As discussed above, *Oncopeltus* provides an excellent outgroup for the Holometabola, the best studied insect clade, both in terms of insect development and molecular evolution. As a group of sequentially segmenting insects, it provides a good comparison with long germ band insects such as *Drosophila*. Parental RNAi provides an easy and quick way to study gene function, and RNA *in situ* hybridization has been established, allowing examination of the expression patterns of genes of interest. All these features make *Oncopeltus* an ideal research subject for biological studies.

3.1.4 The embryonic development of Oncopeltus fasciatus

The embryonic development of *Oncopeltus fasciatus* had been studied by Butt and his student Paz (Butt, 1949; Paz, 1958). According to Butt and Paz, the development of *Oncopeltus* is summarized as follows:

(1) Early embryogenesis: When eggs are kept at 25°C, the male and female pronuclei fuse near the middle of the yolk mass within the first half-hour after egg laying (AEL). After pronuclear fusion, the nuclei go through several rounds of synchronous divisions without concomitant cell divisions, forming a syncytium. The nuclei and their associated cytoplasm then migrate towards the outside of the egg, and reach the periphery at ~15 hours AEL. The nuclei in the periphery increase in number and cell membranes begin to form within the next two hours, forming the blastoderm.

(2) Blastoderm formation: From 17-32 Hours AEL, cell numbers increase in the blastoderm embryo (Figure 3.3 A.). Many mitotic figures can be seen, and cell number continues to increase until they are closely packed and cuboidal in shape at 32 hours AEL. At ~24 hours AEL, a clump of cells (germ cells) appears at the posterior pole of the egg. From 30 hours AEL, the blastoderm begins to thicken on the ventral side of the egg and to thin out on the opposite side. The thickened area will form the germ band, the thinner area will form the serosa. According to Liu et al. (Liu and Kaufman, 2004a), at the end of the blastoderm stage (34-36 hours AEL, Figure 3.3 B), only six segments have been specified. These six segments correspond to the mandibular through third thoracic segments (Figure 3.3 C, D1, and D2). Thus,

in keeping with its status as a sequentially segmenting insect, only the anterior segments are specified at the end of the blastoderm stage.

(3) Germ band invagination: At ~35 hours AEL, a slight depression appears at the caudal end of the rudiments of the germ band,. The depression progresses rapidly, forming a slender tube-like invagination in the yolk. The wall on the ventral side of the tube migrates into the yolk to form the germ band.

(4) Germ band extension: The germ band begins to extend after invagination. As germ band invagination continues, the tip of the germ band eventually reaches the anterior pole of the egg and the head of the resulting germ band stage embryo reaches the posterior of the egg.

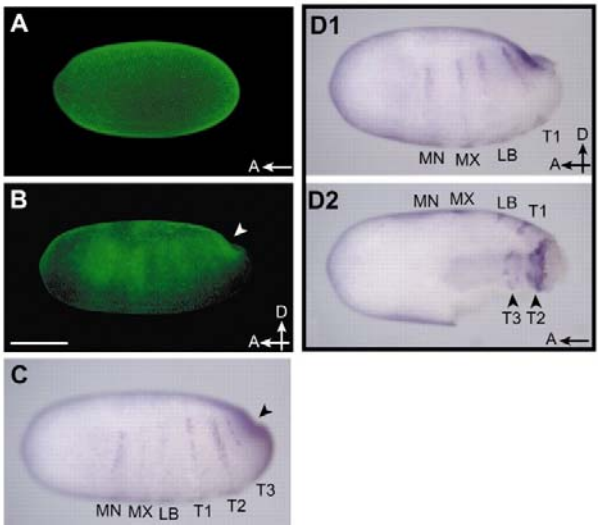

Figure 3.3 *Oncopeltus* embryogenesis and engrailed expression. (A) Early blastoderm at 24-28 hours. (B) Late blastoderm at 36-40 hours. Arrowhead indicates site of germ band invagination. (C) Late blastoderm at 36-40 hours hybridized with engrailed probe (purple color). (D1) Embryo stained for engrailed at slightly later stage than in C. Note that only the four anterior *en* stripes are now visible on the blastoderm surface. (D2) Same embryo as in D1, with `dorsal' region of blastoderm dissected away to reveal developing germ band and rotated to view dorsal aspect. (E-H) Dissected germ bands stained for engrailed, showing growth of the posterior and

the sequential addition of the abdominal segments. Anterior is towards the top. From Liu et al (Liu and Kaufman, 2004a)

Section 3.2 Annotation of genes of interest in *Oncopeltus fasciatus*

3.2.1 Background

Because of the advantages of *Oncopeltus* as a research model, described above, several labs are using *Oncopeltus* for a variety of different developmental and applied studies. For example, Alves et al. studied the immune response of *Oncopeltus* after infection with *Phytomonas serpens* (Alves e Silva et al., 2013); de Almeide et al. studied the role of an *Oncopeltus* Laminin-like protein in pest-host interactions (de Almeida Dias et al., 2012); Moore et al. studied ecological interactions of *Oncopeltus* under poor quality food feeding (Moore and Attisano, 2011); taking advantage of *Oncopeltus*'s unique phylogenetic position, Weisbrod was able to reconstruct the evolution of the terminal patterning system (Weisbrod et al., 2013); and, as summarized above (Section 3.1.3), many developmental studies have been carried out (Ben-David and Chipman, 2010; Chesebro et al., 2009; Erezyilmaz et al., 2009; Hughes and Kaufman, 2000; Liu and Patel, 2010).

A handful of genes have been characterized in *Oncopeltu*s to date. These genes were isolated by virtue of similarity to genes in other species by techniques such as RT-PCR, 5' and 3' RACE (Erezyilmaz et al., 2009; Liu and Patel, 2010), which are time consuming and labor intensive. Having sequence of the complete transcriptome and/or genome of *Oncopeltus* will allow researchers to more rapidly access specific gene sequences and to study gene regulatory networks. In 2011, Ewen-Campen et al. used 454 pyrosequencing technology to sequence mRNAs from *Oncopeltus* ovaries

and embryos (Ewen-Campen et al., 2011). They were able to identify 10,775 unique genes from their mRNA data. Those data provide a useful resource for molecular research. However, since those data were generated with 454 pyrosequencing, which can recover fewer genes and shorter assembled contigs than Illumina (Luo et al., 2012), several of the genes that were of interest for my studies were not found in the 454 data. For example, I was unable to identify orthologs of *E75A, Scr, Dfd , even-skipped and giant,* all of which had been isolated by other researchers using RT-PCR and RACE (Erezyilmaz et al., 2009; Hughes and Kaufman, 2000; Liu and Kaufman, 2005; Liu and Patel, 2010). I was also unable to find sequences corresponding to *paired, odd-paired, sloppy paired and runt* from the 454 transcriptome data. This suggested that the available RNA-seq data may be missing a lot of genes.

Fortunately, *Oncopeltus* was accepted by the i5k Insect and other Arthropod Genome Sequencing Initiative as one of the insect species whose genome would be sequenced at the early stages of the i5K project (i5k Consortium, 2013). As the genome sequence itself devoid of content, in order to make use of it, the genome must be annotated. The first draft of annotation of the *Oncopeltus* genome was carried out by a group of researchers from all over the world during the spring of 2014. Since our lab is interested in using *Oncopeltus* as a model organism for our studies, we participated in this effort.

As mentioned in the first chapter, Ftz-F1 is an orphan nuclear receptor involved in segmentation in *Drosophila*. In *Oncopeltus*, a different nuclear receptor functions as a pair-rule gene: *E75A*. Erezyilmaz et al. found *E75A* in *Oncopeltus* is expressed in a pattern resembling typical PRGs, and *E75A* RNAi caused loss of odd-numbered

parasegments (Erezyilmaz et al., 2009). These findings peaked my interest in the role nuclear receptors play in embryonic development. The nuclear receptor (NR) family of transcription factors is one of the largest transcription factor families, functioning in diverse biological processes such as homeostasis, proliferation, reproduction, development and disease (Chambon, 2005; Gurnell and Chatterjee, 2004; Mangelsdorf et al., 1995; Margolis et al., 2005). In insects and other arthropods, NRs play important roles in a regulatory cascade, initiated by ecdysone that controls metamorphosis (Hill et al., 2013; Rewitz et al., 2013). Founding NR family members were ligand-activated receptors characterized by a common structure and regulated by the binding of small molecules. The NR family was greatly expanded by the identification of a large number of orphan receptors for which no cognate ligand was known but which possess a typical NR structure (Nuclear Receptors Nomenclature, 1999). While ligands have since been found for many such orphan receptors, many others remain orphaned (Benoit et al., 2006; Escriva et al., 2000; Giguere, 1999; Ingraham and Redinbo, 2005; Mangelsdorf et al., 1995; Repa and Mangelsdorf, 2000; Shiau et al., 2001; Sluder et al., 1999). It has been proposed that the ancestral NR functioned in the absence of ligand, with the ability to bind ligand being acquired (Laudet, 1997) during evolution as a mechanism of modulating receptor activity in different biological contexts (Laudet, 1997). More detailed information can be found in Chapter2. Both ligand-activated and orphan NRs share a similar, modular protein structure comprised of a variable N-terminal region, a highly conserved zinc-finger DNA binding Domain (DBD), a second variable hinge region, and a less conserved Ligand Binding Domain (LBD) that includes the AF-2 domain (Figure 3.4).

Figure 3.4 Nuclear receptors share a common structure: An AF1 (Activation Function 1), DBD (DNA Binding Domain), Hinge, LBD (Ligand Binding Domain), and AF2 (Activation Function 2) located within LBD. Modified from Germain et al (Germain et al., 2006)

For ligand-regulated NRs, the role of ligand binding is to induce a conformational change that repositions the activation domain-2 (AF-2 domain), releasing co-repressors and creating a surface for co-activator binding, thus switching the NR between repressive and activating states (Glass and Rosenfeld, 2000) . Orphan NRs are activated by protein-protein interactions and post-translational modifications with, at least in several cases, the LBD folding in an active conformation in the absence of ligand (Tremblay et al., 1999b; Yoo et al., 2011a).

3.2.2 Methods

To identify NRs in the *Oncopeltus* genome, a combination of tblastn/blastp results, mRNA-seq data (Illumina short reads) and GENESCAN prediction was used. The only two exceptions were E75A, for which the annotation was based on a published cDNA (Erezyilmaz et al., 2009) and Ftz-F1, for which we have carried out 5' and 3' RACE to identify the full length sequence. For other cases, and when there was no or limited experimental data to rely on, annotation was based on sequence similarity to known members of the NR family (tblastn), and GENESCAN data were used as reference to determine both exon-intron boundaries and variable regions.

For the annotation of Runt, neither BLAST nor gene prediction methods revealed an mRNA sequence which includes the most important VWRPY motif (Aronson et al., 1997). To find the VWRPY motif, more than 25kbps of sequence was checked manually. A 50bp sequence, in which the VWRPY motif was found, was used to design a reverse primer. This reverse primer was then used together with a forward primer from the nearest upstream exon, which is identified based on homology. RT-PCR reactions were carried out with cDNA made from *Oncopeltus* embryos. The PCR product was isolated and sent out for sequencing

3.2.3 Results

As mentioned above, our lab participated in the group annotating the first draft of the *Oncopeltus* genome. We were assigned to annotate the nuclear receptor family genes. Since we are also interested in pair-rule genes, we voluntarily annotated some of the pair-rule genes in the *Oncopeltus* genome. Here I summarize results on the nuclear receptors. In the following sections, I combine my annotation and experimental results on *Oncopeltus* pair-rule orthologs.

In this first release of the *Oncopeltus* genome, I identified 16 nuclear receptors. This compares to 18 nuclear receptors in *Drosophila*, the insect in which these genes have been best characterized (King-Jones and Thummel, 2005). The two *Drosophila* NRs not found in the *Oncopeltus* genome are ERR and HR83. The details of each annotated NRs are listed in Table3.1. In the table, "NRNC Nomenclature" refers to the official gene name given by NRNC, (Nuclear Receptors Nomenclature, 1999). Under "Location in the Genome", the scaffold, where a NR was found, is indicated. "Number of exon" indicates how many exon each gene has in the annotation;

“Protein Length” shows the number of amino acids encoded by each predicted gene. “Experimental data support” shown if any experiment data is available for each gene; “DBD/LBD, % identity to Dm ortholog” indicates the identity of DBD/LBD of each gene when compared to the orthologs in *Drosophila melanogaster* .

Nuclear receptor	NRNC Nomenclature	Experimental data support	Location in the Genome	Number of exons	Protein Length	DBD/LBD, % identity to Dm ortholog
E-75A	NR1D3	5' and 3' RACE (Erezyilmaz et al., 2009)	Scaffold 434, 3rd exon in Scaffold 8681	6	690	97/63
E78	NR1E1	Illumina mRNA-seq	Scaffold 97	4	245	98/45
HR3	NR1F4	Illumina mRNA-seq	Scaffold 9	7	318	94/57
EcR	NR1H1	Illumina mRNA-seq	scaffold 2309	9	452	87/68
HR96	NR1J1	Illumina mRNA-seq	Scaffold 221	7	402	74/41
HNF4	NR2A4	Illumina mRNA-seq	Scaffold 1726	6	267	98/53
USP	NR2B4	Illumina mRNA-seq	Scaffold 14	8	401	90/38
HR78	NR2D1	Illumina mRNA-seq	DBD in Scaffold 2485, LBD in Scaffold 4234	10	499	92/33
Tailless	NR2E2	None	Scaffold 416, fifth exon located in 16304	6	342	75/36
Dissatisfacion	NR2E4	None	Scaffold 21	4	306	91/66
HR51	NR2E3	None	Scaffold 1675	8	490	62/61
Seven up	NR2F3	Illumina mRNA-seq	Scaffold 908	5	412	100/94
HR38	NR4A4	Illumina mRNA-seq	Scaffold 2708, part of LBD located in scaffold 7911	6	392	96/*
FTZ-F1	NR5A3	5' and 3' RACE *	Scaffold191, 1st exon in 848, 8th in 13623	8	564	98/74
HR39	NR5B1	Illumina mRNA-seq	Scaffold 422	13	588	90/74
HR4	NR6A2	Illumina mRNA-seq	Scaffold 327	11	466	95/48

Table 3.1 List of identified *Oncepeltus* NRs. * Isolation of *Of-ftz-f1* is described in Section 3.3

I found that most of the NRs have very long introns. Thus, for almost all the NRs I annotated, the coding regions were much larger than those generated by the automated genome annotation program MAKER (Cantarel et al., 2008). (For example, see Figure 3.5).

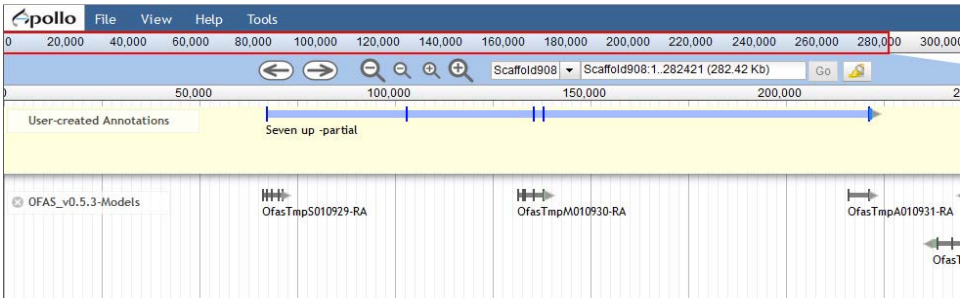

Figure 3.5 Seven up annotation and gene models from gene prediction. Seven-up gene covers a genomic region of 156Kbps, with 5 exons (Blue blocks). Three gene models were predicted in this region and each of them is much shorter (Light blue

blocks, named as OfasTmpS010929-RA, OfasTmpS010930-RA and OfasTmpS010931-RA).

Of the 16 nuclear receptors that were found in the *Oncopeltus* genome, Seven Up retains the highest degree of similarity to its *Drosophila* ortholog, with its DBD 100 percent identical to that of the *Drosophila*, and its LBD is 94 percent identical to that of the *Drosophila*. In scaffold 1872, there is a processed NR pseudogene, most likely an ortholog of HNF4. This psuedogene would encode a protein 470aa; the DBD of the predicted protein shares 79% identity with that of *Drosophila* HNF4, and the LBD shares 60% identity with that of *Drosophila* HNF4. For the 16 NRs identified, all appear to be present in single copy in the *Oncopeltus* genome. This is similar to the situation in other insects, including *Drosophila, Anopheles, Aedes, Tribolium* and *Apis.* Finally, no novel NRs were identified in *Oncopeltus* that are not also found in *Drosophila* and other insects.

In addition to these NRs, I identified three pair-rule genes in the *Oncopeltus* genome: *ftz, runt* and *hairy* (see section 3.3 in this chapter).

Two insulin-like peptides (ILPs) were also been found and annotated in the genome. Finally, I identified *buttonless*, which is a homeobox gene reported to be missing in the *Oncopeltus* genome by other annotator (personal communication).

3.2.4 Discussion

Next generation sequencing provides us with a powerful tool for genome sequencing. It dramatically increased the speed and throughput capacities of genome sequencing, while annotation a genome is time consuming and many annotated genomes contain a lot of errors (Klimke et al., 2011; Salzberg, 2007). From my own

experience, it is clear that combining all available data and methods in annotation improved the quality of the annotation. Of the 16 NRs we annotated, four of them were also annotated by other annotators. For three out of those four annotations, our annotations included more exons. And our annotations are the ones that accepted as the only annotation for those three genes in the final version of annotation. For one out of those four annotations, our result is in agreement with the annotation from other annotator.

The i5K project also provided RNA-seq data from juvenile and adult *Oncopeltus*. I found that the RNA-seq data, when available, are very helpful in determining intron-exon boundaries. As shown in Figure 3.6 A, at one position (red dotted line indicated by red arrow) all the Illumina reads (blue sticks) stopped, most of them have big gaps (light blue) when aligned to the genome. This is strong evidence there is an intron-exon boundary. However, the RNA-seq data were generally not useful in determining the 5' and 3' ends of mRNAs, as is doesn't show unanimous cutoffs at the ends. An example of this is shown in Figure 3.6 B.

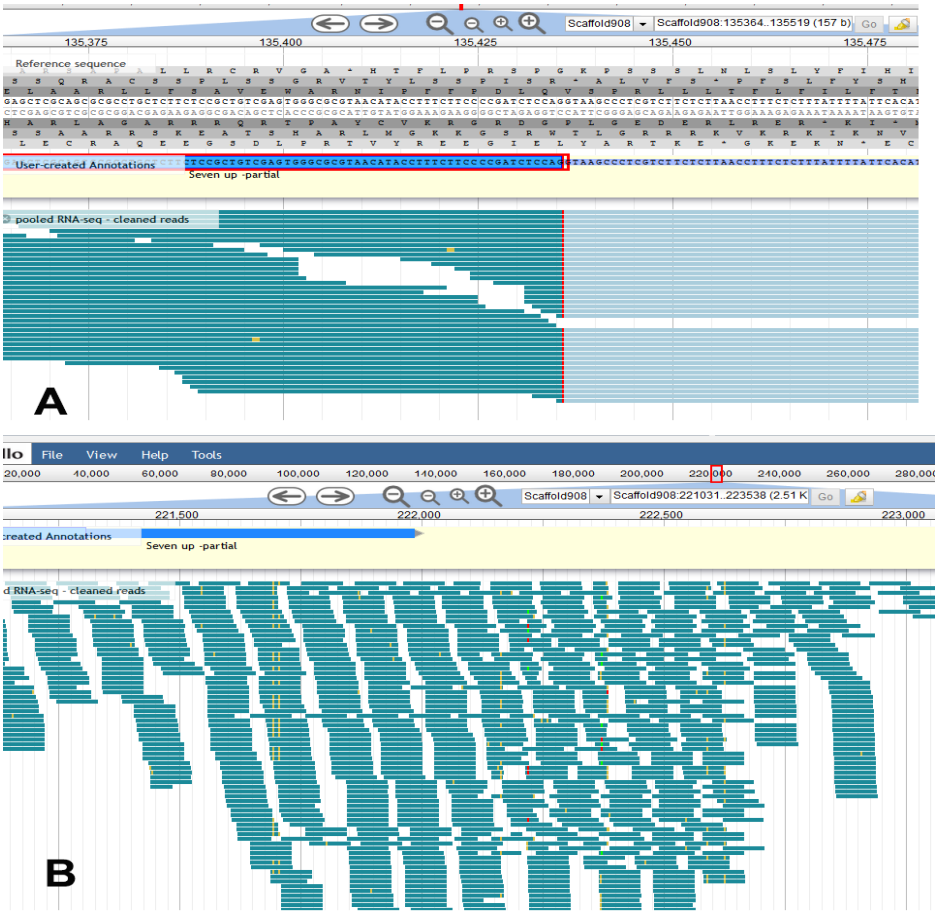

Figure 3.6 RNA-seq A screenshot of the gene "seven up" region. (A) A zoom in view of the third exon. (B) Part of the 3' end of seven-up gene annotation, the RNA-seq data spread in a large region without a unanimous termination site.

When mRNA-seq data are not available, annotation of a gene is mainly dependent on homology. The sequences of the DBD and LBD are highly conserved both in terms of length and sequence for family members across many species and could be predicted reliably. However, since the N-terminal region and the hinge region are variable in terms of both sequence and length, these regions can't be predicted by homology and were annotated using GENESCAN data, which is less accurate. Thus, the sizes of the NRs predicted here may be significantly different from the true genes.

Section 3.3 Orthologs of Drosophila *pair-rule genes in* Oncopeltus fasciatus

3.3.1 Background

As I discussed in Chapter one, most comparative studies of the pair-rule genes (PRGs) were done in holometablous insects. To better understand how PRGs have changed during the radiation of insects, more work needs to be done outside of the holometablous insects. As I also discussed in the section 3.1, *Oncopeltus fasciatus* is located in a unique phylogenetic position, within the Hemipteroid Assemblage, the sister group to the holometabolous insects. In addition, many molecular tools have been developed for this species. These features make *Oncopeltus* a good choice for studies of PRG expression and function. The only PRG studied in this clade when I began my work was *even-skipped*. I set out to identify other embryonic regulatory genes in *Oncopeltus* and to study the function of a selected group of PRG orthologs in this species.

3.3.2 Methods

Insect husbandry and embryo collection

Laboratory colonies of *Oncopeltus fasciatus* were bought from Carolina Biological Supply Company (Burlington, North Carolina) and reared according to the provider's recommendation. Briefly, *Oncopeltus* were kept in transparent plastic cages (14.5 x 8.5 x 10 inches), and organic raw sunflower seeds and water saturated cotton balls were provided as food and water sources. The cages were changed every week or two, depending on the cleanliness of the cages. The cages were kept in the laboratory with a temperature of ~25°C, without extra efforts to control humidity or light cycle. Dry cotton balls were placed in the cages for egg collecting. The eggs were collected

every four hours from the dry cotton balls with brushes. Once collected, the eggs were kept in an incubator with a temperature of 25°C and 16 hour light: 8 hour dark photoperiod, before fixation at the appropriate stage.

Isolation of genes of interest

For isolation of *Of-ftz,* two forward primers and one reverse primer were used to amplify the homeobox. The two forward primers used are ftzdegenS1 5'AAR CGN WSN CGN CAR ACN TAY TCN, and ftzdegenS2 5'AAR CGN WSN CGN CAR ACN TAY AGY. Both are degenerate primers corresponding to the conserved N-terminal of *ftz* homeodomain sequence KRS(/T)RQTYS(/T); the reverse primer used was designed by Heffer (Heffer and Pick, 2011). Three reverse primers for 5' RACE and two forward primers for 3' RACE were designed and used to isolate the 5' and 3' end sequences of *Of-ftz,* along with primers provided by the kit (RLM-RACE Kit, Life Technologies). The sequences of the three reverse primers are offtzp1 5'-TAC GTC TGT CTC TTC CGC TTC G-3', offtzp2 5'-GGA GAT ATC TTG TCA GGC GGA AT-3' and offtzp3 5'-CGC CAG GTC TAT TCT ACG TTT CC-3'. The sequence of the two forward primers are offtz3in 5'-TCT CCC GAG GAA ACG TAG AA-3, offtz3out 5'-GAA AAG GAA TTC CGC CTG AC-3'. For identification of *Of-ftz* sequences from Roche 454 sequencing data (Ewen-Campen et al., 2011), raw RNA-seq data were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/sra?term=SRP002610). The downloaded "sra" format data were converted into "fastq" format using the SRA toolkit provided by NCBI. Fastq data were then converted into "fasta" format with python v2.7. Local BLAST

was carried out using the BLAST+ package (Altschul et al., 1990) with the *Of-ftz* homeobox sequence as subject.

For isolation of *Of-ftz-f1,* sequence of the LBD was isolated using degenerate primers corresponding to the conserved region of Ftz-F1 and designed by Heffer (Heffer and Pick, 2011; Heffer, 2012). Primers used for 5' RACE are offf5inNew 5'- CCT CTG GGT CTT GTC GAT GT -3' and offf5outNew 5'- ACT TTT GAA AGC GGC AGA AA -3'. Primers used for 3' RACE are offf3inNew 5'- CTC TTA TCC CTC GGC CTT CT -3', offf3outNew 5'- ATG GTC GGA TAT GCT GGT TC -3'.

For isolation of *Of-hairy, Dm*-Hairy protein sequence was used to BLAST the *Oncopeluts* genome. Scaffolds that had BLAST hits were further analyzed. Briefly, upstream and downstream sequences surrounding the BLAST hit were used to do BLASTX and GENESCAN analysis. Possible protein sequences were used for reciprocal BLAST searches in order to determine the true ortholog. The scaffold which contained the most probable ortholog was compared to data provided by i5K, including RNA-seq data (pooled Illumina and 454 data), BLASTX-Arthropoda data and protein2genome-Arthropoda data. Final annotation was based on all these comparisons.

Embryo Fixation

Embryos were brushed off cotton balls and put into 2 mL eppendorf tubes, with no more than 50 microliters of embryos per tube. The fixation protocol was based on a protocol kindly shared by Dr. Chipman's lab (Ben-David and Chipman, 2010). In brief, 600ul tap water was added to each tube of embryos which was placed in boiling

water for three minutes and then placed on ice for six minutes. After the water was removed, 600ul of heptane and 600ul 4% PFA in PBS were added. Gentle shaking brought the embryos to the interface. Tubes were shaken vigorously on a Vortex mixer for 20 minutes. After shaking, the heptane and PFA were removed and the embryos were rinsed once with heptane, then twice with methanol. The embryos together with methanol were then put into wells on depression concave slides. The embryos were monitored under a dissection microscope. Within about one minute, the eggshells began to pop open. Within 30 seconds most of the eggshells had an opening. For the ones that don't have an opening, the unopened shells were poked and peeled with forceps to create a big hole on the shell. The embryos were then transferred to 75%, 50%, and 25% methanol in PBST for 2-3 minutes each and then to PBST for 2-3 minutes. The rehydrated embryos were fixed with 4% PFA in PBST for 90 minutes on a nutator. The fixed embryos were then washed three times with methanol and stored in methanol at -20 °C for future use.

