ABSTRACT

Title of Dissertation: EVOLUTION OF THE *HOX* GENE *FUSHI TARAZU* IN ARTHROPODS

Alison M. Heffer, Doctor of Philosophy, 2012

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

Homeotic (*Hox*) genes are important in determining regional identity in virtually all metazoans, and are conserved throughout the animal kingdom. In *Drosophila melanogaster*, *fushi tarazu* (*ftz*) is located within the *Hox* complex and contains a Hox-like DNA-binding homeodomain, but functions as a pair-rule segmentation gene. At some point(s) during evolution, *ftz* has undergone three specific changes thought to contribute to its new segmentation function in *Drosophila*: 1) The gain of an LXXLL motif allowed for interaction with a new co-factor, Ftz-F1; 2) The degeneration of the YPWM motif decreased the ability to interact with the homeotic co-factor Exd; 3) *ftz* expression switched from *Hox*-like to seven stripes in *Drosophila*. Here I isolated *ftz* sequences and examined expression from arthropods spanning 550 million years of evolutionary time to track these changes in *ftz*. I found that while the LXXLL motif required for segmentation was stably acquired at the base of the holometabolous insects, the YPWM motif degenerated independently many times in arthropod lineages, and these 'degen-YPWMs' vary in their homeotic potential. Additionally, *ftz* expression in a crustacean is in a weak *Hox*-like pattern, suggesting a model in which different *ftz* variants could arise in nature and not be detrimental to organismal development. Given my findings that *ftz* sequence and expression is so dynamic, I investigated the features that may be preventing *ftz* fossilization in arthropod genomes. I tested

the hypothesis that a broadly conserved role of *ftz* in the developing central nervous system (CNS) retains *ftz* in arthropod genomes. This model predicts that the homeodomain, but not variable co-factor interaction motifs, is required for Ftz CNS function. Evidence supporting this model was obtained from CNS-specific rescue experiments in *Drosophila*. Additionally I examined the expression and function of *ftz* and *ftz-f1* in the short-germ beetle *Tribolium castaneum*. I found that both genes are expressed in pair-rule patterns, and preliminary results suggest that *ftz-f1* is important for proper segmentation and cuticle deposition, and *ftz* function may be partially redundant with *ftz-f1*. Taken together, these findings show that variation of a pleiotropic transcription factor is more extensive than previously imagined, and suggest that evolutionary plasticity may be widespread among regulatory genes.

Evolution of the *Hox* gene *fushi tarazu* in arthropods

By

Alison M. Heffer

Dissertation 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
2012

Advisory Committee:

Dr. Leslie Pick, Advisor
Dr. Eric Haag
Dr. Jeffrey Shultz
Dr. Louisa Wu
Dr. Thomas Kocher

© Copyright by

Alison M. Heffer

2012

Dedication

To Ross, Jack, and my parents

Acknowledgements

I would first like to thank my advisor, Leslie Pick, for her excellent mentorship during my time as a graduate student. Leslie has always been there for guidance, to listen to my new ideas, and for support both in my academic and personal life. She has made my time as a graduate student very enjoyable and fulfilling. I have grown immensely as a scientist while in her lab, and learned many valuable lessons under her good mentorship. Thank you for your constant support and enthusiasm.

I would also like to thank the members of the Pick lab, both past and present: Amanda Field, Yong Lu, Ray Anderson, Jie Xiang, Georgeta Crivat, Zhaoqing Jin, Caroline Li, Hua Zhang, and Julie Hou. Your scientific help and personal friendships have been invaluable to me during my time at Maryland. I would like to extend a special thanks to my two undergraduate students, Molly Bloom and Margo Faust, who helped with my graduate work as well. I would also like to thank the members of the Wang Lab and fellow MOCB students for your scientific help and conversations, as well as for your friendship.

Lastly, I would like to thank my family for their support and never-ending encouragement during my time as a graduate student. My husband, who has been my biggest supporter, has stood by me through everything and has always been supportive. He is a constant reminder of the importance of family-work balance. My parents have always been supportive of my goals and dreams in life, and I thank them for their unconditional love.

Table of Contents

Dedication……………………………………………………………………........………..ii
Acknowledgements………………………………………………………........……...iii
Table of Contents....………...……………………………………………........……....iv
List of Tables..…...…………...………………………………………………........…...vi
List of Figures......……………………………………………………………........…...vii

Chapter 1: *Introduction*
1.1 Evo-devo at a glance..1
1.2 Homeosis and the discovery of *Hox* genes..2
1.2.1 *Hox* gene clustering, duplication, and divergence..3
1.2.2 *Hox* genes encode regulatory transcription factors...6
1.3 How have *Hox* genes changed during evolution?...8
1.3.1 Changes in *Hox* gene expression: examples of cis-regulatory evolution.................10
1.3.1a The *Hox* gene *fushi tarazu (ftz)* has undergone dramatic changes in expression pattern during arthropod evolution.......................................10
1.3.1b *Hox3/zen* divergence and its co-option into a role in extra-embryonic membrane formation...11
1.3.1c The case of Ubx and leg morphology...13
1.3.1d Scr expression varies among different insects...15
1.3.1e New expression domains suggest novel Antp functions in butterflies.......17
1.3.1f Cis-regulatory changes in vertebrate *Hox* complexes................................18
1.3.2 Changes in Hox downstream target gene regulation without changes in *Hox* expression: cis-regulatory changes in target genes or novel protein functions........19
1.3.2a Scr is a key player in insect morphological evolution...............................19
1.3.2b Many ways to make a wing: the role of Ubx in wing development...........21
1.3.3 Changes in Hox protein function...23
1.3.3a Ubx and abdominal appendages...24
1.3.3b Escape from colinearity enabled variation in Hox protein potential: "*ftz*-ing" around during insect evolution..24
1.3.3c Bcd acquires a new function in higher insects 25
1.3.3d Hox protein changes and the evolution of placental mammals.................26
1.3.4 Post-transcriptional regulation of *Hox* genes..27
1.3.5 *Hox* mechanism conclusions and emerging themes...29

Chapter 2: *Rapid isolation of gene homologs across taxa: Efficient identification and isolation of gene orthologs from non-model organism genomes, a technical report*
[Heffer and Pick, *Evodevo*, 2011]
2.1 Introduction...31
2.2 Results
2.2.1 RIGHT Methodology...32
2.2.2 RIGHT isolation of homeobox and nuclear receptor genes.....................................36
2.3 Conclusions...39

Chapter 3: *Surprising flexibility in a conserved Hox transcription factor over 550 million years of evolution*
[Heffer et. al, *PNAS*, 2010]

3.1 Abstract..41
3.2 Introduction..42
3.3 Results
3.3.1 *ftz* gene diversity in the arthropod tree of life..45
3.3.2 The LXXLL was stably acquired at the base of Holometabola...............................50
3.3.3 The YPWM motif 'flickers' in arthropod phylogeny...51
3.3.4 'Degen-YPWMs' vary in homeotic potential...53
3.3.5 Loss of *Hox*-like expression in crustaceans...57
3.4 Discussion..60
3.4.1 Model for regulatory transcription factor flexibility..62
3.5 Materials and Methods..64

Chapter 4: *Variation and constraint in Hox gene evolution*

4.1 Abstract..66
4.2 Introduction..66
4.3 Results
4.3.1 *ftz* CNS expression is conserved over 550 million years of arthropod evolution.....69
4.3.2 Ftz-F1 and Exd are not co-expressed with Ftz in the CNS......................................70
4.3.3 Cofactor interaction motifs in Dm-Ftz are dispensable for CNS function..............72
4.3.4 The DNA-binding homeodomain is required for CNS function.............................74
4.3.4 Ftz function in the CNS is independent from other *Hox* genes...............................77
4.4 Discussion..78
4.5 Methods...79

Chapter 5: *Investigating the roles of ftz and ftz-f1 in the shot-germ beetle Tribolium castaneum*

5.1 Introduction..82
5.2 Results
5.2.1 *ftz* and *ftz-f1* have pair-rule expression in *Tribolium*..84
5.2.2 *ftz* and *ftz-f1* embryonic RNAi effects on cuticle formation....................................86
5.2.3 Investigating the roles of *Tribolium ftz* and *ftz-f1* in segmentation.........................89
5.3 Discussion..91

Conclusions and Future Directions...93

Appendix I: Ftz and Ftz-F1 sequences..96

Appendix II: Detailed Materials and Methods...116

References...124

List of Tables

TABLE 1-1: A new *zen* expression pattern in insects allowed for cooption into an early developmental pathway involved in extra-embryonic membrane development.

TABLE 3-1: *ftz* genes generally contain small introns and encode short linker regions between the YPWM motif and homeodomain.

List of Figures

FIGURE 1-1: *Hox* complexes in bilaterians are highly conserved, despite duplication and divergence of *Hox* paralogs or clusters.

FIGURE 1-2: Four mechanisms underlie *Hox* regulatory evolution.

FIGURE 1-3: Variations in Ubx expression in developing insect legs have contributed to morphological diversity.

FIGURE 2-1: Overview of RIGHT technique used to isolate homologous genes from large gene families.

FIGURE 2-2: Isolation of *ftz* homologs using RIGHT.

FIGURE 3-1: The *ftz* genes from diverse arthropods display remarkable flexibility.

FIGURE 3-2: Ftz orthologs have similar homeodomain sequences but vary in their cofactor interaction motifs and protein lengths.

FIGURE 3-3: Conservation of YPWM motif and homeodomain in other Hox proteins.

FIGURE 3-4: Degenerate YPWM motifs retain varying degrees of homeotic potential.

FIGURE 3-5: Expression levels of 'degen-YPWM' transgenes are similar.

FIGURE 3-6: Fading *Hox*-like *ftz* expression in the crustacean *Artemia salina.*

FIGURE 3-7: The modularity of *ftz* CREs and protein motifs allows for extensive variation in *ftz* throughout arthropods.

FIGURE 4-1: *ftz* CNS expression is conserved across arthropods that exhibit great diversity in Ftz sequence composition.

FIGURE 4-2: Ftz is not co-expressed with known cofactors in the CNS.

FIGURE 4-3: Strategy to test the role of labile and stabile Ftz protein motifs in *Drosophila* CNS function.

FIGURE 4-4: The homeodomain is required for Ftz function in the CNS while cofactor interaction motifs are dispensable.

FIGURE 4-5: Ftz CNS expression does not overlap with Antp or Scr.

FIGURE 4-6: Ftz does not activate Antp expression in the developing CNS.

FIGURE 5-1: *ftz* and *ftz-f1* are expressed in pair-rule patterns in the short-germ beetle *Tribolium*.

FIGURE 5-2: The effects of *Tc-ftz-f1* embryonic RNAi on cuticle formation.

FIGURE 5-3: Hatching percentages of injected *Tribolium* embryos.

FIGURE 5-4: *ftz* and *ftz-f1* expression during embryogenesis as determined by RT-PCR.

FIGURE 5-5: *Tribolium ftz-f1* exhibits a role in segmentation.

FIGURE 5-6: Effects of injecting *ftz* and *ftz-f1* dsRNA on developmental timing.

Chapter 1: Introduction

[modified from Heffer and Pick, *Annual Reviews of Entomology*, 2012; Pick and Heffer, *Annals of the New York Academy of Sciences*, 2012]

1.1 Evo-devo at a glance

How is it that some insects have two wings used for flight and others have four? How did the jumping legs of crickets and grasshoppers become disproportionately larger than their other legs? Evolutionary developmental biology, or evo-devo, has emerged as a rapidly growing field in biology that addresses these types of questions by studying the basic processes directing organismal development and how they have changed during evolution to promote diversity in body form. Evo-devo encompasses studies of variation in both phenotype and genotype, including embryonic development (Carroll et al., 2005), morphological novelties (Lynch and Wagner, 2008), homology (Hall, 2003; Cracraft, 2005), and developmental plasticity (Moczek, 2010), with the larger goal of discovering molecular mechanisms underlying biological diversity.

One of the core concepts discovered in evo-devo is that organisms possess a "genetic toolkit", or basic collection of genes that control development, which is remarkably conserved throughout the animal kingdom (Carroll et al., 2005; Shubin et al., 2009). Many toolkit genes encode transcription factors, which function as sequence-specific DNA binding proteins that activate or repress expression of downstream target genes involved in the formation of specific body structures. A central question in evo-devo raised by this observation is: how can one genetic toolkit produce diverse body plans? An emerging hypothesis in the field is that changes in gene products and/or changes in the expression patterns of these genes allows them to be 're-wired' or co-opted for use in different developmental pathways with a highly conserved group of transcription factors re-organizing regulatory connections to control development of diverse

organisms (Levine and Davidson, 2005; Hoekstra and Coyne, 2007; Wray, 2007; Lynch and Wagner, 2008; Wagner and Lynch, 2008; Stern and Orgogozo, 2009). Evo-devo studies have contributed to biologists' understanding of organismal development by exploring modes of development in diverse animal systems, and to our understanding of the molecular underpinnings of the evolution of development.

1.2 Homeosis and the discovery of *Hox* genes

Homeobox-containing (*Hox*) genes are fundamental components of the genetic toolkit of metazoans, most widely recognized for their role in determining segment identity (Carroll et al., 2005). A century before these genes were cloned, rare mutations were observed in nature, such as insects with legs replacing antennae (Bateson, 1894). Bateson coined the term 'homeosis' to describe these aberrations, where "something has been changed into the likeness of something else" (Bateson, 1894). Though it would be almost a century before *Hox* genes were isolated, sequenced, and the genetic mechanisms underlying homeotic mutations studied, it was apparent to Bateson that changes could occur during development and that these might play a role in body plan evolution (Bateson, 1894).

One of the most famous examples of a homeotic transformation is the 4-winged fruit fly studied by Ed Lewis: here, the third thoracic (T3) segment, which normally lacks wings, is replaced by a second thoracic-like (T2) segment with a perfect pair of wings (Lewis, 1978; Duncan, 1987; Lewis, 1998). In another startling example – the *Antennapedia* (*Antp*) mutation – the antennae of the fly are replaced with a perfect pair of legs – the exact legs that would normally develop on the T2 segment (Gehring, 1966; Postlethwait and Schneiderman, 1971; Denell, 1973; Duncan and Kaufman, 1975; Kaufman et al., 1980; Lewis et al., 1980; Denell et

al., 1981; Schneuwly et al., 1986; Schneuwly et al., 1987b). Through years of study of these homeotic genes, it became clear that the normal or wild type function of these genes is to determine the unique identities of individual segments. For example, *Antp* normally specifies the unique identity of the T2 segment, including its specific leg. When Antp is mis-expressed in the developing head, it does its job of patterning the T2 leg, but it does it in the wrong place, giving an adult fly with legs where the antennae should be (Schneuwly et al., 1987a). Similarly, other homeotic genes specify other unique identities – for example, *Sex combs reduced* (*Scr*) specifies the identity of the leg on the first thoracic segment (T1) which, in males, bears specialized structures known as sex combs. Loss-of-function mutations in *Scr* thus lead to loss of T1-identity, evidenced by loss of sex combs (Kaufman et al., 1980; Lewis et al., 1980; Struhl, 1982; Mahaffey and Kaufman, 1987; LeMotte et al., 1989).

1.2.1 *Hox* gene clustering, duplication, and divergence

Elucidation of the genetic basis of insect homeotic mutations began in the mid-1900s through studies of the emerging model organism, the fruit fly *Drosophila melanogaster* (Lewis, 1963; Lewis, 1978). Using polytene chromosome mapping, it was found that mutations causing transformations of posterior body segments (e.g. transformation of haltere to wing) clustered in one region of the third chromosome, named the *Bithorax* complex (BX-C; (Lewis, 1978)). Mutations resulting in homeosis of anterior segments (e.g. transformation of antenna to leg) mapped to another cluster on the third chromosome, the *Antennapedia* complex (ANT-C; (Kaufman et al., 1980)). In addition to their chromosomal clustering, Lewis observed 'co-linearity' among these genes: their linear organization along the chromosome correlated with the region of function along the anterior-posterior axis of the animal (Lewis, 1978). *Hox* genes

located at the 3' end of the *Hox* complex (e.g. *labial* and *proboscipedia*) affect body structures in the anterior part of the embryo while genes at the 5' end of the complex (e.g. *Abd-B*) affect the posterior region of the animal. After these genes were cloned and expression patterns analyzed, it was quickly realized that their co-linear action reflects their anterior-posterior order of expression along the embryonic body axis (Lewis, 1978; Bender et al., 1983; Wakimoto et al., 1984; Akam, 1987).

The chromosomal clustering and co-linearity of *Hox* genes are conserved outside of *Drosophila*, in both invertebrates and vertebrates (Figure 1-1). Insects have maintained one *Hox* cluster (split into the ANT-C and BX-C in *Drosophila*), which is thought to be similar in gene composition to the ancestral *Hox* complex in Urbilateria (Grenier et al., 1997; de Rosa et al., 1999; Cook et al., 2001). A single cluster has been maintained outside of vertebrates, as polychaetes (Irvine et al., 1997; Frobius et al., 2008), onychophorans (Grenier et al., 1997), and sea urchins (Cameron et al., 2006) all have one *Hox* cluster. In vertebrates, there have been *Hox* cluster duplications and paralog-specific gene losses and gains (Figure 1-1). Mammals have 4 *Hox* clusters (HoxA-D; (Scott, 1993)), and teleosts have as many as 8 (Amores et al., 1998; Crow et al., 2006). This duplication of entire *Hox* gene clusters has led to multiple copies of these genes enabling diversification of function of individual paralogs, loss of paralogs because of redundancy, and additional gene duplications in some lineages (Wagner et al., 2003). This is evidenced in mammalian *Hox* clusters, which have undergone two rounds of replication to generate four clusters (McGinnis and Krumlauf, 1992; Scott, 1992; Duboule, 1994). Within each cluster, most genes are conserved, but some have been lost (e.g., *Hoxb10* and *Hoxc3*) while others expanded (e.g., the posterior *Hox* genes, represented only by *Abd-B* in *Drosophila*, have expanded in vertebrate lineages). These *Hox* cluster duplications are thought to be important in

the radiation of different lineages and the presence of evolutionary novelties (Wagner et al., 2003; Crow et al., 2006). The one *Hox* cluster present in insects provides an optimal system to examine gene function, as loss- and gain-of-function analyses are not complicated by the presence of multiple *Hox* paralogs and functional redundancy.

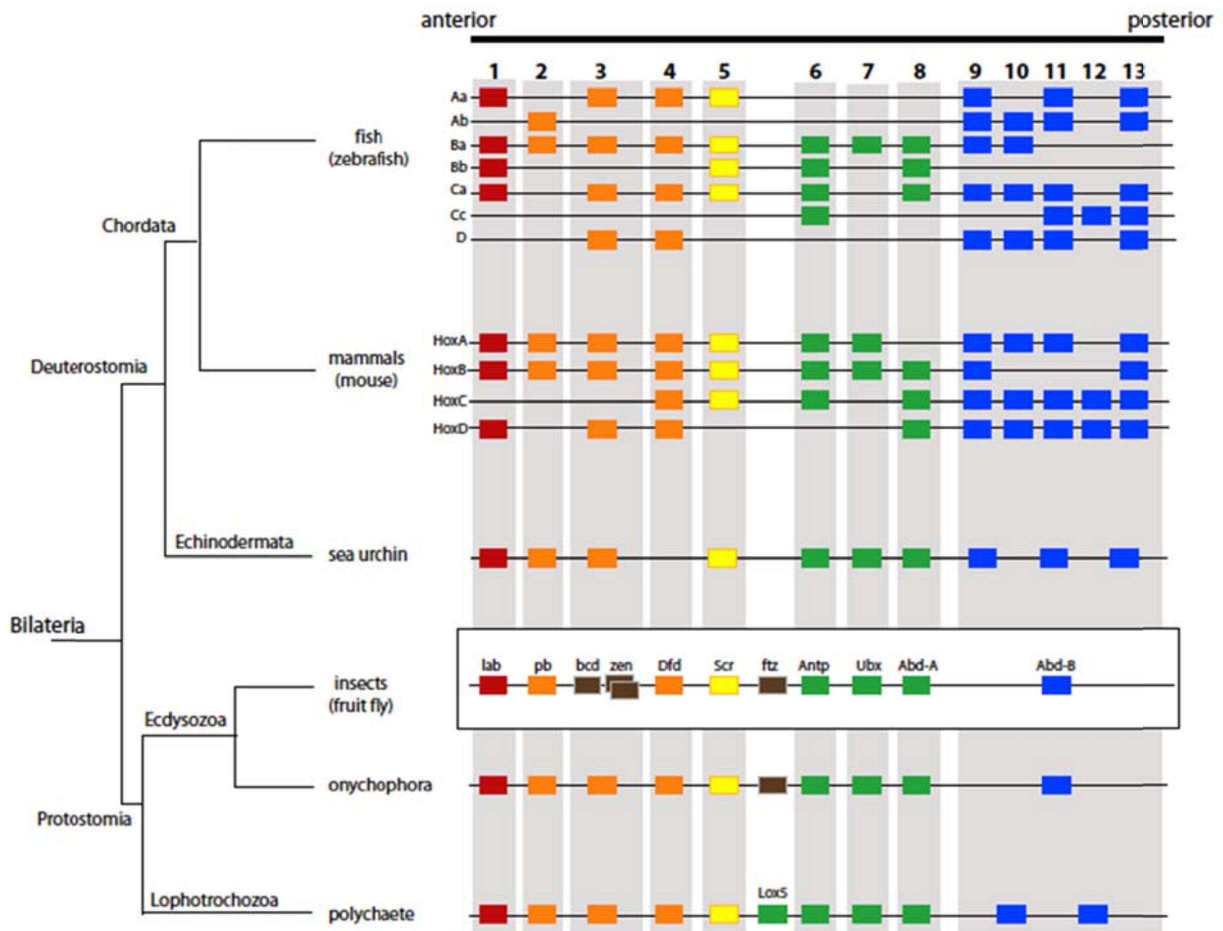


FIGURE 1-1. *Hox* complexes in bilaterians are highly conserved, despite duplication and divergence of *Hox* paralogs or clusters. While lophotrochozoans, ecdysozoans, and some deuterostomes have a single *Hox* cluster, there have been lineage-specific duplications of the *Hox* cluster in chordates. Mammals have 4 *Hox* clusters, with loss of some paralogs within clusters. Fish lineages such as zebrafish have up to 8 clusters. The *Drosophila Hox* complex highlighted in this figure is representative of insects.