Whole mount in situ hybridization

Step one**:** The *in situ* hybridization protocol was modified from that previously used for *Oncopeltus* (Ben-David and Chipman, 2010). The fixed embryos were removed from -20 °C and passed through 75%, 50% and 25% methanol/PBST for rehydration. After rehydration, the embryos were pre-hybridized with hybridization buffer (50% formamide, 5XSSC, 0.1% Tween-20, 50µg/ml yeast tRNA, 5% Dextran, 50µg/ml heparin) at 55°C for 1-4 hours. After pre-hybridization, probe was added to the embryos (1ul probe per 100ul hybridization buffer), and incubated at 55 °C

overnight (16-18 hours). For Step two, the protocol differs for blastoderm and germ band stage embryos.

Step two: For the germ band stage embryos, the next day the probe was removed and the embryos were washed twice with hybridization buffer at 55 °C for 30 minutes each, followed by one wash with 2XSSC at 55 °C, one wash with 2XSSC at room temperature for 30 minutes, and one wash in 0.2XSSC at room temperature for 30 minutes. Follow step three from here.

For blastoderm stage embryos, the next day, after removing the probe, the embryos were washed once with hybridization buffer at 55 °C for 30 minutes, and once with hybridization buffer at room temperature for 30 minutes followed by two washes with 2XSSC at room temperature for 30 minutes each and one wash in 0.2XSSC at room temperature for 30 minutes. Follow step three from here.

Step Three: The embryos were rinsed three times with PBST. After rinsing, the embryos were incubated with 10% sheep serum in PBST for 1-4 hours at room temperature to block non-specific binding. Embryos were next incubated with anti-DIG-AP antibody (1:1500, Roche) at 4 °C overnight. After incubation, the antibody was removed and the embryos were washed five times in PBST for 20 minutes each. For detection, the embryos were washed with staining buffer (100mM NaCl, 50 mM MgCl2, 100mM Tris pH 9.5, 0.1% Tween 20) three times for five minutes each. Staining was carried out using NBT/BCIP solution (Roche) diluted in staining buffer. The color reaction usually takes about two hours, so after one and a half hours, the embryos were checked under a dissection microscope every ten minutes until the desired color had developed. The color reaction was stopped by adding PBST to the

staining solution. The embryos were then washed with PBST three times for five minutes each to remove the staining solution from the embryos. The embryos were washed with 50% methanol in PBST for five minutes and 100% methanol for another five minutes. The embryos were then treated with 100% ethanol for 30 – 120 minutes. The embryos were washed with 50% methanol in PBST for five minutes, and washed with PBST three times for five minutes each. Blastoderm stage embryos were transferred to depression concave slides and germ band stage embryos were dissected away from the yolk and mounted on to slides in 90% glycerol. Photographs were taken under a dissection microscope (Leica M420, 16-20X).

Double-Strand RNA (dsRNA) preparation

Primers were designed to amplify 200 -400 bp of *Oncopeltus* genes of interest with T7 promoter sequences added to the 5' end of both forward and reverse primers. The primer sequences are listed in the Appendices. PCR was carried out with cDNA that had been made from 0-7 day old embryos using the manufacturer's recommended standard conditions (Reverse Transcription system, Promega). The PCR products were separated on an agarose gel and sent out for sequencing (Genewiz) to confirm that the correct gene was amplified. The purified PCR product (Gel Extraction Kit, Qiagen) was use as a template for *in vitro* transcription using MEGAscript® T7 transcription (Life Technologies) Kit following the manufacturer's recommendations. The final product was treated with DNase from the transcription kit to get rid of the DNA template. In order to anneal the in vitro transcription products' single stranded RNAs, in vitro transcription products were heated to 94°C for five minutes and slowly cooled by decreasing temperature 0.8°C every minute, until 45°C was reached,

in a PCR machine (TPersonal, Biometra). The annealed double-strand RNA was precipitated with 1/10 volume of 3M sodium acetate (pH, 5.2), 2X volume of ethanol, and then dissolved in 10-20ul injection buffer (0.1mM NaH_2PO_4, 5mM KCl, pH 6.8), and stored in -20°C. The concentration of double-strand RNA was measured with a spectrophotometer.

RT-PCR

In order to compare expression levels of *Of-ftz* between eggs laid by dsRNA injected females and eggs laid by uninjected females, total RNA was extracted from 4-72 hours eggs (about 100ul in volume) AEL with TRIzol® reagent (Life Technologies), following standard protocols. For all samples, 1ug of RNA was used to make cDNA with a QuantiTect Rev. Transcription Kit (Qiagen). RT-PCR was carried out with primers designed to amply a 203bp *Of-ftz* sequence downstream of the homeobox .

In order to compare expression levels of *Of-ftz-f1* and *Of-runt* in the ovaries from dsRNA inject females and that from uninjected females, ovaries were dissected out from injected females one to two weeks after injection. Ovaries were also dissected out from two to three weeks old uninjected females. Primers to amply a 217bp region between Orange domain and VWRPY of *Of-runt* and 318bp variable region between DBD and LBD of *Of-ftz-f1* were designed.

For relative gene expression comparison, *Of-actin* 4 was isolated by using TBLASTN to BLAST the mRNA-seq data published by Ewen-Campen (Ewen-Campen et al., 2011). *Of-actin* primers was designed to amplify a 179bp conserved

region. All primer sequences and *Of-actin* 4 can be found in the Appendices. A separate round of RT-PCR was carried out using the same cDNA from the above and *Of-actin* 4 primers.

5ul PCR products from *Of-actin 4* and 10ul PCR products from *Of-ftz* , *Of-ftz-f1* and *Of-runt* were loaded onto a 2% agarose gel and electrophoresis was carried out.

dsRNA injection

For parental dsRNA injection, modifications of the protocol of Chesebro et al. were used (Chesebro et al., 2009). Briefly, newly hatched females were separated from the main cage, and put into a new cage with about equal numbers of 2-3 week old adult males. After five to seven days, the females were picked out and used for injection. I also used virgin females for *Scr* dsRNA injection,. No differences in RNAi effects were observed between virgin and pre-mated females (data not shown). Each female was injected with about 5ul of dsRNA at a concentration of 2ug/ul into the abdomen. After injection, each female was kept in a single 150mm petri dish, with a water saturated cotton ball and sunflower seeds as water and food source. About 24 hours after injection, one male *Oncopeltus* was added to each petri dish, and a dry cotton ball was added at the same time for egg collecting. The dry cotton ball was checked twice a day for eggs. Once eggs were found on the cotton ball, that cotton ball was removed together with the eggs and a new cotton ball was added. The eggs were kept in a new petri dish in an incubator with a temperature of 25° C and 16 hour light: 8 hour dark photoperiod.

For embryonic dsRNA injection, embryos younger than 4 hours after egg lay (AEL) were used. The embryos were first aligned along a piece of Scotch® Double Sided Tape on a slide. dsRNA solution with 1/10 volume of food grade green dye was back loaded into glass needles. Injection was carried out with pressures between 60 and 80 psi using a Pneumatic PicoPump PV-820 (World Precise Instrument). After injection the embryos were kept in a humid petri dish at 25°C.

For both parental RNAi and embryonic RNAi, 1/10 volume of food grade green dye dissolved in injection buffer was used as a control.

3.3.4 Results

A putative Oncopeltus orthologs Of-ftz has unusual sequence features and may not have a function in embryonic development

As mentioned above, the *ftz* gene has changed during the radiation of insects from an ancestral homeotic gene to a pair-rule segmentation gene in *Drosophila*. In order to determine the role of *ftz* in an outgroup of the holometabolous insects, the gene was isolated from the genome of *Oncopeltus*. Using primers pairs designed by Heffer (Heffer et al., 2010), I first tried to amplify the homeobox of *Of-ftz*. These primers had been used successfully to isolate *ftz* orthologs from several other species, including the more distant crustacean *Artemia salina* (Heffer et al., 2010). I tried many times with different PCR conditions, different DNA templates and different primers combinations. However, I still could not get a PCR product, although I was able to amplify other genes such as *Of-ftz-f1* (see below). I then re-designed the forward primer. By aligning Ftz sequences from a few species, I found that the forward primer designed by Heffer corresponded to amino acids 2 to 8 at the N-

terminal arm of the homeodomain. The amino acid at position 9 is also conserved, usually either a threonine or a serine. I decided to extend the forward primer at the 3' end to include codons for position 9. By extending the forward primer at 3' end, I increased the possibility that my primer would bind to corresponding genome sequences even if there were mismatch(es) between my primer and genome sequences. In order to decrease the degree of primer degeneracy, two primers were designed, with one corresponding to a threonine at position 9, the other corresponding to a serine at position 9. These two new forward primers were then used along with the same reverse primer designed by Heffer. Using forward primer ftzdegenS2 and reverse primer ftzDEGEN6 I still could not get a PCR product. Using forward primer ftzdegenS1 and ftzDEGEN6 as reverse primer, I amplified a PCR product of approximately 160bp. After sequencing, it was confirmed to be a sequence of a homeobox. With this sequence information, 5' RACE and 3' RACE were done to isolate full length *Of-ftz* mRNA from 0-6 days embryos. The 3'RACE product included the remaining 48 nucleotides of the homeobox and sequence downstream of homeobox including a 3' UTR. The 5'RACE product included an additional ~ 100bp upstream of the homeobox. Combining all these data, a gene of a total of 443 nucleotides was revealed (Appendices). The homeobox of the gene could encode a full length homeodomain. Reciprocal BLAST and comparison to Ftz homeodomains from other species indicate it is an ortholog of Ftz. Specifically, amino acids in homeodomain positions 1, 4, 6, 7, 37 and 56, considered to be diagnostic amino acids (Telford, 2000), suggested it is a Ftz ortholog. *Of*-Ftz matches five diagnostic amino acids in other Ftz orthologs. There is one more diagnostic amino acid located in HD

at position 4. At this position, all previously reported Ftz orthologs have either a serine (S) or a threonine (T). However, the *Of-Ftz* I isolated, encodes a HD protein with a lysine (K) at position 4. Amino acids in HD at position 1, 4, 6 and 7 have been found to determine the functional specificity of homeotic proteins (Furukubo-Tokunaga et al., 1993). Serine and threonine have a neutral side chain, lysine has a positively charge side chain, a change from serine/threonine to lysine may have a big impact on the function of the protein. In order to better find potential amino acid variations in *Of*-Ftz, I collected four more Ftz sequences from four hemipteroid assemblage insects, and aligned them with MUSCLE. In order to compare the Ftz HD variation I saw in hemipteroid assemblage insects, I also arbitrarily picked five Ftz HDs from five different orders，ranging from diptera to crustacea. These alignments are shown in Figure 3.7. As shown in Figure 3.6 A, All other Ftz orthologs from hemipteroid assemblage have conserved amino acid at position 4, 6, and 7, except *Oncopeltus* Ftz, which has a lysine instead of serine or threonine. The C terminal part of the HDs are conserved as indicated by overwhelming blue amino acids.

In addition, for these Hemipteroid assemblage insect Ftz proteins, none of them have an LXXLL motif (data not shown). Figure 3.7 B shows Ftz homeodomains alignment from five species in five order, including four insects and one crustacean. The columns that have BLOSUM62 scores (Henikoff and Henikoff, 1992) higher than three are labeled in blue, scores higher than 0.2 was indicated by light gray, scores lower than 0.2 are not colored. Comparing Figure 3.7 A and B, I noticed that the Ftz proteins from the Hemipteroid Assemblage got lower BLOSUM62 scores

(more grey and un-colored amino acids) than Ftz proteins from different orders. This means that the Ftz homeodomains in hemipteroid assemblage insects have more un-conserved variations than those in other orders. Within hemipteroid assemblage insects, *Ap*-Ftz and *Of*-Ftz have the least conversed homeodomains.

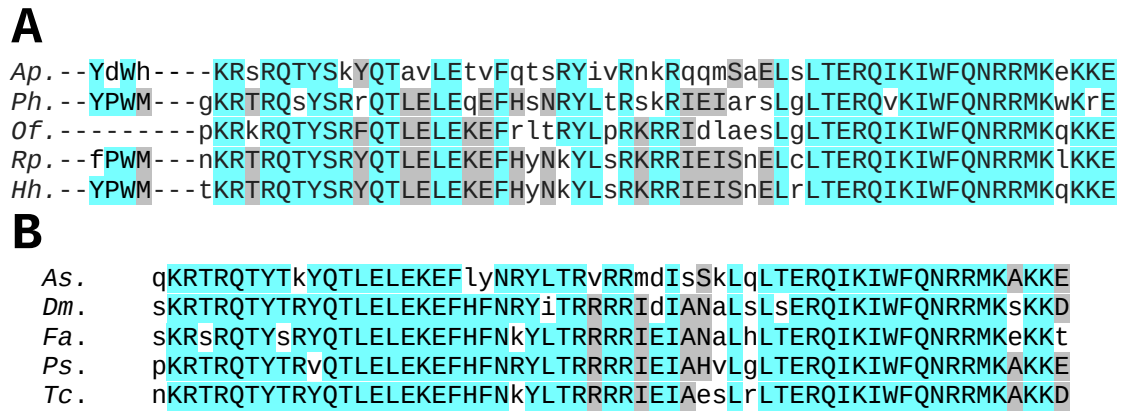

Figure 3.7 Alignment of putative Ftz proteins. (A) Ftz homeodomains and YPWM motifs from five species in the hemipteroid assemblage. The YPWM motifs vary in the lineage. (B) Ftz homeodomains from five species in five different orders different colors are used to indicate the average BLOSUM62 score. light blue>=3, light gray>=0.2, no color otherwise. *Ap. Acyrthosiphon pisum , Ph. Pediculus humanus corporis, Of. Oncopeltus faciatus , Rp. Rhodnius prolixus, Hh. Halyomorpha halys. As. Artemia salina, Dm. Drosophila melanogaster, Fa. Forficula auricularia, Ps. Pedetontus saltator, Tc. Tribolium castaneum*

Examination of the full length *Of-ftz* sequence (Fig. 3.8, *ftz* 5'RACE) revealed only 55 bp upstream of the homeobox. There is no start codon. Further, there is one stop codon in each reading frame upstream of the homeobox (Figure 3.8, red). This suggests that this *Of-ftz* isoform is not translated into protein. However, the sequence isolated encodes a full length homeodomain with no stop codons, which suggests that, if the protein were translated, it could have some function. To determine whether the sequence isolated was just a non-functional isoform of *Of-ftz*, I tried to isolate other possible isoforms. I designed three reverse primers for 5'RACE. Since I expect all potential *ftz* isoforms to share the same homeobox, all these three reverse primers are

located in the homeobox. Using two primers for each 5'RACE, I had three combinations of primer pairs. It is possible there is an isoform that is only expressed in a narrow time window. Narrowing down the time window during mRNA extraction may increase the relative concentration of that isoform, thus increasing the chances of isolating this isoform. I also noticed that *Oncopeltus* rarely lay eggs at night. Therefore, a 24 hour continuous collection contains few eggs laid during that night. I thus did eight hours collections during the day and extracted mRNAs every 8 hours, such that I have mRNAs from 8-16 hours, 16-24 hours, 24-32 hours, 40-48 hours, 48-56 hours, 64-72 hours, until 88-96 hours AEL. 1^{st}, 2^{nd} and 3^{rd} instar nymphs were put in one pool, and 4^{th} and 5^{th} instar nymphs in another pool for mRNA preparation. 5'RACE was carried out with these primers and mRNA combinations. Three to five different annealing temperatures during PCR amplification were also tried with different primer-mRNA combinations. All the mRNAs collected gave a short 5'RACE product (70-100bp depending on primer combinations), under some of the conditions. None of them gave a 5' RACE product of different length. This means that none of the experiments revealed a different isoform. I tried a Northern blot hoping to find a different isoform, but I didn't get a signal, probably due to low concentration of the RNA.

While I was thinking of trying some different approaches, Ewen-Campen kindly shared their *Oncopeltus* mRNA-seq data with us before publication (Ewen-Campen et al., 2011). The *ftz* gene Ewen-Campen sent me is 788 nucleotides long (named *ftz* E-C in figure 3.8) and is longer than the sequence I isolated by RACE. *ftz* E-C has a shorter 3' UTR and a longer 5' UTR than my sequence *ftz* 5'RACE. In the 5' UTR

region, *ftz* E-C includes 300 nucleotides not present in *ftz* 5'RACE. Intron/exon junctions GT and AT were found at the ends of this 300 nucleotide sequence (Figure 3.8, green). These features suggest that *ftz* E-C contains an unprocessed intron. In this putative intron, nine, seven and seven stop codons can be found in each reading frame respectively (data not shown). Both *ftz* 5' RACE and *ftz* E-C share identical homeobox sequences (Figure 3.8, highlighted in blue) and stop codons in each reading frame (Figure 3.8, red). Thus, even if the *Of-ftz* sequence in the RNA-seq data from Ewen-Campen is a processed mRNA, it may not be translated into protein.

Finally, since short sequences are more readily amplified than longer sequences in PCR reactions, I tried to clone the 5'RACE products without a second PCR amplification. A total of twenty colonies were sent out for sequencing: nine colonies have a sequence that is clearly part of *Of-ftz*. Five out of nine sequences matched *ftz* 5'RACE and four out of nine sequences matched *ftz* E-C. No other *ftz* isoforms and no additional or different sequence variants were found, while I see some nucleotides substitution in the putative intron region, none of these substitutions would result in an open reading frame in the putative intron region.

```
ftz 5'RACE ----------------TTTGAACCCACAGCGTAAGGATTGAGATCCCTTAGCCGCCCGAA----
ftz E-C    TCTTGCTCTGGGCGGGATTTGAACCCACAGCGTAAGGATTGAGATCCCTTAGCCGCCCGAAGTGA

                                                P  K  R  K  R  Q  T  Y  S  R  F
ftz 5'RACE(-------------)----ATCCAAGTAGGCCGAAGCGGAAGAGACAGACGTATTCTCGATTCC
ftz E-C    (292nts omitted)CCAGATCCAAGTAGGCCGAAGCGGAAGAGACAGACGTATTCTCGATTCC
```

Figure 3.8 Alignment of the *Of-ftz* sequences. Only the first part of the sequence is included in this alignment, since the only difference between these two isoforms is located in this part. Stop codons are labeled in red, stop codons in the 292 nucleotide(nts) omitted sequence are not shown. GT-AG is labeled in green. Partial Homeodomain and homeobox are highlighted in blue.

In order to determine whether *Of-ftz* is expressed in a pattern reminiscent of a *Drosophila* PRG, *in situ* hybridization was carried out using as probe complimentary to sequence 3' of the homeobox, including partial coding sequence and 3' UTR. In early blastoderm stage embryos, *Of-ftz* was expressed in two broad and diffuse stripes in the central region of the embryo (Figure 3.9 A). Later, as development progresses, the embryo anlagen moves posteriorly and internally towards the yolk. As these shifts occur, the two *Of-ftz* stripes moved towards the anterior of the embryos, and the posterior stripe became narrow and sharp (Figure 3.9 B). At germ band invagination, one broad, diffuse stripe was evident with two weaker, narrower stripes more posterior (Figure 3.9 C). At the germ band extension stage, the two weak and narrow stripes (Figure 3.9 D, green arrows) were barely visible outside of the embryos, while the broad stripe became more intense (Red arrow in Figure 3.9 D). As the germ band extended inside, these stripes moved toward the end of the embryo until they disappeared (Figure 3.9 D from top to bottom). No specific expression was detected in the germ band elongated embryos (Figure 3.9 E).

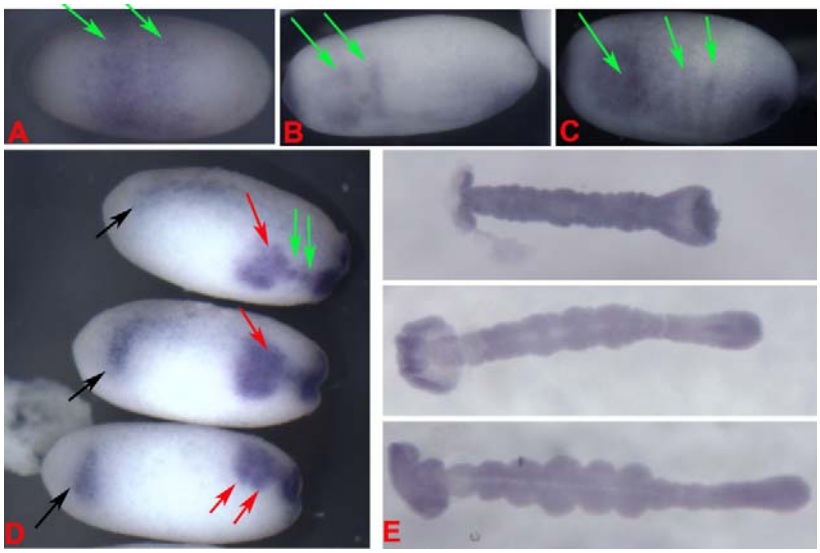

Figure 3.9 *Of-ftz* expression in embryos. (A) Early blastoderm embryo. *Of-ftz* is expressed in two broad stripes. Green arrows point to the stripes; (B) Later blastoderm embryo, stripes move towards anterior, narrower and sharper than early stripes; (C) Germ band invagination embryo, one broad stripe and two weak and narrow stripes can be seen. (D) Germ band extension embryos. Black arrows indicate germ bands beneath the yolk; two weak stripes barely can been seen (green arrows); as embryos growth, the broad stripe tends to split into two (red arrows); all the stripes move towards the posterior as embryos grow; (E) Germ band embryos were dissected out from the embryos shown in D, no staining can been seen. Embryos are aligned such that the anterior is to the left, and posterior is to the right.

To test the function of *Of-ftz*, parental RNAi using dsRNA targeting the 3' part of homeobox region (120bp) and part of sequence downstream of homeobox (118bp) was carried *Oncopeltus* females injected with this dsRNA laid eggs and the embryos died within two days after egg laying. Two more dsRNA targeting *Of-ftz* were made. One targets a 199bp region located in the region upstream of homeobox, within the putative intron. The other targets a 201bp region downstream of homeobox , including part of the CDS and 3'UTR. When I used these two dsRNAs for injection, eggs laid by injected females developed normally with no obvious phenotype. Since the 3' end part of the *Of-ftz* homeobox is very conserved, it is possible that a dsRNA targeting the homeobox would have off-target effects on other homeobox-containing

genes. Since the two dsRNAs that did not include the homeobox resulted in no observable phenotype, we suggest that *Of-ftz* does not have a function in *Oncopeltus* embryonic development.

RT-PCR suggests pRNAi knockdown Of-ftz expression in embryos

Since it is possible that the injection of *Of-ftz* dsRNA did not knockdown the expression of *Of-ftz*. I did RT-PCR with 4-72 hours AEL embryos laid by dsRNA injected females and uninjected females (Figure 3.10). In embryos laid by both injected and uninjected females, the expression level of *actin* are similar, while in the dsRNA injected embryos, *ftz* expression levels decreased dramatically.

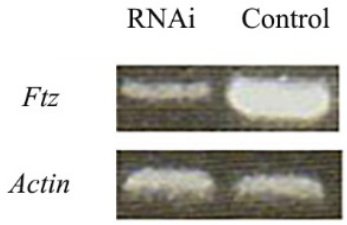

Figure 3.10 *Of-ftz* pRNAi knockdown expression of *Of-ftz* in 4-72 hours AEL embryos. Samples from *Of-ftz* dsRNA injected females are labeled as RNAi, samples from uninjected females are labeled as Control. *Actin* expression levels are similar. *ftz* expression level in RNAi sample decrease dramatically.

Sequence and segmental expression of Of-ftz-f1.

Full length *Of-Ftz-f1* was isolated and is 2081 nucleotides in length. It has a 1962 nucleotide Coding DNA Sequence (CDS), which encodes a protein that is 564 amino acids long (Appendices).

In early blastoderm embryos, *Of-ftz-f1* is expressed in a broad central region of the embryos with a diffuse stripe at the anterior (Figure 3.11 A). At the germ band invagination stage (Figure 3.11 B), six stripes can be seen (green arrows), with a

diffuse background expression in the anterior region. Since *Of-en* had been found to expressed in six stripes (Figure 3.2 C) at germ band invagination (Liu and Kaufman, 2004a), the six stripe expression pattern of *Of-ftz-f1* suggests that *Of-ftz-f1* is expressed in every segment at germ band invagination.

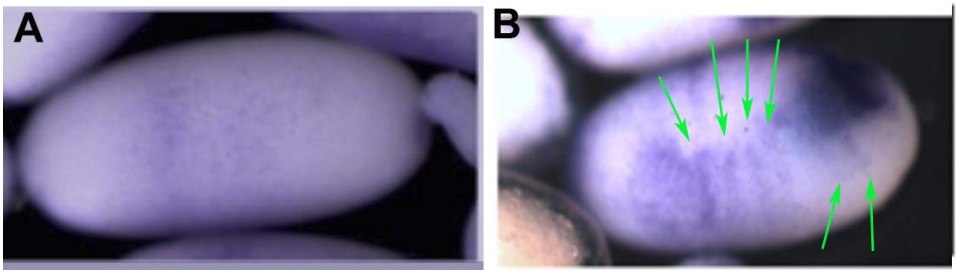

Figure 3.11 Expression of *Of-ftz-f1* at blastoderm stage. (A) At early blastoderm, *Of-ftz-f1* is expressed in a broad central region of the embryo. (B) At germ band invagination, six stripes can be seen (green arrows).

In germ band stage embryos, *Of-ftz-f1* is expressed in striped pattern (Figure 3.12). In early germ band stages, it is expressed in two stripes (Figure 3.12 A and B). Later only one stripe can be seen (Figure 3.12 C and D). From these experiments, it was not possible to determine in which segment(s) *Of-ftz-f1* is expressed. However, it seems that in the late germ band, *Of-ftz-f1* is only expressed in the growth zone, where new segments are generated.

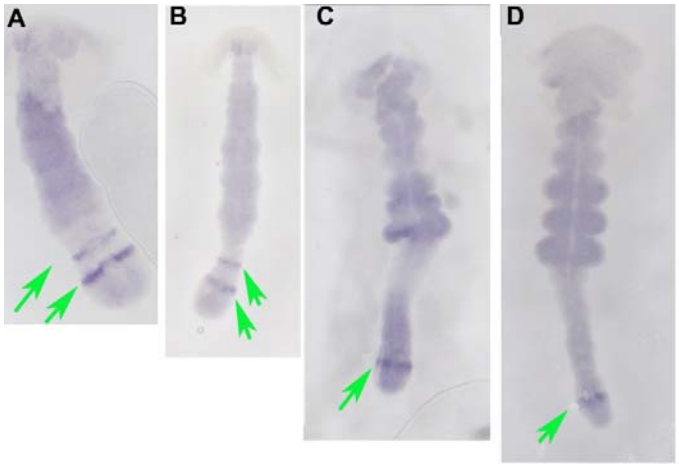

Figure 3.12 Expression of *Of-ftz-f1* during germ band elongation. (A) In 44-48 hour germ band stage embryos, two stripes can be seen. (B) Two stripe expression pattern persists as the germ band grows. (C) One stripe can be seen in 56-60 hour germ band embryo. (D) One stripe can be seen in a 72-76 hour germ band embryo. Green arrows point to stripe(s).

In order to determine the register of the *Of-ftz-f1* stripes, I carried out double *in situ* hybridization with *Of-en* and *Of-ftz-f1*. Since I was unable to work out a way to stain them with two different colors, I had to use one color to stain *Of-en* and *Of-ftz-f1*. I first did *in situ* hybridization with *Of-en* only to determine its expression pattern (Figure 3.13). In agreement with previous reports, *Of-en* was expressed in the posterior portions of each segment and was not expressed in the growth zone region (Liu and Kaufman, 2004a). As the germ band grows, more and more segments are added and *Of-en* stripes appear (Figure 3.13 A to D).

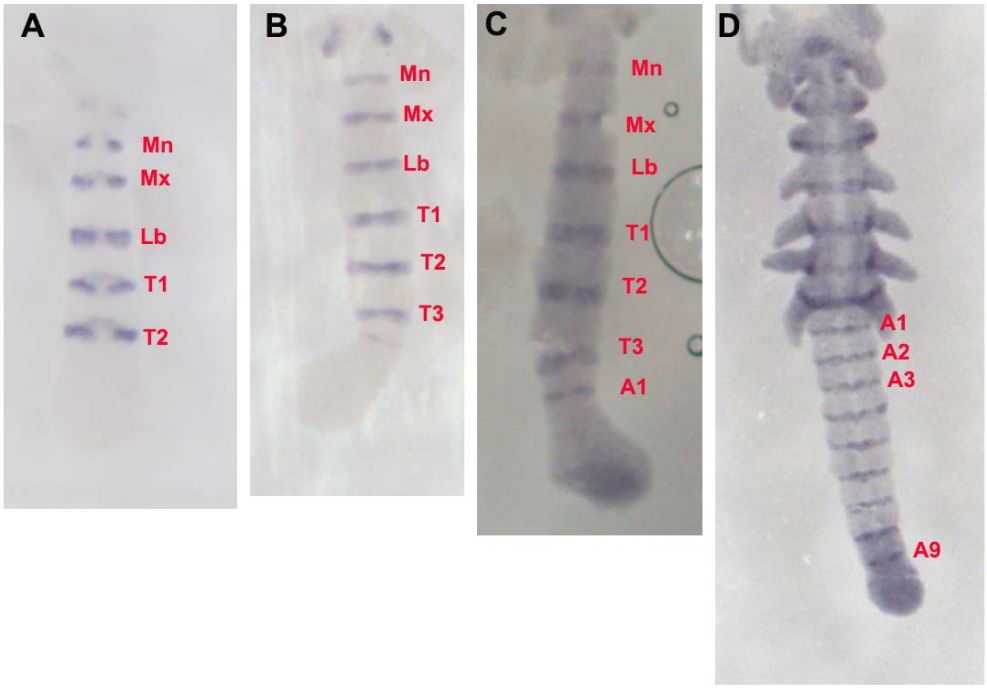

Figure 3.13 Expression of *Of-en* in the growing germ band. (A) *Of-en* is expressed in five stripes in 28-32 hour embryo. (B) Six stripes can be seen as embryos grow. (C) Seven strips can be seen in a 48-52 hour embryo. (D) A 72-76 hours embryo with nine abdominal segments. Abbreviations: T, Thoracic segment, A, Abdominal segment, Mn, Mandible, Mx, Maxillae, Lb, Labium

I then did *in situ* hybridization with both *Of-ftz-f1* and *Of-en* probes (Figure 3.14). The first *Of-ftz-f1* stripe was observed in the T3 segment (Figure 3.14 A, green arrow). Then a stripe appeared in A1 from growth zone as the T3 stripe moved anterior (Figure 3.14 B, green arrow). As the A1 stripe moves out of growth zone, the T3 stripe disappeared and a new A2 stripe emerged in growth zone. After the A2 stripe, all other posterior stripes (A3-A10) only appeared in the growth zone, leading to a one stripe pattern. as seen in Figure 3.11 C and D and Figure 3.14 D, E and F. The last *Of-ftz-f1* was observed when the *en* A9 stripe appeared. After the A10 stripe of *en* appeared, no expression of *Of-ftz-f1* was observed. These findings suggest that after the A5 segment forms, *Of-ftz-f1* stripes only appear in the newly formed

segments in the growth zone region, and they are seldom seen to leave the growth zone region.

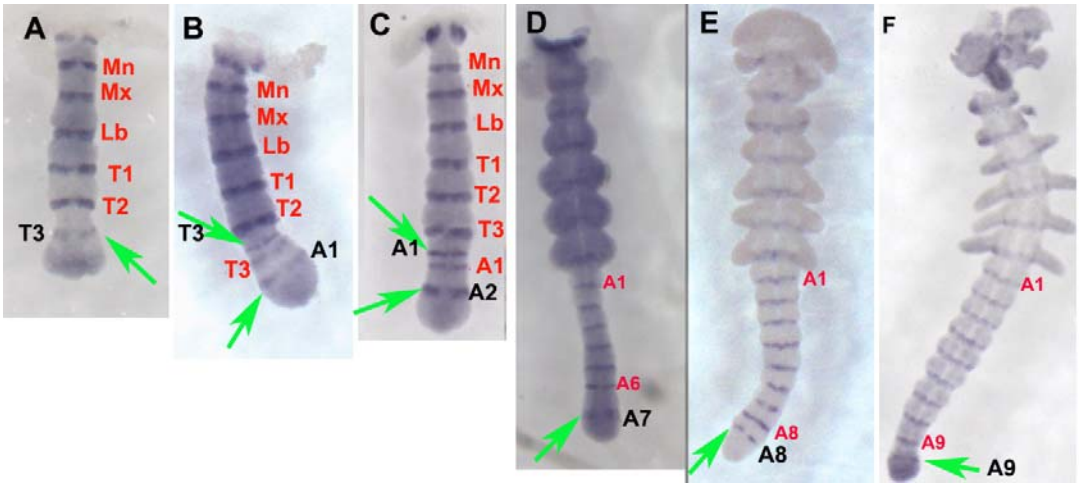

Figure 3.14 *Of-ftz-f1* and *Of-en* double *in situ* hybridization. (A) The first *Of-ftz-f1* stripe is seen in T3. (B) Later, an *Of-ftz-f1* A1 stripe emerges from growth zone and the T3 segment stripe moves out of growth zone. (C) *Of-ftz-f1* A2 stripe appears and the A1 stripe moves out of growth zone; the T3 stripe has disappeared. (D) *Of-ftz-f1* A7 stripe emerges in growth zone; A6 expression has disappeared. (E) *Of-ftz-f1* A9 stripe emerges in the growth zone. (F) *Of-ftz-f1* A10 stripe emerges in the growth zone, without A9 stripe. Segments expressing *Of-en* are labeled in red; segments expressing *Of-ftz-f1* are labeled in black. Green arrows point to *Of-ftz-f1* stripes. Abbreviations: T, Thoracic segment, A, Abdominal segment , Mn, Mandible, Mx, Maxillae, Lb, Labium.

Knockdown of Of-ftz-f1 with RNAi

To determine the function of *Of-ftz-f1*, RNAi was used. For parental RNAi, a total of 28 *Oncopeltus* one week old females were injected with *Of-ftz-f1* dsRNA. Two of them died within 24 hours, all others lived for 4-5 more weeks after injection. Of these 26 females, none of them laid any eggs in their whole life. I then tried to inject 20 females that were about two weeks old. These females had started laying eggs for one week without disruption. After injection they all stopped laying eggs. I then dissected their ovaries out and compared them with normal ovaries. As shown in

Figure 3.15 A, a normal ovary contains a few ovarioles. As the oocytes move from anterior to posterior, they grow in size gradually and oocytes of different in sizes can be seen. When they reach the posterior end of the ovariole, most of them are mature and have a yellow shiny color (Figure 3.15 A arrows). In the *Of-ftz-f1* dsRNA injected females, the oocytes stopped growth right before maturation and all the ovarioles contained oocytes of about the same in size (Figure 3.15B). No shiny, yellow mature oocytes were evident. It is clear that the oogenesis is blocked by *Of-ftz-f1* RNAi.

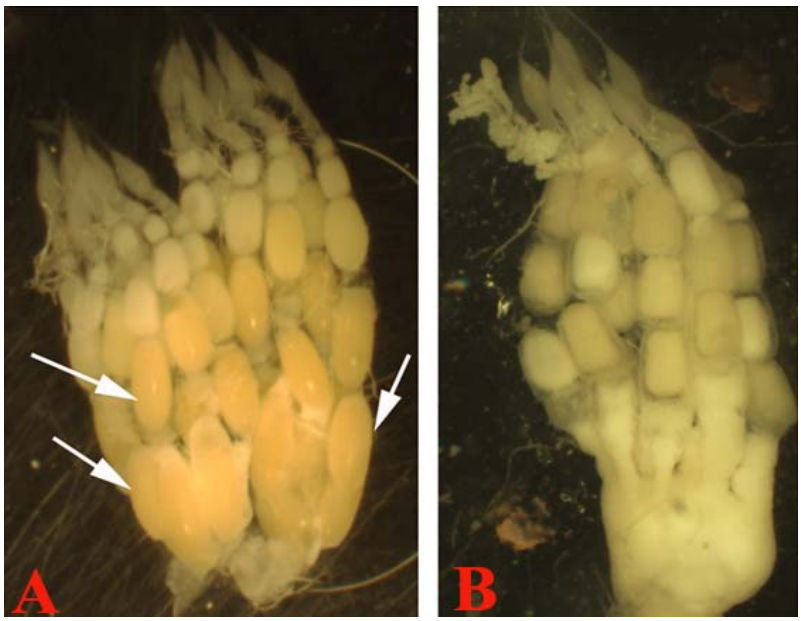

Figure 3.15 *Of-ftz-f1* RNAi blocks oogenesis. (A) Ovary from a normal female. Oocytes of different sizes can be seen, including mature oocytes (arrows). (B) Ovary from an *Of-ftz-f1* dsRNA injected female. Oocytes are similar in size to each other and no mature oocyte can be seen.

In order to test whether the oogenesis blocking phenotype is related to knockdown of *Of-ftz-f1* expression, RT-PCR was used to determine the relative concentration of *Of-ftz-f1* mRNA in ovaries from injected and uninjected females (Figure 3.16). In both injected and uninjected females, *actin* expression levels in were similar., while

in dsRNA injected females, Of-*ftz-f1* expression levels in ovaries decreases to barely detectable levels.

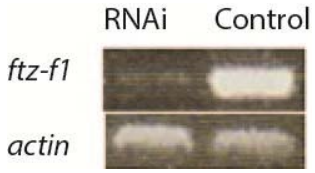

Figure 3.16 *Of-ftz-f1* pRNAi knockdown of *Of-ftz-f1* in ovary. Samples from *Of-ftz-f1* dsRNA injected females are labeled as RNAi, samples from uninjected females are labeled as Control. *Actin* expression levels were similar in both samples but *Of-ftz-f1* expression levels decreased in RNAi-treated ovaries.

Since we are interested in *Of-ftz-f1* function in embryonic development, I tried to injected the dsRNA into eggs (eRNAi). It is very hard to do injection with *Oncopeltus* eggs. There is no way that I could get rid of the chorion without killing the embryo. The chorion remained intact after treatment with undiluted bleach for 24 hours. I therefore had to do injection with the chorion on the embryos. Another property of the embryos is that the pressure inside of the chorion is very high, making the injection very difficult to do. I needed to increase the holding pressure to 20-30 PSI to prevent cytoplasm from flowing back into the injection needle, which caused the dsRNA solution flow out of the needle slowly before or after injection. The injection pressure is 60-80 PSI, which sometimes blew the embryos into pieces. Most of the solution injected into the embryos will be squeezed out due to high pressure inside of the embryos. I practiced injection for a few thousand embryos using injection buffer with 1/10 food grade green dye. During my practices the survival rate ranged from about 30% to 50% (survival here means the embryos developed into pre-hatch stage, when all the segments can be seen). In these embryos, I never saw an embryo with obvious segmentation defects (data not shown).

In one *Of-ftz-f1* dsRNA injection, 12 out of 12 embryos injected at the same time hatched with same phenotype. They had a head with eyes (Figure 3.17 B, red arrows pointed to eyes) and thoracic segments with three pair of legs (Figure 3.17 B, black arrows pointed to legs). The head and thoracic segments apparently did not fully develop. They did not have the reddish color that can be seen in normal embryos (Figure 3.17 A). Once put under a cover slide, the head and thoracic parts burst into pieces (Figure 3.17 B), which indicates a defect in cuticle development. Further, their abdomens were severely affected. Normal embryos have ten abdominal segments, while the RNAi treated embryos had only two or three segments (Figure 3.17 C and D). There is one macrochaete (Figure 3.17 A black arrows) on each side of every abdominal segment (Erezyilmaz et al., 2009). In the *Of-ftz-f1* RNAi embryos, only one macrochaete was found (Figure 3.17 C and D, black arrows). Judging from the color of the abdomen, there may be one more segment anterior to the segment with a macrochaete and one more segment posterior to the segment with a macrochaete. If this is ture, it means the other two segments are missing the anterior part of the segment, where the macrochaete located. This suggests *Of-ftz-f1* may have pair-rule function in some of the abdominal segments.

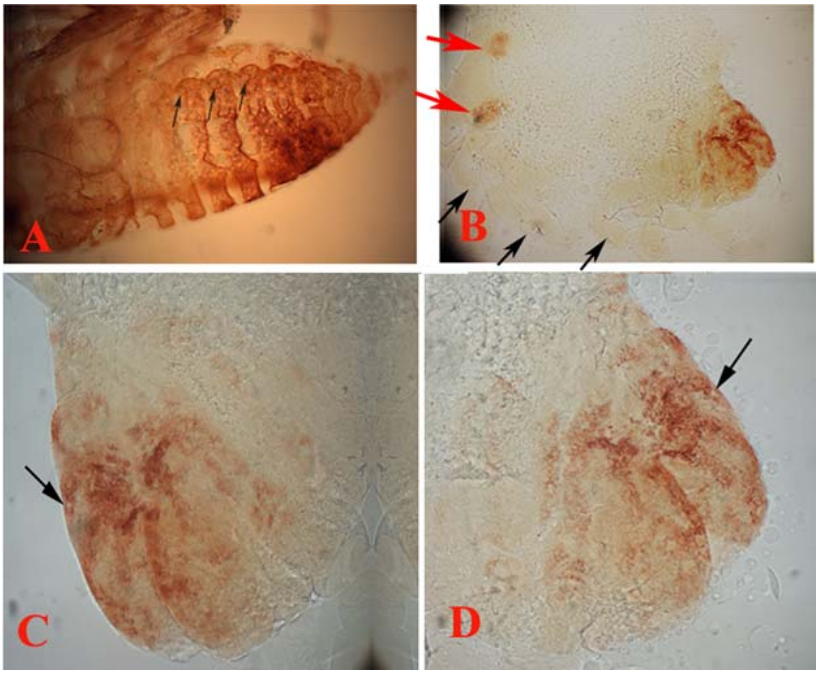

Figure 3.17 RNAi suggests that *Of-ftz-f1* functions in segmentation. (A) A normal embryo, showing mainly the abdomen of the embryo, with 10 abdominal segments. Black arrows point to macrochaetes; (B) An *Of-ftz-f1* RNAi embryo with head, thoracic and abdominal segments. Eyes are indicated by red arrows, leg are indicated by black arrows; (C) and (D) abdominal segment of two other RNAi-treated embryos; macrochaetes indicated by black arrow.

When I tried to repeat this experiment, I never got the same phenotype. Once, with injection of 512 embryos, 272 embryos developed to pre-hatched stage. These embryos were fixed and observed under microscope for phenotypes. Most of them had normal segments. Nine of them had segmentation defects. I got two embryos with exactly the same phenotype as described above; one embryos with only one pair of legs (Figure 3.18 A); four embryos with a shortened posterior part (Figure 3.18 B); one embryo had segmental fusions on the ventral side (Figure 3.18 C); and one embryo had a shrunken abdomen (Figure 3.18 D). For the one with one pair of legs, the identity of the legs (red arrows) could not be determined. The antenna and proboscis were present (yellow and white arrows in Figure 3.18 A). Since two

thoracic segments were deleted, this resembles a pair-rule phenotype. The cuticle of the embryo did not develop well, with-light color, and it was easy to break the body. The embryos with shortened posterior segments had either five or six normal abdominal segments, with one abnormal segment fused by a few segments (Figure 3.18 B). For the embryo with a ventral fusion, two segments were seen at the doral-lateral part of the abdomen, converging together on the ventral side, which indicates a partial deletion of these two segments only ventrally. Notice that all normal segments are dark-colored on the dorsal-lateral side, and light-colored on the ventral side. For the embryo with a shrunken abdomen, the segments next to the shrunken parts were dark colored and the light-colored part was lost. This indicates a total deletion of the ventral part of these two segments.

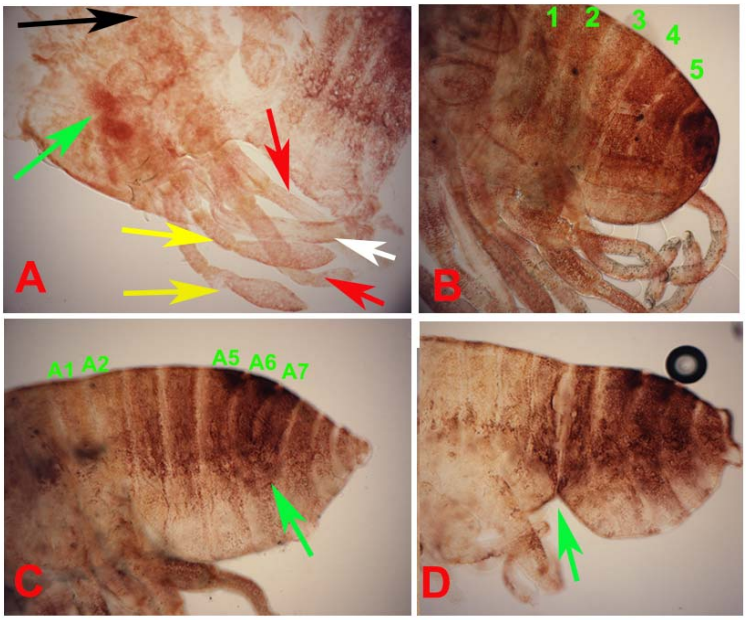

Figure 3.18 RNAi suggests that *Of-ftz-f1* functions in segmentation. (A) Embryo with one pair of legs. Green arrow, two eye dots; yellow arrows, antenna; white arrow, proboscis; red arrows, legs. (B) Embryo with posterior segmental defects. Numbers correspond to five normal abdominal segments, after the 5th segment, ~3 segments fused. (C) Embryo with partial segmental fusion. Ten

segments can be seen, some of which are labeled by numbers. A6 and A7 fused together in the ventral region, green arrow. (D) Embryo with a shrunken abdomen.

Isolation and function of additional PRG orthologs from Oncopeltus

Since in hemiptera insects, *Of-eve* is the only PRG whose expression and function had been studied, knowing the function of other PRG orthologs will help us better understand how PRG evolve in insects. I therefore tried to isolate all other *Oncopeltus* PRG orthologs with the information I got from the i5K project. Some of the PRGs belong to big superfamilies with many homolgues having similar sequences. For example, Prd homolgues are too similar to each other in their conserved domains, and diagnostic motifs were not found in the genome. In addition, the available genome sequence is only a first draft and there were genome quality issues that may have prevented identification of some PRGs. Therefore, I was able to identify only two more PRG orthologs, *Of-runt and Of-hairy*, from the *Oncopeltus* genome. Below I describe the expression pattern and function of these two genes.

Isolation of Of-hairy.

As I described in the Method section, I used *Oncopeltus* genome information and RT-PCR to isolate *Of-hairy* sequence and reciprocal BLAST to determine orthology of the sequence I isolated. An *Of-h* sequence of 580 nucleotides was isolated, with a 555 bp full length CDS, which encodes a protein of 185 amino acids (Appendices). As shown in Figure 3.19, the *Of*-Hairy protein has a full length Helix-Loop-Helix domain (Figure 3.19 A) and a conserved domain which has limited potential to form two amphipathic helices-named - the orange domain (Figure 3.19 B) by Dawson et al. (Dawson et al., 1995). *Of*-H includes a 4-amino acid WRPW motif (Figure 3.19 B to

the right) located at the C-terminus of the protein as do all other Hairy family proteins (Bier et al., 1992; Delidakis and Artavanis-Tsakonas, 1992; Rushlow et al., 1989).

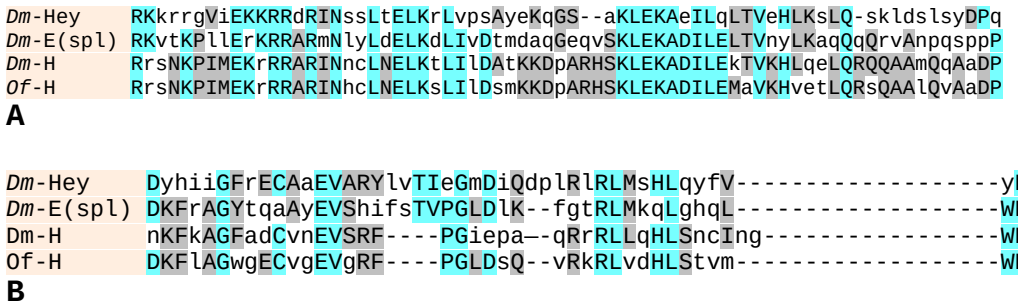

Figure 3.19 Alignment *Of*-H and *Drosophila* homologs Hey, E(spl) and H. (A) Alignment of HLH domain. (B) Alignment of the orange domain and WRPW motif. Colors are use to indicate BLOSUM62 score. light blue>=3, light gray>=0.2, no color otherwise.

Using a probe complimentary to 179nts of the *Of-h* mRNA that codes for the C-terminal of *Of*-H variable region, two rounds of *in situ* hybridization were tried with 18 hours to 96 hours AEL *Oncopeltus* embryos. I did not detect any specific expression pattern in these embryos. Using dsRNA that targets the same region, two rounds of dsRNA injection were done with a total of 21 females. All the *Of-h* dsRNA injected females laid eggs which hatched into normal fertile offspring. These findings suggest that the *Of-h* orthologue isolated does not play a role in embryonic development.

Of-runt RNAi disrupts oogenesis, possibly causing apoptosis

For the gene *runt,* Sanger sequence data was used to extend the annotation. The sequencing result revealed two additional exons that were not found by any other methods (Figure 3.20). The final annotation of *runt* includes four exons (Figure 3. 20). Only the first and second exon can be predicted based on homology. The second and

third introns are both longer than 12,000 base pairs, thus making them difficult to be predicted by software. Both exon three and exon four encode variable regions of Runt.

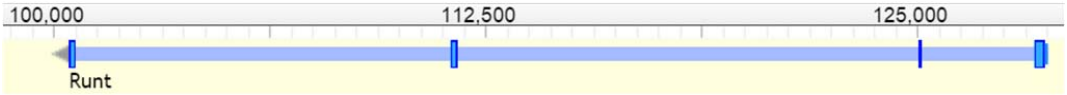

Figure 3.20 Gene Structure of *runt.* The final annotation shown runt has four exons. Numbers on top of the bar indicate the position in the scaffold; exons are labeled as dark blue boxes; introns are labeled in light blue.

A full length *Of-runt* CDS of 843nts was isolated, which could encode a protein with 280 amino acids. It has a VWRPY motif. Three other RUNX family members in *Drosophila* were identified (Appendices).

Of-runt is expressed in a broad region at the anterior the blastoderm embryos (Figure 3.18, green arrow). No expression was detected in the germ band embryos.

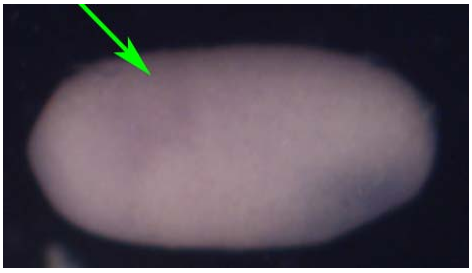

Figure 3.21 *Of-runt* expression in blastoderm embryo. *Of-runt* is expressed in a broad band in the head region (green arrow)

To test the function of *Of-runt*, dsRNA was prepared to target a 219 nt sequence which encodes the variable region located between the Runt domain and the WRPY motif. 18 *Oncopeltus* females were injected with *Of-runt* dsRNA. All of them started laying eggs two days after injection. They continued to lay eggs for 3-4 days, and the eggs hatched normally. After ~ 4 days, the injected females started to lay eggs that were flat in shape, with shrunken embryos inside of the eggshells (Figure 3.22 A,

green arrows). They never developed to germ band stage. After laying one or two batches of eggs like this, injected females started to lay some amorphous yellow masses (Figure 3.22 B, red arrow) together with severely shrunken embryos (Figure 3.22 B, green arrow); later, they laid only amorphous yellow masses (Figure 3.22 C). All the injected females died within three weeks, with most of them died within two weeks (14 out of 18). In contrast, *Of-ftz-f1* dsRNA injected females lived for more than 4 weeks after injection. These results suggest that *Of-runt* is necessary for viability of early embryos.

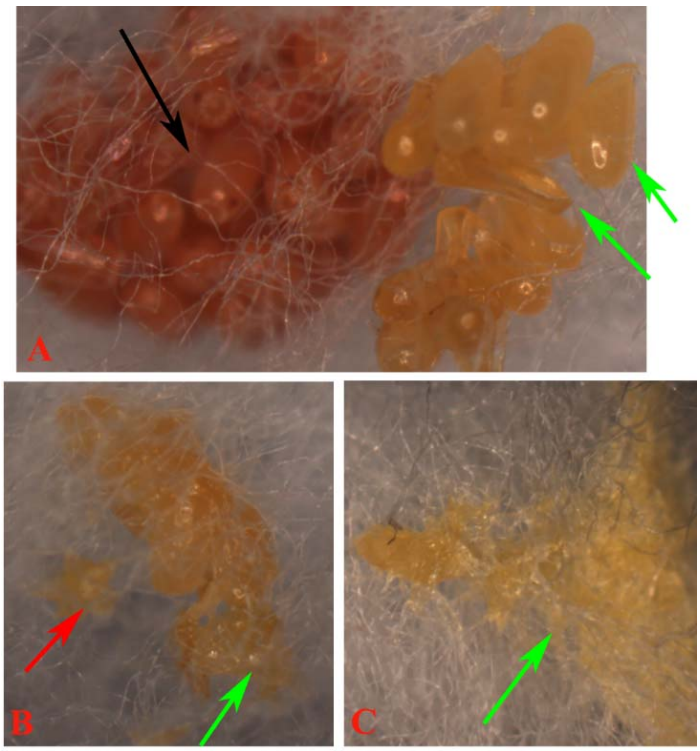

Figure 3.22 *Of-runt* RNAi disrupts oogenesis. (A) Shrunken eggs (green arrows) laid after normal eggs (black arrow). (B) Shrunken eggs (green arrow) laid with amorphous mass (red arrow). (C) Only amorphous mass (green arrow) was laid later.