After their discovery in *Drosophila*, *Hox* clusters were identified in other insects, including honeybees (Walldorf et al., 1989), beetles (Stuart et al., 1991), grasshoppers (Ferrier and Akam, 1996), mosquitoes (Devenport et al., 2000; Powers et al., 2000), and moths (Yasukochi et al., 2004). While *Drosophila Hox* genes are split into two clusters on the same chromosome, *Hox* clusters in these other insects retain the presumed ancestral single cluster (Figure 1-1), with the exception of *labial*, which is located at the opposite end of the chromosome in *Bombyx* (Yasukochi et al., 2004). Thus, while the *Hox* cluster itself appears to be under evolutionary constraint, some cases of split complexes retain function (Struhl, 1984). In addition, there have been instances of gene duplication and divergence within insect *Hox* clusters. For example, *Bombyx* harbor a tandem duplication of twelve homeobox genes between *pb* and *zen/Hox3* that appears to be unique to this lineage (Chai et al., 2008) and *Drosophila* and *Tribolium* carry independent duplications of *Hox3*/*zen* (Brown et al., 2002; Schmidt-Ott, 2005).

1.2.2 *Hox* genes encode regulatory transcription factors

In the 1980s, *Hox* genes were cloned from *Drosophila* (McGinnis et al., 1984b; Scott and Weiner, 1984), and it was a shocking discovery when a region of these genes was detected in evolutionarily distant species, including beetles, earthworms, and humans (McGinnis et al., 1984a). This conserved region of 180 base-pairs was coined the 'homeobox', which encodes a 60 amino acid 'homeodomain,' so-named because of their discovery in homeotic genes (McGinnis et al., 1984a; McGinnis et al., 1984b; Scott and Weiner, 1984). Hox proteins bind DNA via their homeodomains and function as transcription factors that regulate gene expression by binding to specific DNA sequences in cis-regulatory regions of a number of genes (Gehring, 1985; Gehring and Hiromi, 1986; Lawrence, 1992; McGinnis and Krumlauf, 1992; Gehring et

al., 1994; McGinnis, 1994; Pearson et al., 2005). As such, they serve as master regulators or selector genes to initiate developmental programs by activating the expression of downstream or realizator genes involved in growth and differentiation of particular body structures (Garcia-Bellido, 1975; Lawrence, 1992). For example, and broadly speaking, *Sex-combs reduced* (Scr) would bind to cis-regulatory regions controlling genes involved in T1 identity and regulate their transcription, and Antp would regulate genes involved in T2 identity. This explains the ability of *Hox* genes to regulate entire developmental programs and provided insight into the molecular underpinnings of homeotic transformation (Carroll et al., 2005).

Due to similarities in homeodomains, the DNA binding sequences recognized by different Hox proteins are very similar, and yet each Hox protein has a unique and specific role in vivo (the so-called "Hox Paradox"; (Mann, 1995)). One way in which Hox proteins achieve specificity is through interaction with different DNA-binding partners or co-factors, which modulate Hox binding preference for certain sites in the genome such that each Hox protein regulates a discrete set of target genes ((Hayashi and Scott, 1990; Ebner et al., 2005; Mann et al., 2009; Slattery et al., 2011); see below). Several *Drosophila* and mammalian Hox proteins interact with the homeotic co-factor Extradenticle (Exd/Pbx), which increases DNA-binding specificity in vivo (Johnson et al., 1995; Sprules et al., 2003). Hox functional specificity is also influenced by residues at the amino-terminal end of the homeodomain and by other protein motifs that modulate cofactor interactions and/or transcriptional activity ((Gibson et al., 1990; Lin and McGinnis, 1992; Furukubo-Tokunaga et al., 1993; Zeng et al., 1993; Zhao et al., 1996; Galant et al., 2002; Ronshaugen et al., 2002; Tour et al., 2005); see below).

1.3 How have *Hox* genes changed during evolution?

Hox genes are generally considered to be evolutionarily constrained since mis-expression during development results in homeotic transformations. However, small changes in timing and location of expression have been found to promote morphological diversity. Some of these changes impacted the regulation and expression of *Hox* genes (cis-regulatory changes) (Carroll et al., 2005; Prud'homme et al., 2007; Carroll, 2008) while others impacted Hox protein activity (protein coding changes) (Schmidt-Ott and Wimmer, 2004; Lynch and Wagner, 2008; Heffer et al., 2011). Still other changes occurred downstream of the *Hox* genes themselves, particularly in the regulatory regions of targets, which can be gained or lost in a lineage-specific fashion, thereby changing the biological role of a *Hox* gene without changes in its expression. In all cases, gene regulatory networks (GRNs) (Britten and Davidson, 1969; Levine and Davidson, 2005; Davidson and Erwin, 2006) regulated by *Hox* genes are altered, although the mechanisms underlying this alteration are different. We have classified these mechanisms into four categories, diagrammed in Figure 1-2: (1) Changes in *Hox* gene expression, (2) Changes in Hox downstream target gene regulation without change in *Hox* expression, (3) Changes in Hox protein function through changes in protein coding sequence, and (4) Post-transcriptional regulation of *Hox* gene function. Here, we focus on key examples from the literature that demonstrate each mechanism. For several of these case studies, direct links between Hox GRN changes and morphological evolution have been nicely demonstrated. For others, the challenge will be to determine the functional impact of Hox GRN changes.

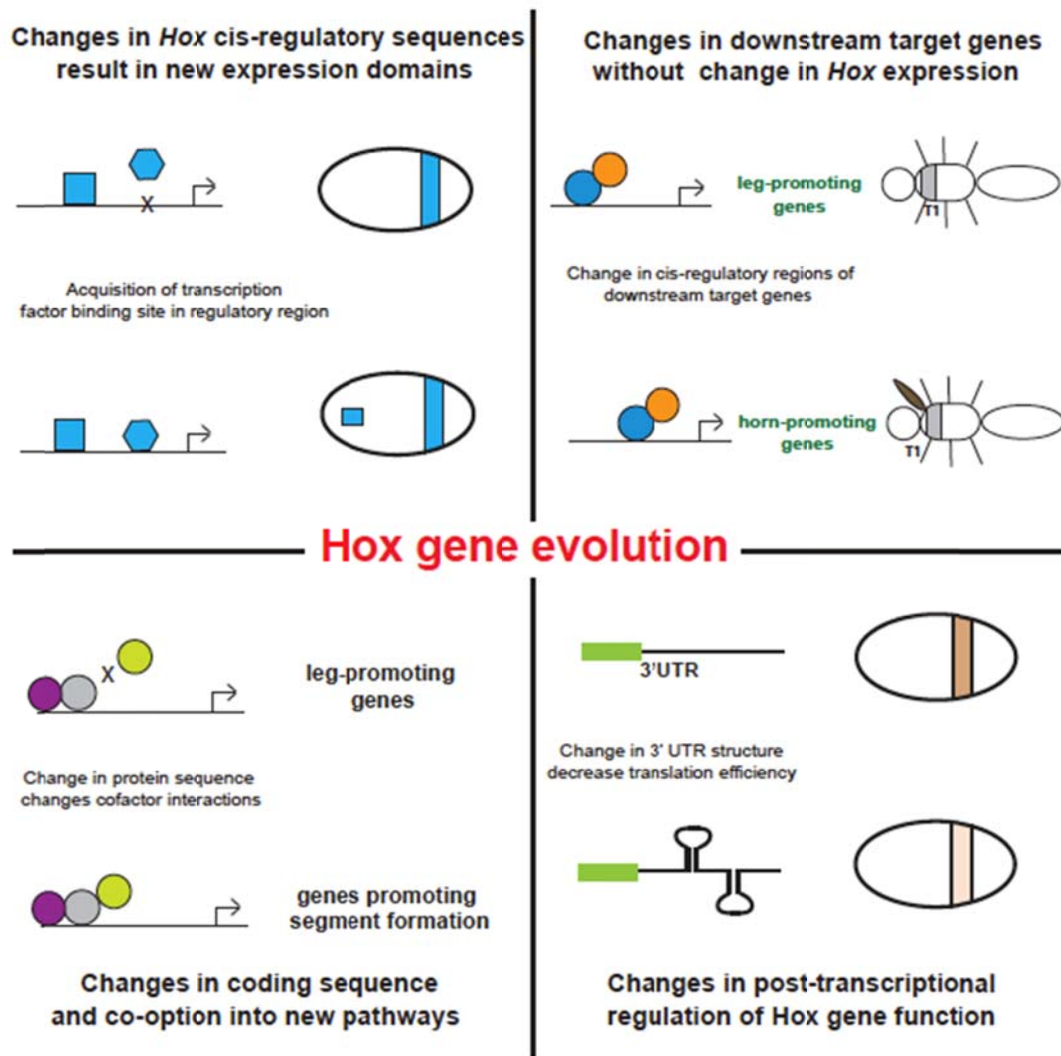


FIGURE 1-2. Four mechanisms underlie *Hox* regulatory evolution. A schematic of the four evolutionary mechanisms discussed in this review is shown. (Upper left) Changes in *cis*-regulatory regions. For example, a transcription factor (blue square) binds to a *cis*-regulatory element and directs *Hox* expression in a specific domain. A binding site for a new transcription factor (blue hexagon) is acquired resulting in a novel expression domain. (Upper right) Changes in downstream target genes without a *Hox* expression change. For example, a Hox protein (blue circle) and its cofactor (orange circle) promote the activation of leg-specification genes. A change in cis-regulatory sequences of target genes occurs, such that the Hox protein and its partner now bind to and activate horn-specification genes. The expression of the *Hox* gene itself is unchanged. (Lower left) Changes in coding sequence and cooption into new pathways. For example, two transcription factors (purple grey circles) bind to DNA and activate genes that promote leg formation. If amutation occurs in the Hox protein such that it can now interact with a new cofactor (green circle), a new function may be acquired. (Lower right) Changes in posttranscriptional regulation of a *Hox* gene. For example, one 3'UTR structure (indicated here as a straight line) is efficient in promoting translation of this protein, whereas another 3'UTR structure (hairpins) results in less-efficient translation and a decrease in protein. Note: mechanisms shown in the upper right and lower left panels each result in change in Hox function without impacting the expression of the Hox gene. Thus, when a *Hox* gene pattern is unchanged but new downstream functions are acquired, either mechanism may be at play.

1.3.1 Changes in *Hox* gene expression: examples of cis-regulatory evolution

Cis-regulatory changes in Hox GRNs have perhaps received the most attention in the literature, in keeping with the cis-regulatory hypothesis that postulates that genes involved in pattern formation and morphogenesis are highly constrained at the protein level but diversify due to changes in cis-regulatory elements (Carroll et al., 2005; Prud'homme et al., 2007; Carroll, 2008). This path of cis-regulatory evolution is thought to be favored because it increases flexibility while decreasing potentially negative consequences. Since most regulatory genes are pleiotropic – acting in different tissues and/or at different times during development – changes are permitted that alter expression of the regulatory gene in only specific body regions, without affecting expression in other regions, thereby limiting the impact of such changes to only a subset of overall gene activity (Stern and Orgogozo, 2008; Stern and Orgogozo, 2009). This cis-regulatory flexibility is explained in part by the modularity of cis-regulatory elements and the relative ease with which transcription factor binding sites can be gained and lost (Howard and Davidson, 2004; Davidson, 2006; Levine, 2010; Wittkopp, 2010). Below we discuss examples of *Hox* evolution due to cis-regulatory changes, including examples of dynamic changes in a rapidly evolving *Hox* gene, small variations in *Hox* expression domains, and acquisition of novel *Hox* expression patterns (Figure 1-2, upper left panel).

1.3.1a The *Hox* gene *fushi tarazu (ftz)* has undergone dramatic changes in expression pattern during arthropod evolution

Ancestrally, *fushi tarazu* (*ftz*) was likely expressed as a typical *Hox* gene, co-linearly with its neighbors in the *Hox* complex (Figure 1-1; (Telford, 2000)). This *Hox*-like pattern is retained in extant species, including chelicerates (mite) (Telford, 2000) myriapods (millipede and

centipede) (Hughes and Kaufman, 2002a; Janssen and Damen, 2006), and a crustacean (water flea) (Papillon and Telford, 2007). Yet, in *Drosophila*, *ftz* is not expressed in a *Hox*-like pattern. Rather, it is expressed in a pair-rule pattern of 7 stripes in the primordia of the alternate segmental regions missing in *ftz* mutants (Nusslein-Volhard and Wieschaus, 1980; Wakimoto and Kaufman, 1981; Hafen et al., 1984; Kuroiwa et al., 1984; Scott and Weiner, 1984; Wakimoto et al., 1984; Carroll and Scott, 1985). Expression of *ftz* in stripes is crucial for its pair-rule function: loss of stripe expression or ectopic expression of *ftz* outside the stripe domain are lethal (Struhl, 1985). The dramatic change in *ftz* expression pattern from *Hox*-like to stripes was thought to have occurred in a basal insect lineage because striped *ftz* expression was observed in the firebrat *Thermobia* (Hughes et al., 2004). However, striped expression was not observed in the grasshopper, *Schistocerca* (Dawes et al., 1994). Thus, either striped expression was lost in an orthopteran lineage or, stripes were gained independently in basal insects (firebrat) and holometabolous insects (beetle, honeybee, and fruit fly), where all *ftz* genes examined are expressed in stripes (Brown et al., 1994; Dearden et al., 2006). In this thesis I found additional changes in *ftz* expression (see Chapter 2).

1.3.1b *Hox3/zen* divergence and its co-option into a role in extra-embryonic membrane formation

zen is another rapidly evolving homeotic gene that has diverged in function from its *Hox3* homolog and taken on a new role in extra-embryonic membrane formation, presumably before the emergence of winged insects (Panfilio and Akam, 2007). *zen* has retained *Hox*-like expression in arthropods such as chelicerates (Damen and Tautz, 1998; Telford and Thomas, 1998; Abzhanov et al., 1999), myriapods (Hughes and Kaufman, 2002b; Janssen and Damen, 2006), a crustacean (Papillon and Telford, 2007), and basal insect (Hughes et al., 2004). In

contrast to *Hox3*, *zen* is expressed much earlier in embryogenesis in many insects, in the developing amnion and serosa (summarized in Table 1-1; (Schmidt-Ott, 2005; van der Zee et al., 2005; Panfilio and Roth, 2010)). Despite differences in extra-embryonic membrane formation between insects that retain separate amnion and serosal membranes and those with a fused amnioserosa (Frank and Rushlow, 1996; Lamka and Lipshitz, 1999; Schmidt-Ott, 2005), *zen* is expressed in these developing membranes in most insects examined; however, detailed analysis revealed small variations in expression patterns of *zen* orthologs (Table 1-1; (Rushlow et al., 1987a; Rushlow et al., 1987b; Falciani et al., 1996; Dearden et al., 2000; Goltsev et al., 2004; Hughes et al., 2004; van der Zee et al., 2005; Dearden et al., 2006; Panfilio et al., 2006; Rafiqi et al., 2008)). Additionally, RNAi studies confirmed that *zen* is required for extra-embryonic membrane formation in diverse insects (Rushlow et al., 1987a; van der Zee et al., 2005; Panfilio et al., 2006; Rafiqi et al., 2008; Panfilio, 2009). In sum, like *ftz*, *zen* is a divergent *Hox* gene that has been co-opted for an earlier embryonic function in insects. It would be interesting to know the function of *zen* in the basal insect *Thermobia*, where it has both *Hox*-like and extra-embryonic expression patterns. These studies would provide further clarification as to when *zen* acquired its early function in extra-embryonic membrane development and reveal whether it retained ancestral Hox-like functions while taking on new biological roles.

TABLE 1-1: A new *zen* expression pattern in insects allowed for cooption into an early developmental pathway involved in extra-embryonic membrane development.

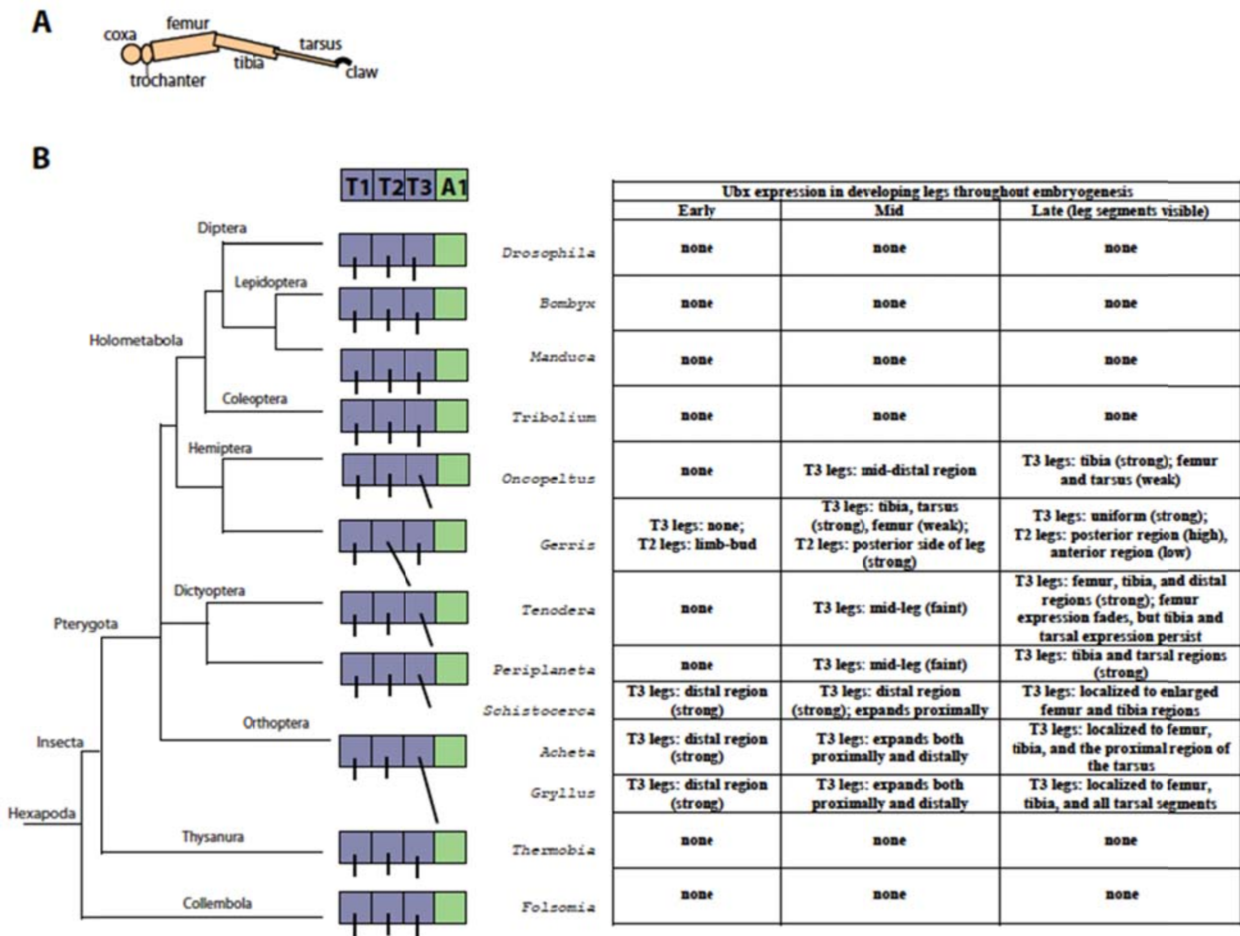
	Zen sequence		Zen expression		
	YPWM	Homeodomain	*Hox*-like	Serosa	Amnion
Chelicerates	+	+	+		
Myriapods	+	+	+		
Crustacea	+	+	+		
Folsomia	+	+	?	?	?
Thermobia	+	+	+	-	+
Schistocerca	-	+	-	+	-
Oncopeltus	-	+	-	+	-
Tribolium	-	+	-	+	+ (only *zen2*)
Apis	-	+	-	+	+
Megasalia	-	+	-	+	-
Drosophila	-	+	-	+ (amnioserosa)	

1.3.1c The case of Ubx and leg morphology

All insects have a pair of legs on each of the prothoracic (T1), mesothoracic (T2), and metathoracic (T3) segments. However, despite this conserved body plan, there is great diversity in insect leg morphology. While the three pairs of legs are uniform in length and size in some insects (e.g. (Mahfooz et al., 2004)), in others one leg is longer relative to the other legs. These changes are thought to have evolved as adaptations to different environments. Studies of the developmental basis for these variations in body plan revealed a strong correlation between leg-length and expression of *Ultrabithorax* (*Ubx*). In all cases examined where there was differentiation in one leg pair length relative to the others, variation in both the timing and domain of expression of *Ubx* during early development was found to differ in the elongated legs versus non-elongated legs (Figure 1-3). Mahfooz and colleagues (Mahfooz et al., 2004; Mahfooz et al., 2007) reported that during embryogenesis of several orthopterans and dictyopterans, Ubx expression is specifically localized to the leg segments that are larger than other leg segments in

the nymph and adult. For example, in nymphal grasshoppers, the femur and tibia segments of the T3 jumping leg are enlarged relative to other leg segments, which correspond to the regions where Ubx expression was detected in the embryo. Crickets have a hindleg similar to grasshoppers, but the tarsal segment is also elongated relative to the other legs. This leg morphology is reflected by differences in Ubx expression: crickets showed Ubx tarsal staining while grasshoppers did not. Similar expression patterns were also seen in mantis and cockroach T3 legs, which are elongated, but not as drastically as the legs used for jumping in orthopterans; in dictyopterans this corresponded to Ubx expression later in embryogenesis, suggesting timing of Ubx expression is also important in determining leg length. Together, these studies correlate increases in Ubx expression with increased growth of leg structures, suggesting that changes in the expression patterns of Ubx promote morphological diversification.

Studies of water striders (hemipterans) have gone one step further by analyzing expression as well as function, using RNAi. In these water striders, the T2 leg is much longer than the T1 and T3 legs. Khila and colleagues (Khila et al., 2009) found that early during embryogenesis Ubx was expressed in the T2 leg, but not the T3 leg. Later in development, Ubx was also strongly expressed throughout the developing T3 leg. *Ubx*-RNAi revealed a dual role in the developing legs of these hemipterans: first, it promotes growth of the T2 leg, as knocking-down gene expression resulted in shorter T2 legs, and second, Ubx acts to shorten the T3 leg, for embryos had a much longer T3 leg when *Ubx* was depleted. In this case, Ubx has opposing functions in the T2 and T3 developing legs. In conclusion, many studies have correlated changes in the timing or domain of Ubx expression with variation in leg morphology and while these changes are sometimes subtle, they likely have adaptive significance.



	Ubx expression in developing legs throughout embryogenesis		
	Early	Mid	Late (leg segments visible)
Drosophila	none	none	none
Bombyx	none	none	none
Manduca	none	none	none
Tribolium	none	none	none
Oncopeltus	none	T3 legs: mid-distal region	T3 legs: tibia (strong); femur and tarsus (weak)
Gerris	T3 legs: none; T2 legs: limb-bud	T3 legs: tibia, tarsus (strong), femur (weak); T2 legs: posterior side of leg (strong)	T3 legs: uniform (strong); T2 legs: posterior region (high), anterior region (low)
Tenodera	none	T3 legs: mid-leg (faint)	T3 legs: femur, tibia, and distal regions (strong); femur expression fades, but tibia and tarsal expression persist
Periplaneta	none	T3 legs: mid-leg (faint)	T3 legs: tibia and tarsal regions (strong)
Schistocerca	T3 legs: distal region (strong)	T3 legs: distal region (strong); expands proximally	T3 legs: localized to enlarged femur and tibia regions
Acheta	T3 legs: distal region (strong)	T3 legs: expands both proximally and distally	T3 legs: localized to femur, tibia, and the proximal region of the tarsus
Gryllus	T3 legs: distal region (strong)	T3 legs: expands both proximally and distally	T3 legs: localized to femur, tibia, and all tarsal segments
Thermobia	none	none	none
Folsomia	none	none	none

FIGURE 1-3. Variations in Ubx expression in developing insect legs have contributed to morphological diversity. A) A schematic showing the five segments and claw of an insect leg. B) (Left) Examples where shifts or variation in Ubx expression are correlated with leg diversity. The thoracic segments (blue) and first abdominal segment (green) of arthropods is shown, with relative leg lengths. (Right) Ubx expression during early, mid, and late embryogenesis correlates with differences in leg morphology seen in nature and depicted on the left.

<u>1.3.1d Scr expression varies among different insects</u>

Several recent studies highlight the evolutionary flexibility of the *Hox* gene *Scr*. In *Drosophila*, *Scr* is expressed in the T1 segment primordial, and plays a role in patterning the labial appendages and the T1 segment, the latter role including suppression of wing development

on this segment (Mahaffey and Kaufman, 1987; Riley et al., 1987; LeMotte et al., 1989; Carroll et al., 1995). *Dm*-Scr also cooperates with other *Hox* genes to impact the formation of a dorsal ridge that demarcates a tagmatic boundary between the insect head and thorax (Rogers and Kaufman, 1996; Rogers et al., 1997). Previous studies of *Scr* expression and function were carried out in the beetle *Tribolium* (*Tc*-Scr) and the milkweed bug *Oncopeltus* (*Of*-Scr), where expression and function were found to be similar to *Drosophila* (Hughes and Kaufman, 2000; Curtis et al., 2001; DeCamillis et al., 2001; Shippy et al., 2006; Chesebro et al., 2009). Popadic's group has extended these earlier studies by examining Scr expression patterns in six ametabolous and hemimetabolous insects (Passalacqua et al., 2010). In all hemimetabolous species, Scr protein was found to accumulate in the head; however, variations in Scr expression were observed between species. These included shifts in the domains within the head that Scr was detected, and variability as to whether and/or where Scr was expressed within the developing T1 leg primordial. Interestingly, in the basal insect *Thermobia*, no Scr protein was detected in T1, although *Scr* RNA expression was found (Popadic et al., 1998; Passalacqua et al., 2010) (see also Posttranscriptional regulation of Hox genes). The functional consequences of these variations in expression were examined using RNAi to knockdown *Scr* expression in the cockroach, *Periplaneta* (Hrycaj et al., 2010). Scr was found to be required for proper development of the labial palps, as seen in other insects. RNAi knockdown also resulted in an ectopic supernumerary segment between the head and first thoracic segment; this phenotype is similar to that observed in *Tribolium* (Shippy et al., 2006). Late RNAi effects revealed Scr's role in wing suppression, as seen also in other insects including hemipterans such as the milkweed bug (Chesebro et al., 2009) and treehoppers (Prud'homme et al., 2011) (see below), and in holometabolous insects such as horned beetles (Wasik et al., 2010) (see below) and fruit flies

(Carroll et al., 1995). Interestingly, and in contrast to *Drosophila*, in neither the cockroach nor horned beetle did *Scr* RNAi affect the external morphology of T1 legs (Hrycaj et al., 2010; Wasik et al., 2010). This observation led the authors to suggest that expression in T1 primordia preceded the function of Scr in T1-leg identity specification, as Scr is expressed in T1 in the cockroach. An alternate possibility that remains to be investigated is that leg identity was altered in more subtle ways that were not assessed in these experiments.

1.3.1e New expression domains suggest novel Antp functions in butterflies

A striking example of a gain of a novel *Hox* expression mode that is correlated with an evolutionary novelty was reported for butterfly eyespots (Saenko et al., 2011). In the nymphalid butterfly *Bicyclus anynana*, a new expression pattern of *Antp* was observed. While still retaining its ancestral *Hox*-like expression pattern, *Antp* was also found to be expressed in a new domain in the organizing center of the eyespots. Previously, several highly conserved developmental genes, such as *Distalless (Dll)* and *Engrailed*, were shown to be co-opted for eyespot specification in butterflies (Weatherbee et al., 1999; Brunetti et al., 2001). Interestingly, *Antp* expression in the eyespot organizer region is earlier than these other regulatory genes, suggesting that it may play a critical role in initiating eyespot formation. This novel *Antp* expression pattern was also seen in several species closely related to *Bicyclus*, but was not found in *Junonia coenia*, a species with morphologically similar eyespots that diverged from *Bicyclus* ~ 90 million years ago. Future work will be needed to uncover the mechanisms that led to activation of *Antp* in this new expression domain in a certain lineages and to test the hypothesis that *Antp* indeed functions as an eyespot regulator, thereby linking the new expression pattern to morphological diversification.

1.3.1f Cis-regulatory changes in vertebrate *Hox* complexes

In a landmark study some years ago, Capecchi's group showed that the coding regions of *Hox* paralogs were functionally interchangeable in mice, thereby demonstrating that cis-regulatory change played a dominant role in the diversification of the *Hox* genes present in different clusters in vertebrates (Greer et al., 2000). Recent work has extended these studies to elucidate the underlying evolutionary mechanisms. In one example, using a novel approach in which a targeted translocation was induced in the mouse genome (Wu et al., 2007), Duboule's group placed the *HoxC* gene cluster under the regulatory control of the *HoxD* genomic locus and tested its ability to rescue *HoxD* loss-of-function phenotypes, which include defects in digit formation (Tschopp et al., 2011). Their studies showed the *HoxC* cluster was largely able to rescue *HoxD* mutants, providing a compelling example of the importance of regulatory evolution within *Hox* complexes. Thus, after the duplication of *Hox* complexes in vertebrates, redundancy permitted diversification of highly related paralogs. This diversification appears to have occurred primarily at the level of cis-regulatory change, with the Hox proteins themselves retaining ancestral and shared properties.

Differences in *Hox* gene regulation do not only apply across *Hox* complexes within a given species, but are also thought to be responsible for morphological differences between species. An important new study demonstrated that variations in *Hox* expression between birds and mammals in sensory systems that detect pain, touch and other external stimuli, result from differences in expression of *Hoxd1* (Guo et al., 2011). In mice, but not chick, the growth factor NGF induces expression of *Hoxd1*. Mice lacking *Hoxd1* develop altered neuronal circuitry that resembles that seen in chick. Conversely, mis-expression of *Hoxd1* in the chick induced an axonal patterning similar to that seen in the mouse. These studies thus revealed a novel role for a

Hoxd1 GRN in wiring of the sensory system in vertebrates. Importantly, they implicate a change in *Hoxd1* expression, and define its origin – a switch in responsiveness to growth factor signaling - as the causal switch in an important functional difference between species. Together, these studies provide nice examples of how changes in the expression of *Hox* genes may have driven the evolution of novel morphologies.

1.3.2 Changes in Hox downstream target gene regulation without changes in *Hox* expression: cis-regulatory changes in target genes or novel protein functions

In some evolutionary scenarios, new biological functions of a *Hox* gene have been observed without corresponding changes in the expression pattern of that *Hox* gene. In such cases, the change in phenotype may be a result of changes in the cis-regulatory regions of downstream target genes (Figure 1-2, upper right panel; (Carroll et al., 2005)) or changes in the Hox protein that alter its regulatory specificity (see mechanism III below, Figure 1-2 lower left panel, (Lohr et al., 2001; Lohr and Pick, 2005)). For all of the examples discussed in this section, future studies are required to distinguish between these mechanisms.

1.3.2a Scr is a key player in insect morphological evolution

Two striking examples of the genetic basis of morphological evolution both result from novel functions of the *Hox* gene *Scr*. The first example links Scr to the horns of dung beetles, which differ dramatically in size and shape. These horns, which develop in the pre-pupal stage as epidermal outgrowths of the head and/or prothorax and then undergo remodeling during the pupal stage, are diverse and dramatic in appearance in adults (Moczek et al., 2007; Moczek, 2008). Their location suggested that the *Hox* gene *Scr* might be involved in their patterning. To

examine this, *Scr* orthologs were isolated from two horned beetle species and expression and function were assessed by RNAi knock-down (Wasik et al., 2010). The expression pattern of Scr was *Hox*-like: both mRNA and Scr protein were expressed in patterns similar to those seen in other insects. Interestingly, RNAi experiments revealed a novel role for Scr in growth of the pronotal horns. These effects differed somewhat between the two species, which also showed sex differences in response to Scr RNAi, suggesting variations in Scr function and its interaction with sex determination pathways across species. Together, these results demonstrate that a morphological novelty in this group of dung beetles, which is not seen in the numerous other insect lineages that express Scr in the same domain, results from changes in Scr function. Scr was co-opted into at least one new developmental pathway, allowing it to acquire a new function with distinct effects on morphology without noticeable change in its typical *Hox*-like expression pattern or loss of its “traditional” *Hox* roles in body patterning.

A second elegant example of changing Scr function without change in its expression pattern was found for treehopper helmets (Prud'homme et al., 2011). Treehoppers are a large group of diverse hemipteran insects that share a novel helmet structure, which manifests in a remarkable array of appearances. Similar to the logic for beetle horns discussed above, the location of the helmet suggested a potential role for Scr. In their recent study, Prud’homme et al. (2011) showed that Scr expression in treehoppers is similar to that in other insects (*Hox*-like), but helmets, novel wing-like structures, have been allowed to develop and diversify on T1 because Scr has lost the ability to repress genes necessary for wing development. One of these genes is *nubbin* (*nub*): *nub* is necessary for wing development in *Drosophila* and is absent from T1(Cifuentes and Garcia-Bellido, 1997). However, in the treehopper *Publilia modesta*, Nub was detected in the developing helmet in a pattern similar to that of developing wings. This suggests

that *Pm*-Scr has lost the ability to down-regulate *nub* expression, as well as other wing specification genes, and this was likely a critical step in the morphological evolution of this structure. The authors further showed that ectopic expression of treehopper *Scr* repressed wing formation in *Drosophila*, suggesting that change in the function of Scr does not explain its loss of ability to repress wing-realizator genes. These results leave open two possible mechanistic explanations: first, that changes in the cis-regulatory regions of target genes rendered them unresponsive to Scr-repression; second, that changes in an Scr-cofactor interaction altered its regulatory specificity such that wing-realizator genes were no longer negatively regulated. Although the former is likely (cis-regulatory changes in *Hox* targets), further experiments are required to distinguish between these mechanisms (e.g., see approach of (Gompel et al., 2005; Prud'homme et al., 2006; Wittkopp et al., 2008)). Interestingly, Miko and colleagues recently suggested that helmets are not likely to be homologous to wings, based on detailed morphological comparions (Miko et al., 2012). Irrespective of how this disagreement is resolved, the incredible morphological diversity of treehopper helmets and the rapid progress in identifying the patterning genes controlling its development make this an exciting system for working out molecular mechanisms leading to development and differentiation of complex and evolutionarily plastic body structures.

1.3.2b Many ways to make a wing: the role of Ubx in wing development

Most modern insects have two pairs of wings on the thorax: one pair on each of the T2 and T3 segments (Angelini and Kaufman, 2005). Insects such as dragonflies and damselflies have two pairs of very-similar wings, a situation reminiscent of the ancestral state of winged insects (Weatherbee et al., 1999); however, wing-pair morphology differs in many extant insects.

Cases discussed below implicate *Hox* genes in the diversification of distinct fore- and hindwing morphologies seen in several different insect lineages.

In most insects, Ultrabithorax (Ubx) is expressed in the developing T3 segment where it impacts hindwing morphology (Lewis, 1978; Carroll et al., 2005). *Drosophila* and other dipterans have two sets of wings that differ in appearance: the wing found on T2 is important for flight, while the hind-wing has been modified to a balancing structure called a haltere. In *Drosophila*, loss-of-function Ubx mutations result in transformation of T3 towards T2 with the haltere transformed toward forewing (Lewis, 1978). Conversely, mutations that cause ectopic expression of Ubx in the developing wing transform wing tissue into haltere tissue (Gonzalez-Gaitan et al., 1990). These results suggest that Ubx plays two roles in *Drosophila* wing formation on T3: one in promoting haltere development and a second role in suppressing forewing development. The latter role has been examined at the molecular level by Weatherbee et al. who found that Ubx negatively regulates target genes involved in forewing formation, such as *wingless*, *spalt-related*, *vestigial*, *Serum Response Factor*, and *achaete-scute* (Weatherbee et al., 1998). Ubx directly binds the cis-regulatory regions of forewing promoting genes such as *spalt*, resulting in silencing of gene expression in the haltere (Galant et al., 2002).

In Lepidoptera, fore- and hindwings differ, but in contrast to the dipteran haltere, the hindwing is fully developed. While Ubx is expressed in the developing hindwing in the butterfly *Precis coenia* (Warren et al., 1994), it does not repress the forewing promoting genes, suggesting loss of Ubx-responsiveness to Ubx target genes in this species (Weatherbee et al., 1999). In Coleoptera, both the T2 and T3 segments have wings, but different from most insects, the wings on T3 most resemble typical hindwings used for flight, while the T2 wings are modified to sclerotized elytra, or wing covers (Tomoyasu et al., 2005). Mutations in the *Tc-Scr* homolog

Cephalothorax result in elytron-like structures on T1, suggesting that *Cx* suppresses wing formation, as is the case in *Drosophila*, and several hemipterans. However, the role of Ubx in wing identity in *Tribolium* appears to differ from that in *Drosophila*. RNAi targeting the *Tribolium Ubx* gene *Ultrathorax* (*Utx*) transformed the hind-wing to an elytron and small elytra-like appendages appeared on the first abdominal segment (Tomoyasu et al., 2005). An examination of several genes involved in wing development revealed that several genes have different expression patterns in the T2 elytron and T3 hind-wing developing regions, all of which are altered in *Ubx* RNAi experiments (Tomoyasu et al., 2005). These results suggest that *Ubx* functions in the beetle to promote the development of hindwings by repressing genes involved in elytra formation – a taxon-specific role that required re-organization of gene regulatory connections. In sum, Ubx shares a role in hindwing development in diverse insects, but its specific role in this process can change.

1.3.3 Changes in Hox protein function

Hox genes are generally considered to be highly conserved and evolutionarily constrained at the level of protein activity. This conclusion comes largely from trans-species experiments in which *Hox* genes from distant taxa were expressed in *Drosophila* and demonstrated conserved function (Malicki et al., 1990; McGinnis et al., 1990; Malicki et al., 1993; Zhao et al., 1993; Zhao et al., 1996). However, evidence is continuing to emerge that changes in Hox protein sequence occur and that these changes can lead to changes in the functional properties of Hox proteins by altering their regulatory specificity (Figure 1-2, lower left panel; (Mann and Carroll, 2002; Hsia and McGinnis, 2003; Lynch and Wagner, 2008; Wagner and Lynch, 2008)).

1.3.3a Ubx and abdominal appendages

An important example of this type of change was provided several years ago by groups studying the role of *Hox* genes in patterning abdominal appendages. In contrast to insects, crustaceans have appendages on posterior segments (Galant and Carroll, 2002; Ronshaugen et al., 2002). The limb-less insect abdomen is thought to be explained in part by the ability of Ubx to repress the target gene *Dll* (Vachon et al., 1992). In constrast to Dm-Ubx, Ubx proteins from an onychophoran and a crustacean did not repress *Dll* when expressed in *Drosophila*. It was further shown that insect Ubx proteins have taken on a role in *Dll* repression via the acquisition of a repressor domain, which is missing or non-functional in non-insect Ubx proteins. This suggests that the acquisition of a repressor function, due to specific changes in Ubx protein sequences in insect lineages, contributed to the evolution of the limbless abdomen in insects.

1.3.3b Escape from colinearity enabled variation in Hox protein potential: "*ftz*-ing" around during insect evolution

The case of *fushi tarazu* (*ftz*) provides a compelling example of a *Hox* gene that has changed function during evolution. *ftz* pair-rule segmentation function differs from neighboring homeotic *Hox* genes, which specify the identity of body regions (see above), and changes in Ftz protein sequence have contributed to this change in function. *Dm*-Ftz interacts with an obligate co-factor, the orphan nuclear receptor Ftz-F1 (Guichet et al., 1997; Yu et al., 1997) and together activate downstream target gene expression to promote segment formation (Florence et al., 1997; Guichet et al., 1997; Yu et al., 1997; Yussa et al., 2001; Hou et al., 2009). The interaction between Ftz and Ftz-F1 is dependent upon a nuclear receptor coactivator-like LXXLL motif in Ftz that binds the AF-2 domain of Ftz-F1 (Schwartz et al., 2001; Suzuki et al., 2001; Yussa et al.,

2001). We examined the homeotic and segmentation potential of Ftz proteins from the beetle *Tribolium* (*Tc*-Ftz) and grasshopper *Schistocerca* (*Sg*-Ftz) by ectopic expression in *Drosophila* (Lohr et al., 2001). Antenna-to-leg transformations were seen with both *Tc*-Ftz and *Sg*-Ftz but not *Dm*-Ftz (Lohr et al., 2001), suggesting the beetle and grasshopper proteins retain homeotic potential, while *Dm*-Ftz has lost the potential to carry out homeotic functions even when expressed in a homeotic fashion. *Tc*-Ftz also displayed segmentation potential similar to that of *Dm*-Ftz, whereas *Sg*-Ftz only showed marginal segmentation potential (Lohr et al., 2001). These functional properties correlate with cofactor interaction motifs: *Dm*-Ftz lacks the YPWM motif present in most Hox proteins and required for interaction with Exd, has an LXXLL motif required for Ftz-F1 interaction, and displays only segmentation potential; *Tc*-Ftz, has both a YPWM and LXXLL motif and homeotic and segmentation potential in vivo; *Sg*-Ftz, has only a YPWM motif and exhibits mostly homeotic potential. In conclusion, protein changes were important for a switch in Ftz function from a *Hox*-like to pair-rule segmentation gene in insects. While functional studies in more insects are needed to study biological roles of diverse Ftz proteins, *Drosophila* has provided an excellent model system for testing hypotheses about sequences changes required for a Hox protein to switch function during evolution.

1.3.3c Bcd acquires a new function in higher insects

In higher Diptera, a duplication of *zen* produced *bcd*, a novel *Hox* gene which took on a unique role in anterior patterning due to its expression at the embryonic anterior pole and unique modifications of its protein sequence, including a novel DNA binding specificity and the ability to bind to RNA and thus regulate translation (Hanes and Brent, 1989; Lynch and Desplan, 2003; Schmidt-Ott and Wimmer, 2004; McGregor, 2005; Lemke et al., 2008). Recently the sequence,

expression and function of the *bcd* gene from a 'lower' fly have been studied (Lemke et al., 2010). *Episyrphus bicoid* (*Eb-bcd*) is localized to the anterior pole of the embryo, as is *Dm-Bcd*. However, while *Eb*-Bcd protein is similar to *Dm*-Bcd, being a clear ortholog with a similar homeodomain, it lacks several of the sequence motifs that are important for *Dm*-Bcd function, suggesting differences in the biochemical properties of the two proteins. Further, RNAi experiments showed that *Eb*-Bcd is the major anterior determinant in this fly and that it is responsible for additional aspects of patterning, such as regulation of gap gene expression. This latter role of *Eb*-Bcd function is shared among several different maternal gene products in *Drosophila*. Thus, variations in the protein sequence and biological functions of Bcd were observed, despite a shared expression pattern. Future experiments will determine whether *Eb*-Bcd gained additional functions, which are carried out by different genes in *Drosophila*, or *Dm*-Bcd lost some of the regulatory potential of a shared Bcd ancestor.

1.3.3d Hox protein changes and the evolution of placental mammals

Perhaps the best-documented example of a change in Hox protein sequence implicated in the evolution of a morphological novelty is mammalian HoxA11, which underwent a period of adaptive evolution in the stem lineage of placental mammals to take on roles in the establishment and maintenance of pregnancy (Lynch et al., 2004). It was previously shown that HoxA11 is a transcriptional activator of prolactin, a gene critical for establishment of pregnancy in mammals. HoxA11 from placental mammals activated prolactin gene expression, while HoxA11 from birds or non-placental mammals (opossum, platypus) did not. This functional difference in HoxA11 was explained by changes in HoxA11 protein that allowed for interaction with a new partner, Foxo1a (Lynch et al., 2008; Lynch et al., 2009). Recently, Wagner's group tested the

biochemical basis of this species-specific protein-protein interaction. Co-immunoprecipitation experiments were carried out with proteins from extant species as well as reconstructed ancestral HoxA11 and Foxo1A proteins (Brayer et al., 2011). Foxo1a interacted with HoxA11 from placental mammals (human, opossum) and with ancestral eutherian, therian and mammalian HoxA11. However, Foxo1a failed to interact with extant bird (chicken) HoxA11. To determine whether changes in HoxA11, Foxo1a, or both, facilitated the acquisition of this protein-protein interaction, binding of human HoxA11 to a range of Foxo1a proteins was examined; all Foxo1a proteins interacted with human HoxA11 showing that changes in Foxo1a were not necessary for the functional switch. Rather, changes in HoxA11 protein permitted a new interaction with Foxo1a. Thus, interaction between HoxA11 and Foxo1a only occurs in mammals, despite the fact that both proteins are present in outgroups. Interestingly, this interaction arose in a mammalian stem lineage, before HoxA11/Foxo1a acquired the ability to regulate prolactin gene expression – a feature that arose later, in placental mammals. It will be of great interest to see what the original role was of the HoxA11/Foxoa1 pair, prior to its recruitment for prolactin regulation, what the factors were that allowed and selected for this gain-of-function change in placental mammals and to what extent this single switch in function of a transcriptional regulator explains the emergence of placental development.

1.3.4 Post-transcriptional regulation of *Hox* genes

Hox genes may also be regulated post-transcriptionally such that the expression and function of Hox proteins is modulated without impacting cis-regulatory or coding sequences (Figure 2, lower right panel). This regulatory mechanism has not received much attention in the past, but is highlighted by recent findings.