To determine whether the dsRNA injections knocked down *Of-runt* levels, RT-PCR was used to determine the relative concentration of *Of-runt* mRNA in ovaries from injected and uninjected females (Figure 3.23). In both injected and uninjected females, *actin* expression level in ovaries were similar. In dsRNA injected females, Of-*runt* expression levels in ovaries decreased to very low levels.

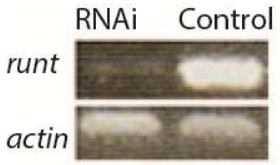

Figure 3.23 *Of-runt* pRNAi knockdown expression of *Of-runt* in ovary. Samples from *Of-runt* dsRNA injected females are labeled as RNAi, samples from uninjected females are labeled as Control. *Actin* expression levels are similar. *Of-runt* expression levels decreased after RNAi treatment .

3.3.5 Discussion

With the goal of assessing the role of PRG orthologs in the Hemipteroid Assemblage, I attempted to isolate and study the function of *Oncopeltus* PRGs. In total, I identified four clear PRG orthologs in *Oncopeltus*. *Of-hairy* does not appear to have a segmentation function in *Oncopeltus.* More experiments may need to be done to confirm this negative result. Similarly, *Of-ftz* may be non-functional in this species (see below). This result itself is not surprising, since three PRG orthologs did not have a segmentation function in *Tribolium* (Choe et al., 2006). However, both *Of-ftz-f1* and *Of-runt* appear to be required in different stages of embryonic development.

Of-ftz may be non-functional

Two isoforms of *Of-ftz* were identified. While each maintains a complete ORF for the homeodomain, neither includes an upstream initiator methionine and stop codons

are present in all three reading frames upstream of the homeodomain for both. This suggests that *Of-ftz* is either not be translated into protein, or a different start codon is used to generate a very short peptide including basically only the homeodomain. It is also possible that there are other isoforms of *Of-ftz* that we failed to find. However, I tried many different approaches and still did not isolate another isoform. In addition, Even-Campen et al. did RNA-seq with Roche 454 (Ewen-Campen et al., 2013) and the i5K project did RNA-seq with Illumina, but neither of these projects identified an additional isoform.

pRNAi with *Of-ftz* dsRNA did not result in abnormal embryos, suggesting that *Of-ftz* does not have a function in *Oncopeltus* embryonic development. This is not the first time *ftz* was shown to have no function in embryonic development, as a similar phenomenon was observed for *Tribolium Tc-ftz* (Brown et al., 1994a; Stuart et al., 1991). It is interesting that I did not detect *Of-ftz* expression in the CNS, while Ftz expression in CNS seems to be maintained in all other arthropods examined (See chapter 1 for more detail about Ftz CNS expression). These findings suggest that *ftz* has lost function in *Oncopeltus* or lineages leading to *Oncopeltus*. It will therefore be of great interest to examine *ftz* function in related hemimetabolous insects.

ftz-f1 may have a conserved role in segmentation in divergent insects

Of-ftz-f1 is expressed in segmental pattern in *Oncopeltus* embryos. Preliminary RNAi data suggest that it has function in segmentation. pRNAi of *Of-ftz-f1* blocked oogenesis. In contrast to this, in *Drosophila,* Ftz-F1 does not have a function in oogenesis. Embryos derived from germline clones for at least six different *ftz-f1* alleles produce eggs with pair-rule mutant phenotypes (Guichet et al., 1997; Guichet

et al., 1996; Hacker et al., 2003; Luschnig et al., 2004; Yu et al., 1997). In *Tribolium, Tc-ftz-f1* RNAi had been found to block oogenesis (Xu et al., 2010). Thus, the function of Ftz-F1 in oogenesis is conserved in *Tribolium and Oncopeltus* and may be more broadly conserved in basally branching insects.

My experiments using eRNAi for *Of-ftz-f1* were not definitive. In *Of-ftz-f1* RNAi embryos with less severe phenotypes, it was not possible to determine if the defects had pair-rule or segmental patterns. The one with one pair of legs did suggest a pair-rule function. Some *Of-ftz-f1 e*RNAi embryos also had either shrunken abdomens or ventral fusions, which indicates the ventral part of the segments are more sensitive than dorsal portions to the knockdown of *Of-ftz-f1*.

The *Drosophila* and *Tribolium ftz-f1* orthologs function at several stages of development. Both function as pair-rule segmentation genes and both play important roles in cuticle development (Heffer et al., 2013; Ruaud et al., 2010). The cuticle defects in the head and thoracic segment of *Of-ftz-f1* RNAi embryos suggests *Of*-Ftz-F1 may have similar rules. Further, in *Oncopeltus* dsRNA injected embryos, most of the abdominal segments were missing in the most severely affected embryos, with only three abdominal segments evident morphologically. This suggests that *Of-ftz-f1* may function as a PRG in some of the abdominal segments, despite its segmental expression pattern. In *Drosophila, Dm-ftz-f1* is expressed ubiquitously in blastoderm embryos but it depends on its co-factor Ftz to regulate alternate *en* stripes and functions as a PRG (Guichet et al., 1997; Yu et al., 1997). Since *Of-ftz* RNAi did not effect embryonic development, it is possible that in *Oncopeltus,* the function of *Of*-Ftz-F1 is dependent on another co-factor, which is expressed in a pair-rule pattern. In

Gryllus bimaculatus, Gb-eve was found to partially function as a PRG in posterior segmentation (Mito et al., 2007). In *Oncopeltus, Of*-E75A had been found to have pair-rule function, and it is not expressed from A7 to A10 segments (Erezyilmaz et al., 2009). Which gene(s) take over *Of*-E75A's function in those posterior segments is unknown. Together, these observations suggest that, compared with *Drosophila,* the segmentation network may be more complicated in other insects.

Of-runt may be necessary for oocyte viability

In mammals, there are three Runt homologues *RUNX1, RUNX2* and *RUNX3.* RUNX1 and RUNX2 were found to have oncogenic potential (Reviewed in Ito, 2008; Levanon and Groner, 2008). Recent studies suggest that RUNX proteins play a direct role in inhibiting apoptosis induced by p53 (Ozaki et al., 2013a; Ozaki et al., 2013b; Wu et al., 2013). In the sea urchin, *Sp-runt*-1 deficiency leads to extensive cell death (Dickey-Sims et al., 2005). *Sp-runt*-1 was found to directly regulate the expression of Protein kinase C (PKC) to support cell growth and inhibit apoptosis (Dickey-Sims et al., 2005). In *Oncopeltus,* I found that *Of-runt* disrupts oogenesis. Oocytes were shrinking as they developed, suggesting that in *Of-runt* knockdown oocytes, apoptosis was induced. It has been found that *Drosophila* PRG mutants cause pair-rule phenotypes by inducing apoptosis (Hughes and Krause, 2001). However, I suggest that this kind of apoptosis maybe different from the apoptosis we seen in RUNX proteins mutants. During development, if cells can't be specified correctly, these cells will be removed by cell death (Adachi-Yamada et al., 2005; Adachi-Yamada and O'Connor, 2002, 2004). This phenomenon have been observed in mutations of not only PRGs but also gap genes and segment-polarity genes (Hughes and Krause, 2001;

Klingensmith et al., 1989; Magrassi and Lawrence, 1988; Martinez-Arias and Ingham, 1985; Pazdera et al., 1998; Tepass et al., 1994). Cell death has also been observed in mutations related to imaginal development (Fristrom, 1968). It is possible, in these scenarios, that loss of function of a gene does not trigger cell death directly, but rather, it is the fact that the cells do not take on the fate they are supposed to, triggers the cell death. This could be triggered by alterations in the reception of Dpp signals or changes in the ratio of multiple morphogens (Adachi-Yamada and O'Connor, 2002). The cell death I observed in *Of-runt* knockdown oocytes may be different from the cell death caused by pair-rule gene mutations during embryonic development. *Of*-Runt may directly participate in the apoptosis signaling pathway during oogenesis, as has been shown in mammals and sea urchin.

Chapter 4: Molecular Biology study of the invasive brown marmorated stink bug, *Halyomorpha halys*

Section 4.1 Background

Native to East Asia, *Halyomorpha halys*, commonly known as the Brown Marmorated Stink Bug (BMSB) is an invasive species in the United States. It was first discovered in North America in Allentown, PA in 1996 (Hoebeke and Carter, 2003). According to a recent investigation, BMSB has been found in 41 states in the USA and two Canadian provinces. BMSB is a polyphagous insect, which eats tree fruits, vegetables, legumes, cotton, ornamentals in the field and in nursery crops (Nielson et al., 2008; Panizzi et al., 2000; Zhu et al., 2012). BMSB has caused significant agriculture damage in recent years; in 2011, stone fruit growers considered BMSB to be the single most important pest in the mid-Atlantic region, as more than 90% of their crop was lost due to BMSB (Leskey et al., 2012). BMSB is also a nuisance for homes and businesses, since in the fall BMSB seek overwintering sites in any human-made structures. They congregate inside or outside of buildings, in attics, near windows or doors, on ceilings, and other similar areas. BMSB is resistant to common insecticides, forcing farmers to use broad spectrum insecticides, including pyrethroids to control the population of BMSB (Leskey et al., 2012). The use of broad spectrum insecticides can cause outbreaks of secondary pests such as European red mites, wooly apple aphids and others, since their natural enemies are killed by these insecticides (Leskey et al., 2012).

BMSB has spread in the USA and there are no current sustainable integrated pest management (IPM) solutions. With funding from USDA, a group of more than 60 scientists is working on finding pest management solutions for controlling BMSB. Understanding the development of BMSB from a molecular biology perspective will provide inroads into novel approaches for management and monitoring of BMSB populations.

A promising approach for pest management is the use of RNA interference (RNAi). In brief, when a double strand RNA (dsRNA) is introduced into a cell, it initiates a process that silences the corresponding endogenous mRNA. RNAi is a posttranscriptional gene-silencing mechanism which was first discovered in *Caenorhabditis elegans* (Fire et al., 1998), and subsequently shown to function in fungi, plants, insects and other animals (Mello and Conte, 2004). Two major pathways mediate the RNAi process: the siRNA and miRNA pathways (Carthew and Sontheimer, 2009; Tomari et al., 2007). The miRNA pathway uses endogenous dsRNA that is transcribed from the cells' own genome, and regulates endogenous genes. The siRNA pathway responds to foreign or invasive nucleic acids such as viruses and transposons, thus serving as a defense mechanism against exogenous dsRNAs. The effects of RNAi have been observed to be both cell-autonomous and non-cell-autonomous (Whangbo and Hunter, 2008). In cell-autonomous RNAi, the RNAi only affects the cell that expresses the dsRNA or that is directly exposed to experimentally introduced dsRNA. In non-cell-autonomous RNAi, the RNAi effect spreads from an initiating site to other distant tissues or cells (Fire et al., 1998). For

example, non-cell-autonomous RNAi can be initiated by soaking *C. elegans* in dsRNA solution or by feeding them bacteria expressing dsRNAs (Tabara et al., 1998; Timmons and Fire, 1998). There are two types of non-cell-autonomous RNAi: environmental RNAi, and systemic RNAi (Whangbo and Hunter, 2008). Environmental RNAi refers the phenomena by which environmentally encountered dsRNA triggers RNAi effects; systemic RNAi refers to the spreading of RNAi effects from an initial site to distant tissues.

For non-cell-autonomous RNAi to function, cells must have the ability to uptake the RNAi signals from the outside. Screens for uptake defects in *C. elegans* identified two proteins, SID-1 (Winston et al., 2002) and SID-2 (Winston et al., 2007) which are responsible for systemic distribution of RNAi effects. Orthologs of SID-1 have been found in some insects. For example, SID-1 orthologs were discovered in *Tribolium castaneum* (Tomoyasu et al., 2008), *Bombyx mori* (Tomoyasu et al., 2008), *Aphis gossypii* and *Sitobion avenae* (Xu and Han, 2008), and *Schistocerca americana* (Dong and Friedrich, 2005). To my knowledge, no SID-2 has been found to date in any insect species whose genome has been sequenced.

In animals that carry out non-cell-autonomous RNAi, dsRNA that is injected or supplied in the diet can trigger systemic gene silencing. This phenomenon makes RNAi a very useful tool to knockdown genes in order to study their functions. This approach has been pursued in recent years for several insect model systems. As a well-studied model organism, *Drosophila melanogaster* was the first insect in which RNAi was shown to work. Shortly after Fire et al. revealed RNAi in *C. elegans* (Fire et al., 1998), Kennerdell *et al.* showed that injection of dsRNA corresponding to the

segmentation genes *fushi tarazu (ftz)* or *even-skipped (eve)* into *Drosophila* embryos phenocopied *ftz* and *eve* loss-of-function mutations (Kennerdell and Carthew, 1998). Using RNAi, they found that *Drosophila frizzled* and *frizzled 2* act downstream of *wingless* (Kennerdell and Carthew, 1998). Later, the RNAi technique was used in *Drosophila* cell culture for high-throughput screens. With this system, the function of thousands of genes can be analyzed (Boutros et al., 2004; Kiger et al., 2003). RNAi was also shown to work well in the flour beetle *Tribolium castaneum*: Injection of *Tc-Dfd* dsRNA into *Tribolium castaneum* embryos resulted in embryos that phenocopy the *Dfd* loss-off-function mutant phenotype (Brown et al., 1999). Further studies showed that RNAi works well in all life stages of *Tribolium* (Bucher et al., 2002; Tomoyasu and Denell, 2004). Bucher et al. found that injecting dsRNA into the haemocoel of female *Tribolium* resulted in RNAi effects in the offspring of the injected females. This phenomenon is called parental RNAi (pRNAi) and was first discovered in *C. elegans* (Fire et al., 1998; Timmons and Fire, 1998). In addition to *Tribolium,* pRNAi has been proven to be effective in a few other insects, such as the silk moth *Bombyx mori* (Quan et al., 2002) and the milkweed bug *Oncopeltus* (Liu and Kaufman, 2004). Overall, many experiments have shown that delivering dsRNA into insects by injection is an efficient method to knockdown gene expression.

The success of using RNAi to knock down gene function encouraged many researchers to explore the possibility of using RNAi as tool for protecting crops against insect pests. Most of these researchers choose housekeeping genes to target pests. Delivering dsRNA by injection, which is used broadly in basic research, is not feasible for controlling insect pests. For this purpose, delivering dsRNA through

feeding would be ideal. Baum et al. tested whether RNAi functions in western corn rootworm (WCR, *Diabrotica virgifera*) by separately targeting 25 different genes (Baum et al., 2007). The dsRNAs were sprayed on the surface of the artificial WCR diet. They found that many of the dsRNAs caused larval stunting and mortality at a concentration of 520 ng/cm^2. Further screening revealed that dsRNAs targeting genes encoding vacuolar ATPase (V-ATPase) subunits A, D or E and alpha-tubulin, were effective at a concentrations below 52 ng/cm^2. A transgenic corn line was constructed that expresses both strands of the WCR V-ATPase A gene under the CaMV e35S promoter, such that it would be expressed in the whole plant and assemble into dsRNA in the transgenic corn. When exposed to WCR infection, the transgene effectively protected the corn from root damage. Around the same time Mao et al. analyzed a cDNA library from the midgut of cotton bollworm (*Helicoverpa armigera*) treated with gossypol, and found several gossypol-induced genes (Mao et al., 2007). Gossypol is a metabolite of cotton, which is toxic to most insects. The gossypol-induced genes could play an important role in the tolerance to gossypol in cotton bollworm. Mao et al. tested this hypothesis with one gene, P450 monooxygenase, CYP6AE14. *Arabidopsis thaliana* and *Nicotiana tobacum* transgenic lines were made to express dsRNAs targeting CYP6AE14. When cotton bollworm larvae were fed with leaves from those transgenic plants, a decrease in CYP6AE14 expression level was observed. Four days after feeding on the transgenic plants, the cotton bollworm larvae were transferred to artificial diets. When gossypol was added to the diet, the larvae, which had been feed on transgenic plants for four

days, did not grow, while larvae fed on diet without gossypol grew normally. This indicates that RNAi targeting CYP6AE14 reduced gossypol tolerance in WCR.

More recently, a broad range of insects and target genes have been tested, examining the feasibility of using RNAi for pest control. Gong et al. found that spraying dsRNA targeting the Rieske iron–sulfur protein (RISP) gene on cabbage leaves killed diamondback moth (*Plutella xylostella*) larvae feeding on them (Gong et al., 2013). Their study showed that RNAi targeted to different regions of the RISP gene had different efficiencies. The most efficient dsRNA killed 73% of the larvae, while the least efficient dsRNA killed only 38% of the larvae. They also noted that if the larvae survived after 72 hours of feeding on dsRNA soaked leaves, the larvae would developed normally. Li et al. fed oriental fruit flies (*Bactrocera dorsalis*) with dsRNA expressing-bacteria, similar to what is done routinely for *C. elegans* (Li et al., 2011). They found that dsRNA targeting *v-ATPase-D* did not cause high mortality, different from what Baum et al. found in WCR. Of the four genes they tested (Rpl9, V type ATPase D subunit, the fatty acid elongase *Noa* and GTPase *Rab11*), only *Rab11* caused relatively high mortality but targeting *Rab11* and *Noa* decreased egg production in adult females(Baum et al., 2007). In the whitefly, Upadhyay et al. tried five different genes (*actin ortholog, ADP/ATP translocase, α-tubulin, ribosomal protein L9 (RPL9)* and *V-ATPase*), mixing the dsRNA into the diet. All five dsRNAs had some effect but dsRNAs targeting *RPL9* and *V-ATPase A* were more efficient than the other three in killing whitefly nymphs (Upadhyay et al., 2011). Eastern subterranean termites (*Reticulitermes flavipes*), were fed high doses of dsRNAs targeting two genes (cellulose enzyme gene *Cell-1* and a caste-regulatory hexamerin

storage protein gene *Hex-2*) (Zhou et al., 2008). After feeding the termites for 24 days, they observed the group fitness and mortality of the treated termites were significantly reduced. To test the possibility of controlling peach potato aphid (*Myzus persica*), Mao et al. cloned its *hunchback* gene and created transgenic tobacco plants which express *hunchback* dsRNA. The aphids that fed on the dsRNA-expressing tobacco plants had a smaller population and lower insect biomass compared with aphids grown on control plants (Mao and Zeng, 2014). Overall, these results suggest that transgenic plants expressing specific dsRNAs can efficiently knockdown the expression of target genes in insects that feed on the plants. Spraying dsRNA on the surface of artificial food or leaves can have similar effects. Thus, RNAi is a promising approach for control of pest insects. RNAi can be used to control insect pests that develop resistance to standard insecticides. Specially designed dsRNA could be used to kill a single species without affecting other species. Finally, combining one or more genes as RNAi targets may increase the efficiency of RNAi (Huvenne and Smagghe, 2010; Price and Gatehouse, 2008; Zhang et al., 2013).

When I began this study, there was no molecular information available about BMSB. I sought to establish molecular approaches in BMSB for both basic and applied research. I contributed to the sequencing of the BMSB transcriptome, isolated full length sequences of BMSB v-ATPase gene, partial sequence of BMSB *Scr, engrailed, even-skipped* and *moesin* (membrane-organizing extension spike protein) genes, established embryo *in situ* hybridization and immunohistochemistry, and demonstrated that RNAi is a feasible approach for gene knock down in BMSB.

These tools will be useful for studying genes controlling BMSB embryonic development and for development of specific RNAi tools for pest control.

Section 4.2 Methods

4.2.1 Insect husbandry and embryo collection

Laboratory colonies of BMSB were reared as previously described (Taylor et al., 2014). Briefly, BMSB were collected in soybean fields at the University of Maryland Beltsville Research Farm. The collected BMSB were kept in mesh cages (60×30×35 cm). Potted plants of *Phaseolus vulgaris* were used as the major food source and hiding place for BMSB. Organic green bean pods and raw sunflower seeds were added to the cages to provide extra nourishment. These foods were replaced with fresh ones once or twice a week. Other extra food sources tested for growth were blueberries, apples, grapes and carrots. All foods were certified organic and were washed extensively before placement in cages. We did not notice any difference in BMSB growth with these more expensive extra food sources. BMSB cages were kept at 25°C, RH of 65±5%, and a 16 hour light: 8 hour dark photoperiod. For timed egg collections, cages were checked every four hours for newly laid eggs. The eggs were removed from the cage by hand, and kept in a petri dish, usually still attached to a piece of the leaf they were laid on, under the same environmental conditions described above to the desired time points.

4.2.2 Identification of genes of interest

Assembled BMSB RNA-seq data (Ioannidis et al., 2014) in fasta format was used to create a local BMSB BLAST database using the BLAST+ package (Altschul et al.,

1990). TBLASTN was carried out using full length *Drosophila* protein sequence of Engrailed or Scr as query sequence with the local BMSB BLAST database as subject database. The TBLASTN results were generated in XML format and reviewed with BlastViewer. Sequences of candidate BLAST hits were retrieved using Cdbtools. Reciprocal BLAST with the insect non-redundant protein sequences database was carried out to find orthologs. Predicted BMSB genes were experimentally verified by Reverse Transcription PCR (RT-PCR) followed by Sanger sequencing.

4.2.3 Embryo Fixation

Embryos were collected from cages and put into 2 mL eppendorf tubes, with ~ 20 embryos per tube. The fixation protocol was based on that developed for *Oncopeltus*, kindly shared by Dr. Chipman's lab (Ben-David and Chipman, 2010). In brief, 600ul tap water was added to each tube of embryos which was placed in boiling water for three minutes and then placed on ice for six minutes. After the water was removed, 600ul of heptane and 600ul 4% PFA in PBS were added. Gentle shaking brought the embryos to the interface. Tubes were shaken vigorously on a Vortex mixer for 20 minutes. After shaking, the heptane and PFA were removed and the embryos were rinsed once with heptane, then once with methanol. The embryos together with methanol were then put into wells on depression concave slides and the egg shells were manually removed with forceps under a dissection microscope. The embryos were then passed through 75%, 50% and 25% methanol/PBST gradient rinses for rehydration. The rehydrated embryos were fixed with 4% PFA in PBST for 90 minutes on a nutator. The fixed embryos were then washed three times with methanol and stored in methanol at -20 °C for future use.

4.2.4 Antibody staining

The fixed embryos were removed from -20 °C and passed through a 75%, 50% and 25% methanol/PBST gradient for rehydration. Embryos were then rocked on a nutator in 5% BSA in PBST for 2-3 hours to block non-specific binding. After blocking, the embryos were incubated with a 1:10 dilution of monoclonal anti-Engrailed antibody 4D9 (Developmental Studies Hybridoma Bank) in 5% BSA at 4 °C overnight. The 4D9 antibody was then removed and the embryos were washed three times for twenty minutes each with PBST. The embryos were then incubated with 1:300 Biotinylated Goat Anti-Mouse IgG Antibody (Vector Labs) for two hours at room temperature. The secondary antibody was then removed and embryos were washed with PBST three times for 20 minutes each. After washing, the embryos were incubated one hour with ABC reagent (Vector labs) followed by three 20 minutes washes with PBST. Detection by a color reaction was then carried out using the SigmaFast DAB kit (Sigma Aldrich). Expression was monitored under a dissection microscope and terminated when stripes were evident, usually within 30 minutes. The DAB was then removed and embryos were rinsed three times with PBS. Embryos were post-fixed with 4% PFA for 20 minutes and germband embryos were removed from the yolk using forceps. Germ bands were mounted in 90% glycerol/PBS. Photographs were taken under a dissection microscope (Leica M420, 16-20X).

4.2.5 Whole mount in situ hybridization

The *in situ* hybridization protocol was modified from that used for *Oncopeltus* (Ben-David and Chipman, 2010). Probe was made with *in vivo* transcription using

PCR production as template. The T7 promoter that is needed for *in vivo* transcription was added to the reverse PCR primer and used to amplify PCR template. The primers used are Hh-enF 5'-TACCCTTCTCCGTCGACAAC-3' and Hh-enRT7 5'-TAATACGACTCACTATAGGGAGACGGCCTCTTGTCTTCTTTGT-3' for *engrailed;* and primers for amplifying 3'UTR of *eve* are hhEve1F 5'-GATATACTATTGACTCGCGGCTGA-3' and 5'-TAATACGACTCACTATAGGGAGAACTATCTTCCTGCTATCACTGGT-3' (T7 promotor sequence underlined). PCR products were purified and used as template for *in vivo* transcription to make probes. Embryos were fixed and rehydrated as described above. After rehydration, the embryos were pre-hybridized with hybridization buffer (50%formamide, 5XSSC, 0.1% Tween-20, 50µg/ml yeast tRNA, 5%Dextran, 50µg/ml heparin) at 55 °C for 1-4 hours. After pre-hybridization, probe was added to the embryos (1ul probe per 100ul hybridization buffer), and incubated at 55 °C overnight (16-18 hours). The next day, the probe was removed and the embryos were washed twice with hybridization buffer at 55 °C for 15 minutes each, followed by two washes with 2XSSC at room temperature for 30 minutes and one wash in 0.2XSSC at room temperature for 30 minutes. The embryos were rinsed three times with PBST. After rinsing, the embryos were incubated with 10% sheep serum in PBST for 1-4 hours at room temperature to block non-specific binding. Embryos were next incubated with anti-DIG-AP antibody (1:1500, Roche) at 4 °C overnight. After incubation, the antibody was removed and the embryos were washed five times in PBST for 20 minutes each. For detection, the embryos were washed with staining buffer (100mM NaCl, 50mM $MgCl_2$, 100mM Tris pH 9.5, 0.1% Tween 20) three

times for five minutes each. Staining was carried out using NBT/BCIP solution (Roche) diluted in staining buffer. The embryos were checked under a dissection microscope every ten minutes until the desired color reaction had developed, generally within one hour. The color reaction was stopped by adding PBST to the staining solution. The embryos were then washed with PBST three times for five minutes each to remove the staining solution from the embryos. The embryos were washed with 50% methanol in PBST for five minutes and 100% methanol for another five minutes. The embryos were then treated with 100% ethanol for 30 – 120 minutes. The embryos were washed with 50% methanol in PBST for five minutes, and washed with PBST three times for five minutes each. Blastoderm stage embryos were transferred to depression concave slides to take pictures, germ band stage embryos were dissected out of yolk and mounted on slides with 90% glycerol. Photographs were taken under a dissection microscope (Leica M420, 16-20X).