Several examples were reported recently in which Hox protein is not detectable in regions of embryos where mRNA is found. In a crustacean, the brine shrimp *Artemia*, *Abd-A* mRNA was detected in a *Hox*-like pattern in the trunk region of early embryos but Abd-A protein was undetectable (Hsia et al., 2010). When the shrimp protein, *Af-Abd-A,* was expressed in *Drosophila*, no protein was detected either. Together, this suggests a change in *Abd-A* mRNA that decreased its stability or translation efficacy, or both. The absence of Abd-A protein in the trunk likely contributes to the ability of this species to develop limbs throughout the trunk region – a phenomenon which is repressed by both Ubx and Abd-A in insects that lack abdominal legs (Hsia et al., 2010). Similarly, discrepancies between the mRNA and protein patterns were reported for Scr in both *Thermobia* and *Oncopeltus* (Popadic et al., 1998; Passalacqua et al., 2010).

Additional studies provide evidence for regulation of *Hox* genes at the mRNA level. Studies from the Alonso lab showed that differential 3'UTR formation in *Ubx* generates targets for regulation by different miRNAs and that this occurs in a developmental and tissue-specific fashion (Thomsen et al., 2010). Building upon this, Patraquim et. al (2011) compared sequences of *Hox* gene 3'UTR sequences from the 12 sequenced drosophilid genomes. They found that these 3'UTRs are evolving (as would be expected for any nucleotide sequence), but while the sequences differ greatly, the topology of these regions appears to be under strong selective pressure, suggestive of functional constraint. The changes seen in 3'UTR sequence include changes in potential regulatory sites for miRNAs (Patraquim et al., 2011). Finally, in a recent study in millipedes, a *Ubx* antisense transcript was found that is expressed in a pattern complementary to *Ubx* coding RNA, suggesting that the antisense RNA is negatively regulating the transcription or stability of *Ubx* sense RNA (Janssen and Budd, 2010). Although the

mechanism remains to be determined, this scenario is reminiscent of the noncoding RNAs at the *bithoraxoid* (*bxd*) locus in *Drosophila,* which repress *Ubx* expression in *cis* by transcriptional interference (Petruk et al., 2006).

Taken together, these findings document additional levels at which evolution tinkers with *Hox* function. In this context, observations from the Hogness lab in the early 1980's may have been visionary, as they suggested, long before RNAi was a common tool of molecular geneticists, that "The elements of the *bxd* region might make RNA products that interact with the *Ubx* RNA or with other small RNA molecules involved in processing *Ubx* RNA." (Bender et al., 1985; Hogness et al., 1985).

1.3.5 *Hox* mechanism conclusions and emerging themes

Clearly, there is evolutionary flexibility in *Hox* genes, despite the fact that these genes were thought to be highly constrained because of their essential roles in embryonic patterning. These genes, which were once thought to be highly static building blocks of the animal body plan, are in fact changing, and change is occurring at many levels. In the examples reviewed here, both gain and loss of Hox activity has been observed. In some cases, *Hox* genes have been co-opted into new developmental pathways during evolution without loss of "traditional" *Hox*-like functions (e.g., co-option of Scr into regulation of beetle horns or treehopper helmets). In others, redundancy of duplicate genes has allowed for subfunctionalization or neofunctionalization (e.g., in the case of *Antp* and *ftz*, *Antp* has maintained the traditional *Hox*-like roles, freeing up *ftz* to diverge). Still, in other cases, new expression domains have emerged, due to cis-regulatory changes in the *Hox* genes, or changes in upstream activators (e.g., expression of Antp in the eyespot primordia of butterflies). In some cases, variations in *Hox*

expression patterns have yet to be correlated with specific morphological divergence, but their variation, while in some ways subtle, is much more extensive than previously imagined (e.g., Scr expression within the head and thorax of diverse insects). Finally, the importance of post-transcriptional control of Hox function has emerged, making extrapolation about function from *in situ* hybridization data even more difficult. In sum, there are many examples of evolutionary flexibility in *Hox* genes, and these changes occur at all four mechanistic levels discussed here. Much remains to be learned about these and other as yet undiscovered mechanisms underlying

y Ĉ y Ī y y Ĉ y y

Chapter 2

Rapid isolation of gene homologs across taxa: Efficient identification and isolation of gene orthologs from non-model organism genomes, a technical report
[Published: Heffer and Pick, *Evodevo*, 2011]

2.1 Introduction

One focus of evolutionary biologists is to understand how changes in regulatory and coding regions of genes contribute to species evolution and adaptation (Schlosser and Wagner, 2004; Carroll et al., 2005). This requires sequence comparisons across distantly related taxa as well as among closely related species. A major limitation in studying molecular evolution is the amount of comprehensive sequence data available to track these changes in genes and their networks. Standard approaches include comparisons across widely divergent model organisms, comparison of gene sequences that have been deposited in databases, and comparisons of whole genome sequences. This can result in an incomplete matrix of information about the lineages of particular gene families, making it difficult to trace steps leading to functional changes in regulatory and coding sequences. Additionally, the sequence conservation of duplicated and diverged genes within gene families (Ohno, 1970; Lynch and Force, 2000) poses a challenge: How can we identify a particular member of a gene family without isolating and screening through closely-related homologs? Here we report a strategy to efficiently isolate genes from genomic DNA that can be used to obtain sequence information from un-sequenced genomes and non-model organisms not easily reared in the laboratory. Rapid Isolation of Gene Homologs across Taxa (RIGHT) is based on the fact that homologous genes (both paralogs and orthologs) generally show conservation of at least one domain, even if other parts of the sequence are under weaker selective pressure. For example, the Hox proteins have retained the conserved DNA binding domain after duplication and divergence (McGinnis and Krumlauf, 1992; Gehring et al.,

2009). While not forging fundamentally new technology, this approach combines and modifies existing procedures to facilitate the rapid isolation of genes, allowing sampling of a large number of taxa.

2.2 Results

2.2.1 RIGHT Methodology

RIGHT methodology utilizes degenerate polymerase chain reaction (PCR) and gene-specific amplified-length fragment polymorphism (AFLP) to allow for rapid gene isolation. First, degenerate primers are designed to amplify a small region of less than 200 bases of the conserved domain characteristic of the gene family (Figure 2-1, Step 1). One primer is derived from the unique signature motif in the homolog of interest, while the other can be shared with other family members. A variation of touchdown PCR is used that is optimized for these degenerate primers. Nested PCR is done to ensure that the correct gene-family member is amplified. The PCR product is run on an agarose gel, purified and sequenced. The product can be positively identified by characteristic residues in the homolog of interest, along with BLAST searches against other species. Together, this allows for increased confidence that the isolated gene product corresponds to the gene of interest. From this short DNA sequence obtained from degenerate PCR, gene-specific primers are designed for subsequent reactions.

Sequence up- and downstream of the conserved region (obtained in Step 1, Figure 2-1) is next isolated by modifications of AFLP and TE-display techniques (Vos et al., 1995; Beeman and Stauth, 1997; Casa et al., 2000; Hawthorne, 2001; Biedler et al., 2003; Subramanian et al., 2007) that allow selective amplification of only the gene sequence of interest. Traditional AFLP

uses restriction enzymes to digest genomic DNA followed by ligation of adapters of known sequence to DNA ends. Adapter-specific primers are used in subsequent PCRs to amplify DNA fragments, which are then separated on a gel and analyzed. RIGHT uses the basic idea of AFLP up to the amplification step; however, instead of amplifying DNA fragments using adapter sequences as both primers (which generates many fragments), an adapter-specific primer is used as one primer and a gene-specific primer (derived from degenerate PCR used in Figure 2-1, Step 1) as the other primer. Thus, only a sequence from the gene of interest is isolated. The digestion of genomic DNA and ligation of adapters is done in a single step (Figure 2-1, Step 2). Adapter sequences are designed to anneal to, but destroy, restriction sites in order to avoid re-digestion in this combined restriction/ligation reaction. Several different restriction digests are set up in parallel to provide different-length PCR templates covering the gene of interest. This is also beneficial because restriction site locations are not known for genomes that have not been sequenced. The digestion/ligation is followed by two rounds of nested PCR (Figure 2-1, Step 3), which functions to increase specificity of primer binding and the amount of product. After the PCR product is amplified and sequenced, new gene-specific primers are designed at the sequence ends to repeat PCRs with a different restriction digest/ligation as template in order to extend the sequence. By repeating this process, one can “walk” along the genomic sequence to isolate the entire coding sequence (Figure 2-1, Step 4).

In most cases only one clear product was observed after nested PCR; however, occasionally there were several. In these situations, either all products were sequenced or products were re-amplified using the same primers or another nested set to reduce the number of products. In cases where multiple bands persisted, it was usually due to restriction sites that were very close together in the genome and almost all of the sequenced regions overlapped. After a

new sequence has been isolated, its continuity is always checked by PCR with primers at extreme opposite ends of the sequence that has been obtained to make sure the sequence being isolated is contiguous with that upstream and/or downstream (Figure 2-1, Step 5). This is very important because, although infrequent, ligation may occur between genomic DNA fragments in Step 2. As demonstrated, RIGHT provides efficiency and saves time when compared to other protocols. This combination is a powerful method for obtaining full gene sequence information, including coding and regulatory regions.

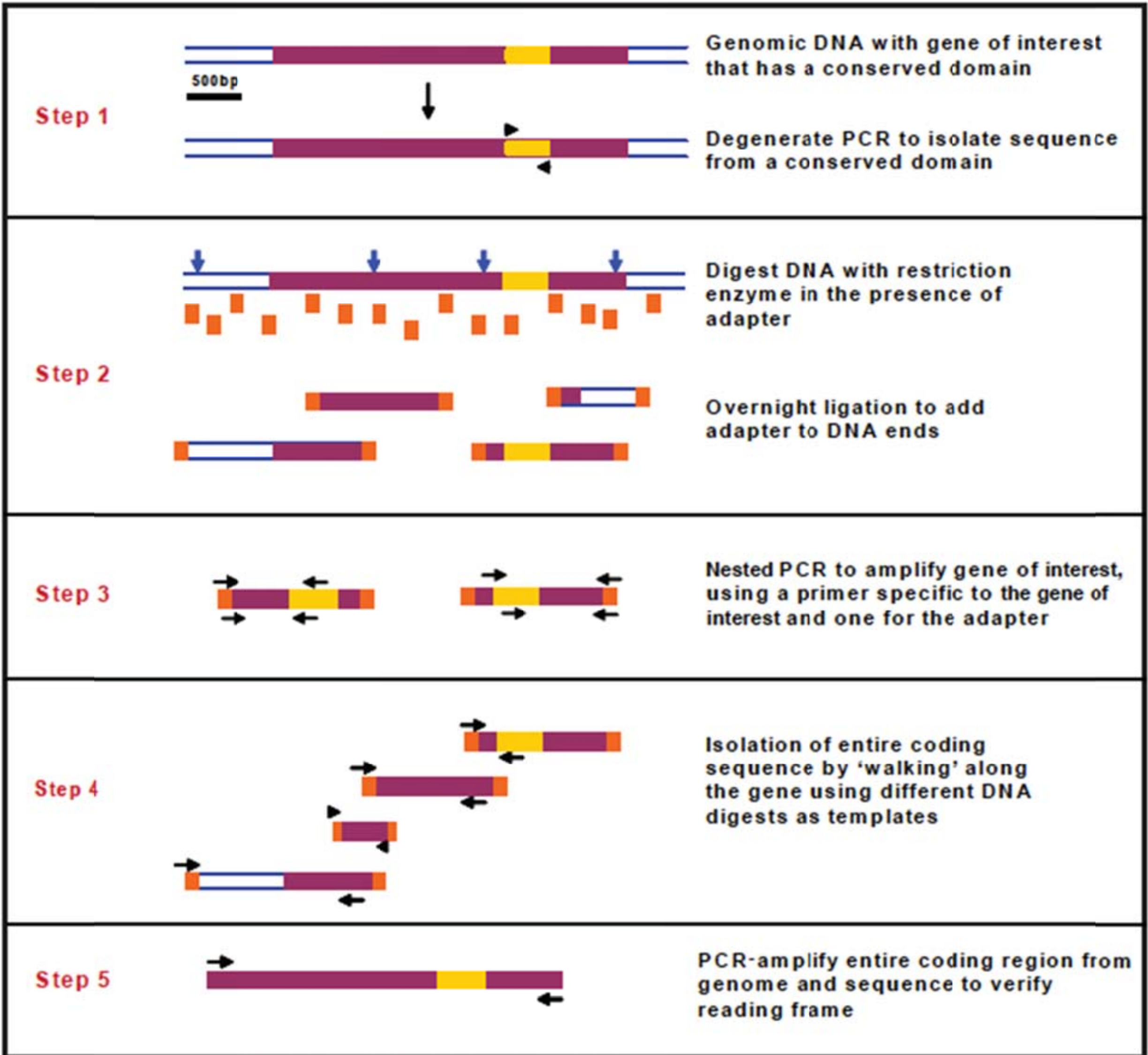


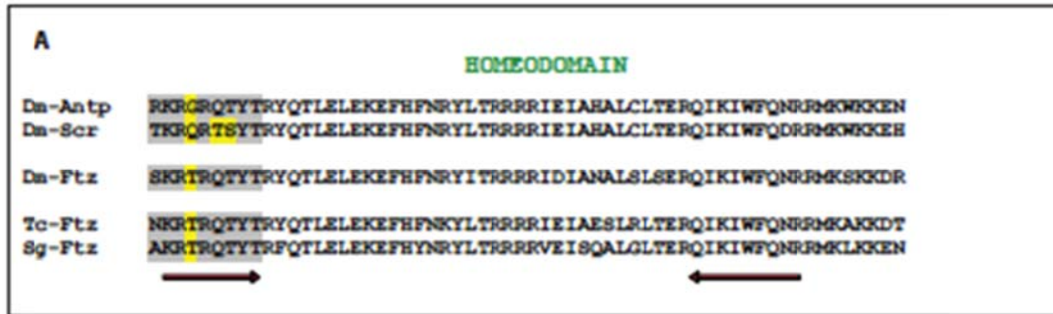
FIGURE 2-1. Overview of RIGHT technique used to isolate homologous genes from large gene families. All steps are described in the text. Oligonucleotides that were annealed to make the adapter destroyed the restriction site. All reverse primers were ordered with the 5' end phosphorylated, or were phosphorylated before annealing with an appropriate enzyme. For example, *MseI* digest/ligation: F-5' GACGATGAGTCTTGAGTTCAGTCTGTA, R-5'PhosTATACAGACTGAACTCAAGACTCATC; *XhoI*: F-5' GACGATGAGTCTT-GAGTTCAGTCTGTA, R-5'PhosTCGATACAGACTGAACTCAAGA CTCATC.

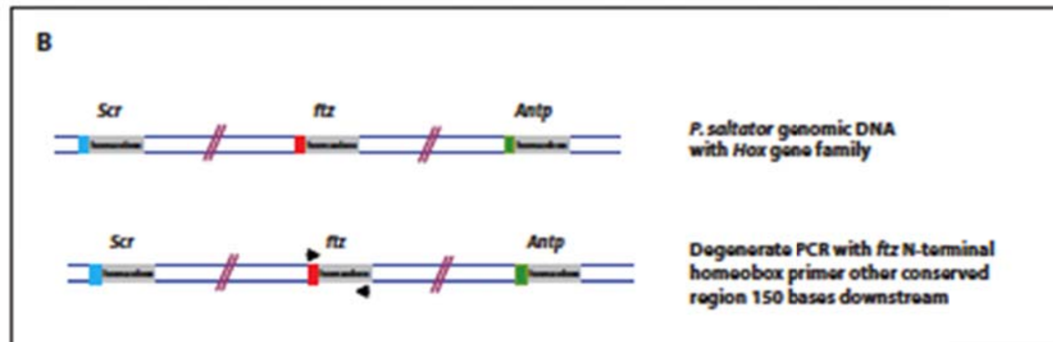
2.2.2 RIGHT isolation of homeobox and nuclear receptor genes

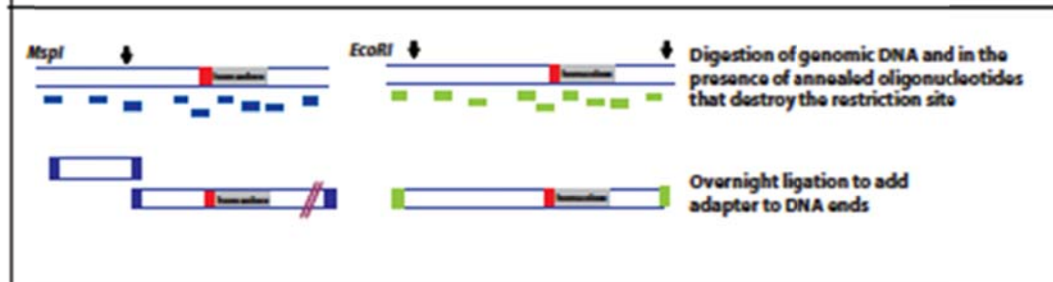
RIGHT has been used successfully in our laboratory to efficiently isolate specific members of several large gene families. The technique was first developed to isolate a rapidly-evolving member of the *Hox* gene family, *fushi tarazu* (*ftz)* (Telford, 2000; Lohr et al., 2001; Lohr and Pick, 2005). First, degenerate PCR primers were designed based upon signature residues encoding the amino-terminal end of the DNA-binding homeodomain and another highly conserved region with low degeneracy approximately 150 bases downstream (Figure 2-2A, arrows). Ftz homologs were positively identified by characteristic residues in the homeodomain. Next, gene-specific AFLP was carried out using a combination of unique restriction/ligation templates for PCR with one *ftz*-specific primer and one adapter-specific primer (Figure 2-2B). All products were sequenced and gene-sequence continuity verified by PCR with genomic DNA and primers designed to the extreme 5' and 3' ends of sequence that had been isolated. Full-length *ftz* sequences, including putative introns, were isolated by genomic walking until translation initiation and stop codons were identified. Using RIGHT allowed us to isolate *ftz* genes from diverse arthropods representing approximately 550 million years of evolutionary divergence (Heffer et al., 2010), including the dermapteran *Forficula auricularia* and archaeognathan *Pedetontus saltator* (Figure 2-2C). Additionally, genomic DNA of two non-model beetles was used for degenerate PCR to obtain the *ftz* homeobox, and in combination with RACE on embryonic cDNA, full-length *ftz* sequences were obtained (Heffer et al., 2010). To date, we have isolated 2 full-length and 10 partial *ftz* genes from a range of non-model organisms using RIGHT.

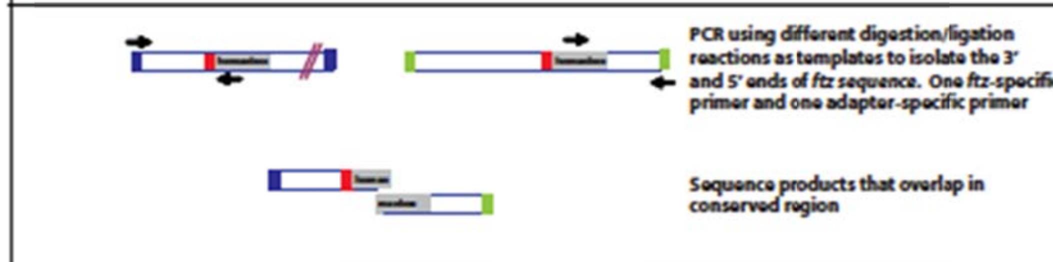
In addition to *ftz*, we isolated other homeobox-containing genes such as *extradenticle* (*exd*) and the orphan nuclear receptor *ftz-f1* from multiple species with great success

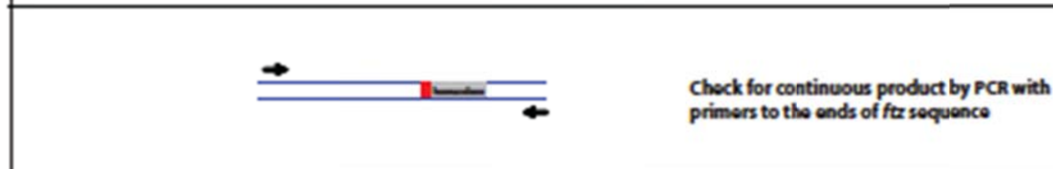
(unpublished). RIGHT was used to isolate partial *exd* sequences from *Thermobia domestica* (firebrat), *Callosobruchus maculatus* (beetle), and *Folsomia candida* (collembolan). In combination with RACE, full-length *exd* coding regions were isolated from these species. Several partial *ftz-f1* sequences were isolated, including *Artemia salina* (brine shrimp), *Folsomia candida* (collembolan), *Thermobia domestica* (firebrat), *Callosobruchus maculatus* (beetle), *Dermestes maculatus* (beetle) *Oncopeltus fasciatus* (milkweed bug), and *Acyrthosiphon pisum* (aphid). As for *exd*, full-length *ftz-f1* sequences have been obtained from many of these organisms in combination with RACE. For this work, as per experimental design, sequences were obtained from species representing key points in arthropod phylogeny to allow for systematic analysis of a small network of functionally related genes from different families (*ftz, ftz-f1, exd*). Thus far, every gene that we have attempted to isolate from any chosen species using RIGHT has been obtained.


A
HOMEODOMAIN
Dm-Antp RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
Dm-Scr TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQDRRMKWKKEH
Dm-Ftz SKRTRQTYTRYQTLELEKEFHFNRYITRRRRIDIANALSLSERQIKIWFQNRRMKSKKDR
Tc-Ftz NKRTRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAESLRLTERQIKIWFQNRRMKAKKDT
Sg-Ftz AKRTRQTYTRFQTLELEKEFHYNRYLTRRRRVEISQALGLTERQIKIWFQNRRMKLKKEN


B
Scr
ftz
Antp
P. saltator genomic DNA
with Hox gene family
Degenerate PCR with ftz N-terminal
homeobox primer other conserved
region 150 bases downstream


MspI
EcoRI
Digestion of genomic DNA and in the
presence of annealed oligonucleotides
that destroy the restriction site
Overnight ligation to add
adapter to DNA ends


PCR using different digestion/ligation
reactions as templates to isolate the 3'
and 5' ends of ftz sequence. One ftz-specific
primer and one adapter-specific primer
Sequence products that overlap in
conserved region


Check for continuous product by PCR with
primers to the ends of ftz sequence

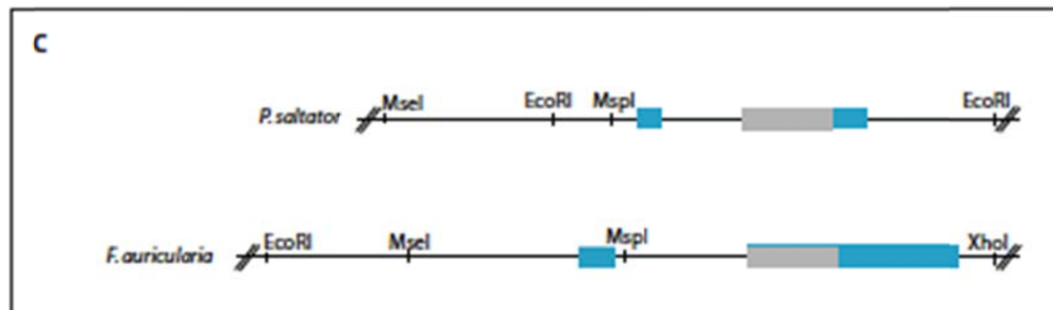

C
P. saltator
MseI
EcoRI
MspI
EcoRI
F. auricularia
EcoRI
MseI
MspI
XhoI

FIGURE 2-2. Isolation of *ftz* homologs using RIGHT. A) Homolog-specific residues guide degenerate PCR design. Different Hox proteins have different N-terminal regions (grey shaded region, with differences highlighted in yellow) that can be used for isolation of one family member. The arrows indicate the regions used for degenerate primer design to isolate *ftz*. The forward degenerate primer makes use of the signature motifs in the N-terminal region, allowing for specific amplification of one member of the *Hox* gene family. (*Drosophila melanogaster* Antp: Dm-Antp; *D. melanogaster* Scr: Dm-Scr; *D. melanogaster* Ftz: Dm-Ftz; *Tribolium castaneum* Ftz: Tc-Ftz; *Schistocerca gregaria* Ftz: Sg-Ftz). B) Isolation of *ftz* from genomic DNA of non-model organism. A schematic of one application of our approach to isolate new homologs is shown. C) Degenerate PCR was used to isolate the *ftz* homebox of P.salt and F.auri, and full-length *ftz* sequences were obtained using different restriction digests/ligations and subsequent PCRs. For *P. salt*, three fragments were obtained by RIGHT and sequenced after degenerate PCR identified the *ftz* homeobox; fragment sizes are (from left to right): *MseI-EcoRI* 320 bp, *EcoRI-MspI* 114 bp, *MspI-EcoRI* 945 bp. For *F. auri*, three fragments were also obtained by RIGHT and sequenced after degenerate PCR identified the *ftz* homeobox; fragment sizes are (from left to right): *EcoRI-MseI* 273 bp, *MseI-MspI* 383 bp, *MspI-XhoI* 875 bp. Homeobox regions are shown in grey, and other coding regions in blue.

2.3 Conclusions

The ability to isolate homologous genes from diverse taxa will empower studies of molecular evolution of genes, families and gene networks. In the past, these approaches were limited by absence of genomic information. Even though genome sequencing is now practical for a larger number of species, it is unlikely to make a dent in the millions of species on Earth. Similarly, investments are being made in developing new model systems, to expand on the standard fly, mouse and worm systems. However, the investment to bring a new model system up to speed is substantial and it is neither necessary nor practical to fully develop hundreds of genetic model systems. We suggest that these approaches, while enormously important for the field of evo-devo, are not always necessary to answer specific evolutionary questions. RIGHT provides a fast and efficient way to isolate genes, including coding regions and candidate cis-regulatory regions, and overcomes many practical constraints, realistically allowing for the

isolation of 10s if not 100s of genes from families or gene networks to study molecular evolution across divergent taxa or within specific clades. This approach obviates common limitations, such as genome sequence availability or rearing species in the lab. It has been used successfully to isolate specific members of several large gene families, allowing for a comparative analysis over millions of years of evolutionary time.

Chapter 3

Surprising flexibility in a conserved Hox transcription factor over 550 million years of evolution
[Published: Heffer et. al, *PNAS*, 2010]

3.1 Abstract

While metazoan body plans are remarkably diverse, the structure and function of many embryonic regulatory genes are conserved because large changes would be detrimental to development. However, the *fushi tarazu* (*ftz*) gene has changed dramatically during arthropod evolution from *Hox*-like to a pair-rule segmentation gene in *Drosophila*. Changes in both expression and protein sequence contributed to this new function: *ftz* expression switched from *Hox*-like to stripes and changes in Ftz cofactor interaction motifs led to loss of homeotic- and gain of segmentation-potential. Here, we reconstructed *ftz* changes in a rigorous phylogenetic context. We found that *ftz* did not simply switch from *Hox*-like to segmentation function; rather, *ftz* is remarkably labile having undergone multiple changes in sequence and expression. The segmentation LXXLL motif was stably acquired in holometabolous insects, after the appearance of striped expression in early insect lineages. The homeotic YPWM motif independently degenerated multiple times. These "degen-YPWMs" showed varying degrees of homeotic potential when expressed in *Drosophila,* suggesting variable loss of *Hox* function in different arthropods. Finally, the intensity of *ftz Hox*-like expression decreased to marginal levels in some crustaceans. We propose that decreased expression levels permitted *ftz* variants to arise and persist in populations without disadvantaging organismal development. This, in turn, allowed evolutionary transitions in protein function, as weakly expressed 'hopeful gene variants' were co-opted into alternative developmental pathways. Our findings show that variation of a

pleiotropic transcription factor is more extensive than previously imagined, suggesting that evolutionary plasticity may be widespread among regulatory genes.

3.2 Introduction

Developmental regulatory genes encode transcription factors that participate in evolutionarily conserved gene regulatory networks (GRNs) crucial for regional specification and patterning during embryonic development (Frigerio et al., 1986; Gehring et al., 1994; Akam, 1995; Davidson and Erwin, 2006; Gehring et al., 2009). This "toolkit" of regulatory genes controls the development of diverse animals with different types of body plans (Carroll et al., 2005). Further, these genes are pleiotropic, being reutilized within individual animal lineages in different combinations and at different developmental stages (Stern and Orgogozo, 2008). These findings raise two related issues: How do regulatory genes change during evolution to direct the development of diverse animals? How can these changes be tolerated during development, as they are expected to be highly detrimental, reminiscent of Goldschmidt's “hopeful monster” (Goldshmidt, 1940)? The modularity of toolkit genes provides a partial answer to these questions, as it allows for changes in both expression and function in only specific tissues or at specific developmental times (Schlosser and Wagner, 2004). Thus, while coding regions may be similar in diverse taxa, their differential expression resulting from changes in modular cis-regulatory elements (CREs) contributes to morphological diversity throughout Metazoa (Prud'homme et al., 2007; Carroll, 2008). However, this modularity also applies to protein-coding regions, such that changes in coding regions that affect distinct functions of a particular protein also contribute to morphological evolution. These changes may result in gain or loss of

cofactor interaction motifs, post-translational modifications, DNA binding specificity, or other functions (Berry and Gehring, 2000; Lohr et al., 2001; Galant and Carroll, 2002; Mann and Carroll, 2002; Ronshaugen et al., 2002; Hsia and McGinnis, 2003; Schlosser and Wagner, 2004; Lohr and Pick, 2005; Hoekstra and Coyne, 2007; Lynch and Wagner, 2008).

One scenario for changes in developmental networks is gene duplication followed by divergence (Ohno, 1970; Force et al., 2005). The *Hox* genes, which pattern the body plans of most metazoans, provide a prime example of this (McGinnis and Krumlauf, 1992; Wagner et al., 2003; Carroll et al., 2005; Gehring et al., 2009). Duplication events that generated *Hox* clusters in early Bilateria (Telford, 2000) provided opportunities for genes to diverge, partitioning existing functions (subfunctionalization) or acquiring new functions (neofunctionalization) (Force et al., 2005). A dramatic example of neofunctionalization is the *Hox* gene *fushi tarazu (ftz)* which switched function from an ancestral *Hox* gene to a pair-rule segmentation gene, originally identified in *Drosophila melanogaster* (Gibson, 2000; Alonso et al., 2001; Lohr et al., 2001). Initial changes in *ftz* were likely enabled by the relaxation of constraints due to overlap in expression and function between *ftz* and *Antp* and/or *Scr*. This is supported by the finding that Ftz from several insects showed Antp-like functional specificity when expressed in *Drosophila* (Lohr et al., 2001) and sequence comparisons that suggest *ftz* and *Antp* are closely related (Telford, 2000).

We previously showed that changes in two cofactor interaction motifs in Ftz switched its regulatory specificity from a canonical homeotic protein to a segmentation protein, found in *Drosophila*: (1) An LXXLL motif in *Dm*-Ftz confers strong interaction with the orphan nuclear receptor Ftz-F1 and is required for segmentation function (Yu et al., 1997; Schwartz et al., 2001; Yussa et al., 2001); (2) The YPWM motif, present in most Hox proteins is degenerate in *Dm*-

Ftz. The YPWM is required for homeotic function by virtue of interaction with cofactor Extradenticle (Exd), a TALE family homeodomain protein (Johnson et al., 1995; Mann and Chan, 1996; Zhao et al., 1996; Burglin, 1997; Passner et al., 1999). These two changes resulted in gain of segmentation potential and loss of homeotic potential, specializing *Dm*-Ftz for segmentation. Ftz proteins that include an intact YPWM motif, such as grasshopper *Sg*-Ftz and beetle *Tc*-Ftz have homeotic potential when expressed in *Drosophila*, and addition of a YPWM motif to *Dm*-Ftz restored ancestral homeotic function (Lohr and Pick, 2005). In addition to these protein changes, *ftz* expression changed during arthropod evolution from a *Hox*-like domain in an arthropod ancestor (Telford, 2000; Hughes and Kaufman, 2002b; Janssen and Damen, 2006; Papillon and Telford, 2007) to seven pair-rule stripes, seen in modern day drosophilids (Hafen et al., 1984; Maier et al., 1990). Striped expression was also observed in the basal insect *Thermobia* (Hughes et al., 2004) and two other holometabolous insects, the beetle *Tribolium casteneum* and the honeybee *Apis mellifera* (Brown et al., 1994; Dearden et al., 2006) but stripes were absent in a grasshopper *Schistocerca gregaria* (Dawes et al., 1994). This suggests that striped expression was either gained twice in arthropods, in a basal insect lineage and during early radiations of holometabolous insects, or was gained once in basal insects and lost in orthopteran lineages.

Here, we address the question: When and in what order did the changes in *ftz* expression and function occur during arthropod evolution? We find that the LXXLL motif was stably acquired at the base of the holometabolous insects while the YPWM degenerated in sequence and function multiple times independently in various arthropod lineages. While strong *ftz Hox*-like expression is likely ancestral, it has decreased to marginal levels in a crustacean, the brine shrimp *Artemia*, where Ftz lacks an LXXLL and carries a degenerate YPWM motif. We suggest

a mechanism that incorporates both cis-regulatory and coding changes to explain how large variations in an embryonic transcription factor can be tolerated during evolution.

3.3 Results

3.3.1 *ftz* gene diversity in the arthropod tree of life

To identify when during arthropod evolution the segmentation and homeotic cofactor interaction motifs were gained and lost, *ftz* orthologs were isolated and sequenced from organisms at representative points along the phylogenetic path from the base of Arthropoda to *Drosophila melanogaster,* spanning ~550 million years of geological time (Regier et al., 2010). These data were combined with published *ftz* gene sequences and reconstructed sequence information from ongoing genome projects (Figures. 3-1, 3-2, Appendix I). Full-length *ftz* cDNAs were isolated from embryonic RNA of organisms that could be cultured: beetles *Callosobruchus maculatus* (*Cm*) and *Dermestes maculatus* (*Dmac*), thysanuran *Thermobia domestica* (*Td*), collembolan *Folsomia candida* (*Fc*), and the brine shrimp *Artemia salina* (*As*). For the dermapteran *Forficula auricularia* (*Fa*) and archaeognathan *Pedetontus saltator* (*Ps*), putative full-length *ftz* coding regions were isolated from genomic DNA. Although the Ftz homeodomain is similar to that of other Hox proteins, the characteristic nine residues at the amino terminal end of the homeodomain (KR(T/S)RQTYTR) distinguish Ftz from other Hox homologs. *Thermobia*, *Folsomia*, and *Artemia* partial *ftz* homeobox and 3' fragments had been previously identified (Averof and Akam, 1993; Hughes et al., 2004). We used these sequences to design *ftz*-specific primers to isolate full-length sequence. Since there was no *ftz* sequence data in the literature for *Callosobruchus, Dermestes, Forficula* or *Pedetontus*, partial homeobox

sequence was isolated by degenerate PCR, using primers specific for the Ftz homeodomain N-terminal arm and another highly conserved region of the homeodomain (QIKIWFQN). Once *ftz* was positively identified, sequence up- and down-stream of the homeobox was isolated using 5' and 3' RACE or modified-AFLP and genomic walking (Materials and Methods). In combination with *ftz* genes assembled from available genomes, we report nine new Ftz sequences: *Bm*-Ftz (447 amino acids), *Am*-Ftz (713 amino acids), *Cm*-Ftz (368 amino acids), *Dmac*-Ftz (377 amino acids), *Td*-Ftz (369 amino acids), *Fa*-Ftz (191 amino acids), *Ps*-Ftz (134 amino acids), *Fc*-Ftz (161 amino acids), and *As*-Ftz (201 amino acids) (Figure 3-2, Appendix I). Adding these nine new sequences to the eleven previously described yields twenty full-length arthropod *ftz* gene sequences available for analysis (Figure 3-1).

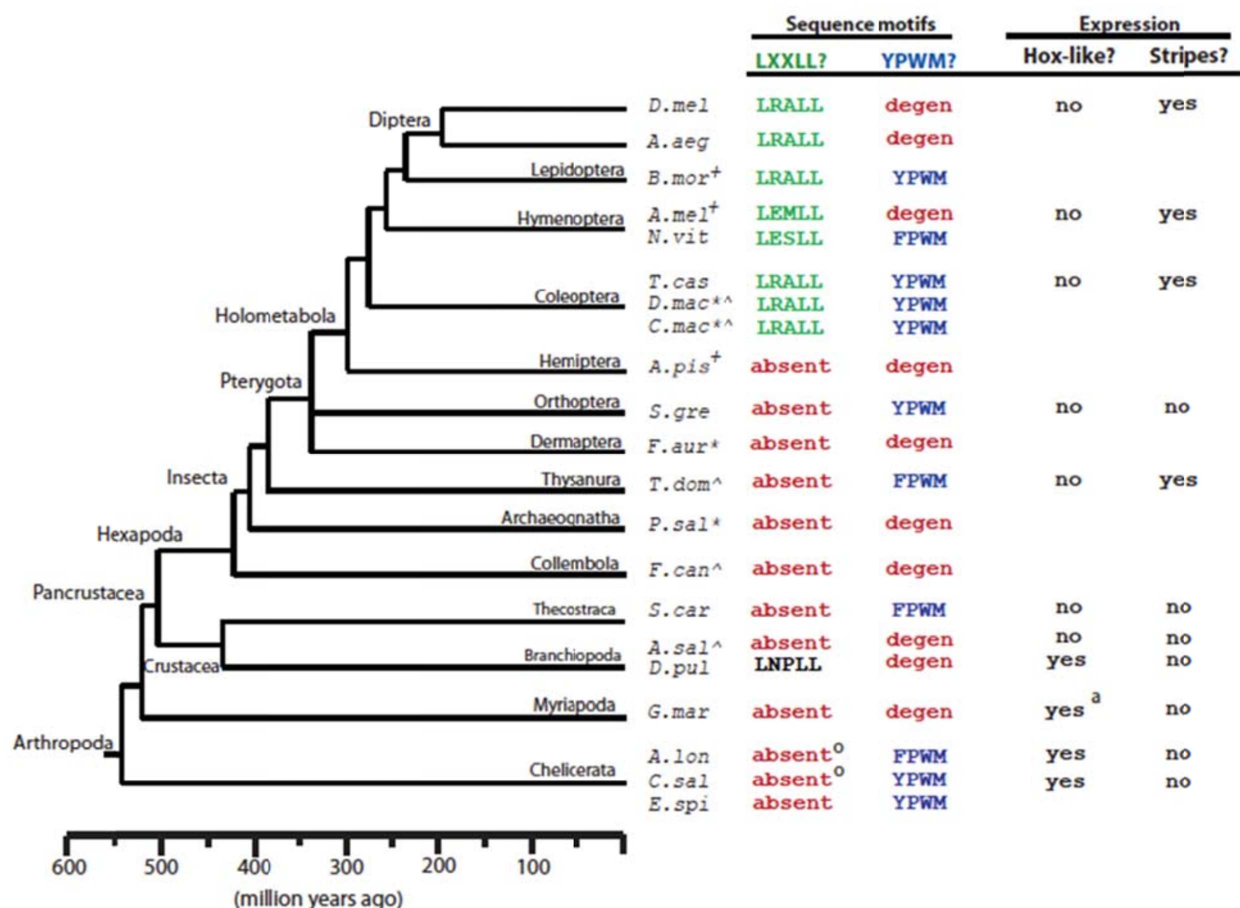


FIGURE 3-1. The *ftz* genes from diverse arthropods display remarkable flexibility. Cladogram of major arthropod taxa is shown with divergence timeline below. The presence of cofactor interaction motifs (LXXLL motif, green; YPWM motif, blue; absent, red) and observed expression patterns (stripes; *Hox*-like) are indicated. [+]*ftz* assembled from genome project contigs [Sources: *B. mor* (69); *A. mel*, Honey bee Genome Sequencing Consortium; *N. vit*, Nas_1.0, 2007; *A. pis*, BCM-HGSC]; ^*ftz* sequence isolated in this study by RACE; **ftz* sequence isolated in this study by modified AFLP from gDNA; [a]Striped expression seen only after segments formed (38); [o]sequence not full length. Other sequences: (*Sg*) (Dawes et al., 1994); (*Dp*) (Papillon and Telford, 2007); (*Sc*) (Mouchel-Vielh et al., 2002); (*La*) (Hughes and Kaufman, 2002b); (*Gm*) (Janssen and Damen, 2006); (*Cs*) (Damen et al., 2005). Partial *ftz* sequences: AAS17755 (*Td*), AAK51915 (*Fc*), CAA49684 (*As*) (36), AAF63162 (*Al*), CAI91292 (*Cs*).

```
                       LXXLL         (Y/F)PWM                          Homeodomain
D.mel[1]   (108)--LRALL-(123)-DFNWSH-(11)-SKRTRQTYTRYQTLELEKEFHFNRYITRRRRIDIANALSLSERQIKIWFQNRRMKSKKDR--(97)
A.gam[2]   (184)--LRALL-(117)-SNSWTQ-(11)-SKRTRQSYSRHQTIELEKEFHFNRYLNRRRRIEIASMLKLTERQIKIWFQNRRMKAKKDN-(120)
A.agy[3]   (191)--LRALL-(109)------------SKRTRQSYSRHQTLELEKEFHSNKYLTRRRRIEVANVLRLTERQVKIWFQNRRMKAKKDK--(63)
B.mor      (174)--LRALL-(120)-YYPWMK-(10)-SKRTRQTYTRFQTLELEKEFHFNKYLSRRRRIEVSHALGLTERQIKIWFQNRRMKAKKDG--(72)
A.mel      (170)--LEMLL-(329)-NYSWLK-(13)-QKRTRQTYTRYQTLELEKEFHFNRYLTRNRRLEIARALHLTERQVKIWFQNRRMKAKKFN-(130)
N.vit[4]    (95)--LESLL-(151)-DFPWMK-(14)-QKRTRQTYTRYQTLELEKEFHFCRYLSRKRRVEIAHSLGLTERQIKIWFQNRRMKAKKDS-(117)
T.cas[5]   (125)--LRALL--(37)-FYPWMK--(9)-NKRTRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAESLRLTERQIKIWFQNRRMKAKKDT--(48)
C.mac      (140)--LRALL--(61)-FYPWMR--(9)-NKRTRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAHTLCLTERQIKIWFQNRRMKAKKGD--(87)
D.mac      (159)--LRALL--(68)-IYPWMK--(9)-GKRTRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAHALCLSERQIKIWFQNRRMKAKKDN--(70)
A.pis[6]          ---------(32)------------GKRSRQTYSKYQTAVLETVFQTSRYIVRNKRQQMSAELSLTERQIKIWFQNRRMKEKKCH-(135)
S.gre[7]          --------(189)-FYPWMK-(11)-AKRTRQTYTRFQTLELEKEFHYNRYLTRRRRVEISQALGLTERQIKIWFQNRRMKLKKEN--(48)
F.aur             ---------(17)-SIPQRK--(5)-SKRSRQTYSRYQTLELEKEFHFNKYLTRRRRIEIANALHLTERQIKIWFQNRRMKEKKTR-(103)
T.dom[8]          --------(220)-YFPWMK--(9)-PKRTRQTYTRFQTLELEKEFHFNKYLTRRRRIEIAHSLGLSERQIKIWFQNRRMKAKKEI--(75)
P.sal             ---------(10)-YIPQMS--(8)-PKRTRQTYTRVQTLELEKEFHFNRYLTRRRRIEIAHVLGLTERQIKIWFQNRRMKAKKES--(51)
F.can[9]          ---------(38)-YPPWLK-(26)-SKRTRQTYTRCQTLELEKEFHFNKYLTRRRRLDLAKMLTLSERQIKIWFQNRRMKAKKEV--(31)
S.car[10]         --------(206)-IFPWMK-(10)-AKRTRQSYTRYQTLELEKEFRTNRYLTRRRRIEIARTVALTERQIKIWFQNRRMKAKKEH-(111)
A.sal[11]         ---------(90)-PYHQMQ-(18)-QKRTRQTYTKYQTLELEKEFLYNRYLTRVRRMDISSKLQLTERQIKIWFQNRRMKAKKEN--(27)
D.pul[12]  (204)--LNPLL--(59)-PGSWMQ--(8)-PKRTRQTYTRYQTLELEKEFHFNRYLTRRRRVEIAHALCLTERQIKIWFQNRRMKAKKET--(86)
G.mar[13]         --------(235)-KDCWKA-(19)-RKRTRQTYTRFQTLELEKEFHSNRYLNRRRRIEIATSLTLTERQVKIWFQNRRMKAKREP--(45)
E.spi[14]         --------(193)-FYPWMK--(8)-PKRTRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALGLSERQIKIWFQNRRMKAKKEN--(35)
```

FIGURE 3-2. Ftz orthologs have similar homeodomain sequences but vary in their cofactor interaction motifs and protein lengths. Residues in the segmentation LXXLL, homeotic YPWM, and homeodomain are shown as identical (gray) or similar (yellow) to the most common amino acid at that position in Ftz. All Ftz proteins share a characteristic nine amino acid N-terminal homeodomain arm (KRT/sRQT/sYT/sR/k). Sequences in red were isolated in this study. Other sequences: National Center for Biotechnology Information accession numbers: 1, NP_477498; 2, NT_078265; 3, CH477233; 4, XP_001603670; 5, NP_001034539; 6, NW_001923321; 7, CAA52160; 8, AAS17755; 9, AAK51915; 10, AAM50460; 11, CAA49684; 12, ABQ22961; 13, CAJ56096; 14, ABD46730.