4.2.6 Double-Strand RNA (dsRNA) preparation

Primers were designed to amplify 300 -500 bp of BMSB genes of interest with T7 promoter sequences added to the 5' end of both forward and reverse primers. The primer sequences are as follow: Hh-ScrFT7 5' – TAATACGACTCACTATAGGGAGAGCAGGACCTGACTACGTCCTC-3' and Hh-SciRT7 5' – AATACGACTCACTATAGGGAGATCCAGCTCCAGCGTCTGGTA-3' (T7 promoter sequences underlined). PCR was carried out with cDNA that had been made from 0-6 day embryos using the manufacturer's recommended standard conditions (Reverse Transcription system, Promega). The PCR products were separated on an agarose gel and sent out for sequencing (Genewiz) to confirm that the correct gene was amplified. The purified PCR product (Gel Extraction Kit, Qiagen) was use as a template for in vitro transcription using MEGAscript® T7 transcription (Life Technologies) Kit following the manufacturer's recommendations. The final product was treated with DNase from the transcription kit to get rid of the DNA template. In order to anneal the in vitro transcription product single stranded RNAs, in vitro transcription products were heated to 94°C for five minutes and slowly cooled by decreasing temperature 0.8°C every minute, until 45°C was reached, in a PCR machine (TPersonal, Biometra). The annealed double-strand RNA was precipitated with 1/10 volume of 3 M sodium acetate (pH, 5.2), 2X volume of ethanol, dissolved in 10-20ul injection buffer (0.1mM NaH_2PO_4, 5mM KCl, pH 6.8), and stored in -20°C. The concentration of double-strand RNA was measured with a spectrophotometer.

Section 4.3 Results:

4.3.1 The BMSB Transcriptome

As a first step towards genomic analysis of BMSB, the transcriptome from all stages of the life cycle of BMSB was sequenced in collaboration with Dr. J. Dunning-Hotopp (Ioannidis et al., 2014). My contribution to this project was the culture and collection of representative animals and isolation of RNA from eggs, 1st and 2nd instar nymph (Figure 4.1)

Figure 4.1. Life Stages of BMSB. The life stages of BMSB are shown starting with eggs followed by 1st instar nymph, 2nd instar nymph, 3rd instar nymph, 4th instar nymph, 5th instar nymph, and an adult in a counter-clockwise spiral outwards and from largest to smallest. The bar in the low left represents 1 cm.

.

4.3.2 Establishment of in situ hybridization and immunohistochemistry to monitor gene expression in BMSB embryos

In *Drosophila*, *engrailed* (*en*) is a segment polarity gene expressed in and specifying the posterior compartment of each segment (Kornberg, 1981; O'Farrell et al., 1985). Similar segmental expression patterns for *en* were observed in a wide range of other species including all insects examined to date as well as more distant arthropods (Patel et al., 1989a; Patel et al., 1989b). For example, in grasshoppers (*Schistocerca americana*), shortly after gastrulation, *en* is expressed in stripes, and the stripes are located in the posterior part of every segment (DiNardo and O'Farrell, 1987; Patel et al., 1989a). Similarly, in some Crustacean, crayfish (*Procambarus clarki*) and lobster (*Homarus americanus*), *en* was found to be expressed in the posterior part of each segment (Patel et al., 1989b). This high degree of conservation of expression makes *en* very useful for determining the identity and number of segments when a new species of arthropod is studied. In order to establish techniques for examining gene expression in BMSB embryos, we utilized the *en* gene –whose expression would also be expected to be observed in segmental stripes in early embryos. This type of clear expression pattern allows one to distinguish false positive patterns from true patterns, as it is easy to tell which staining is background and which is *bona fide* signal.

To carry out *in situ* hybridization, a portion of the *Hh-engrailed* (*Hh-en*) gene was isolated. Two degenerate primers DEGNenF 5'- GARAAYMGNTAYYTNACNGA and DEGenR 5'- RTGRTTRTANARNCCYTGNGC were used to amply *Hh-en* from cDNA prepared from 1-5 day BMSB embryos. A fragment of ~300bp was amplified from the cDNA and sequenced. A sequence of 286bp was acquired.

Around the same time, the BMSB transcriptome data were made available to us (Ioannidis et al., 2014).This sequence was used as subject in a BLASTN query against the BMSB transcriptome database and compared with TBLASTN results (Method). One 566 base pair sequence was shared by both BLASTN and TBLASTN results. This 566bp sequence (Appendices) could encode a 187 amino acid region, which includes the 60 amino acid En homeodomain（Figure 4.2). 56% of the 187 amino acids are identical to *Drosophila* Engrailed and 49% are identical to *Drosophila Invected (Inv)*, which is an *en* related gene (Coleman et al., 1987; Poole et al., 1985). Reciprocal BLAST result suggests that this sequence is an ortholog of *Drosophila* En. The *Hh*-En homeodomain is 85% identical to *Drosophila* En, and 83% identical to *Drosophila* Inv.

A

```
  1   QSECGCEMAL EAERRPSAFH RIELLTCKDA PPAPASLPFS VDNILRPEFG
 51   QAARRPVRLS RSDDEGPSPT KIDPPIPDDP NGPVWPAWVY CTRYSDRPSS
101   GGPRSRRIKR KDKNKEDKRP RTAFSGEQLA RLKTEFSINR YLTERRRQAL
151   ASELGLNEAQ IKIWFQNKRA KIKKASGNRN PLALQLM
```

B

```
Dm-En    eKRPRTAFSsEQLARLKrEFNeNRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKSt
Dm-Inv   DKRPRTAFSGtQLARLKhEFNeNRYLTEkRRQQLSgELGLNEAQIKIWFQNKRAKlKKSS
Hh-En    DKRPRTAFSGEQLARLKtEFsiNRYLTERRRQaLaSELGLNEAQIKIWFQNKRAKIKKaS
```

Figure 4.2 *Hh*-En partial sequence and HD alignment. (A) The partial protein sequence of the putative *Hh*-En. (B) Alignment of Homeodomains of *Dm*-En, *Dm*-Inv and *Hh*-En. From this figure, it is not obvious to me that it is more closely related to En and Inv.

To examine the expression of *Hh-en* in BMSB embryos, I began with protocols used for *Oncopeltus fasciatus*, another member of the Hemipteroid Assemblage. The

basic *in situ* hybridization protocol is similar to that used in the model insect *Drosophila melanogaster*. However, the embryo collection and fixation steps differ, which is described in Methods. The *in situ* hybridization protocol needs to be specially fine tuned for a new species. I tried different time and temperature combinations for the hybridization buffer washing step and SSC washing step, which are the main steps that determine the signal-to-background ratio. One combination that gave a good signal-to-background ratio was found and described in Methods. One observation worth mentioning is that the same protocol works well for both blastoderm stage and germ band stage embryos. In contrast, for *Oncopeltus,* the protocol that works for germ band embryos did not work for blastoderm stage embryos. Using the sequence information I got for *Hh-eve* (check next subsection for detail), I made an RNA probe for *Hh-eve* mRNA.

I encountered some problems when I tried to remove the embryos out of the yolk. The yolk is very sticky, and the embryos are very fragile. Trying to completely remove the yolk from the embryos resulted inevitably in broken embryos. I found that by adding some methanol into the PBST solution, I could decrease the stickiness of the yolk, while at the same time, the embryos were more fragile. 50% methanol in PBST seems to give reasonably less sticky yolk and embryos with acceptable fragileness.

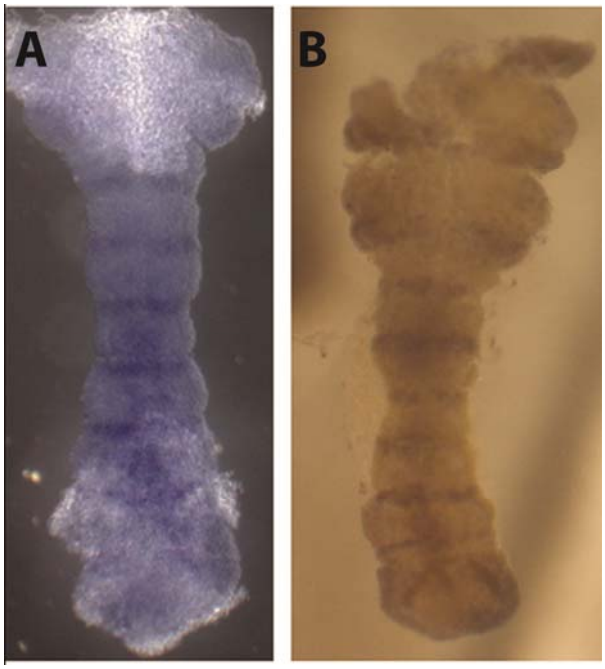

Figure 4.3 Engrailed expression in BMSB. (A) *In situ* hybridization to BMSB germ band embryos using an *Hh–en* probe. Six stripes were detected. The anterior and posterior parts of the embryo remained covered with yolk in this photograph. (B) En antibody staining to an embryo slightly older than that shown in (A). Nine stripes were observed Embryos oriented anterior, top.

As shown in Figure 4.3 A, the *in situ* hybridization technique worked well in BMSB embryos. In the early germ band embryo, *en* mRNA was detected in segmental stripes. In the embryo shown, six stripes of *en* can be seen. Note that it is quite possible that there were more stripes on the embryo, as both the anterior and posterior ends were both covered by the yolk.

A monoclonal antibody raised against *Dm*-En (MAb 4D9), has proven to be a useful tool to examine En in diverse species as it recognizes an epitope located in the variable region of the homeodomain of the En and INV proteins, and does not cross react with other homeodomain proteins (Patel et al., 1989b). To establish immunohistochemical techniques in whole mount BMSB embryos, I utilized this

antibody. I used the same fixation protocol that was used for *Drosophila in situ* hybridization. The fixed embryos were remove from -20°C, re-hydrated in PBST and incubated in 5% BSA for 2-3 hours, Antibody 4D9 was added to new 5% BSA and incubated at 4°C on a nutator overnight. After washing off the antibody, the embryos were incubated with Biotinylated Goat Anti-Mouse IgG Antibody (Vector Labs) for two hours at room temperature. Color reaction was carried out using the ABC kit from Vector lab. As shown in figure 4.3B, EN protein was detected in stripes at the boundary of each segment.

As shown in Figure 4.3 B, immunochemistry staining also worked well in BMSB embryos. In a germ band embryo, En proteins was detected in segmental stripes.

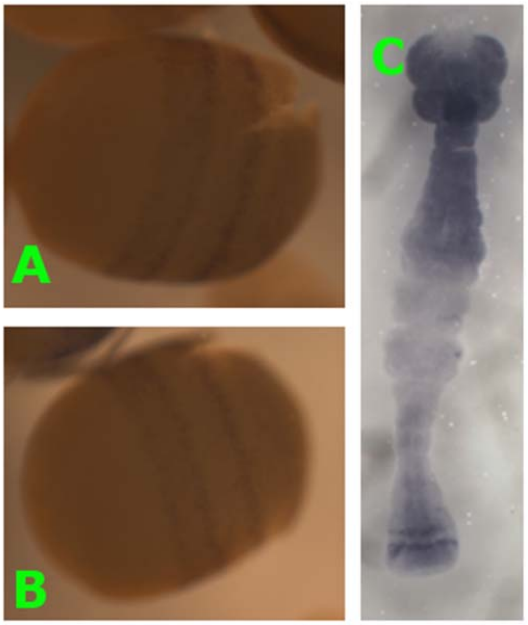

Figure 4.4 *Hh-eve* expression in blastoderm stage and germ band stage embryos. (A) A blastoderm stage embryo with three *Hh-eve* stripes in the central region of the embryo; B) another blastoderm stage embryo with three *Hh-eve* stripes.(C) a germ band embryo with two *Hh-eve* stripes in the growth zone. In A and B embryos are orientated from anterior (left) to posterior (right), in C anterior (top) to posterior (bottom)

To test whether *in situ* hybridization works for other genes expressed in early BMSB embryos, I used BLAST as described in the method to isolate *Hh-eve,* and orthology is confirmed by reciprocal BLAST.The i*n situ* probe that detects the 3'UTR of *Hh-eve* was made use primers listed in Method part. As shown in Figure 4.4 *Hh-eve in situ* hybridization works well in both blastoderm stage embryos (Figure 4.4 A and 4.4 B) and germ band embryos (Figure 4.4C). In Figure 4.4 A and 4.4 B, three *Hh-eve* stripes can been seen in the central region of the BMSB blastoderm stage embryos. In Figure 4.4 C, two *Hh-eve* stripes can be seen in the growth zone region of the BMSB embryos.

In sum, I successfully established embryo collection and fixation protocols for BMSB embryos. Both antibody staining and *in situ* hybridization were carried out in this species and can be used to examine the expression of additional genes.

4.3.2 Parental RNAi in BMSB

As mentioned above, RNAi is emerging as a useful method to knock down gene expression in both plants and animals. Therefore, I designed experiments to determine whether RNAi functions in BMSB. dsRNA sometimes has off-target RNAi effects (Jackson and Linsley, 2010). Tests of mortality caused by dsRNA thus may not test specificity of target gene knock down and, in addition, require a large sample size due to the traumatic nature of the injection itself. Therefore, instead of choosing a gene designed to kill the dsRNA-treated insects, we chose the *Scr* gene which had been shown to have very clear and unique effects in both *Oncopeltus* and the American cockroach (*Periplaneta americana*) (Chesebro et al., 2009; Hrycaj et al., 2010). In these species, knock down of *Scr* resulted in the transformation of proboscis

toward leg. This unique phenotype will allow for a definitive determination of the effectiveness of dsRNA simply by examining the shape of the proboscises of the offspring.

To test the ability of parental RNAi to function in BMSB, the *Hh-Scr* gene was isolated using a similar approach as that used for *Hh-engrailed.* In brief, a fragment of *Hh-Scr* was isolated by degenerate PCR with forward primer scrdegF 5'-CCRCARATHTAYCCRTGGATG-3' and revers primer ScrdegR1 5'-CATRTGGYANGGNACRATRTTCAT-3'. The fragment of *Hh-Scr* gene acquired was used to BLAST the BMSB transcriptome data. This result was compared with TBLASTN queries. A partial *Hh-Scr* was found in the transcriptome data. This partial *Hh-Scr* is 816 bp long and spans 76 bp of 5'UTR through a coding region of 246 amino acids that includes a partial homeodomain of 36 amino acids (Figure 4.4A). It is 100% identical to a partial *Scr* gene (185 amino acids) that was isolated from the southern green stink bug (*Nezara viridula*) (Tian et al., 2011). The partial homeodomain has a *Scr* signature sequence at the N-terminal arm of the homeodomain (highlighted in Figure 4.4 B) and is 100% identical to that of *Drosophila melanogaster Scr*. A 327 bp sequence, including 276 bp upstream of the homeobox and 51 bp in the homeobox , was amplified using RT-PCR with forward primer bMSBScrFT7 5'-TAATACGACTCACTATAGGGAGAGCAGGACCTGACTACGTCCTC-3' and reverser primer bMSBScrRT7 5' – TAATACGACTCACTATAGGGAGATCCAGCTCCAGCGTCTGGTA-3', and used as template to make dsRNA (Methods described in chapter3, section 3.3.2).

A

```
  1   MSSYQFVNSL ASCYQQAGRP SPDGPQSPDY YPQVSYPGCY SPQQYTSGYM
 51   QQSPSGMMDY TQLHGSNHQR LASHLQPLGH PAGPVSPALN NNTVTNLSGS
101   TSCKFADSTT TSSGIASPQD LTTSSSGPGG PRSTPPKTGL HSPSGARAPS
151   APAPTSQTSS SPASSISSSS STTQGTAAKS PAQPGQNPPQ IYPWMKRVHL
201   GQSTVNANGE TKRQRTSYTR YQTLELEKEF HFNRYLTRRR RIEIAH
```

B

```
Hh-Scr   NPPQIYPWMKRVHLGqSTVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAH
Dm-Scr   NPPQIYPWMKRVHLGtSTVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAH
```

Figure 4.4 Sequence of *Hh*-Scr and alignment with *Dm*-Scr. (A) The sequence of *Hh*-Scr. A cDNA encoding a portion of the *Hh*-Scr gene was isolated. This cDNA would encode a protein of 246 amino acids, including a portion of the HD. (B) Alignment of the partial HD and sequences around YPWM motif of *Hh*-Scr and *Dm*-Scr. HD of *Hh-Scr* is indicated by blue. Note 100% conservation of the region around YPWM motif and HD. The homeodomain N-terminal arm Scr signature sequence (Furukubo-Tokunaga et al., 1993) is highlighted in yellow.

Since parental RNAi was shown to function in another member of the Hemipteroid Assemblage, *Oncopeltus fasciatus*, I began with protocols used for that species and modified them as necessary for BMSB. The dsRNA was injected using a Hamilton syringe and needle as described previously for *Oncopeltus fasciatus* (Liu and Kaufman, 2004). dsRNA was loaded into a Hamilton syringe for injection. The BMSB females were anesthetized under CO_2 and the Hamilton needle was punched into abdomen between the third and fourth abdominal sternites (Figure 4.5).

Figure 4.5 Injection of dsRNA. dsRNA was injected into the abdomen of a female BMSB with a Hamilton Syringe. The injection point is located between the 3rd abdomen sternum and 4th abdominal sternum, away from ventral midline. N indicates the needle. Red arrow indicates the top of the needle and the point of injection.

I first injected five BMSB females with *Scr* dsRNA at a concentration of 2ng/ul and 5ul per bug. I injected ten female BMSB with *Scr* dsRNA at a concentration of 3ng/ul and 6ul per bug. After injection, the needle was held at the injection site for approximately one minute to prevent leakage from the injection site. The injected females were kept separately in a new cage for one day before any male BMSB was added.

The eggs laid by the first five injected females did not show any obvious phenotype. For the ten female BMSB which were injected with higher concentrations

of *Scr* dsRNA, three of the ten females died within three days after injection before laying any eggs. Five of them died within the first two weeks, one died within the third week, and only one dsRNA injected BMSB lived more than four weeks after injection. Within the first two weeks, a total of eight egg masses were laid: 143 out of 146 embryos hatched. Of these 143 embryos that hatched into 1st instar nymphs, all had abnormal proboscises (Figure 4.6 B, C). The proboscis of wild type 1st instar nymphs is needle -like in shape and has a sharp tip (Figure 4.6 A). The most severe phenotypes seen in 1st instar nymphs hatched from eggs laid by *Scr* dsRNA-injected females had a bifurcated proboscis, and claws were seen at the tip of the proboscis (Figure 4.6 B). Some 1st instar nymphs had less severe phenotypes; for example, in some cases, instead of being bifurcated, the end of the proboscises expanded into a bat shape but claws were still seen on the tip of the proboscis (Figure 4.6 C). In the third week, two egg masses were laid, (a total of 41 eggs) and all hatched without obvious defects. No eggs could be collected after the third week.

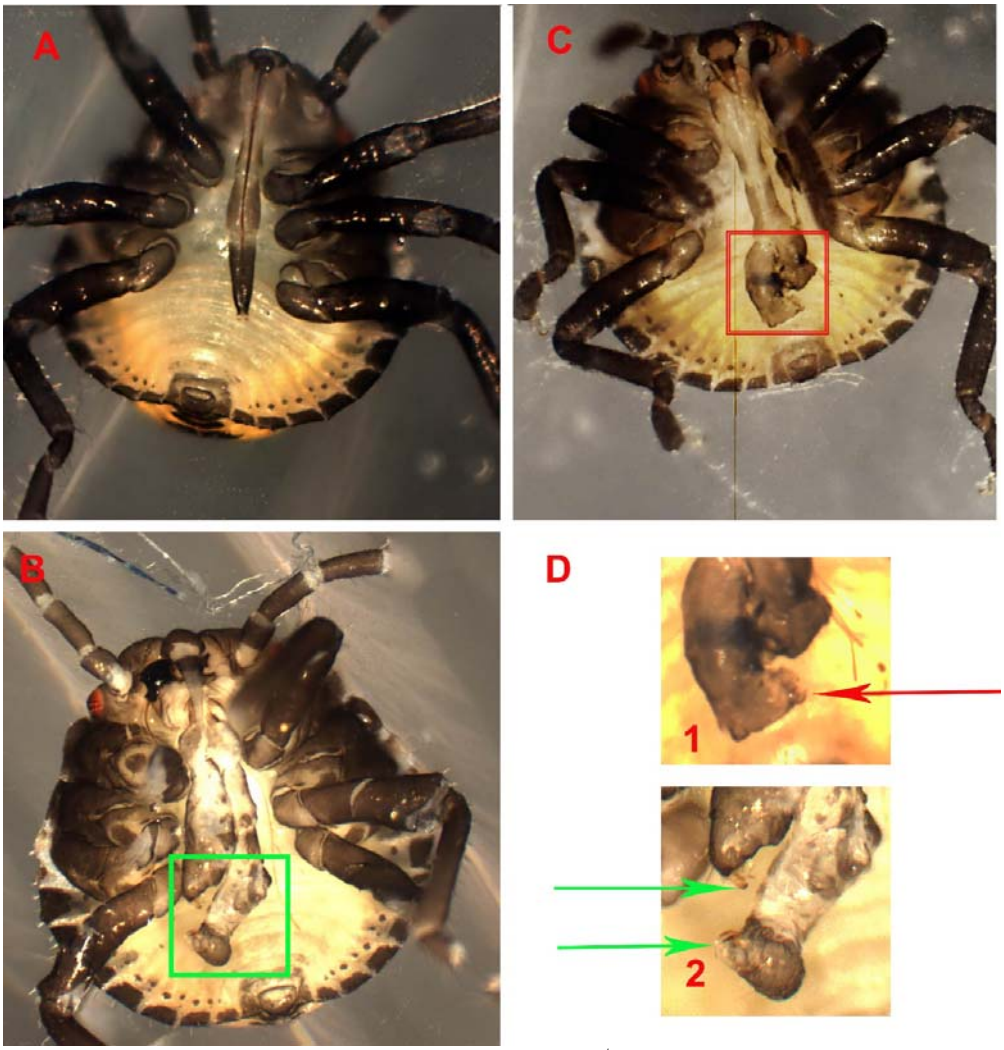

Figure 4.6 ***Scr*** **pRNAi phenotype.** Photos of 1st instar nymphs are shown. (A) Wild type. The proboscis has a needle-like shape with pointed tip; (B) *Scr* pRNAi. 1st instar nymph with severe phenotype has a bifurcated proboscis (green square); (C) 1st instar nymph with less severe phenotype has a bat-shaped proboscis (red square); (D1) tip of the proboscis from figure 4.6 C (red square region), red arrow indicates a claw, (D2) tip of the proboscis from figure 4.6 B (green square region), green arrows indicates claws.

We noticed that the severity of the defects attenuated as time went on. Embryos in the first five egg masses, laid in the first two weeks, showed the most severe phenotypes (bifurcated proboscises); the last three egg masses that were laid within the first two weeks has less severe phenotypes (bat shape proboscises); by the third week, eggs were all normal. Overall, these results show that parental RNAi works

well in BMSB. The phenotype observed with knock down of *Hh*-Scr was similar to that found in other insects (Hrycaj et al., 2010; Passalacqua et al., 2010).

Section 4.4 Discussion

engrailed as a marker for segments

engrailed is a homeobox gene, with a divergent homeobox (Poole et al., 1985). In most arthropods, *en* is expressed in the posterior part of each segment. In *Drosophila, en* was found to specify the identity of the posterior segmental compartments and it is believed that this function is conserved in arthropods (Peel et al., 2006).

In *Oncopeltus fasciatus* (Hemiptera), *en* is expressed in the posterior portion of each segment (Liu and Kaufman, 2004) and RNAi knock down disrupted body segmentation (Angelini and Kaufman, 2004). I showed that in BMSB, *en* is expressed in a similar striped pattern, in the posterior part of every segment. Thus, *Hh-en* will be useful for determining the register of the segments for studies related to BMSB segmentation.

Scr RNAi

In *Drosophila Melanogaster*, *Sex combs reduced* (*Scr*) plays important roles in labial and T1 segment identity specification (Pattatucci et al., 1991; Sato et al., 1985; Struhl, 1982). *Scr* mutations in *Drosophila* cause the transformation of T1 to T2 legs and the transformation of labia into maxillae (Sato et al., 1985; Struhl, 1982; Wakimoto and Kaufman, 1981). In *Oncopeltus*, *Scr* parental RNAi caused transformation of the labia towards leg-like identity (Chesebro et al., 2009; Hughes and Kaufman, 2000). T1 to T2 segment transformation was not observed with parental RNAi, while T1 to T2 leg transformation was observed when nymphs were

injected with *Scr* dsRNA (Chesebro et al., 2009). My *Scr* pRNAi result in BMSB is similar to the parental *Scr* RNAi in *Oncopeltu*s. All abnormal proboscises had claws which is a character of legs, and no T1 to T2 leg transformation was observed. In this experiment, dsRNA was injected into the abdomen of the female BMSB and the eggs laid by the injected females were affected. As mentioned above, this phenomenon is defined as systematic RNAi (Whangbo and Hunter, 2008). An effective systematic RNAi requires the organism to have the ability to uptake the dsRNA into the cells, machinery to amplify the initial RNAi signal (Carthew and Sontheimer, 2009), and to export it to other tissues in the organism (Huvenne and Smagghe, 2010). My results indicate that BMSB has the machinery to amplify the dsRNA signal. The amplified signal can spread to other tissues and be taken up by individual cells in BMSB. Thus, BMSB is a suitable target for RNAi based pest-control experiments. It also indicates that it is possible to use RNAi to knockdown specific genes, which will elucidate the function of these genes in BMSB.

For the *Scr* RNAi experiments in BMSB, the phenotype seemed to attenuate as time went on; the eggs laid after the third week hatched without obvious phenotype. In contrast, the phenotype did not attenuate when I carried out similar experiments with *Oncopeltus* (data not shown). *Scr* dsRNA injected *Oncopeltus* females laid eggs with proboscis defects for five weeks until they were infertile because of aging or until they died; none of the eggs they laid hatched without a defective proboscis. This may suggest that BMSB has a less robust RNAi system than that of *Oncopeltus*. It had been shown that targeting different regions of the same housekeeping gene generated different mortality effects (Baum et al., 2007; Li et al., 2011; Mao et al.,

2007). I tried to use one region in the *Of-Scr* gene for RNAi knockdown, and two regions in *Of-ftz-f1*; none of these showed attenuation of the phenotype during the whole life of the bug. All our dsRNAs were made about the same length (~300bp), so differences in length may not account for the differences seen here.