Arthropod Ftz orthologs differ greatly in size and composition (Figure 3-2; Table 3-1). The putative *Ps*-Ftz and *Fa*-Ftz sequences have very short coding regions upstream of the homeodomain (<30 amino acids), while *As*-Ftz and *Fc*-Ftz have slightly longer protein sequences upstream of the homeodomain (~100 amino acids). Interestingly, Ftz sequences that have an LXXLL motif are much larger (Figure 3-2). Though we cannot positively confirm the coding sequences of *Ps*-Ftz and *Fa*-Ftz because embryonic RNA is not available, we have several reasons to believe these sequences are full-length. First, there are splice donor (GT) and

splice acceptor (AG) sites flanking small introns directly upstream of the homeobox, which are comparable in size to other *ftz* introns (Table 3-1). Second, there are no other open reading frames with a splice donor site ~800bp upstream of the homeodomain. Third, there are several possible transcription initiator and TATA-consensus sequences upstream of the translation start site. Finally, sequence from the aphid genome shows that the predicted *ftz* gene in this organism does not encode an LXXLL or YPWM motif, and has very little coding region upstream of the homeodomain (32 amino acids; Aphid Genome Project).

TABLE 3-1. *ftz* genes generally contain small introns and encode short linker regions between the YPWM motif and homeodomain.

Species	**Intron size (bp)**	**# Residues in linker**
D. mel	150	11
A.gam	59	11
A.aeg	60	?
B.mor	89	10
A.mel	1657	13
N.vit	157	14
T.cas	50	9
D.mac	60	9
A.pis	224	?
S.gre	>679*	11
F.aur	335	5
P.sal	141	8
S.car	150	10
D.pul	?	8

*complete intron sequence not published

3.3.2 The LXXLL was stably acquired at the base of Holometabola

The LXXLL motif in *Dm*-Ftz is necessary for segmentation function and mediates interaction with the cofactor Ftz-F1 (Schwartz et al., 2001; Yussa et al., 2001; Suzuki et al., 2002). Ftz from the flour beetle *Tribolium castaneum* (*Tc*-Ftz) contains an LXXLL motif and displayed segmentation potential when expressed in *Drosophila* (Lohr et al., 2001). We found that Ftz orthologs from *Callosobruchus* and *Dermestes,* long- and intermediate-germ beetles, encode proteins very similar to *Tc*-Ftz, including LRALL sequences and similar flanking amino acids. Ftz sequences assembled from the genomes of the silkworm *Bombyx mori* (*Bm*-Ftz), honeybee *Apis mellifera* (*Am*-Ftz), and mosquitoes *Aedes agypti* (*Aa*-Ftz) and *Anopheles gambiae* (*Ag*-Ftz) all include LXXLL motifs. Interestingly, most of these Ftz proteins share an LRALL sequence. Though the importance of the "RA" in Ftz has not been studied, *Am*-Ftz and *Nv*-Ftz (wasp) have EM and ES substituted at these positions. This suggests the internal residues ('XX') are somewhat flexible, while the three leucine residues required for interaction with Ftz-F1 (Lohr and Pick, 2005) are constrained. Whereas all Ftz proteins isolated to date from holometabolous insects harbor LXXLL motifs (Figure 3-1, green), no other insect *ftz* encodes this motif: *Sg*-Ftz, *Ap*-Ftz, *Fa*-Ftz, *Td*-Ftz, *Ps*-Ftz, *Fc*-Ftz, and *As*-Ftz all lack LXXLL sequences. A Ftz LXXLL motif (LNPLL) appears in one other arthropod, the crustacean *Daphnia pulex* (*Dp*-Ftz). However, although functional experiments will be interesting in the future, as proposed by Papillon and Telford (2007), the appearance of this motif is probably not functional in *Daphnia*, as it is unlikely to participate in segmentation particularly in light of the *Hox*-like expression of *Dp-ftz* (Papillon and Telford, 2007). Together, these findings suggest that the segmentation LXXLL motif was acquired once at the stem of the holometabolous clade, and that it has been stably retained in this lineage.

3.3.3 The YPWM motif 'flickers' in arthropod phylogeny

While the homeodomain is sufficient for binding DNA, a (Y/F)PWM sequence (referred to throughout as 'YPWM motif') found at variable distances upstream of the homeodomain in most Hox proteins is crucial for cooperative binding to Exd/Pbx (Chang et al., 1995; Johnson et al., 1995; Neuteboom et al., 1995; Phelan et al., 1995) and biological specificity *in vivo* (Zhao et al., 1996; Tour et al., 2005). The YPWM motif is found in diverse Antp and Ubx proteins (Figure 3-3) and is considered the ancestral condition for Ftz, represented by a chelicerate (mite, *Al*-Ftz) (Telford, 2000) and Onychophora (Grenier et al., 1997), an arthropod outgroup. Consensus YPWM motifs are also found in Ftz in both holometabolous (beetles *Tc*-Ftz, *Cm*-Ftz, *Dmac*-Ftz) and other insects (grasshopper *Sg*-Ftz, firebrat *Td*-Ftz). However, a degenerate motif (FNWS), with decreased Exd-binding ability and homeotic potential, is found in *Dm*-Ftz (Lohr and Pick, 2005). We found degenerate motifs ('degen-YPWMs') in several other Ftz sequences, including a YPPWLK in *Fc*-Ftz, a YHQM in *As*-Ftz, an IPQM in *Ps*-Ftz and an IPQRK in *Fa*-Ftz (Figure 3-2, Appendix I). These sequences all resemble YPWM, and are considered 'degenerate' rather than completely lost. Additionally, degenerations appear to have occurred independently as each motif has a different sequence. Dollo parsimony, which allows only losses after one initial gain, indicates that the motif degenerated eight times (Figure 3-1: Diplopoda, Branchiopoda, Collembola, Archaeognatha, Dermaptera, Hemiptera, Hymenoptera, Diptera). Alternatively, a strict parsimony analysis, which minimizes the number of total evolutionary events regardless of direction, suggests five losses (Diplopoda, Dermaptera, Hemiptera, Hymenoptera, Diptera) and two gains (Thecostraca, Insecta). We favor the Dollo parsimony analysis, suggesting that this motif independently degenerated multiple times for several reasons. First, in each case, the specific sequence change is different, sometimes

involving changes in amino acid sequence (e.g., FNWS or IPQM), other times involving insertions and amino acid substitutions (e.g., YPPWLK). Second, within multiple taxa, closely related species 'flicker' (Marshall et al., 1994) with respect to YPWM. For example, within Hymenoptera, honeybee Ftz (*Am*-Ftz) has a degenerate YPWM while wasp Ftz (*Nv*-Ftz) retains a consensus YPWM; within crustaceans, brine shrimp Ftz (*As*-Ftz) YPWM is degenerate while barnacle Ftz (*Sc*-Ftz) retains YPWM (Mouchel-Vielh et al., 2002). Third, some losses (e.g., dipterans) may be secondary, occurring after addition of LXXLL, and presumed gain of segmentation function. In sum, whereas the LXXLL motif of Ftz has established itself at the base of the holometabolous insects, the YPWM motif in Ftz proteins shows a complex evolutionary history with a flickering pattern in arthropod phylogeny suggesting that it has been independently lost in multiple lineages.

```
Antennapedia
D.mel  -LYPWMR--RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-
A.mel  -LYPWMR--RKRGRQTYTRYQTLELEKEFHYNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-
T.cas  -LYPWMR--RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-
A.pis  -LYPWMR--RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-
D.pul  -LYPWMR--RKRGRQTYTRFQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-

Sex combs reduced
D.mel  -IYPWMK--TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH-
A.mel  -IYSWMK--VKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH-
T.cas  -IYPWMK--TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH-
A.pis  -IYPWMK--TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH-
D.pul  -IYPWMK--TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHSLCLSERQIKIWFQNRRMKWKKVG-

Ultrabithorax
D.mel  -FYPWMA--RRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEI-
A.mel  -FYPWMA--RRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHSLCLTERQIKIWFQNRRMKLKKEI-
T.cas  -FYPWMA--RRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEI-
A.pis  -FYPWMA--RRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEI-
D.pul  --FYPWMA--RRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEI-
```

FIGURE 3-3. Conservation of YPWM motif and homeodomain in other Hox proteins. The residues flanking the YPWM motif and homeodomain sequences of Antennapedia, Sex combs reduced, and Ultrabithorax are highly conserved. Sequences were obtained from the genomes of *Drosophila melanogaster (D.mel), Apis mellifera (A.mel), Tribolium castaneum (T.cas), Acyrthosiphon pisum (A.pis), and Dapnia pulex (D.pul).* Only nonsynonymous amino acid substitutions are highlighted.

3.3.4 'Degen-YPWMs' vary in homeotic potential

The lability of the YPWM motif through phylogeny reflects surprising evolutionary flexibility in this homeotic cofactor interaction motif. In contrast to what is observed for Ftz, the homeodomains and YPWM motifs encoded by neighboring *Hox* genes are highly conserved. Comparison of Antp, Scr and Ubx from five divergent taxa (*D. mel, A. mel, T. cas, A. pis, D. pul*) revealed only one amino acid change in a YPWM motif of the 15 proteins, and virtually identical homeodomains among orthologous proteins (Figure 3-3). Thus, the changes in Ftz YPWM, as well as the divergence seen within the homeodomains (Figure 3-2), are specific to this protein and not a general feature of Hox proteins. The YPWM motif is required for interaction with a

hydrophobic pocket on the surface of the Exd homeodomain and based upon the mode of action of YPWM in mediating interaction with Exd (Chang et al., 1995; Johnson et al., 1995; Neuteboom et al., 1995; Phelan et al., 1995), the observed deviations from YPWM reported here are all expected to result in loss of interaction with Exd. We therefore asked whether these 'degen-YPWMs' retain homeotic potential *in vivo*. We previously showed that ectopic expression of *Dm*-Ftz in imaginal discs did not cause a homeotic transformation, but rather resulted in antennal truncation due to cell death. In contrast, more homeotic-like *ftz* genes such as *Tc-ftz* resulted in *Antp*-like transformations of antennae to legs accompanied by activation of *Antp*-target genes (Lohr et al., 2001). Additionally, replacement of FNWS in *Dm*-Ftz with YPWM conferred homeotic function to *Dm*-Ftz (Lohr and Pick, 2005). Here, we used a similar strategy to assess the activity of degen-YPWMs from Ftz in other taxa. The homeotic potential of *Dm*Ftz-FNWS (*Drosophila* degenerate motif), *Dm*Ftz-YPPWLK (*Folsomia* degenerate motif), and *Dm*Ftz-YHQM (*Artemia* degenerate motif) were compared to that of a protein that completely lacked a functional motif, *Dm*Ftz-AAAA. All mutations were made in a *Dm*-Ftz background that included a mutation of LRA<u>LL</u> to LRA<u>AA</u> because homeotic effects were found to be stronger when the LXXLL motif was inactivated (Lohr and Pick, 2005). Additionally, the degen-YPWMs tested in this experiment were derived from Ftz proteins lacking LXXLL motifs (Figure 3-2).

Multiple independent transformant lines were established for each construct and modified Ftz proteins were expressed in developing imaginal discs with a *Dll-GAL4* driver (Figures 3-4, 3-5). Transgenic flies expressing *UAS-lacZ* (negative control) had wildtype antennae with three antennal segments (A1-A3) and aristae, demonstrating that phenotypes seen with *ftz* transgenes were specific and not caused by the GAL4 driver (Figure 3-4a). Expression of *Dm*Ftz-AAAA

resulted in antennae with normal A1 and A2 segments, but a malformed A3 segment with extra bristles and a truncated arista (Figure 3-4b, arrowhead). Expression of *Dm*Ftz-FNWS (Figure 3-4c; (Lohr and Pick, 2005) and *Dm*Ftz-YHQM (Figure 3-4d) caused phenotypes similar to *Dm*Ftz-AAAA, suggesting neither the *Drosophila* FNWS nor the *Artemia* YHQM conferred any further homeotic potential to *Dm*-Ftz. In contrast, the YPPWLK motif (*Dm*Ftz-YPPWLK) conferred some homeotic potential (Figure 3-4e), but the transformation was not as strong as that induced by *Dm*Ftz-YPWM (Figure 3-4f): *Dm*Ftz-YPWM transformed antennae to complete legs with five distinguishable segments (Figure 3-4f) while *Dm*Ftz-YPPWLK animals showed only two distal leg segments (Figure 3-4e, arrows) and a malformed A3 segment. Together, these results suggest that the YPWM motif has functionally degenerated independently multiple times and that it has lost function to different extents in different lineages.

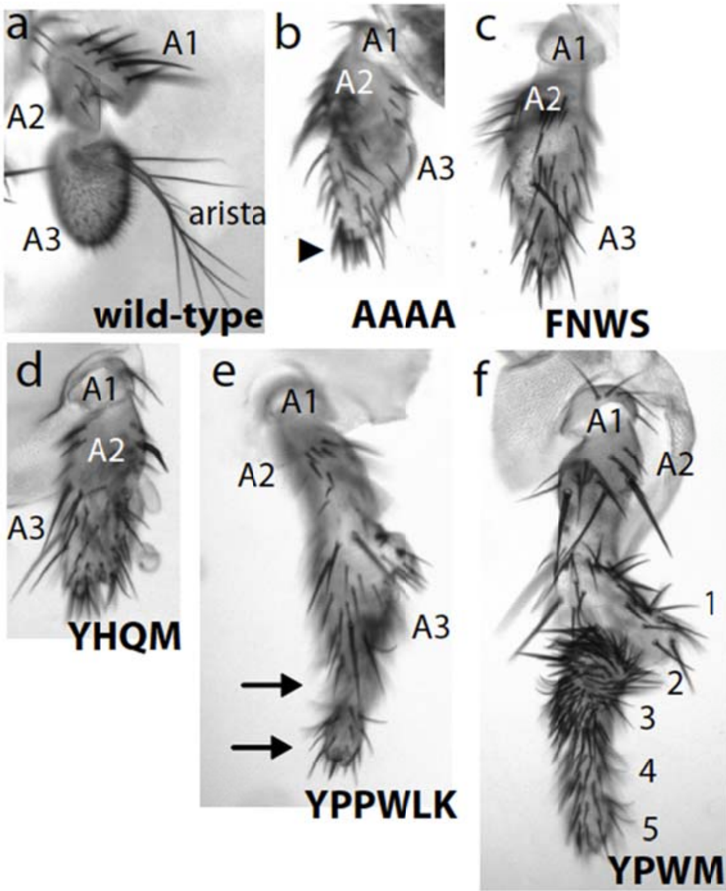


FIGURE 3-4. Degenerate YPWM motifs retain varying degrees of homeotic potential. The *ftz* transgenes carrying examples of natural variation in YPWM motifs were expressed in developing imaginal discs with the *Dll-Gal4* driver. (*A*) Control, expression of *UAS-lacZ* did not cause homeotic transformation of antennae. (*B*) *Dm*Ftz-AAAA animals showed normal A1 and A2, but abnormal A3 segments with bristles (arrowhead) and no aristae. (*C*) *Dm*Ftz-FNWS and (*D*) *Dm*Ftz-YHQM effects were similar to *Dm*Ftz-AAAA. (*E*) *Dm*Ftz-YPPWLK caused transformation of aristae into partial legs with two segments (arrows) and malformed A3 segments. (*F*) *Dm*Ftz-YPWM caused complete transformation of aristae to legs with five segments.

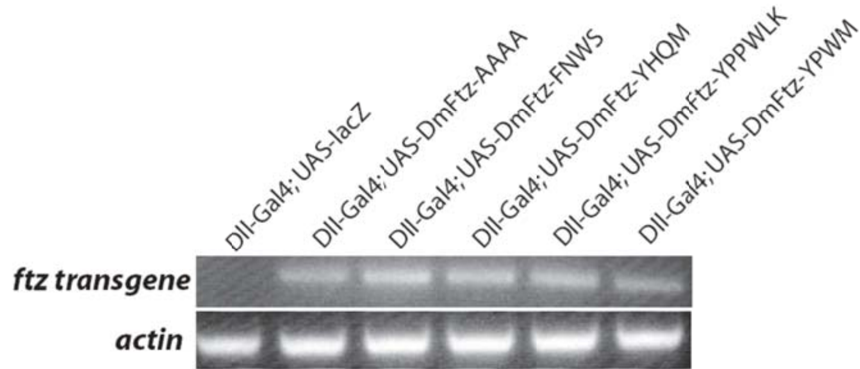


FIGURE 3-5. Expression levels of 'degen-YPWM' transgenes are similar. A *Dll*-Gal4 driver was used to express 'degen-YPWM' transgenes in developing imaginal discs during *Drosophila* development. L1 larvae were collected, and cDNA was made from 1ug of RNA. RT-PCR confirmed a *ftz* transgene product was detected in all 'degen-YPWM' lines, but was absent in the control line (*lacZ*). Actin levels in all samples were similar. For each transgene, only one antennal-to-leg phenotype was seen; there was no variation.

3.3.5 Loss of *Hox*-like expression in crustaceans

In addition to changes in sequence, *ftz* expression has changed during the radiation of arthropods from *Hox*-like to stripes. Our sequence of *Td*-Ftz suggests that the LXXLL motif was acquired after *ftz* was expressed in a striped pattern (Figure 3-1). To investigate how a *ftz* gene encoding neither an LXXLL nor YPWM motif is expressed, we examined *ftz* expression in the brine shrimp *Artemia*. *Artemia* naupliar development is similar to that of short-germ insects, in that upon hatching the antennular, antennal and mandibular segments are present, while the remaining segments are added on sequentially from the growth zone during postembryonic development. As shown in Figure 3-6, *As-ftz* is not expressed in nauplii in a pattern seen for *ftz* in other species. *ftz* expression was undetectable in L1 and L2 nauplii (Figure 3-6a,e,i) and later, *ftz* was weakly expressed throughout the trunk of the L4 nauplius (Figure 3-6i). This pattern was

better visualized when *Artemia* were over-stained (Figure 3-6m,n). Quantitative RT-PCR with L1 and L4 nauplii confirmed *ftz* expression in the trunk was not background or a staining artifact, and strong *ftz* expression in the nervous system later in development (Figure 3-6o) confirmed that the weak expression observed was not due to technical problems with the probe. In other taxa, *ftz* expression has been observed in the growth zone, in stripes, or in a *Hox*-like pattern in segment primorida (Figure 3-1). While we were able to detect expression in these regions using various probes (growth zone: *cad* (Copf et al., 2004), Figure 3-6d,h,l; stripes: *en* (Manzanares et al., 1993), Figure 3-6c,g,k; *Hox*: *Antp* (Averof and Akam, 1995), Figure 3-6b,f,j), none of these patterns were seen for *ftz* in *Artemia* (Figure 3-6a,e,i). Together, the expression and quantitative data suggest that *ftz* has lost *Hox*-like expression and potential to participate in homeosis in *Artemia* nauplii.

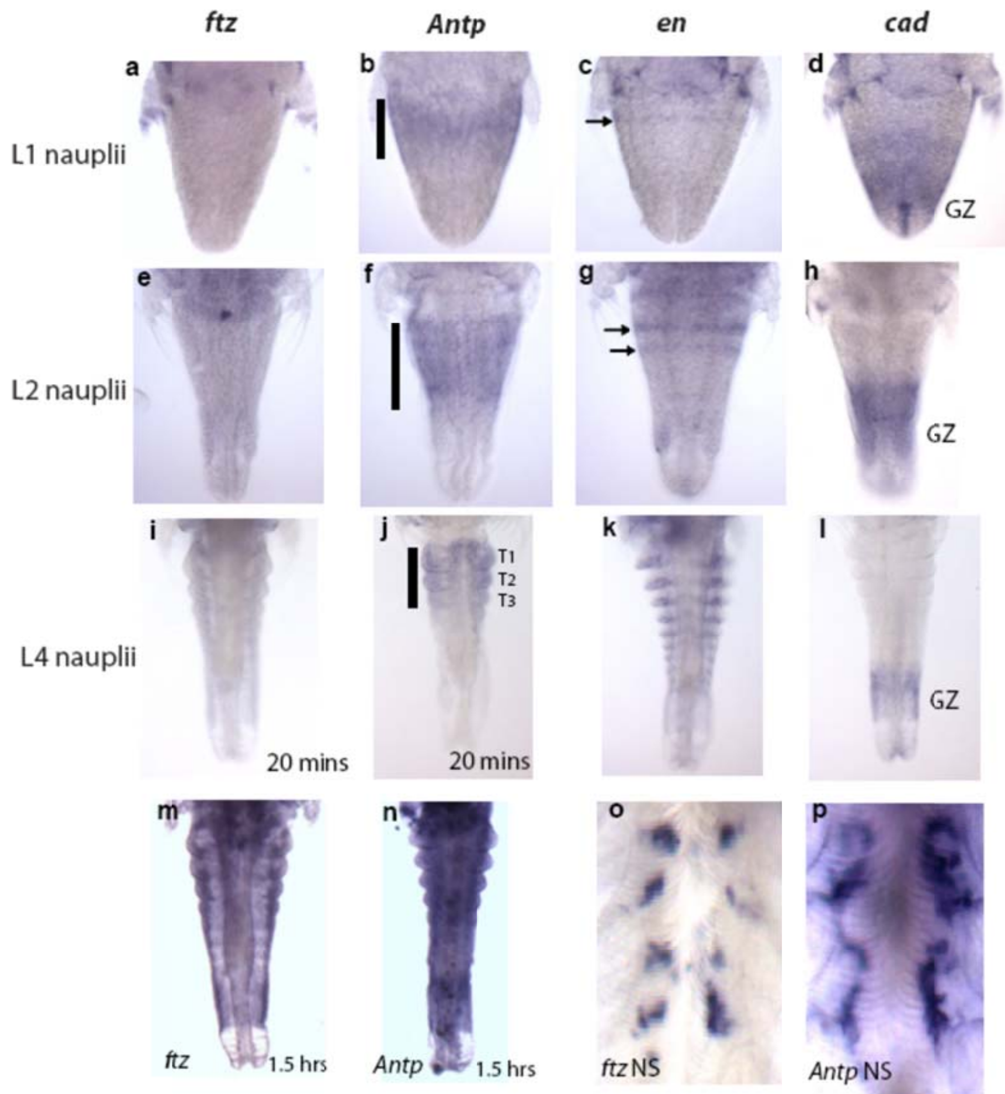


FIGURE 3-6. Fading *Hox*-like *ftz* expression in the crustacean *Artemia salina.* Expression patterns determined by in situ hybridization. (*A*, *E*, *I*, *M*, *O*) *As-ftz*; (*B*, *F*, *J*, *M*, *P*) *As-Antp*; (*C*, *G*, *K*) *As-en*; (*D*, *H*, *L*) *As-cad* in L1 (*A–D*), L2 (*E–H*), or L4 (*I–N*) nauplii. *As-ftz* was not detected in L1 or L2 and was weakly expressed in L4 nauplii. *Antp* was *Hox*-like in thoracic primordia (bar). (*M* and *N*) Overstaining highlights relative weakness of *As-ftz* compared with *As-Antp*; *engrailed* was detected in the posterior of every segment (*C*, *G*, *K*, arrows); *caudal* localized to the growth zone (*D*, *H*, *L*). Both *As-ftz* (*O*) and *As-Antp* (*P*) were detected later in the nervous system.

3.4 Discussion

Embryonic regulatory genes are remarkably conserved across broadly divergent taxa. Their structure and functional specificity are generally thought to be highly constrained, yet *Hox* genes are master regulatory genes that pattern body plans of diverse types of animals (McGinnis and Krumlauf, 1992; Akam, 1995; Gehring et al., 2009). This paradox has raised questions as to how *Hox* GRNs have evolved. One answer to this is provided by the cis-regulatory hypothesis, which posits that changes in the expression of either the *Hox* genes themselves, or cis-regulatory elements of *Hox* gene targets have changed during evolution (Carroll et al., 2005; Carroll, 2008). For example, the loss of limbs in the ancestor of snakes is thought to be due to a shift in *Hox* expression (Cohn and Tickle, 1999) and shifts in the borders of *Ubx/Abd-A* expression correlate with changes in appendage morphology in myriapods and crustaceans (Abzhanov and Kaufman, 2000). Ubx regulates the development of both the butterfly hind-wing and fruit fly haltere, two structures with very different morphology, as a result of evolutionary changes in Ubx-responsive target genes (Weatherbee et al., 1999). In mammals, where duplications have led to four paralogous *Hox* complexes, exchange of the protein coding regions of the paralogs *Hox A3* and *D3* revealed functional equivalence; differential gene function *in vivo* results from differences in gene expression (Greer et al., 2000). However, evolution does not work by one mechanism alone and changes in the coding regions of *Hox* genes have also been correlated with morphological diversification. For example, changes in Ubx protein led to the acquisition of a repression domain in insects that contributed to differences in limb number between crustaceans and hexapods (Grenier and Carroll, 2000; Galant and Carroll, 2002; Ronshaugen et al., 2002) and changes in Hox-A11 altered its regulatory specificity such that it regulates prolactin production, critical for pregnancy, specifically in eutherian mammals (Lynch et al., 2008). More

dramatic perhaps than these are the changes in *Hox3* and *ftz* in arthropods, which have escaped the rules of colinearity and taken on new roles during embryogenesis in different taxa. Duplication of *Hox3* in flies generated the *zen* and *bcd* genes which have novel functions due to shifts in expression patterns and protein sequence (Schmidt-Ott and Wimmer, 2004). Bcd switched DNA-binding specificity due to a single amino acid change in the homeodomain (Hanes and Brent, 1989), and acquired RNA-binding ability (reviewed in Hsia and McGinnis, 2003).

Here we initiated a phylogenetically structured analysis to reconstruct the sequence of events leading to the switch in Ftz function. Because *Hox* genes are thought to be so highly constrained, we began with an assumption that a minimum number of changes (3 total: switch to pair-rule stripes, YPWM degeneration, LXXLL acquisition) would be sufficient to describe the evolutionary trajectory of *ftz*. Thus, our initial goal for the present study was to map the switch points for each of these changes with the expectation that each would map to a distinct branch. Contrary to this expectation, we found that *ftz* has varied multiple times in both coding sequence and expression pattern (Figure 3-1). **(1)** *ftz* expression changed at least three times during arthropod evolution: loss of *Hox*-like expression, gain of striped expression and secondary loss of striped expression. **(2)** The homeotic YPWM motif degenerated independently at least eight times. **(3)** The LXXLL motif was stably acquired in a single "switch" at the base of the holometabolous insects. This acquisition appears be under functional constraint in holometabolous insects, as an LXXLL motif is found in Ftz throughout this taxon. The gain of a striped expression pattern in early hexapod lineages, represented by *Td-ftz* (Hughes et al., 2004), preceded the stable gain of the segmentation LXXLL motif. This 'snapshot' of molecular evolution in progress revealed a surprisingly dynamic pattern of changes in a transcription factor

whose pleiotropic roles during embryonic development would be expected to restrict functional changes. We suggest that deep phylogentic sampling, such as that carried out here, will reveal similar variation in expression and function of other regulatory genes, exemplified by variations in Ubx protein domains from different taxa (see above) and loss of Abd-A expression in *Artemia* (Hsia et al., 2010). These changes in protein motifs and expression beg for a mechanistic explanation as loss- and gain-of-function changes in Hox proteins are deleterious and ectopic expression of transcription factors usually results in lethality, even in the unchallenging environment of a laboratory.

3.4.1 Model for regulatory transcription factor flexibility

How could changes in *ftz* be so pervasive in nature? We propose that cis-regulatory changes that altered *ftz* expression were permissive for changes in protein function, enabling flexibility and variation (Figure 3-7). Decreased *Hox* expression, seen in extant crustaceans (Figure 3-6), presumably due to mutation in a cis-regulatory element directing *Hox*-like expression (*Hox* CRE), removed *ftz* from homeotic pathways, relieving constraints on its homeotic function and allowing degeneration of the YPWM motif and eventual loss of homeotic potential. We propose that reduced levels of *Hox*-like expression, seen in at least two crustaceans, represent a transition state that was permissive for additional changes in *ftz* expression and sequence (Figure 3-7): low levels of gene expression provide a platform for changes that impact protein function because their weak expression dampens activity and thus minimizes impact on existing GRNs. While many protein variants could produce inviable ‘hopeful monsters’ (Goldshmidt, 1940) if expressed at higher levels, at sub-threshold levels they

can provide raw material for co-option of regulatory proteins with novel functions into alternate GRNs. Some 'hopeful gene variants' can endure to take on new and essential roles, exemplified by the pair-rule function of *Dm*-Ftz. A second cis-regulatory change in *ftz* was the acquisition of a striped expression pattern (Stripe CRE). This pattern arose earlier but was stabilized in holometabolous insects where acquisition of an LXXLL motif conferred interaction with the cofactor Ftz-F1, generating a Ftz able to regulate whole new sets of downstream target genes (Yu et al., 1997; Yussa et al., 2001). We suggest that maintenance of stripes in this lineage is in turn explained by the regulatory switch in Ftz (LXXLL acquisition) as interaction with Ftz-F1 allowed for *ftz* autoregulation (Hiromi and Gehring, 1987), thus reinforcing striped expression.

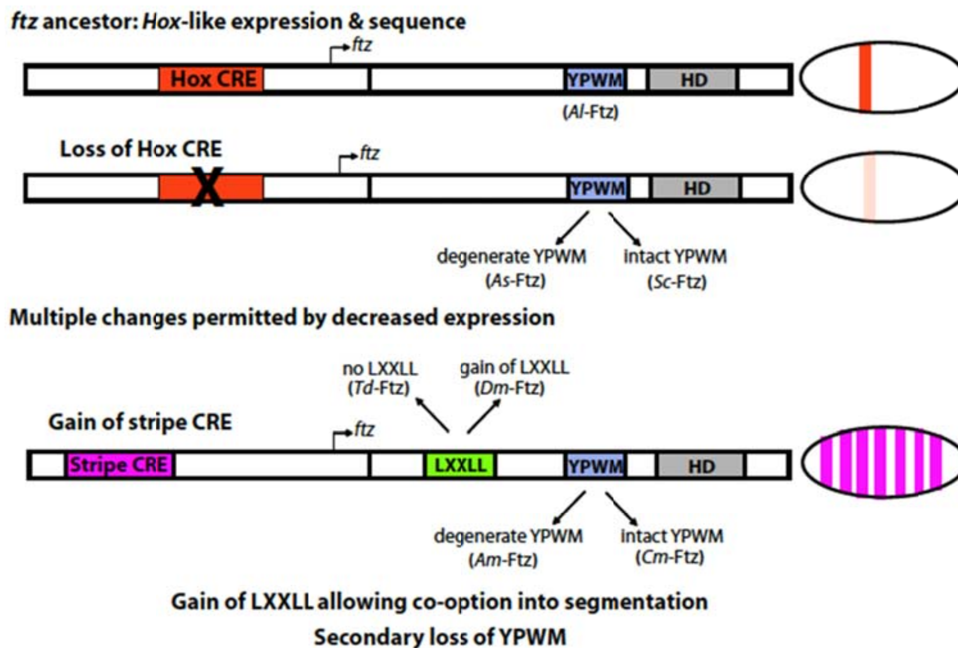


FIGURE 3-7. The modularity of *ftz* CREs and protein motifs allows for extensive variation in *ftz* throughout arthropods. Ancestrally *ftz* was expressed in a *Hox*-like pattern because of a "Hox CRE." CRE mutation weakened *ftz Hox*-like expression. Low expression levels enabled additional protein changes without deleterious consequence. The YPWM motif degenerated and lost homeotic function in multiple lineages. A CRE directing striped expression was gained and *ftz* was coopted into segmentation GRNs when the LXXLL motif was acquired, providing an interaction with the cofactor Ftz-F1.

3.5 Materials and Methods

Arthropod sources and care

Artemia salina were obtained as dehydrated cysts from Carolina Biological and rehydrated in 3% salt water. Once hatched, they were maintained in a salt water solution containing an air source and fed a dilute yeast solution. *Thermobia domestica* were raised at 35°C in a humid incubator, and fed oatmeal and hermit crab food. *Folsomia candida* were kept in petri dishes containing charcoal/plaster of paris and fed dry yeast. *Pedetontus saltator* and *Forficula auricularia* were captured in the field, preserved in >95% EtOH and stored at -80° C before isolation of genomic DNA.

Isolation of ftz sequence by RLM-RACE and modified AFLP

RNA was extracted from 0-4d *Artemia* nauplii, 0-4d *Folsomia* eggs, and 0-9d *Thermobia* eggs using the TRIzol reagent (Invitrogen) and Qiagen RNA extraction kit. Full-length *ftz* cDNAs were obtained by 5' and 3' RLM-RACE (Ambion) and PCR, using primers designed to previously-identified partial *ftz* homeobox regions (NCBI accession numbers: X70079, AF361331, AY456923). Genomic DNA was extracted from *Pedetontus* and *Forficula* using a standard *Drosophila* protocols. Additional sequence was obtained by modified AFLP (Biedler et al., 2003) and genomic-walking . Primer sequences available by request.

Artemia expression analysis

Artemia nauplii were fixed in 4% paraformaldehyde for 2h at room temperature, and taken through a series of PBS/MeOH rinses: 75%, 50%, 25%. After 4 additional washes in 100% MeOH, fixed nauplii were stored at -20°C. Digoxygenin-labeled probes were made with T7/T3 polymerase (NCBI references: *Antp*: AF435786, (Averof and Akam, 1995); *en*: X70939, (Manzanares et al., 1993); *cad*: AJ567452). Expression was examined in *Artemia* using protocols established by others (Manzanares et al., 1993). Nauplii were mounted in 90% glycerol and viewed Leica DMRB microscopy.

Transgenic Drosophila

Mutations to alter the FNWS in *Dm*-Ftz were generated by site-directed mutagenesis as previously described (Lohr and Pick, 2005). Multiple independent transformant lines were generated by Rainbow Transgenic Flies (Newbury Park, CA). Phenotypes shown were observed in at least five independent transgenic lines for each construct, and only one phenotype – that shown - was observed for each transgene. The levels of expression of the transgenes shown (Figure 3-5) were similar, as determined by RT-PCR using cDNA generated from L1 larvae.

Chapter 4

Variation and constraint in Hox gene evolution
[Heffer and Pick, manuscript in preparation]

4.1 Abstract

Embryonic transcription factors are often pleiotropic, having functions in diverse tissues at different times during development. This pleiotropy is thought to increase evolutionary flexibility, as gene expression patterns can be gained or lost in certain tissues without affecting essential function. How protein function can vary during evolution is less clear, as changes in functional motifs are expected to impact function in all tissues. Here we show that protein-interaction motifs important for early embryonic function of the Hox protein Fushi Tarazu (Ftz) that vary extensively, are not required for activity in developing the central nervous system, the tissue which shows the most highly conserved expression pattern in arthropods. Rather, the homeodomain – which is the only region conserved >550 million years of evolutionary time – is required for this function of Ftz. We propose that *ftz* has been maintained in all insect genomes examined to date because of its essential, homeodomain-dependent role in CNS development. This finding is a striking example of mosaic pleiotropy enabling regulatory protein evolution: co-option of Ftz into alternate early developmental pathways was permitted as long as the required variations in expression pattern and protein sequence did not impact CNS function.

4.2 Introduction

Transcription factor expression in certain tissues at particular times during development is crucial for the proper patterning of an organism. Homeotic (*Hox*) genes are a subset of these transcription factors, best known for their role in determining segment identity in virtually all

metazoans (Carroll et. al, 2005). Hox proteins were first characterized and classified based on their DNA-binding homeodomain, a highly conserved region in many transcription factors required for early embryogenesis (McGinnis et al., 1984a; Scott and Weiner, 1984). This sixty amino acid region, which is a helix-loop-helix motif similar to the DNA-binding domains of many bacterial proteins (Brennan and Matthews, 1989), binds to a TAAT consensus sequence both *in vitro* and *in vivo* to activate or repress downstream target genes (Ekker et al., 1991). Hox homeodomain sequences are highly conserved, with the most variation at the N-terminal arm, which has been found to increase functional specificity (Zeng et al., 1993).

Outside of the homeodomain, several functional motifs have been identified in Hox proteins. The YPWM motif, located upstream of the homeodomain and conserved in almost all Hox proteins, is important for interaction with Hox cofactor Exd (Johnson et al., 1995; Passner et al., 1999). The UbdA motif, located directly downstream of the homeodomain is found in Ubx, has been shown to be important for limb repression in the abdomen of insects (Galant and Carroll, 2002; Ronshaugen et al., 2002). The SSYF motif, found at the N-terminal end of many Hox proteins have been shown to be important for transcriptional activation of Scr (Zhao et al., 1996) and Ubx (Tour et al., 2005). Also, the C-terminal end of Dfd has been found to be important in functional identity (Lin and McGinnis, 1992).

fushi tarazu (*ftz*) is a rapidly evolving *Hox* transcription factor that has changed from *Hox*-like to pair-rule segmentation gene during the radiation of arthropods. Previously, we tracked changes in *ftz* sequence and expression that were important for this switch in function over 550 million years of arthropod evolution (Heffer et al., 2010). Specifically, an LXXLL motif necessary for interaction with the co-factor orphan nuclear receptor Ftz-F1 in *Drosophila* was acquired early in holometabolous insect lineages. The homeotic YPWM motif present in

other Hox proteins has degenerated several times independently in arthropod lineages. Relaxed selective evolutionary pressures has produced Ftz proteins that are very diverse from one another: some encode proteins that lack both a segmentation LXXLL and homeotic YPWM motif (like *Artemia* Ftz), while others encode proteins that have one or both of these motifs (*Tribolium* and *Drosophila* Ftz, respectively). All Ftz sequences isolated to date, however, share one similarity: a DNA-binding homeodomain.

In addition to having an early role in segmentation, *ftz* is expressed again later during embryogenesis in the developing central nervous system (CNS) in *Drosophila*. Specifically, *ftz* is expressed in mid-line precursor lineages (dMP2 and vMP2), neuroblast lineages (aCC, pCC, RP1, and RP2), and glial lineages (GP); however loss of function studies have shown that only the RP2 neurons are affected, in that there is transformation to the RP1 neuron (Doe et al., 1988). Others have shown that *ftz* is required for activation of Eve expression and function in the RP2 neurons (McDonald et al., 2003), as mutants lacking the region of the *ftz* promoter required for CNS expression (Doe et al., 1988) do not express Eve in developing RP2 neurons (Doe et al., 1988; McDonald et al., 2003).

Here we investigate the persistence of *ftz* in all arthropod genomes examined thus far, despite the great diversity in protein sequence. Specifically, we address the question: why has the *ftz* locus not become fossilized in any arthropod genomes examined so far spanning over >550 million years of evolution, despite it being apparently non-functional in some organisms? We tested the hypothesis, first suggested by Akam and colleagues (Alonso et al., 2001) that *ftz* has been conserved primarily because of its later role in embryogenesis in the developing nervous system. We find that the LXXLL segmentation motif and degenerated homeotic motif (FNWS) in *Drosophila* Ftz are dispensable for CNS function, but a homeodomain is required for

activation of Eve expression in RP2 neurons in the CNS. Together, these results suggest that Ftz has been retained because of its role in the developing CNS, and sequence diversity has coopted Ftz into earlier developmental pathways in some insects.

4.3 Results

4.3.1 *ftz* CNS expression is conserved over 550 million years of arthropod evolution

Despite diversity in Ftz sequence and early expression during embryogenesis (reviewed in (Pick and Heffer, 2012)), *ftz* expression in the embryonic central nervous system has been documented in a broad range of arthropods, including myriapods (Hughes and Kaufman, 2002b; Damen et al., 2005; Janssen and Damen, 2006), crustaceans (Mouchel-Vielh et al., 2002; Heffer et al., 2010), a few insects (Carroll and Scott, 1985; Brown et al., 1994; Dawes et al., 1994; Hughes et al., 2004), and in a distant lophotrochozoan outgroup, where the Ftz ortholog Lox5 is expressed in the CNS (Kourakis et al., 1997; Telford, 2000). Here, we show that this CNS expression pattern is conserved in arthropods with very diverse Ftz sequences and early expression patterns (Figure 4-1). Ftz from the brine shrimp *Artemia* is 201 amino acids in length, lacking both segmentation LXXLL and homeotic YPWM motifs and is expressed in a marginally detectable homeotic pattern in early nauplii (Heffer et al., 2010). *ftz* from two beetle species – *Tribolium* and *Callosobruchus* – encode proteins that are 290 and 368 residues, respectively; both sequences have LXXLL and YPWM motifs and are expressed in stripes (Brown et al., 1994; Heffer et al., 2011; A.H. unpublished data). *Drosophila* Ftz is 410 amino acids long and includes an LXXLL but no YPWM motif and is expressed in stripes. Outside of these motifs, there are no conserved regions of the protein sequence except the homeodomain. Despite diversity in sequence and expression, conservation of *ftz* expression in the developing

central nervous system appears to be the only constant feature associated with extant *ftz* genes. If this role provides the evolutionary constraint that retains *ftz* in all arthropod lineages, we hypothesize that the protein motifs that vary in Ftz between species, specifically the LXXLL and YPWM motifs, should be dispensable for this conserved function.

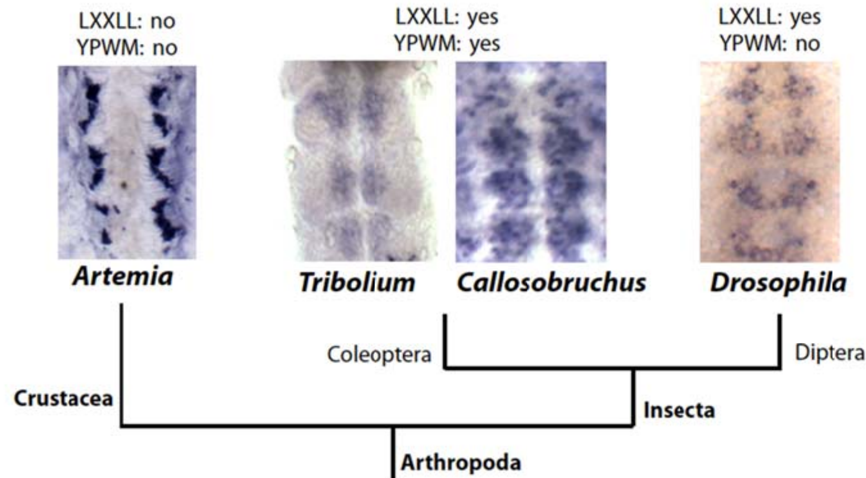


FIGURE 4-1. ***ftz*** **CNS expression is conserved across arthropods that exhibit great diversity in Ftz sequence composition.** Expression of *ftz* was analyzed by in situ hybridization using conspecifc probes in embryos from diverse arthropods, as indicated below. Neuronal expression was detected in *Artemia*, *Tribolium*, *Callosobruchus*, and *Drosophila*, Ftz from these species share a homeodomain but harbor different protein interaction motifs, as indicated above.

4.3.2 Ftz-F1 and Exd are not co-expressed with Ftz in the CNS

During the blastoderm stage of development Ftz interacts with co-factor Ftz-F1, and together Ftz/Ftz-F1 bind composite sites in the regulatory regions of target genes involved in segmentation (Florence et al., 1997; Yussa et al., 2001; Bowler et al., 2006; Hou et al., 2009). If Ftz functions in a similar fashion to regulate target genes in the CNS, a minimal requirement is that Ftz and Ftz-F1 would be co-expressed in this tissue, as they are at earlier developmental stages. As shown in Figure 4-2A, Ftz-F1 was not detected in Ftz^+ neurons. Similarly, *ftz-f1* RNA was not detected from stages 7-12 of embryogenesis (Project, 2012) and Ftz-F1-DNA

complexes were not observed in 4-13hr embryo extracts by gel-shift assays (Ueda et al., 1990). Interestingly, there are no Ftz-F1 consensus binding sequences in the *ftz* neurogenic element, further suggesting activation of *ftz* in the central nervous system is Ftz-F1 independent (data not shown). Other Hox proteins interact with Exd via a YPWM motif located upstream of the homeodomain (Johnson et al., 1995; Passner et al., 1999). Since Ftz retains the "W" residue critical for interaction with Exd, we asked whether Exd was co-expressed with Ftz in the CNS. As shown in Figure 4-2B, although Exd is expressed during the time of Ftz CNS expression, Exd expression was localized to the ectoderm and did not overlap with Ftz^+ neurons. Thus, neither Ftz-F1 nor Exd co-localized with Ftz in the developing CNS. This suggests that Ftz functions in a Ftz-F1- and Exd-independent fashion in the CNS.