In addition, we observed that injected BMSB females died earlier than un-injected male bugs, which were kept in the same cage from the second day of injection. However, general toxicity of the injection remains to be investigated since the dsRNA-injected females and un-injected males were picked randomly from a lab population; some of them maybe older than others, which means some of them may die quicker even without any intervention. We also noticed that the whole abdomen of the injected bugs turned black one week after the injection. This could be a sign of infection. In *Oncopeltus*, the dsRNA-injected females did not have a shorter life time than the un-injected *Oncopeltus* females and we didn't observe the black-abdomen phenomenon in dsRNA-injected *Oncopeltus*. This difference may be caused by the difference between their abdominal structures. In *Oncopeltus* there is a membranous structure between each sternite. During injection, the needle is punched through this membranous structure and when the needle is removed, the hole closes due to the properties of the membranous structure. In BMSB, there is no membranous structure between each sternite. The injection creates a hole at the injection site and that hole never closed. We suspect that the opening made the dsRNA-injected BMSB more susceptible to bacterial infection, which decreased their longevity. If this is true, using something to seal this hole after injection may help to increase longevity.

Chapter 5: Overall conclusions and future direction

Combining results from my studies of PRGs in *Oncopeltus* with other published studies, I conclude that the PRG network is flexible in insects. Some genes that function as PRGs in *Drosophila* do not have pair-rule function in other insets. For example, in *Tribolium Tc-ftz* does not have a function in segmentation (Brown et al., 1994a; Stuart et al., 1991). Similarly, in *Tribolium*, knockdown of *Tc-h* or *Tc-Opa* with RNAi showed no effect on segmetation (Choe et al., 2006). In *Oncopeltus,* my study described in Chapter Four indicates that *Of-ftz* and *Of-hairy* do not have segmentation function. At the same time, some genes, whose orthologs in *Drosophila* are not PRGs, have pair-rule function in other insects. For example, in *Oncopeltus, E75A* was shown to be expressed in pair-rule stripes and RNAi knockdown caused pair-rule segmentation defects (Erezyilmaz et al., 2009). The flexibility we observed in the PRG network is unexpected, given that changes in the function of embryonic regulatory genes are highly deleterious, usually lethal, in the lab, a much less harsh environment than the wild.

Many of the studies in our lab and in my thesis have focused on the partner pair *ftz* and *ftz-f1*, which have been shown to vary extensively in arthropods. In my work, I found that *ftz* may have even lost function completely in *Oncopeltus*, since both two isoforms of *Of-ftz* mRNA identified so far do not have a promising ORF. In addition to this, when I aligned five Ftz HDs from five hemipteroid assemblage insects and compared this to Ftz HD alignments from five other more distantly related species

(Figure 3.7), I found big variation in the Ftz HD in this superoder. *Hh*-Ftz and *Ph-Ftz* HD are conversed, with YPWM motif upstream of the HD. *Of*-Ftz and *Ap*-Ftz are less conversed and don't have a YPWM motif. These findings suggest that there may be functional differences among Ftz orthologs in this superorder. Future studies of the function of these genes which will help us to better understand how *ftz,* and other genes involved in segmentation, evolved.

Further, *ftz-f1* varies in its expression pattern in different insects. In *Drosophila*, Ftz-F1 is present ubiquitously in all somatic cells at the blastoderm stage. I found that in *Oncopeltus*, *ftz-f1* is expressed in stripes that appear to have single segment periodicity at the blastoderm stage. Although functional studies were inconclusive, this result suggests that *ftz-f1* may have function in *Oncopeltus* segmentation. Future studies are needed to firmly elucidate the function of both *ftz* and *ftz-f1* in *Oncopeltus* and other hemipteran insects.

Another point highlighted by my work is the fact that expression pattern is not always predictive of function. In both *Oncopeltus* and *Tribolium,* Ftz orthologs were found to be expressed in stripes but neither appears to function in segmentation (Chapter Three; Brown et al., 1994a; Stuart et al., 1991). In contrast, some *Drosophila* PRGs do not have a stripy expression patterns but do function as PRGs, impacting the development of alternate body segments. Examples of this are *Dm-ftz-f1* and *Dm-Opa.* Thus, showing that a gene is expressed in stripes is not sufficient to conclude that it functions in segmentation. Further, these findings highlight the importance of co-factor interactions in regulating the activity of pair-rule proteins. In *Drosophila,* Ftz-F1 and Ftz interact with each other and mutually depend on each

other for their segmentation function. *Dm-ftz-f1* is not expressed in stripes but its function is constrained by the stripy expression of *Dm-ftz* (Guichet et al. 1997; Yu et al., 1997). Given the fact that in both *Oncopeltus* and *Tribolium, ftz-f1* orthologs are expressed in stripy patterns and have segmetation function (data in Chapter three, Heffer et al., 2013), their co-factor(s) in these species do not need to be expressed in stripes. This could also be true in other insects and for other PRGs, which could lend some flexibility to the expression patterns of PRGs during evolution.

When I think about long germ band and sequentially segmenting insects, I expect that there will be differences in how PRGs function between these two different modes of development. In sequentially segmenting insects, since most of the posterior segments are added sequentially during development, I would expect that knockdown of any PRG would delete the entire segments posterior to the first sequentially added segment that PRG specifies. This is partially true as RNAi knockdown of *Tc-eve, Tc-Odd* and *Of-eve* generated asegmental embryos (Choe et al. 2006; Liu and Kaufman, 2005). However, RNAi knockdown of some PRGs (including *Tc-ftz-f1*, *Tc-odd* and *Of-ftz-f1)* in some sequentially segmenting insects caused *Drosophila*-like pair-rule phenotypes (Choe et al., 2006; Heffer et al., 2013). One suggestion is that the genes that cause asegmental phenotypes are primary PRGs. But why do primary PRGs in *Drosophila* not cause the same kind of phenotypes? I think the answer is to be found in the way that PRG expression is generated. As I mentioned in Chapter One, in *Tribolium,* a segmentation clock was found, and expression of *Tc-odd* was found to be oscillating (Sarrazin et al., 2012). This phenomenon may explain the differences among the function of different PRGs in sequentially segmenting insects. The PRGs

that cause asegmental phenotypes are the genes that participate in the regulation and generation of the segmentation clock. Blocking the function of these genes blocks the segmentation clock; thus no new stripes can be generated. The PRGs that cause pair-rule phenotypes are the genes that are regulated by the segmentation clock. Blocking the function of these genes does not block the segmentation clock and posteriorly located segments are still able to form because of signals from the segmentation clock. Tests of this hypothesis would need tools to closely study what happened within the growth zone. These tools include transgenic techniques, techniques to generate specific mutations, live imaging techniques with deep field depth, and more.

Appendices

Gene sequences that I isolated for this thesis..

BMSB *Scr*, partial, length=738

ccttagtcgagtcttgttcaagtcctgtccggggtgagaccgtgatgtgagtggaacctctgatagctgtgacaagatgagctcgtac
cagttcgtcaattcgctggcctcgtgctaccaacaggccggccgcccgtcccccgatggcccccagagcccagactactacccccaa
gtcagctaccccggttgttactctccgcaacaatatacttcaggctacatgcaacagagtccttccgggatgatggactacacccagc
tgcacggcagcaatcaccagcgtcttgcttcacatttgcagcccctagggcatccagcgggtccggtatcgcctgctctcaacaaca
acacagtgaccaacctgtcaggatcgacctcctgcaagtttgccgacagtaccacgacaagttccgggatcgcctccccgcaggac
ctgactacgtcctctagcggtcctgggggcccaaggagcacgcctccgaagaccgggctccactcgccgtctggagcgcgcgcccc
gtcggctccggctccaacatcccagacatcttcttcaccggcttcttcaatatcctcatcatcgtccacgacacagggaactgccgca
aagagtcccgcccaaccgggacagaacccgccgcagatatatccatggatgaagagagttcatcttggacaaagtacagttaacg
ctaatggggagacgaagaggcaaaggacctcctacacccggtaccagacgctggagctggagaaggagttccacttcaacaggta
cctgacgaggcggaggaggatagagatcgcccacga

BMSB *en*, partial, length=556

actcaatccgaatgcgggtgtgagatggcgttggaagcggagcggcgaccctcggctttccacaggatcgagttgctgacttgcaa
ggatgcacccccggcaccggcatcgttacccttctccgtcgacaacatcctccggccggagttcggccaagccgcccggcgtccggt
caggctatcgaggtccgacgacgaaggaccaagtcccaccaagatcgacccccccaataccggacgatcccaatggacccgtctgg
cctgcttgggtctactgcaccagatactccgaccggcccagttcaggaggtcctcgctcgaggaggataaagaggaaggacaaga
acaaagaagacaagaggccgcggaccgccttcagtggcgaacagctggccaggctaaagaccgagttcagtataaacaggtacc
tgacggagagacggagacaagccctggcttccgaactaggcctcaacgaggctcagatcaagatctggttccagaacaagagggc
caagatcaagaaggcctccggcaacaggaaccccctcgccctccagttaatggc

BMSB *eve,* length=1401, full CDS=663

agttgattcagttctgtgaaagatggagtttggatatactattgactcgcggctgagccagcagatgctgaagtccaacggagttccg
tcacaaattgttgtcgatctcttgccacccgcataccagaagtctttggctccggagagcaacgcggaatcgcccaccagaacagaa
gaaataaaacaacaagaaagtaccaatgtggtccctgaaggcaacataaggcggtataggactgcctttaccagagagcagttaa
acagactggagaaggaattttataaagaaaattatgtttctagaccaaggcggtgtgaactggcaactcaactgcagttaccggaa
gccacgattaaggtgtggtttcaaaatagaaggatgaaagacaaacgacaaagaatggcaatggcctggccctacgcgatgtata
ccgatccaacaatagcggcgtcgctgctcacctccttacattacccttacgcccccctattaccacccgcggtactatttgcccccgccg
ccgccacctcctgcaccggtctgcgatggatcatccagctgccgctgcggcatcgtcaactgcgtcgcctcttaccctagccctccatt
gacaacatccccagtcactgttgagcaaaagaaaccccctgtttcaaccatataaggagcatgtcatcgagaaggcgtagcaatcttt
cgccgaatttttcctatgtaaataagttcatgcggatccgctttttcgggttcggttgttatttaatttccttttgaaagagatatcaagt
ataaaaacacacacagacatatatatatgtatgtatgtatgttttaagttaaatttacagaatgttctggactacttatggactaagta
attctgatttttttttttaaatctataaagaaacataaccagtgatagcaggaagatagtccgggtggatcttcctccgatctagttcat
agtcaactttcttcttccaatatgaccagtgataatttagatataatgcaatggtaccaaaaccaagtgaaaatgttcgtatccagtg
acaaaatgaacattatcgtatgataataatgaaactaacccccaaaagctactggtagctaaggtgctatatgggtgatttgacatt
gtttattgtatctagattaaaaattaaaactggttttatttttttgtgaaattggaattttctctttacaagtgggtcttctatacatcaat

gtttagttttgacataatttcataaatttagtttttctttttggtttctgtttttaaactaacttttttttgtagtgtatgtacacgtaaagcac
acttgttaagaaaattcttctaggaaataacagtgataaaatttttatatttatataaatagtttttattatatt

Of-ftz Isoform A, Ewan-Campen's data , length=788

ttagttagggtttcggtcttgcaagctagaagaattaaaaactgatttgtaaaaaaaaatccatgactctggacaatcttttattta
tagataagaagataatcatatttctttatagataagtcttgctctgggcgggatttgaacccacagcgtaaggattgagatccct
tagccgcccgaagtgagggtttacataacatggacagttgtaaaatgatgaatttatttttattatcatgttgttgaagtatctac
tcaaatatactattgtctatattataacttcaccagcttgcactgtattagtaaatgtatctcttgtgagataaactatttttaaatttt
ttatttttattaaagttatcaataattagtggttgttcattattttattatgcttatgactgatgaacatttaaattaagtattaaatttttactt
taaaattgttaatgtgcatattttacttttcggtttccagatccaagtaggccgaagcggaagagacagacgtattctcgattcc
agactttggaactagaaaaggaattccgcctgacaagatatctcccgaggaaacgtagaatagacctggcggagagcctg
ggcctgacagagagacagatcaagatctggttccagaacaggaggatgaagcagaagaaggaacaggagagcaaacc
catccagtccgaggacctcttggtgcaggcgacgaggaatcccttcccggatatagcgtccttcgtgcaggatctacttccg
acagtgaacccgctccaggc

Of-ftz Isoform B Yong's 5' 3' RACE, length=443

Tttgaacccacagcgtaaggattgagatcccttagccgcccgaaatccaagtaggccgaagcggaagagacagacgtat
tctcgattccagactttggaactagaaaaggaattccgcctgacaagatatctcccgaggaaacgtagaatagacctggcg
gagagcctgggcctgacagagagacagatcaagatctggttccagaacaggaggatgaagcagaagaaggaacagga
gagcaaacccatccagtccgaggacctcttggtgcaggcgacgaggaatcccttcccggatatagcgtccttcgtgcagg
atctacttccgacagtgaacccgctccaggcttaactgctcatcccgaagaggacatgaggactgaggactataacctcttc
gctgagcacgtcgctagataactagactaaaaaaaaaaaaaagc

Six *Of-ftz* partial sequences isolated from Ewan-campen's raw data, listed in fasta format

>noNorAll.1030880 GEQE5QV02JTNX1 length=597

ttagttgttgtagtaatatactactaataatactattagtctattattataacttactacctacgacttcgtntacttagttactttaagt
taaaattagttattacttacttagttgtagtattaaactaatttaaatttttttttaatttttttattttaaagtttaattctaataattaagtgg
ttgttctatatttattattgcttatgtactgatgaacatttaaattaaagtataaatttactttaaaattagttaatagtgcataatttactt
tggtttcctagatccaagtaggccgaagcgaagagacagacgattctcgattccagactttggaactagaaaaggaattac
cgcctgacaagatatctcccgaggaaacgtagaatagacctggcggagagcctgggcctgacagagagacagatcaag
atctggttccagaacaggaggatgaagcagaagaaggaacaggagagcaaacccatccagtccgaggacctcttggtg
caggcgacgaggaatcccttcccggatatagcgtccttcgtgcaggatctacttccgacagtgaacccgctccaggcagtc

>noNorAll.1267703 GEQE5QV02HHPA6 length=630

cttaagttacggtttggttaagtaagttacttacctacctaaactaacttaataactttacgttcgttactatttaattaaacttcttaa
cctaaccgcttaggtcactacgtattagttaaattagtaagttntaccttgttgtagagttaaaacatttaaaatttttttttaatttttttt
tattttaaagttttaaattctaataatttaagtaggtttggttctatatttttaattattagcttattgactgatgaacatttaattaaagta
taaatttactttaaaattagttaatagtgcatatttactttggtttcctaganccagtaaggccgaagcggaagagacagacgta
tctcgattccagactttggaactagaaaggaatcgcctgacaagatatntcccgaggaaacgtagaatagacctggcgga
gagcctgggcctgacagagagacagatcaagatctggttccagaacaggaggatgaagcagaagaaggaacaggaga
gcaaaccccatccagtccgaggacctcttggtgcaggcgacgaggaatcccttcccggatatagcgtccttcgtgcagga
tctacttccgacagtgaacccgctccaggcagtc

>NorAll.156560 GESJTKM01BVLD2 length=644

gttaaaaaattaaacgtaagtttagttaaaaaaattaaaacctaaactagtactacttggtaacgaataatttacctttttattaagtt
ataaagtaagaagtaaattactactaattacttttatttaagtagtaaagtacttgtcgtctaggcgggagttgtaaccgtacaac
gcgtaaggagtagtagatccttacgtccgcccgaacgtagagggttacgtatacaatggacagttgtaaaattagatgaatt
agtttttttatttattctatgttgttgtaagtatctactcaaatataactattagtctatattataacttacaccagcttgcactgtattag
taaatgtatctcttntgagataaacatttaaaatttttttatttttattaaagttatcaataattagtggttgttcattattttattatgcttat
gactgatgaacatttaaattaagtattaaatttttacttttaaaattgttaatgtgctttattttacttttggtttccagatccaagtaggc
cgaagcggaagagacagacgtattctcgattccagactttggaactagaaaaggaattccgcctgacaagcccggcctcg
tggccactctgcgttgataccactgcttagtc

>NorAll.415171 GESJTKM02GRK1B length=669

ttttttaacgttattagttaggttttgttaacggttaaagttacttaactttacctaaacttaaattaacttaactttagttacgttattaa
ctttattaaactttacttaacctaacgacgttagtcgtactacgttattaggtttaaaattaagttattacttacttgttgtagtagtta
aactaatttaaaatttttttaaatttttttaattttaaagtttaaatttataataattaataggttggtttcgtattattttatttattagtcttatt
gtactgatgtaacattaaatttaaagttattaaatttacttttaaaatttagtttaatagttgtcatatttacttttaggttttcctacgat
ccaacgtaggcagaagcggaagagacagacgattctcgattccagactttggaactagaaaaggaaattaccgcctgaca
agatatctcccgaggaaacgtagaatagacctggcggagagcctgggcctgacagagagacagatcaagatctggttcc
agaacaggaggatgaagcagaagaaggaacaggagagcaaacccatccagtccgaggacctcttggtgcaggcgacg
aggaatcccttcccggatatagcgtccttcgtggcactctgcgttgataccactgcttagtc

>noNorAll.516047 GEQE5QV01BYSSH length=554

tttactatgttntttgagtatctatcaatatactattagtctatattataacttatacctaagcttgtcactgtatttgtaaattagtatnt
cttgttgtagntaaanatttaaatttttattttttattaagttatnataattatgttgttcatatttattatgcttatgactgatgaacattta
aattaagtattaaatttactttaaaattgttaatgtgcatatttactttcggtttccagatccaagtaggccgaagcggaagagac
agacgtattctcgattccagactttggaactagaaaaggaaattccgcctgacaagatatctcccgaggaaacgtagaatag
acctggcggagagcctgggcctgacagagagacagatcaagatctggttccagaacaggaggatgaagcagaagaag
gaacaggagagcaaacccatccagtccgaggacctcttggtgcaggcgacgaggaatcccttcccggatatagcgtcctt
cgtgcgggatctacttccgacagtccccggccttcgtggcagtc

>noNorAll.657448 GEQE5QV01ES05F length=623

Ttttattntatntgttntagatctaatcaatataactattgtctattattattaaactttactacctaacgcgtttagtctactggtaac
ttaggttaaaattagtattactcttnttagtagattaaacaattttaaatttttttaatttttttatttaagtttaattctanaattagttagtt
agtttcattattttatttattagcttattgacttagatgtaacatttaattaaagttataaattttactttaaaatttagtttaattagtgtca
taattttaacttttttaggtttttcctacgatccaactaggccgaagcggaagagacagacgtattctcgatttaccagaaccttttt

ggaactagaaaaggaaattaccgcctgacaagatatctcccgnggaaacgtagaatagacctggcggagagcctgggc
ctgacagagagacagatcaagatctggttccagaacagaggatgaagcagaagaaggaacaggagagcaaacccatc
cagtccgaggacctcttggtgcaggcgacgaggaatcccttcccggataaagcgtccttcgtgcaggatctacttccgaca
agtgaacccgctcccccgagtc

Of-ftz-f1 sequence, length=2081, full length CDS length=1962

gaagtgcaagtgaagtggcctagagacntacattttgtgggacgttcgatttttaactgtgattctctgctgtagagccttggc
atagaaaagagcccagggctttatggtgttgagaaggtacgcgcttactggaacaggacaagcccaggggaaataaaat
gcgggaagacgagctgcccgtcagcgttccgaccagttctgcccacgaaactagcacagatctggatccattagtatcaa
acgatcaaaataattcgacagagcttcaagtgtcttactcaacgaattcttcagatgcatactctgtcgaaggtaaggtgcttc
gggccgtatctgacgatgctgaagagtgttcctaccagattattttgtgcgggggggagcaagccgcagccttccagcccg
tcctccagaacgtacccgaattgccggacaccaaggagggcatagaggagctctgtccggtctgcggagacaaggtgcc
tgggtaccactacgggctcctcacctgcgagtcctgcaagggcttcttcaaacggaccgttcagaacaagaaagtgtacac
ttgcgtcgccgagaggagctgccacatcgacaagacccagaggaagaggtgtcctttctgccgctttcaaaagtgcctgg
atgtcggaatgaaactcgaagctgttcgagcggacaggatgcgaggcggtaggaacaaattcggcccgatgtacaagcg
cgaccgagcgaggaagctgcagatgatgcgacagaggcagatggcgctgcagacgctgcggaactcggccggctaca
gcagcacgagcggcgaaggcgtctccctcagctacgcgggggccggctccaacttcggctcgctccagatcaagcagg
agatccagatcccgcaagtctcgtcgctcacctcgagccccgactccagtcccagtcccatcaccgtcgccctcggccag
caccagcagtcccagggccagggcctgtaccaggcggccgtcctaccggcccagcagctccaggtcgagcccaggca
ggcccagcagtcccaggtcttcgagcagggcaagtcctggtccgccccgagccccccgccccacaagtccttccctac
gagggctcgtcctccggcaacagccacaacaaggtggcgcccatgatccgggacttcgtccaggcgatagacgaccag
gagtggcagacctccctctacaacctcctacagaatcagacctacaatcagtgtgaggtggacttatttgaacttatgtgtaa
agtgttggaccaaaatcttttctcacaggttgattgggcacgaaattccacatttttcaaagacctaaaggtggatgaccaaat
gaaattattacagcactcatggtcggatatgctggttctggatcacatccaccaaaggatgcacaataacctcccagatgag
atgaccctcccaaatggacaaaagtttgacctcttatccctcggccttctcggggtgcctagtcttgccgagccctttgccga
gatcacctcacagcttcaggaccttaagttcgatctctctgactacatctgcgtaaagttccttctcctgcttaatccagatgtgc
ggaatctgatgaataggaagcatgtccaggagggccatgaccaagtccagcaagcccttcatgattactgtctcggatcat
atccacaggtccaggacaagttcaacaagctgctcctaatcttgccagaaattcatggggttgccgtgaggggagaggaa
cacctctatatgaagcactgcaatggtggagcacctactcaaacacttcttatggagatgctccatgcaaaaagaaagtagg
atcataggaaaacagataggagtagatgaaaaggcgagctcaccaacctgaagtgcaatatgagtggactggaggggat
gatgtgttcataagtttaatctcgagttacctagcgtgtccacattgaaacggctcatacggctctgtgtgtaaaggactcttctt
cctctantgttanctttgtatgtgttgtgcgttgcgcatgttcgtcttataaggacta

Of-hairy length=580bp, full length CDS length=555

tacgcgatggtcactggaagtcttgctccgtcagctgcctccaggcctcccgactctaacagaaggagtaacaagcccatc
atggagaaacggcggagggctcggatcaaccattgcctcaatgagctcaagtccctcatcctggactccatgaagaagga
cccggcgcgtcactcaaagctggagaaggcagacatacttgagatggcggtgaagcacgttgagacgctccagcggtc

ccaggcggcactccaagtagctgccgacccaggagtactggataagttcctggcaggctgggggaatgtgttggagag
gttggcaggttccctggtcttgactctcaggtgagaaaaaggctggttgaccatctctcaacagttatggaaagacgaagtg
aggcaactcctgtagttactctggtccctactaggctggctagtggggagatagcgcttcttgttccttgccaagcagcacct
ccgttacctcagcctcctaaacttcagcagcctatttcctttgttgctgataaggaggagaaaccttggagaccctggtgaca
cggccac

Four RUNX family proteins sequences in fasta format

>O*f*-Run

MHLPGVTTEMWGEMYGGLCSALSEHHGELIQTGSPSVLCSALPTHWRSKSLPTAFK
VIALDDVSDGTVVTIKAGDEYCAELRSTAVMKQVAKFDLRFVGRSGRGKSFSLTIVI
GSSPHQMATYSKAIKVTVDGPREPRSKSFHYMTGGGAPLGPLGFSLLPPGWLDAAYL
TAYWPDYLRRPPAPPLDLVKVPPLPGHHPQTPTVLPPSPPKESCSPPAIVTPGSELLPPP
LPLLPPLLPPQASAFKPPPISPPKEHTTKTKPTSKATVWRPY

>*Of*-RuxA

MHLTPAVVEGGGPLAEVYVKMTSDILAERTLGDFLAEHPGELVRTGSPLFVCTGLPS
HWRSKTLPVGFKVVALGEILKTDLKRDITYTEYMYAIAELYWLSLTRACWKFQRKS
FTLTITVSSSPPQVATYTKAIKVTVDGPREPRSKTICEGIQGEILR

>*Of*-lz

MTSDGKKRRRGGGGEAAAASDLWWTERVVVEAQQEHPGELVRTGSPYFLCSALPT
HWRSKTLPVGFKVVALGEVMDGTVVTIRAGDEYCAELRSTAVMKQVAKFDLRFVG
RSGRGEELHTDHHSRLLSPASGHIHQGHQGDCGRTSGTQVQPSVSAIQSSRPGSEAVP
RTPSTFWLQSPTQAAGSPGPLQVAPQPPREHGE

>*Of*-RunxB partial

ELLAERAIESLLAEHPGELVRTGSPHVVCTVLPSHWRSNKTLPVGFKVVALGEVMDG
TVVTIRAGNDENYCAELRKELRNSTAVMKNQVAKFNDLRFVGRSGRGE

Of-actin sequence, length=3172, first 1200nts with high confidence

gactctcctcctggagaccaacggctcaatcctccttcactcccgtcaattgtgtgatagtgtacgcattggctctcgcctcctt
cgtccagtaaagcttatttaactaacccaaaacaaaatgtgtgacgaagaagttgctgcactcgttgttgacaatggatccgg
aatgtgcaaggctggcttcgctggcgatgatgcaccaagggctgtcttcccatcaattgttggaaggccaaggcatcaggg
agtgatggtcggtatgggacagaaagacagctatgtcggtgatgaggcccaaagcaagagaggtatcctcaccctaaaat
accctattgaacacggtatcgtcactaactgggatgacatggagaagatctggcatcatacattctacaatgaactcagagtt

gcccctgaagaacatcctgtcctccttacagaagctccactcaaccctaaggccaaccgagaaaagatgacccagatcat
gtttgagaccttcaacacccccgccatgtatgtcgctatccaggcagtactgtccctatacgcatctggacgtactactggtat
tgtgttggactcaggtgatggtgtctcccacactgtccccatctacgaaggttatgccctgccccacgctatcctcaggttgg
acctggctggccgagatctcacagattacctcatgaagattcttaccgagaggggatactccttcaccaccactgccgaac
gagaaattgtcagggacatcaaggaaaaactttgctatgttgcccttgactttgaacaggaaatggctacggctgccagctc
cagctcccttgaaaagtcctatgaactccctgatggtcaggttatcacaattggcaacgagaggttccgatgccctgaagcc
ctcttccagccctcgttccttggtatggaagcttgtggaatccatgagactacttacaattccatcatgaagtgtgatgtagaca
tcaggaaggacttgtatgccaacacagtcctctctggaggtaccaccatgtaccctggtattgccgataggatgcagaaag
aaattactgcccttgctccctcaacaatgaagatcaagatcatcgctccccccgagaggaaatactctgtatggattggagg
atctatccttgcttctctttctaccttccagcagatgtggatctcaaagacaggaatacgacgagtctggcccatccattgttca
caggaagtgcttctaagcagtttgttgcattttctcgtctcactccccttttacatattctctttcgggagacagcttattcttttgtat
tatactgtttattatgtttgtttatttgattttaaatattattcagcgacttactaagttaaattgcagttgctcattgagttgatctgcaa
tttcttccattatttaaatccattccatataattggtcttatctccatgtaccatccccctttttttatgcccattttataaagggcttatt
ttttcaaccgttttctattgtactagttatctttataagaaaatattttatataatatatataaatatatttaaaaaagacactttgcattc
caatttttttccatgtggctcaaaatgtgcttactttctaagtttgcgattttgaagatttagaaagcgtaggtttgtcttttgtattaag
atcatgttgtctgttggttcctttaatggacatttgagcctgtgaattcggtgacacctatgtatgcactaggccaaagtattattt
attcccatcagcagaactggaattttagattatgggaataataaattataatttattgtacttctctcaaaaaaaaaggggatcatg
gtgctggagttggagaaagggatggtgttctcaatgatatggatatgatctcagggacattgggaaaggcttttggcaatatt
ggtggctatgttgcctcgagcgcgaccctcattgacatgatccgtagttatgctgctggatttatattcactacttctcttccacc
aactgttcttgctggtgctctaaaatcgattgaaatactttcctctgatgaagggcgccatttaagaaaacgccaccaagaaa
atgtagcttatttaaaaaatttattgctagaaagaggttttcctgttgagcctactccatctcatatcatacccatctatattggtga
tccctttatttgtactaaaatatcagatgaacttttgaaacaaaagggacattacattcaggccattaactatcctactgtaccaa
aaggagaagaaaaattgcgtctagctcctacacctttcataccatttctatgatggacgcgtttgttgatgacttcactgaaatt
tggcatagcctgggactaaaacatgctcagccatgtcaacattgcaagaaaatctttgatgtttgccaccaaaatgaatttttct
gtgaaatgccaaattgtcctcaagtaatggctcactaatcttaattttacttaaatatagtggtggctatagtagaaaggatacta
gctccaggtttcttttaatgtaataggatattaaaattctgacagatatgcacaagtcaactcatagatttttaaatcatgttatattt
ctcataatttaatgatttgttttgttcggttgtgtattaaatttttaaacatgattttggttaaccatttggtttgttacttaagttatccta
agatttatttattgaatttcctgaataaacaatgtaaaattctgattttccttttaaagtatgttaaatatttaatttaatgcaaactctttt
attaatgttttatcaagtccagaattggccagaactcttctgatactgaagttagatttaattaaattgtgacatttttttaatctatgtt
atcatggttttaaatatatttttttaatcaagtaaatgtgatcttgttaccagttctcaagaaagtgcctgagaggtttagaatttttat
ttataaaagttatttatggggaaaaagtttggaattacaattattaaatttattctgaatacaaattgttatttgattgtcgtatttcttg
ttaatttaaaacatgtattgttgttgtaaaatgtggaagaataaagcttatgtggaagcattaatccttcctctcaaa

Primers used for making dsRNA in *Oncopeltus*

Of-ftz RNAi primers

offtz1exT7F 5' TAATACGACTCACTATAGGGAGAAAATCCATGACTCTGGACAATCT
offtz1intrT7F 5' TAATACGACTCACTATAGGGAGAAACTTCACCAGCTTGCAC TGT

offFtz T7 5 'TAATACGACTCACTATAGGGGAAAAGGAATTCCGCCTGAC
offtzR T3 5' AATTAACCCTCACTAAAGGGGATGAGCGTTAAGCCTGGAG

ofFz3UT7F 5' TAATACGACTCACTATAGGGAGAAGAACAGGAGGATGAAGCAG
ofFz3UT7R 5' TAATACGACTCACTATAGGGAGATTATCTAGCGACGTGCTCAG

Of-ftz RT-PCR primers

ofFz3UF AGAACAGGAGGATGAAGCAG
ofFz3UR TTATCTAGCGACGTGCTCAG

Of-actin RT-PCR primers

of-actin ATGGTCGGTATGGGACAGAA
of-actinR TGTTCTTCAGGGGCAACTCT

Of-ftz–f1

off1RT7 5' TAATACGACTCACTATAGGGAGAGCCCAATCAACCTGTGAGAA
off1FT7 5' TAATACGACTCACTATAGGGAGAGTCGGAATGAAACTCGAAGC
Offf1F 5' GTCGGAATGAAACTCGAAGC
Off1R 5' GCCCAATCAACCTGTGAGAA

Of-hairy

ofHairyFT7 5' TAATACGACTCACTATAGGGAGACGAAGTGAGGCAACTCCTGT
OfHairyRT7 5' TAATACGACTCACTATAGGGAGAGTGTCACCAGGGTCTCCAAG

Of-runt

ofRuntFT7; 5' TAATACGACTCACTATAGGGAGAAGGTCACCACCCTCAGACAC
ofRuntR1T7; 5' TAATACGACTCACTATAGGGAGATTGGTGGTATGCTCCTTTGG
ofRunF 5' AGGTCACCACCCTCAGACAC
ofRuntR 5' GCCTACCAACAGCAAAGCA

Abbreviations used in the text

Al	*Archegozetes longisetosus*
AEL	After Egg Lay
Am	*Apis mellifera*
Ap	*Acyrthosiphon pisum*
As	*Artemia salina*
BLAST	Basic Local Alignment Search Tool
BLOSUM	BLOcks SUbstitution Matrix
BM	*Bombyx mori*
BMSB	Brown Marmorated Stink Bug
Ca	*Clogmia albipunctata*
CNS	Central Nervous System
DBD	DNA Binding Domain
Dm	*Drosophila melanogaster*
eRNAi	embryonic RNA interference
en	*engrailed*
eve	*even-skipped*
Fa	*Forficula auricularia*
ftz	*fushi tarazu*
ftz-f1	*fushi tarazu factor-1*
Gb	*Gryllus bimaculatus*
h	*hairy*
HD	Homeodomain
Hh	*Halyomorpha halys*
odd	*odd-skipped*
opa	*odd-paired*
ORF	Opening Reading Frame
pRNAi	parental RNA interference

prd	*paired*
Ph	*Pediculus humanus corporis*
Ps	*Pedetontus saltator*
Tc	*Triboliu castanem*
LBD	Ligand Binding Domain
Ms	*Manduca sexta*
Nv	*Nasonia vitripennis*
Of	*Oncopeltus fasciatus*
PCR	Polymerase Chain Reaction
PBST	Phosphate Buffered Saline Tween-20
PFA	formaldehyde solution made from Paraformaldehyde
PRG	Pair-rule gene
Rp	*Rhodnius prolixus*
runt	*runt*
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SSC	Saline-Sodium Citrate
slp	*sloppy paired*

References

Adachi-Yamada, T., Harumoto, T., Sakurai, K., Ueda, R., Saigo, K., O'Connor, M.B., Nakato, H., 2005. Wing-to-Leg homeosis by spineless causes apoptosis regulated by Fish-lips, a novel leucine-rich repeat transmembrane protein. Mol Cell Biol 25, 3140-3150.

Adachi-Yamada, T., O'Connor, M.B., 2002. Morphogenetic apoptosis: a mechanism for correcting discontinuities in morphogen gradients. Dev Biol 251, 74-90.

Adachi-Yamada, T., O'Connor, M.B., 2004. Mechanisms for removal of developmentally abnormal cells: cell competition and morphogenetic apoptosis. J Biochem 136, 13-17.

Akam, M., 1989. *Drosophila* development: making stripes inelegantly. Nature 341, 282-283.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403-410.

Alves e Silva, T.L., Vasconcellos, L.R., Lopes, A.H., Souto-Padron, T., 2013. The immune response of hemocytes of the insect Oncopeltus fasciatus against the flagellate Phytomonas serpens. PLoS One 8, e72076.

Anderson, D.T., 1973. CHAPTER 1 - INTRODUCTION—EMBRYOLOGY AND PHYLOGENY, in: Anderson, D.T. (Ed.), Embryology and Phylogeny in Annelids and Arthropods. Pergamon, pp. 1-5.

Andrioli, L.P., Oberstein, A.L., Corado, M.S., Yu, D., Small, S., 2004. Groucho-dependent repression by sloppy-paired 1 differentially positions anterior pair-rule stripes in the *Drosophila* embryo. Dev Biol 276, 541-551.

Aranda, M., Marques-Souza, H., Bayer, T., Tautz, D., 2008. The role of the segmentation gene hairy in Tribolium. Dev Genes Evol 218, 465-477.

Aronson, B.D., Fisher, A.L., Blechman, K., Caudy, M., Gergen, J.P., 1997. Groucho-dependent and -independent repression activities of Runt domain proteins. Mol Cell Biol 17, 5581-5587.

Bai, J., Uehara, Y., Montell, D.J., 2000. Regulation of invasive cell behavior by taiman, a *Drosophila* protein related to AIB1, a steroid receptor coactivator amplified in breast cancer. Cell 103, 1047-1058.

Bao, R., Friedrich, M., 2008. Conserved cluster organization of insect Runx genes. Dev Genes Evol 218, 567-574.

Batterham, P., Crew, J.R., Sokac, A.M., Andrews, J.R., Pasquini, G.M., Davies, A.G., Stocker, R.F., Pollock, J.A., 1996. Genetic analysis of the lozenge gene complex in *Drosophila* melanogaster: adult visual system phenotypes. J Neurogenet 10, 193-220.

Baumgartner, S., Noll, M., 1990. Network of interactions among pair-rule genes regulating paired expression during primordial segmentation of *Drosophila*. Mech Dev 33, 1-18.

Ben-David, J., Chipman, A.D., 2010. Mutual regulatory interactions of the trunk gap genes during blastoderm patterning in the hemipteran Oncopeltus fasciatus. Dev Biol 346, 140-149.

Benedyk, M.J., Mullen, J.R., DiNardo, S., 1994. odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in *Drosophila* embryos. Genes Dev 8, 105-117.

Benoit, G., Cooney, A., Giguere, V., Ingraham, H., Lazar, M., Muscat, G., Perlmann, T., Renaud, J.P., Schwabe, J., Sladek, F., Tsai, M.J., Laudet, V., 2006. International Union of Pharmacology. LXVI. Orphan nuclear receptors. Pharmacol Rev 58, 798-836.

Bern, C., Kjos, S., Yabsley, M.J., Montgomery, S.P., 2011. Trypanosoma cruzi and Chagas' Disease in the United States. Clin Microbiol Rev 24, 655-681.

Bialecki, M., Shilton, A., Fichtenberg, C., Segraves, W.A., Thummel, C.S., 2002. Loss of the ecdysteroid-inducible E75A orphan nuclear receptor uncouples molting from metamorphosis in *Drosophila*. Dev Cell 3, 209-220.

Bier, E., Vaessin, H., Younger-Shepherd, S., Jan, L.Y., Jan, Y.N., 1992. deadpan, an essential pan-neural gene in *Drosophila*, encodes a helix-loop-helix protein similar to the hairy gene product. Genes Dev 6, 2137-2151.

Bond, N.D., Nelliot, A., Bernardo, M.K., Ayerh, M.A., Gorski, K.A., Hoshizaki, D.K., Woodard, C.T., 2011. ssFTZ-F1 and Matrix metalloproteinase 2 are required for fat-body remodeling in *Drosophila*. Dev Biol 360, 286-296.

Broadus, J., McCabe, J.R., Endrizzi, B., Thummel, C.S., Woodard, C.T., 1999. The *Drosophila* beta FTZ-F1 orphan nuclear receptor provides competence for stage-specific responses to the steroid hormone ecdysone. Mol Cell 3, 143-149.

Brown, S.J., Denell, R.E., 1996. Segmentation and dorsoventral patterning in Tribolium. Seminars in Cell & Developmental Biology 7, 553-560.

Brown, S.J., Hilgenfeld, R.B., Denell, R.E., 1994a. The beetle *Tribolium castaneum* has a fushi tarazu homolog expressed in stripes during segmentation. Proc. Natl. Acad. Sci. USA 91, 12922-12926.

Brown, S.J., Parrish, J.K., Beeman, R.W., Denell, R.E., 1997. Molecular characterization and embryonic expression of the even-skipped ortholog of Tribolium castaneum. Mech Dev 61, 165-173.

Brown, S.J., Patel, N.H., Denell, R.E., 1994b. Embryonic expression of the single Tribolium engrailed homolog. Dev Genet 15, 7-18.

Butt, F.H., 1949. Embrylogy of the milkweed bug, *Oncopeltus fasciatus* (Hemiptera) Cornell university Agricultural experiment station, Memoir 283, 41-43.

Cadigan, K.M., Grossniklaus, U., Gehring, W.J., 1994a. Functional redundancy: the respective roles of the two sloppy paired genes in *Drosophila* segmentation. Proc Natl Acad Sci U S A 91, 6324-6328.

Cadigan, K.M., Grossniklaus, U., Gehring, W.J., 1994b. Localized expression of sloppy paired protein maintains the polarity of *Drosophila* parasegments. Genes Dev 8, 899-913.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., Yandell, M., 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research 18, 188-196.

Carroll, S.B., DiNardo, S., O'Farrell, P.H., White, R.A.H., Scott, M.P., 1988a. Temporal and spatial relationships between segmentation and homeotic gene expression in *Drosophila* embryos: distributions of the *fushi tarazu*, *engrailed*, *Sex combs reduced*, *Antennapedia*, and *Ultrabithorax* proteins. Genes Dev. 2, 350-360.

Carroll, S.B., Laughon, A., Thalley, B.S., 1988b. Expression, function, and regulation of the hairy segmentation protein in the *Drosophila* embryo. Genes Dev 2, 883-890.

Carroll, S.B., Scott, M.P., 1985. Localization of the *fushi tarazu* protein during *Drosophila* embryogenesis. Cell 43, 47-57.

Carroll, S.B., Scott, M.P., 1986. Zygotically active genes that affect the spatial expression of the fushi tarazu segmentation gene during early *Drosophila* embryogenesis. Cell 45, 113-126.

Chambon, P., 2005. The nuclear receptor superfamily: a personal retrospect on the first two decades. Mol Endocrinol 19, 1418-1428.

Chapman, R.F., 1998. The insects : structure and function, 4th ed. Cambridge University Press, Cambridge, U.K. ; New York.

Chen, G., Courey, A.J., 2000. Groucho/TLE family proteins and transcriptional repression. Gene 249, 1-16.

Chesebro, J., Hrycaj, S., Mahfooz, N., Popadic, A., 2009. Diverging functions of Scr between embryonic and post-embryonic development in a hemimetabolous insect, Oncopeltus fasciatus. Dev Biol 329, 142-151.

Choe, C.P., Brown, S.J., 2007. Evolutionary flexibility of pair-rule patterning revealed by functional analysis of secondary pair-rule genes, paired and sloppy-paired in the short-germ insect, Tribolium castaneum. Dev Biol 302, 281-294.

Choe, C.P., Miller, S.C., Brown, S.J., 2006. A pair-rule gene circuit defines segments sequentially in the short-germ insect Tribolium castaneum. Proc Natl Acad Sci U S A 103, 6560-6564.

Chou, T.-B., Perrimon, N., 1996. The autosomal FLP-DFS technique for generating germline mosaics in *Drosophila* melanogaster. Genetics 144, 1673-1679.

Cline, T.W., 1986. A female-specific lethal lesion in an X-linked positive regulator of the *Drosophila* sex determination gene, Sex-lethal. Genetics 113, 641-663.

Consortium, i.K., 2013. The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. Journal of Heredity 104, 595-600.

Copeland, J.W., Nasiadka, A., Dietrich, B.H., Krause, H.M., 1996. Patterning of the *Drosophila* embryo by a homeodomain-deleted Ftz polypeptide. Nature 379, 162-165.

Coulter, D.E., Swaykus, E.A., Beran-Koehn, M.A., Goldberg, D., Wieschaus, E., Schedl, P., 1990. Molecular analysis of odd-skipped, a zinc finger encoding segmentation gene with a novel pair-rule expression pattern. Embo j 9, 3795-3804.

Coulter, D.E., Wieschaus, E., 1988. Gene activities and segmental patterning in *Drosophila*: analysis of odd-skipped and pair-rule double mutants. Genes Dev 2, 1812-1823.

Damen, W.G., 2002. Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. Development 129, 1239-1250.

Damen, W.G., Janssen, R., Prpic, N.M., 2005. Pair rule gene orthologs in spider segmentation. Evol Dev 7, 618-628.

Davis, G.K., Jaramillo, C.A., Patel, N.H., 2001. Pax group III genes and the evolution of insect pair-rule patterning. Development 128, 3445-3458.

Davis, G.K., Patel, N.H., 2002. Short, long, and beyond: molecular and embryological approaches to insect segmentation. Annu Rev Entomol 47, 669-699.

Dawes, R., Dawson, I., Falciani, F., Tear, G., Akam, M., 1994. *Dax*, a locust Hox gene related to fushi-tarazu but showing no pair-rule expression. Development 120, 1561-1572.

Dawson, S.R., Turner, D.L., Weintraub, H., Parkhurst, S.M., 1995. Specificity for the hairy/enhancer of split basic helix-loop-helix (bHLH) proteins maps outside the bHLH domain and suggests two separable modes of transcriptional repression. Mol Cell Biol 15, 6923-6931.

de Almeida Dias, F., Souza dos Santos, A.L., Santos Lery, L.M., Alves e Silva, T.L., Oliveira, M.M., Bisch, P.M., Saraiva, E.M., Souto-Padron, T.C., Lopes, A.H., 2012. Evidence that a laminin-like insect protein mediates early events in the interaction of a Phytoparasite with its vector's salivary gland. PLoS One 7, e48170.

Dearden, P.K., Wilson, M.J., Sablan, L., Osborne, P.W., Havler, M., McNaughton, E., Kimura, K., Milshina, N.V., Hasselmann, M., Gempe, T., Schioett, M., Brown, S.J., Elsik, C.G., Holland, P.W., Kadowaki, T., Beye, M., 2006. Patterns of conservation and change in honey bee developmental genes. Genome Res 16, 1376-1384.

Delidakis, C., Artavanis-Tsakonas, S., 1992. The Enhancer of split [E(spl)] locus of *Drosophila* encodes seven independent helix-loop-helix proteins. Proc Natl Acad Sci U S A 89, 8731-8735.

Dickey-Sims, C., Robertson, A., Rupp, D., McCarthy, J., Coffman, J., 2005. Runx-dependent expression of PKC is critical for cell survival in the sea urchin embryo. BMC Biology 3, 18.

DiNardo, S., O'Farrell, P.H., 1987. Establishment and refinement of segmental pattern in the *Drosophila* embryo: spatial control of engrailed expression by pair-rule genes. Genes & Development 1, 1212-1225.

Doe, C.Q., Hiromi, Y., Gehring, W.J., Goodman, C.S., 1988. Expression and function of the segmentation gene fushi tarazu during *Drosophila* neurogenesis. Science 239, 170-175.

Dormand, E.L., Brand, A.H., 1998. Runt determines cell fates in the *Drosophila* embryonic CNS. Development 125, 1659-1667.

Duffy, J.B., Gergen, J.P., 1991. The *Drosophila* segmentation gene runt acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene Sex-lethal. Genes Dev 5, 2176-2187.

Duffy, J.B., Kania, M.A., Gergen, J.P., 1991. Expression and function of the *Drosophila* gene runt in early stages of neural development. Development 113, 1223-1230.

Erezyilmaz, D.F., Kelstrup, H.C., Riddiford, L.M., 2009. The nuclear receptor E75A has a novel pair-rule-like function in patterning the milkweed bug, Oncopeltus fasciatus. Developmental Biology 334, 300-310.

Escriva, H., Bertrand, S., Laudet, V., 2004. The evolution of the nuclear receptor superfamily. Essays Biochem 40, 11-26.

Escriva, H., Delaunay, F., Laudet, V., 2000. Ligand binding and nuclear receptor evolution. Bioessays 22, 717-727.

Ewen-Campen, B., Jones, T.E., Extavour, C.G., 2013. Evidence against a germ plasm in the milkweed bug Oncopeltus fasciatus, a hemimetabolous insect. Biol Open 2, 556-568.

Ewen-Campen, B., Shaner, N., Panfilio, K.A., Suzuki, Y., Roth, S., Extavour, C.G., 2011. The maternal and early embryonic transcriptome of the milkweed bug Oncopeltus fasciatus. BMC Genomics 12, 61.

Feir, D., 1974. Oncopeltus Fasciatus: A Research Animal. Annual Review of Entomology 19, 81-96.

Fernandez-Marcos, P.J., Auwerx, J., Schoonjans, K., 2011. Emerging actions of the nuclear receptor LRH-1 in the gut. Biochim Biophys Acta 1812, 947-955.

Fisher, A., Caudy, M., 1998. The function of hairy-related bHLH repressor proteins in cell fate decisions. Bioessays 20, 298-306.

Fitzpatrick, V.D., Percival-Smith, A., Ingles, C.J., Krause, H.M., 1992. Homeodomain-independent activity of fushi tarazu polypeptide in *Drosophila* embryos. Nature 356, 610-612.

Fleig, R., 1990. Engrailed expression and body segmentation in the honeybee Apis mellifera. Roux's archives of developmental biology 198, 467-473.

Florence, B., Guichet, A., Ephrussi, A., Laughon, A., 1997. Ftz-F1 is a cofactor in Ftz activation of the *Drosophila* engrailed gene. Development 124, 839-847.

Forero, D., 2008. The systematics of the Hemiptera

Revista Colombiana de Entomología 34, 21.

Forman, B.M., 2005. Are those phospholipids in your pocket? Cell Metab 1, 153-155.

Frasch, M., Hoey, T., Rushlow, C., Doyle, H., Levine, M., 1987. Characterization and localization of the even-skipped protein of *Drosophila*. Embo j 6, 749-759.

Frasch, M., Levine, M., 1987. Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in *Drosophila*. Genes Dev 1, 981-995.

Fristrom, D., 1968. Cellular degeneration in wing development of the mutant vestigial of *Drosophila* melanogaster. J Cell Biol 39, 488-491.

Fujioka, M., Jaynes, J.B., Goto, T., 1995. Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression. Development 121, 4371-4382.

Furukubo-Tokunaga, K., Flister, S., Gehring, W.J., 1993. Functional specificity of the Antennapedia homeodomain. Proc Natl Acad Sci U S A 90, 6360-6364.

Gauthier, S.A., VanHaaften, E., Cherbas, L., Cherbas, P., Hewes, R.S., 2012. Cryptocephal, the *Drosophila* melanogaster ATF4, is a specific coactivator for ecdysone receptor isoform B2. PLoS Genet 8, e1002883.

Gergen, J.P., Butler, B.A., 1988. Isolation of the *Drosophila* segmentation gene runt and analysis of its expression during embryogenesis. Genes Dev 2, 1179-1193.

Germain, P., Staels, B., Dacquet, C., Spedding, M., Laudet, V., 2006. Overview of nomenclature of nuclear receptors. Pharmacol Rev 58, 685-704.

Giguere, V., 1999. Orphan nuclear receptors: from gene to function. Endocr Rev 20, 689-725.

Gilbert, S.F., 2010. Developmental biology, 9th ed. Sinauer Associates, Sunderland, Mass.

Glass, C.K., Rosenfeld, M.G., 2000. The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev 14, 121-141.

Goltsev, Y., Hsiong, W., Lanzaro, G., Levine, M., 2004. Different combinations of gap repressors for common stripes in Anopheles and *Drosophila* embryos. Dev Biol 275, 435-446.

Grossniklaus, U., Cadigan, K.M., Gehring, W.J., 1994. Three maternal coordinate systems cooperate in the patterning of the *Drosophila* head. Development 120, 3155-3171.

Grossniklaus, U., Pearson, R.K., Gehring, W.J., 1992. The *Drosophila* sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. Genes Dev 6, 1030-1051.

Guichet, A., Copeland, J.W.R., Erdelyi, M., Hlousek, D., Zavorszky, P., Ho, J., Brown, S., Percival-Smith, A., Krause, H.M., Ephrussi, A., 1997. The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors. Nature 385, 548-552.

Guichet, A., Erdelyi, M., Zavorzsky, P., Ephrussi, A., 1996. Maternal effect pair-rule gene. Program and Abstracts. 37th Annual *Drosophila* Research Conference, San Diego, 1996, 185.

Gupta, B.P., Rodrigues, V., 1995. Distinct mechanisms of action of the Lozenge locus in *Drosophila* eye and antennal development are suggested by the analysis of dominant enhancers. J Neurogenet 10, 137-151.

Gurnell, M., Chatterjee, V.K., 2004. Nuclear receptors in disease: thyroid receptor beta, peroxisome-proliferator-activated receptor gamma and orphan receptors. Essays Biochem 40, 169-189.

Gutjahr, T., Frei, E., Noll, M., 1993a. Complex regulation of early paired expression: initial activation by gap genes and pattern modulation by pair-rule genes. Development 117, 609-623.

Gutjahr, T., Patel, N.H., Li, X., Goodman, C.S., Noll, M., 1993b. Analysis of the gooseberry locus in *Drosophila* embryos: gooseberry determines the cuticular pattern and activates gooseberry neuro. Development 118, 21-31.

Hacker, U., Nystedt, S., Barmchi, M.P., Horn, C., Wimmer, E.A., 2003. piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in *Drosophila*. Proc Natl Acad Sci U S A 100, 7720-7725.

Hafen, E., Kuroiwa, A., Gehring, W.J., 1984. Spatial distribution of transcripts from the segmentation gene *fushi tarazu* during *Drosophila* embryonic development. Cell 37, 833-841.

Han, W., Yu, Y., Su, K., Kohanski, R.A., Pick, L., 1998. A binding site for multiple transcriptional activators in the *fushi tarazu* proximal enhancer is essential for gene expression in vivo. Mol. Cell. Biol. 18, 3384-3394.

Heffer, A., Grubbs, N., Mahaffey, J., Pick, L., 2013. The evolving role of the orphan nuclear receptor ftz-f1, a pair-rule segmentation gene. Evolution & development 15, 406-417.

Heffer, A., Pick, L., 2011. Rapid isolation of gene homologs across taxa: Efficient identification and isolation of gene orthologs from non-model organism genomes, a technical report. EvoDevo 2, 7.

Heffer, A., Pick, L., 2013. Conservation and variation in Hox genes: how insect models pioneered the evo-devo field. Annual review of Entomology 58, 161-179.

Heffer, A., Shultz, J.W., Pick, L., 2010. Surprising flexibility in a conserved Hox transcription factor over 550 million years of evolution. Proc Natl Acad Sci U S A.

Heffer, A.M., 2012. Evolution of the Hox gene fushi tarazu in arthropods. University of Maryland, College Park, Ann Arbor, p. 154.

Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89, 10915-10919.

Hill, R.J., Billas, I.M., Bonneton, F., Graham, L.D., Lawrence, M.C., 2013. Ecdysone receptors: from the Ashburner model to structural biology. Annu Rev Entomol 58, 251-271.

Holmgren, R., 1984. Cloning sequences from the hairy gene of *Drosophila*. Embo j 3, 569-573.

Hooper, K.L., Parkhurst, S.M., Ish-Horowicz, D., 1989. Spatial control of hairy protein expression during embryogenesis. Development 107, 489-504.

Hou, H.Y., Heffer, A., Anderson, W.R., Liu, J., Bowler, T., Pick, L., 2009. Stripy Ftz target genes are coordinately regulated by Ftz-F1. Developmental biology 335, 442-453.

Hughes, C.L., Kaufman, T.C., 2000. RNAi analysis of Deformed, proboscipedia and Sex combs reduced in the milkweed bug Oncopeltus fasciatus: novel roles for Hox genes in the hemipteran head. Development 127, 3683-3694.

Hughes, C.L., Kaufman, T.C., 2002. Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede. Development 129, 1225-1238.

Hughes, C.L., Liu, P.Z., Kaufman, T.C., 2004. Expression patterns of the rogue Hox genes *Hox3/zen* and *fushi tarazu* in the apterygote insect *Thermobia domestica*. Evol Dev 6, 393-401.

Hughes, S.C., Krause, H.M., 2001. Establishment and maintenance of parasegmental compartments. Development 128, 1109-1118.

Ingham, P., Gergen, P., 1988. Interactions between the pair-rule genes runt, hairy, even-skipped and fushi tarazu and the establishment of periodic pattern in the *Drosophila* embryo. Development 104, 51-60.

Ingham, P.W., 1988. The molecular genetics of embryonic pattern formation in *Drosophila*. Nature 335, 25-34.

Ingham, P.W., Baker, N.E., Martinez-Arias, A., 1988. Regulation of segment polarity genes in the *Drosophila* blastoderm by fushi tarazu and even skipped. Nature 331, 73-75.

Ingham, P.W., Hidalgo, A., 1993. Regulation of wingless transcription in the *Drosophila* embryo. Development 117, 283-291.

Ingham, P.W., Howard, K.R., Ish-Horowicz, D., 1985. Transcription pattern of the *Drosophila* segmentation gene hairy. Nature 318, 439-445.

Ingraham, H.A., Redinbo, M.R., 2005. Orphan nuclear receptors adopted by crystallography. Curr Opin Struct Biol 15, 708-715.

Ito, Y., 2008. RUNX genes in development and cancer: regulation of viral gene expression and the discovery of RUNX family genes. Adv Cancer Res 99, 33-76.

Iwaki, D.D., Johansen, K.A., Singer, J.B., Lengyel, J.A., 2001. drumstick, bowl, and lines are required for patterning and cell rearrangement in the *Drosophila* embryonic hindgut. Dev Biol 240, 611-626.

Janssen, R., Damen, W.G., 2006. The ten Hox genes of the millipede Glomeris marginata. Dev Genes Evol 216, 451-465.

Jaynes, J.B., Fujioka, M., 2004. Drawing lines in the sand: even skipped et al. and parasegment boundaries. Dev Biol 269, 609-622.

Jimenez, G., Paroush, Z., Ish-Horowicz, D., 1997. Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. Genes Dev 11, 3072-3082.

Johnson, F.B., Parker, E., Krasnow, M.A., 1995. Extradenticle protein is a selective cofactor for the *Drosophila* homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. Proceedings of the National Academy of Sciences of the United States of America 92, 739-743.

Jost, W., Yu, Y., Pick, L., Preiss, A., Maier, D., 1995. Structure and regulation of the fushi tarazu gene from *Drosophila* hydei. Development Genes and Evolution 205, 160-170.

Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., Kluding, H., 1984. Mutations affecting the pattern of the larval cuticle in*Drosophila* melanogaster. Wilhelm Roux's archives of developmental biology 193, 283-295.

Kagoshima, H., Shigesada, K., Satake, M., Ito, Y., Miyoshi, H., Ohki, M., Pepling, M., Gergen, P., 1993. The Runt domain identifies a new family of heteromeric transcriptional regulators. Trends Genet 9, 338-341.

Kilchherr, F., Baumgartner, S., Bopp, D., Frei, E., Noll, M., 1986. Isolation of the paired gene of *Drosophila* melanogaster and its spatial expression during early embryogenesis. Nature 321, 493-499.

King-Jones, K., Thummel, C.S., 2005. Nuclear receptors--a perspective from *Drosophila*. Nat Rev Genet 6, 311-323.

Klimke, W., O'Donovan, C., White, O., Brister, J.R., Clark, K., Fedorov, B., Mizrachi, I., Pruitt, K.D., Tatusova, T., 2011. Solving the Problem: Genome Annotation Standards before the Data Deluge. Stand Genomic Sci 5, 168-193.

Klingensmith, J., Noll, E., Perrimon, N., 1989. The segment polarity phenotype of *Drosophila* involves differential tendencies toward transformation and cell death. Dev Biol 134, 130-145.

Klingler, M., Gergen, J.P., 1993. Regulation of runt transcription by *Drosophila* segmentation genes. Mech Dev 43, 3-19.

Kobayashi, M., Goldstein, R.E., Fujioka, M., Paroush, Z., Jaynes, J.B., 2001. Groucho augments the repression of multiple Even skipped target genes in establishing parasegment boundaries. Development 128, 1805-1815.

Kraft, R., Jackle, H., 1994. *Drosophila* mode of metamerization in the embryogenesis of the lepidopteran insect Manduca sexta. Proc Natl Acad Sci U S A 91, 6634-6638.

Krylova, I.N., Sablin, E.P., Moore, J., Xu, R.X., Waitt, G.M., MacKay, J.A., Juzumiene, D., Bynum, J.M., Madauss, K., Montana, V., Lebedeva, L.A., Suzawa, M., Williams, J.D., Williams, S.P., Guy, R.K., Thornton, J.W., Fletterick, R.J., Wilson, T.M., Ingraham, H.A., 2005. Structural analyses

reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. Cell 120, 343-355.

Kuroiwa, A., Hafen, E., Gehring, W.J., 1984. Cloning and transcriptional analysis of the segmentation gene fushi tarazu of *Drosophila*. 37, 825-831.

Lan, Y., Fujioka, M., Polsgrove, R., Miskiewicz, P., Morrissey, D., Goto, T., Weir, M., 1998. Plasticity of *Drosophila* paired function. Dev Genet 23, 45-55.

Laudet, V., 1997. Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. J Mol Endocrinol 19, 207-226.

Lawrence, P.A., 1981. The cellular basis of segmentation in insect. Cell 26, 3-10.

Levanon, D., Groner, Y., 2008. Structure and regulated expression of mammalian RUNX genes. Oncogene 23, 4211-4219.

Lewis, E.B., 1978. A gene complex controlling segmentation in *Drosophila*. Nature 276, 565-570.

Lewis, R.A., Wakimoto, B.T., Denell, R.E., Kaufman, T.C., 1980. Genetic Analysis of the Antennapedia Gene Complex (Ant-C) and Adjacent Chromosomal Regions of *DROSOPHILA* MELANOGASTER. II. Polytene Chromosome Segments 84A-84B1,2. Genetics 95, 383-397.

Li, Y., Choi, M., Cavey, G., Daugherty, J., Suino, K., Kovach, A., Bingham, N.C., Kliewer, S.A., Xu, H.E., 2005. Crystallographic identification and functional characterization of phospholipids as ligands for the orphan nuclear receptor Steriodogenic Factor-1. Mol. Cell 17, 4911-4502.

Lifschytz, E., Falk, R., 1968. Fine structure analysis of a chromosome segment in *Drosophila* melanogaster. Analysis of x-ray-induced lethals. Mutat Res 6, 235-244.

Lindsley, D.L., Zimm, G.G., 1992. The genome of *Drosophila* melanogaster. Academic Press, San Diego.

Liu, P., Kaufman, T., 2004a. hunchback is required for suppression of abdominal identity, and for proper germband growth and segmentation in the intermediate germband insect Oncopeltus fasciatus. Development 131, 1515-1527.

Liu, P.Z., Kaufman, T.C., 2004b. Kruppel is a gap gene in the intermediate germband insect Oncopeltus fasciatus and is required for development of both blastoderm and germband-derived segments. Development 131, 4567-4579.

Liu, P.Z., Kaufman, T.C., 2005. even-skipped is not a pair-rule gene but has segmental and gap-like functions in Oncopeltus fasciatus, an intermediate germband insect. Development 132, 2081-2092.

Liu, P.Z., Patel, N.H., 2010. giant is a bona fide gap gene in the intermediate germband insect, Oncopeltus fasciatus. Development 137, 835-844.

Liu, W., 2012. Functional analyses in the silkworm, Bombyx mori, support a role for Notch signaling in appendage development but not the groucho-dependent pair-rule process. J Exp Zool B Mol Dev Evol 318, 651-662.

Lohr, U., Pick, L., 2005. Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of *Drosophila melanogaster*. Curr Biol 15, 643-649.

Lohr, U., Yussa, M., Pick, L., 2001. *Drosophila fushi tarazu*: a gene on the border of homeotic function. Current Biology 11, 1403-1412.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., Konstantinidis, K.T., 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 7, e30087.

Luschnig, S., Moussian, B., Krauss, J., Desjeux, I., Perkovic, J., Nuesslein-Volhard, C., 2004. An F1 genetic screen for maternal-effect mutations affecting embryonic pattern formation in *Drosophila* melanogaster. Genetics 167, 325-342.

Macdonald, P.M., Ingham, P., Struhl, G., 1986. Isolation, structure, and expression of even-skipped: a second pair-rule gene of *Drosophila* containing a homeo box. Cell 47, 721-734.

Maderspacher, F., Bucher, G., Klingler, M., 1998. Pair-rule and gap gene mutants in the flour beetle Tribolium castaneum. Dev Genes Evol 208, 558-568.

Magrassi, L., Lawrence, P.A., 1988. The pattern of cell death in fushi tarazu, a segmentation gene of *Drosophila*. Development 104, 447-451.

Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., Evans, R.M., 1995. The nuclear receptor superfamily: the second decade. Cell 83, 835-839.

Mannervik, M., Levine, M., 1999. The Rpd3 histone deacetylase is required for segmentation of the *Drosophila* embryo. Proc Natl Acad Sci U S A 96, 6797-6801.

Margolis, R.N., Evans, R.M., O'Malley, B.W., 2005. The Nuclear Receptor Signaling Atlas: development of a functional atlas of nuclear receptors. Mol Endocrinol 19, 2433-2436.

Martinez-Arias, A., Ingham, P.W., 1985. The origin of pattern duplications in segment polarity mutants of *Drosophila* melanogaster. J Embryol Exp Morphol 87, 129-135.

Martinez-Arias, A., White, R.A.H., 1988. Ultrabithorax and engrailed expression in *Drosophila* embryos mutant for segmentation genes of the pair-rule class. Development 102, 325-338.

McKenna, N.J., Cooney, A.J., DeMayo, F.J., Downes, M., Glass, C.K., Lanz, R.B., Lazar, M.A., Mangelsdorf, D.J., Moore, D.D., Qin, J., Steffen, D.L., Tsai, M.J., Tsai, S.Y., Yu, R., Margolis, R.N., Evans, R.M., O'Malley, B.W., 2009. Minireview: Evolution of NURSA, the Nuclear Receptor Signaling Atlas. Mol Endocrinol 23, 740-746.

Mito, T., Kobayashi, C., Sarashina, I., Zhang, H., Shinahara, W., Miyawaki, K., Shinmyo, Y., Ohuchi, H., Noji, S., 2007. even-skipped has gap-like, pair-rule-like, and segmental functions in the cricket Gryllus bimaculatus, a basal, intermediate germ insect (Orthoptera). Dev Biol 303, 202-213.

Mito, T., Okamoto, H., Shinahara, W., Shinmyo, Y., Miyawaki, K., Ohuchi, H., Noji, S., 2006. Kruppel acts as a gap gene regulating expression of hunchback and even-skipped in the intermediate germ cricket Gryllus bimaculatus. Dev Biol 294, 471-481.

Monte, D., DeWitte, F., Hum, D.W., 1998. Regulation of the human P450scc gene by Steroidogenic Factor a is mediated by CBP/p300. J. Biol. Chem. 273, 4585-4591.

Moore, P.J., Attisano, A., 2011. Oosorption in response to poor food: complexity in the trade-off between reproduction and survival. Ecol Evol 1, 37-45.

Mouchel-Vielh, E., Blin, M., Rigolot, C., Deutsch, J.S., 2002. Expression of a homologue of the *fushi tarazu* (*ftz*) gene in a cirripede crustacean. Evol Dev 4, 76-85.

Mullen, J.R., DiNardo, S., 1995. Establishing parasegments in *Drosophila* embryos: roles of the odd-skipped and naked genes. Dev Biol 169, 295-308.

Nakao, H., 2010. Characterization of Bombyx embryo segmentation process: expression profiles of engrailed, even-skipped, caudal, and wnt1/wingless homologues. J Exp Zool B Mol Dev Evol 314, 224-231.

Nasiadka, A., Krause, H.M., 1999. Kinetic analysis of segmentation gene interactions in *Drosophila* embryos. Development 126, 1515-1526.

Noll, M., 1993. Evolution and role of Pax genes. Curr Opin Genet Dev 3, 595-605.

Nuclear Receptors Nomenclature, C., 1999. A unified nomenclature system for the nuclear receptor superfamily, Cell, United states, pp. 161-163.

Nüsslein-Volhard, C., Kluding, H., Jurgens, G., 1985. Genes affecting the segmental subdivision of the *Drosophila* embryo. Cold Spring Harbor Symp. Quant. Biol. 50, 145-154.

Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in *Drosophila*. Nature 287, 795-801.

Nüsslein-Volhard, C., Wieschaus, E., Kluding, H., 1984. Mutations affecting the pattern of the larval cuticle in*Drosophila* melanogaster. Wilhelm Roux's archives of developmental biology 193, 267-282.

Ortlund, E.A., Lee, Y., Solomon, I.H., Hager, J.M., Safi, R., Choi, Y., Guan, Z., Tripathy, A., Raetz, C.R., McDonnell, D.P., Moore, D.D., Redinbo, M.R., 2005. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. Nat Struct Mol Biol 12, 357-363.

Osborne, P.W., Dearden, P.K., 2005. Expression of Pax group III genes in the honeybee (Apis mellifera). Dev Genes Evol 215, 499-508.

Ozaki, T., Nakagawara, A., Nagase, H., 2013a. RUNX Family Participates in the Regulation of p53-Dependent DNA Damage Response. Int J Genomics 2013, 271347.

Ozaki, T., Wu, D., Sugimoto, H., Nagase, H., Nakagawara, A., 2013b. Runt-related transcription factor 2 (RUNX2) inhibits p53-dependent apoptosis through the collaboration with HDAC6 in response to DNA damage. Cell Death Dis 4, e610.

Papillon, D., Telford, M.J., 2007. Evolution of Hox3 and ftz in arthropods: insights from the crustacean Daphnia pulex. Dev Genes Evol 217, 315-322.

Paroush, Z., Finley, R.L., Jr., Kidd, T., Wainwright, S.M., Ingham, P.W., Brent, R., Ish-Horowicz, D., 1994. Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. Cell 79, 805-815.

Patel, N.H., Ball, E.E., Goodman, C.S., 1992. Changing role of even-skipped during the evolution of insect pattern formation. Nature 357, 339-342.

Patel, N.H., Condron, B.G., Zinn, K., 1994. Pair-rule expression patterns of even-skipped are found in both short- and long-germ beetles. Nature 367, 429-434.

Paz, J.d.l., 1958. Studies on the embryogeny ofOncopeltus fasciatus (Dallas) (Hemiptera: Lygaeidae). Cornell University.

Pazdera, T.M., Janardhan, P., Minden, J.S., 1998. Patterned epidermal cell death in wild-type and segment polarity mutant *Drosophila* embryos. Development 125, 3427-3436.

Peel, A.D., 2008. The evolution of developmental gene networks: lessons from comparative studies on holometabolous insects. Philos Trans R Soc Lond B Biol Sci 363, 1539-1547.

Peel, A.D., Chipman, A.D., Akam, M., 2005. Arthropod segmentation: beyond the *Drosophila* paradigm. Nat Rev Genet 6, 905-916.

Pepling, M.E., Gergen, J.P., 1995. Conservation and function of the transcriptional regulatory protein Runt. Proc Natl Acad Sci U S A 92, 9087-9091.

Pick, L., A., S., Affolter, M., Schmidt-Glenewinkel, T., Gehring, W.J., 1990. Analysis of the ftz upstream element: germ layer-specific enhancers are independently autoregulated. Genes & Dev. 4, 1224-1239.

Pick, L., Shultz, J., Anderson, W.R., Woodard, C.T., 2006. The Ftz-F1 family: orphan nuclear receptors regulated by novel protein-protein interactions, in: Taneja, R. (Ed.), Nuclear Receptors in Development. Elsevier.

Repa, J.J., Mangelsdorf, D.J., 2000. The role of orphan nuclear receptors in the regulation of cholesterol homeostasis. Annu Rev Cell Dev Biol 16, 459-481.

Rewitz, K.F., Yamanaka, N., O'Connor, M.B., 2013. Developmental checkpoints and feedback circuits time insect maturation. Curr Top Dev Biol 103, 1-33.

Robin, C., Lyman, R.F., Long, A.D., Langley, C.H., Mackay, T.F., 2002. hairy: A quantitative trait locus for *Drosophila* sensory bristle number. Genetics 162, 155-164.

Rohr, K.B., Tautz, D., Sander, K., 1999. Segmentation gene expression in the mothmidge Clogmia albipunctata (Diptera, psychodidae) and other primitive dipterans. Dev Genes Evol 209, 145-154.

Rosenberg, M.I., Brent, A.E., Payre, F., Desplan, C., 2014. Dual mode of embryonic development is highlighted by expression and function of Nasonia pair-rule genes. eLife 3.

Ruaud, A.F., Lam, G., Thummel, C.S., 2010. The *Drosophila* nuclear receptors DHR3 and betaFTZ-F1 control overlapping developmental responses in late embryos. Development 137, 123-131.

Rushlow, C.A., Hogan, A., Pinchin, S.M., Howe, K.M., Lardelli, M., Ish-Horowicz, D., 1989. The *Drosophila* hairy protein acts in both segmentation and bristle patterning and shows homology to N-myc. Embo j 8, 3095-3103.

Sablin, E.P., Krylova, I.N., Fletterick, R.J., Ingraham, H.A., 2003. Structural basis for ligand-dependent activation of the orphan nuclear receptor LRH-1. Mol. Cell 11, 1575-1585.

Salzberg, S.L., 2007. Genome re-annotation: a wiki solution? Genome Biol 8, 102.

Sander, K., 1976. Specification of the Basic Body Pattern in Insect Embryogenesis, in: J.E. Treherne, M.J.B.a.V.B.W. (Ed.), Advances in Insect Physiology. Academic Press, pp. 125-238.

Sander, K., Lohs-Schardin, M., Baumann, M., 1980. Embryogenesis in a *Drosophila* mutant expressing half the normal segment number. Nature 287, 841-843.

Sarrazin, A.F., Peel, A.D., Averof, M., 2012. A segmentation clock with two-segment periodicity in insects. Science 336, 338-341.

Schwartz, C.J.E., Sampson, H.M., Hlousek, D., Percival-Smith, A., Copeland, J.W.R., Simmonds, A.J., Krause, H.M., 2001. FTZ-Factor 1 and Fushi Tarazu interact via conserved nuclear receptor and coactivator motifs. EMBO J. 20, 510-519.

Shiau, A.K., Coward, P., Schwarz, M., Lehmann, J.M., 2001. Orphan nuclear receptors: from new ligand discovery technologies to novel signaling pathways. Curr Opin Drug Discov Devel 4, 575-590.

Sluder, A.E., Mathews, S.W., Hough, D., Yin, V.P., Maina, C.V., 1999. The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. Genome Res 9, 103-120.

Sommer, R.J., Tautz, D., 1993. Involvement of an orthologue of the *Drosophila* pair-rule gene hairy in segment formation of the short germ-band embryo of Tribolium (Coleoptera). Nature 361, 448-450.

St Pierre, S.E., Ponting, L., Stefancsik, R., McQuilton, P., 2014. FlyBase 102--advanced approaches to interrogating FlyBase. Nucleic Acids Res 42, D780-788.

Stuart, J.J., Brown, S.J., Beeman, R.W., Denell, R.E., 1991. A deficiency of the homeotic complex of the beetle *Tribolium*. Nature 350, 72-74.

Suzuki, T., Kawasaki, H., Yu , R.T., Ueda, Umesono, K., 2002. Segmentation gene product Fushi tarazu is an LXXLL motif-dependent coactivator for orphan receptor FTZ-F1. Proc. Natl. Acad. Sci. U S A 98, 12403-12408.

Telford, M.J., 2000. Evidence for the derivation of the *Drosophila* fushi tarazu gene from a Hox gene orthologous to lophotrochozoan Lox5. Curr Biol 10, 349-352.

Tepass, U., Fessler, L.I., Aziz, A., Hartenstein, V., 1994. Embryonic origin of hemocytes and their relationship to cell death in *Drosophila*. Development 120, 1829-1837.

Tremblay, J.J., Marcil, A., Gauthier, Y., Drouin, J., 1999a. Ptx1 regulates SF-1 activity by an interaction that mimics the role of the ligand-binding domain. EMBO J 18, 3431-3441.

Tremblay, J.J., Marcil, A., Gauthier, Y., Drouin, J., 1999b. Ptx1 regulates SF-1 activity by an interaction that mimics the role of the ligand-binding domain. Embo j 18, 3431-3441.

Tsai, C., Gergen, J.P., 1994. Gap gene properties of the pair-rule gene runt during *Drosophila* segmentation. Development 120, 1671-1683.

Ueda, H., Sonoda, S., Brown, J.L., Scott, M.P., Wu, C., 1990. A sequence-specific DNA-binding protein that activates fushi tarazu segmentation gene expression. Genes & Dev. 4, 624-635.

Urs, A.N., Dammer, E., Kelly, S., Wang, E., Merrill, A.H., Jr., Sewer, M.B., 2007. Steroidogenic factor-1 is a sphingolipid binding protein. Mol Cell Endocrinol 265-266, 174-178.

Wakimoto, B.T., Kaufman, T.C., 1981. Analysis of larval segmentation in lethal genotypes associated with the Antennapedia gene complex in *Drosophila* melanogaster. Developmental Biology 81, 51-64.

Wakimoto, B.T., Turner, F.R., Kaufman, T.C., 1984. Defects in embryogenesis in mutants associated with the antennapedia gene complex of *Drosophila* melanogaster. Dev Biol 102, 147-172.

Waltzer, L., Bienz, M., 1999. A funciton of CBP as a transcriptional co-activator during Dpp signalling. EMBO J. 18, 1630-1641.

Wang, W., Zhang, C., Marimuthu, A., Krupka, H.I., Tabrizizad, M., Shelloe, R., Mehra, U., Eng, K., Nyguyen, H., Settachatgul, C., Powell, B., Milburn, M.V., West, B.L., 2005. The crystal structures of human steroidogenic factor-1 and lever receptor homologue-1. Proc Natl Acad Sci U S A 102, 7505-7510.

Weisbrod, A., Cohen, M., Chipman, A.D., 2013. Evolution of the insect terminal patterning system--insights from the milkweed bug, Oncopeltus fasciatus. Dev Biol 380, 125-131.

Wilson, M.J., Dearden, P.K., 2012. Pair-Rule Gene Orthologs Have Unexpected Maternal Roles in the Honeybee (Apis mellifera). PloS one 7, e46490.

Wu, D., Ozaki, T., Yoshihara, Y., Kubo, N., Nakagawara, A., 2013. Runt-related transcription factor 1 (RUNX1) stimulates tumor suppressor p53 protein in response to DNA damage through complex formation and acetylation. J Biol Chem 288, 1353-1364.

Xu, J., Tan, A., Palli, S.R., 2010. The function of nuclear receptors in regulation of female reproduction and embryogenesis in the red flour beetle, Tribolium castaneum. Journal of insect physiology 56, 1471-1480.

Xu, X., Xu, P.X., Amanai, K., Suzuki, Y., 1997. Double-segment defining role of even-skipped homologs along the evolution of insect pattern formation. Dev Growth Differ 39, 515-522.

Yoo, J., Ko, S., Kim, H., Sampson, H., Yun, J.H., Choe, K.M., Chang, I., Arrowsmith, C.H., Krause, H.M., Cho, H.S., Lee, W., 2011a. Crystal structure of Fushi tarazu factor 1 ligand binding domain/Fushi tarazu peptide complex identifies new class of nuclear receptors. J Biol Chem 286, 31225-31231.

Yoo, J., Ko, S., Kim, H., Sampson, H., Yun, J.H., Choe, K.M., Chang, I., Arrowsmith, C.H., Krause, H.M., Cho, H.S., Lee, W., 2011b. Crystal structure of Fushi tarazu factor 1 ligand binding domain/Fushi tarazu peptide complex identifies new class of nuclear receptors. The Journal of biological chemistry 286, 31225-31231.

Yu, Y., Li, W., Su, K., Han, W., Yussa, M., Perrimon, N., Pick, L., 1997. The nuclear hormone receptor FTZ-F1 is a cofactor for the *Drosophila* homeodomain protein Ftz. Nature 385, 552-555.

Yu, Y., Pick, L., 1995. Non-periodic cues generate seven ftz stripes in the *Drosophila* embryo. Mechanisms of Development 50, 163-175.

Yu, Y., Yussa, M., Song, J., Hirsch, J., Pick, L., 1999. A double interaction screen identifies positive and negative *ftz* gene regulators and Ftz-interacting proteins. Mech. Dev. 83, 95-105.

Yussa, M., Lohr, U., Su, K., Pick, L., 2001a. The nuclear receptor Ftz-F1 and homeodomain protein Ftz interact through evolutionarily conserved protein domains. Mech. Dev. 107, 39-53.

Yussa, M., Lohr, U., Su, K., Pick, L., 2001b. The nuclear receptor Ftz-F1 and homeodomain protein Ftz interact through evolutionarily conserved protein domains. Mech. Dev. 107, 39-53.

Zhu, J., Chen, L., Sun, G., Raikhel, A.S., 2006. The Competence Factor βFtz-F1 Potentiates Ecdysone Receptor Activity via Recruiting a p160/SRC Coactivator. Molecular and Cellular Biology 26, 9402-9412.