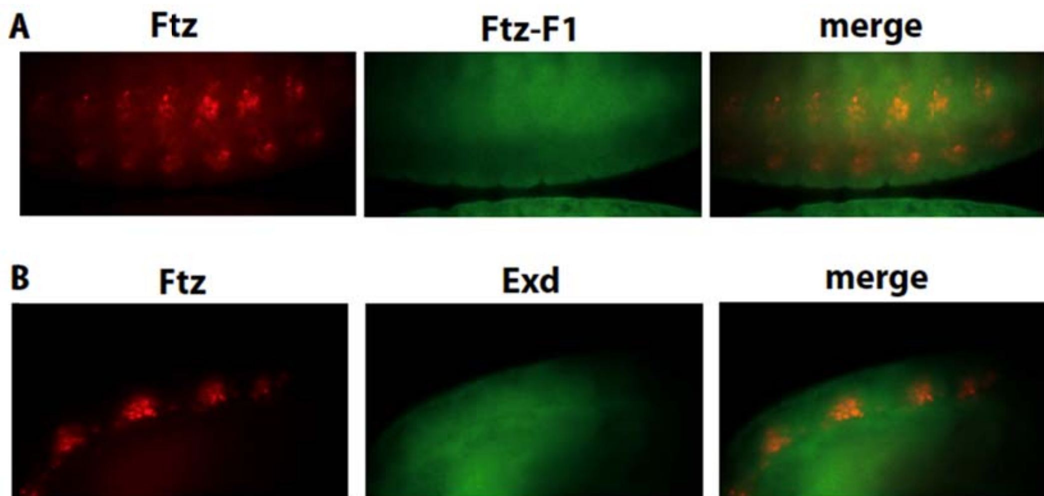


FIGURE 4-2. Ftz is not co-expressed with known cofactors in the CNS. Ftz expression was compared to that of known Hox cofactors, Ftz-F1 and Exd by immunohistochemical staining of whole mount *Drosophila* embryos. (A) Ftz (red) expression in a cluster of cells in every segment of the developing CNS. Only background staining was observed with anti-Ftz-F1-antibody (green) at this time of development. Merge of these images shows Ftz and Ftz-F1 do not co-localize in the CNS. (B) Ftz (red) does not overlap with Exd (green), which was expressed in the nuclei of ectodermal cells and not the CNS. Merge of images from shows expression of Ftz and Exd in different cell layers. Exd was expressed in a different cell layer (out of view in this figure).

4.3.3 Cofactor interaction motifs in Dm-Ftz are dispensable for CNS function

The protein motifs in Ftz that mediate interaction with known cofactors, LXXLL and YPWM, show a high degree of variation in arthropod lineages. If the evolutionary constant function of Ftz is in the CNS, then these motifs should be dispensable for that role, as many Ftz proteins lack one or both motif (Heffer et al., 2010). To test this, we made use of a *Drosophila* line carrying a rescue transgene that lacks the *ftz* neurogenic element, *ftzK* (Hiromi et al., 1985; Doe et al., 1988). Embryos carrying *ftzK* in a *ftz*^{9H34} background have normal segmentation but no *ftz* CNS expression (Doe et al., 1988). In the absence of Ftz CNS function in these animals, RP2 neurons fail to develop during neurogenesis, as evidenced by lack of Even-skipped (Eve) expression ((Doe et al., 1988); Figure 4-3A). To test whether Ftz protein motifs are necessary for CNS function, we generated a series of transgenes containing the *ftz* neurogenic element (NE), *ftz* basal promoter, and Ftz wild type coding sequence (NE-Ftz), or coding sequences with mutations in motifs known to be important for Ftz function (Figure 4-3B). The LRALL motif was changed to LRAAA (NE-FtzLRAAA); the FNWS in Dm-Ftz was changed to AAAA (NE-FtzAAAA); and several mutations were made in the homeodomain: 1) the N-terminal arm of the Ftz homeodomain (<u>S</u>KR<u>T</u>RQTY) – which defines this group of homeodomain proteins (Duboule, 1994; Telford, 2000) – was changed to that of Antp (<u>R</u>KR<u>G</u>RQTY; NE-FtzNTAntp); 2) The entire Ftz homeodomain was swapped for that of Antp (NE-FtzAntpHD); 3) the entire Ftz homeodomain was deleted (NE-FtzΔHD). Together, these lines would allow us to evaluate the importance of the *Drosophila* segmentation motif (LRALL), degenerated homeotic motif (FNWS), and homeodomain in proper RP2 formation.

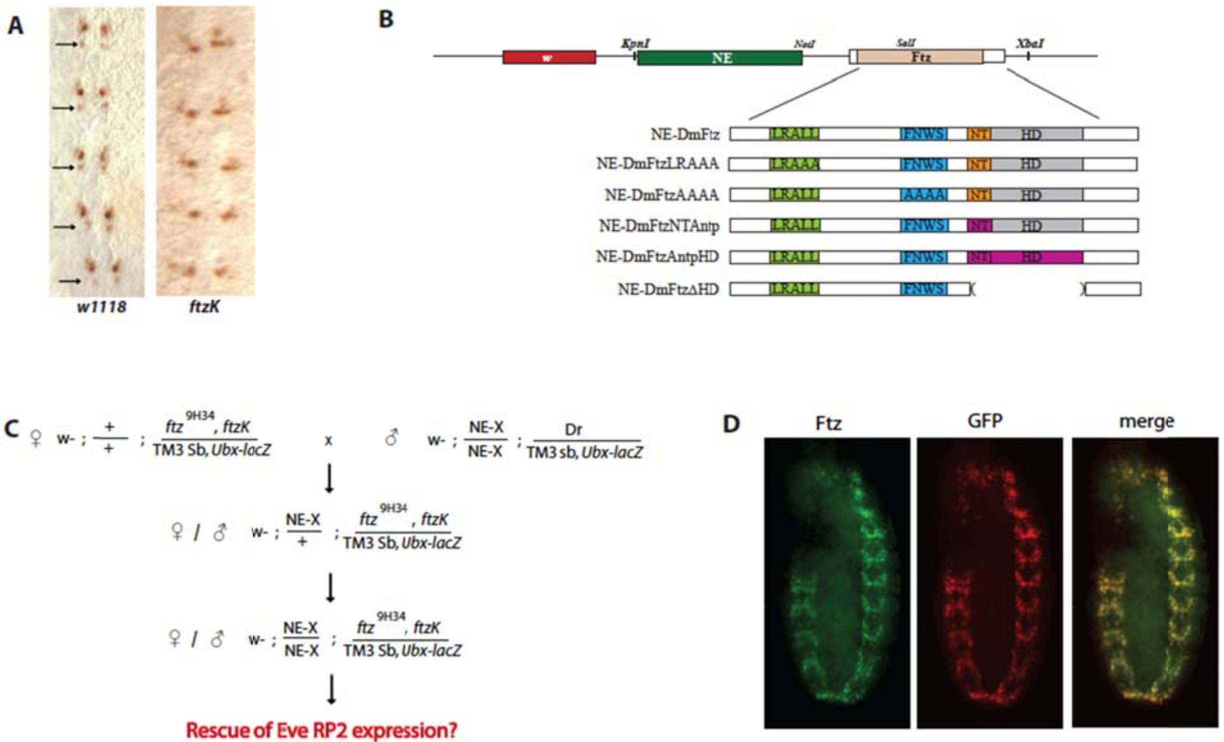


FIGURE 4-3. Strategy to test the role of labile and stabile Ftz protein motifs in *Drosophila* CNS function. A) Eve expression was monitored with an anti-Eve antibody in stage 10-12 embryos. Eve expression marks the RP2 neurons of wildtype embryos (left panel, arrows). Eve was not detected in *ftz* mutants expressing *ftzK*, which rescues *ftz* expression and function in segmentation but lacks the neurogenic element (NE) that drives CNS expression (Doe et al., 1988). B) Schematic of constructs designed to test function of Ftz motifs in the CNS. All transgenes included the 2.2kb NE, *ftz* basal promoter, *ftz* coding region, and ~200bp downstream of the stop codon. Transgenes included the following changes: the LRALL motif required for segmentation and interaction with Ftz-F1 (bright green) was mutated to LRAAA in NE-FtzLRAAA, the degenerated homeotic FNWS motif (blue) was changed to AAAA in NE-FtzAAAA, the N-terminal arm of Ftz (orange) was changed to that of Dm-Antp (purple) in NE-FtzNTAntp, the Ftz homeodomain (grey) was swapped with that of DmAntp (purple) in NE-FtzAntpHD, and the homeodomain was deleted from NE-FtzΔHD (shown with empty parenthesis). C) The crossing scheme used to test CNS functional rescue is shown. These crosses were carried out with multiple independent transformant lines for each of the transgenes shown in (B), indicated as NE-X. In the final cross, virgins and males homozygous for NE-X and carrying *ftzK* on a ftz^{9H34} chromosome were self-crossed and embryos were tested for rescue of Eve RP2 neuron expression. D) The Ftz NE is sufficient to drive transgene expression in Ftz+ neurons. NE-GFP lines were double-stained with α-Ftz (green) and α-GFP (red) antibodies. Expression of these proteins overlapped (merge, yellow), showing that the regulatory region used was sufficient to test for rescue.

As shown in Figure 4-4A, NE-Ftz rescued Eve RP2 expression in approximately 60% of embryos (n=188). NE-FtzLRAAA rescued Eve RP2 expression in ~45% of the embryos (n=66), which is consistent with the finding that Ftz-F1 is not expressed during stages at the time during embryonic development when *ftz* is expressed in the central nervous system (Figure 4-4A; see above). NE-FtzAAAA also showed rescue of Eve expression in RP2 neurons (~50% rescue, n=57; Figure 4-4A). Curiously, Dm-Ftz has a degenerate YPWM sequence (FNWS) that still includes the "W" residue important for interaction with a hydrophobic binding pocket on the surface of Exd (Johnson et al., 1995; Passner et al., 1999). However, changing the FNWS motif to AAAA had no drastic effect on the ability of the NE-FtzAAAA transgenes to rescue Eve RP2 expression, suggesting this motif is not required for Ftz function in the central nervous system. Taken together with the expression data above, these results suggest that Ftz function in the CNS is independent of Ftz-F1 and Exd.

4.3.4 The DNA-binding homeodomain is required for CNS function

In contrast to the motifs described above, the homeodomain was absolutely required for Ftz CNS function, as NE-FtzΔHD showed virtually no rescue of Eve RP2 expression (~2.5% of embryos, n=110, Figure 4-4A). However, when the N-terminal arm of the Ftz homeodomain was swapped with that of Antp (NE-FtzNTAntp), rescue levels were similar to that of NE-Ftz, suggesting that any specificity encoded in the N-terminal arm of the homeodomain was not important in Ftz CNS function (56%, n=46). To further test the extent of homeodomain specificity in Ftz CNS function, the entire Ftz homeodomain was replaced with that of Antp (NE-FtzAntpHD). Although levels were somewhat reduced, Ftz protein with the Antp HD effectively rescued Eve RP2 expression there was rescue of Eve RP2 expression (38%, n=82).

These results demonstrate that the homeodomain is required for Ftz activation of Eve in RP2 neurons. Further, they suggest that neither the N-terminal region nor any other specific residues of the Ftz homeodomain are uniquely required for function in the central nervous system.

When the Ftz and Antp homeodomain sequences are aligned, only three of the sixty amino acids are non-synonymous substitutions between the two (Figure 4-4B; highlighted): one residue in the N-terminal arm (residue 4), one in the linker between helix 2 and 3 (residue 39), and one in the third helix (residue 56). Interestingly, all three of these non-synonomous substitutions are residues that could be phosphorylated in the Ftz homeodomain, but not in the Antp homeodomain. However, since the Antp homeodomain was able to rescue Eve expression in the RP2 neurons, this suggests that phosphorylation of these sites is not crucial for Ftz homeodomain function. Also, an alignment of all Ftz homeodomain sequences collected from arthropods that have documented *ftz* expression in the CNS suggests that residue 39 has no constraints, as its identity varies greatly in nature, and residue 56 can have either a hydrophobic or hydrophilic nature, as long as the side-chain is small (Figure 4-4B). Therefore, we cannot confidently predict a residue (or residues) required for Ftz-specific function in the CNS.

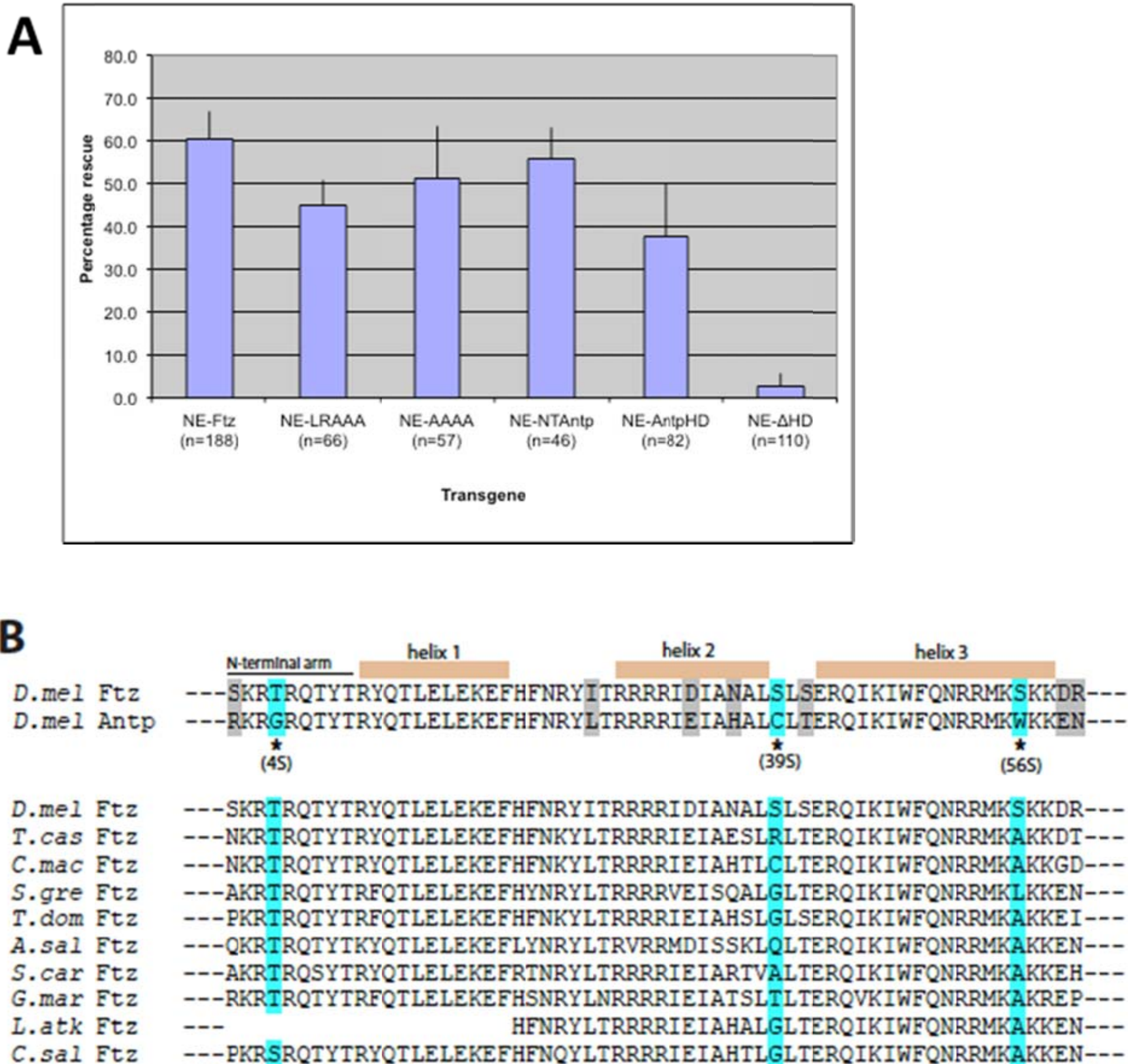


FIGURE 4-4. The homeodomain is required for Ftz function in the CNS while cofactor interaction motifs are dispensable. A) Rescue of Eve expression in RP2 neurons by different transgenes. For each construct indicated, the percentage of embryos that displayed Eve RP2 neuronal staining is shown. Error = 1 standard deviation. B) Homeodomain alignments highlighting differences in sequence. Dm-Ftz and Dm-Antp homeodomain sequences are 95% similar and 83% identical, with only 3 non-synonomous substitutions at the amino acid level (top). Alignment of Ftz homeodomains from species where CNS expression has been reported. All residues highlighted in blue are non-synonymous substitutions and those in grey are synonymous (bottom).

4.3.5 Ftz function in the CNS is independent from other *Hox* genes

Like many *Hox* genes, *Dm-Antp* and *Dm-Scr* are also expressed in the developing CNS. However, Antp and Scr expression in the CNS does not overlap with Ftz (Figure 4-5), suggesting the roles each has in the CNS are distinct. We found that Antp and Scr expression in the CNS not only appears later than Ftz expression, but is in a different set of neurons; both of these other Hox proteins' expression also persists later in the developing than Ftz does (Figure 4-5). Since Ftz expression in the CNS is slightly earlier, it was possible that *ftz* acts upstream of other *Hox* genes and could activate their CNS expression, potentially explaining the ability of NE-FtzAntpHD to rescue Ftz CNS function. However, Dm-Antp expression is the same as wild-type in the *ftzK* line (Figure 4-6). Together, these results suggest Ftz has a role in the developing CNS distinct from other *Hox* genes, and Ftz doesn't activate *Antp* in the CNS.

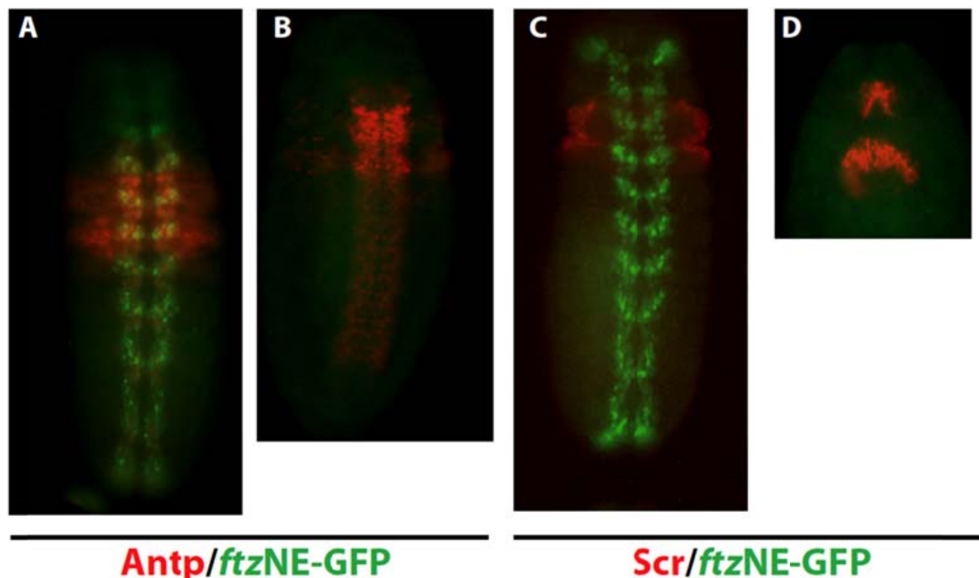


FIGURE 4-5. Ftz CNS expression does not overlap with Antp or Scr. Ftz (green) expression in the CNS does not overlap with Antp (red, A) or Scr (red, C). Later in development when Antp (B) and Scr (D) are expressed in a subset of neurons, Ftz expression is no longer detected.

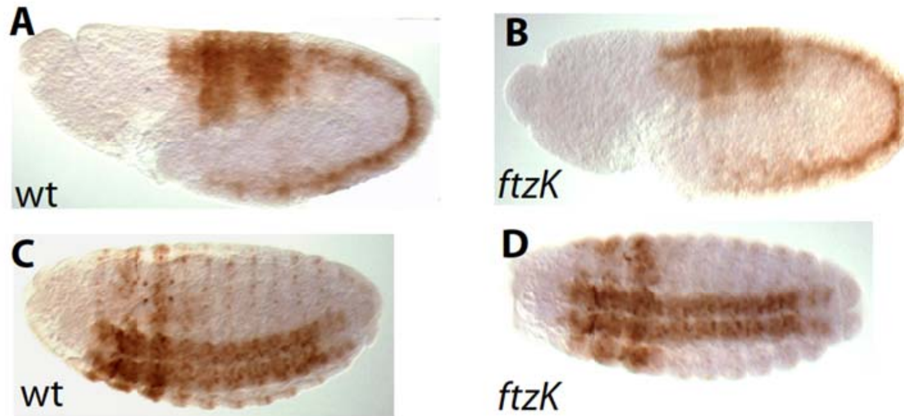


FIGURE 4-6. Ftz does not activate Antp expression in the developing CNS. Antp expression in *ftzK* mutants is identical at early (A and B) and later (C and D) stages of CNS development, and is indistinguishable in wild type (wt, A,C) and *ftzK*-expressing (*ftzK*, B,D) embryos, suggesting loss of *ftz* CNS expression does not affect Antp expression.

4.4 Discussion

Pleiotropy is often discussed with regards to different cis-regulatory elements directing expression in different tissues at different times during development. Here, we add an additional level to this pleiotropy, where to our knowledge, for the first time, we show that a transcription factor can act at different times during development in different tissues through the use of specific motifs or domains. While these motifs are important for one function of the protein, they are dispensable for others. In the case of *ftz*, the LXXLL motif is required for segmentation but is not needed for Ftz function in the developing nervous system (Figure 4-4), whereas the homeodomain is required for CNS function. Interestingly, this is the only domain conserved across all Ftz proteins isolated to date, as it is present in *ftz* genes with very diverse protein sequences (Heffer et al., 2010). Interestingly, the Ftz homeodomain is much less conserved across arthropod Ftz sequences when compared to other Hox homeodomains (Heffer et al., 2010). While the functional consequences of this homeodomain sequence diversity are not

known, it can be suggested that it is the structure of the homeodomain and not its sequence that are important for CNS function. Supporting this is our result that a Ftz transgene with an Antp homeodomain could rescue Eve RP2 neuron expression (Figure 4-4). While the percentage rescue of Ftz with an Antp homeodomain was slightly lower and perhaps due to differences in binding affinities for different regulatory sites, this homeodomain substitution was still able to rescue Eve RP2 expression.

Hox gene evolution is generally thought to be very constrained because precise timing and location of expression is needed for proper segment identity during development, and mis-expression has detrimental consequences to development. Here we show that a rapidly evolving *Hox* gene has escaped many of these evolutionary constraints imposed on other *Hox* genes while being co-opted into earlier developmental pathways, but has likely been retained in arthropod genomes because of the constraints of the DNA-binding homeodomain in CNS function. Thus, Ftz partitions it functions during embryogenesis through pleiotropy of both cis-regulatory sequences as well as protein motifs. This "double pleiotropy" not only allows a transcription factor to function in multiple tissues during development, but also provides a template for evolution to act upon.

4.5 Methods

Ftz rescue transgenes

The ~2.2kb fragment containing the Neurogenic Element (NE, (Hiromi et al., 1985)), extending from the XbaI to BalI restriction sites in the 10 kb genomic region sufficient for rescue of *ftz* mutants (Hiromi et al., 1985) was inserted into pCasper4. The *Drosophila ftz* 5' UTR, coding region, 3'UTR, and ~200bp downstream of the poly-adenylation signal were inserted

downstream of the NE in pCasper4 using standard techniques. Mutations made to the Ftz coding region were done using site-directed mutagenesis (primer sequences available upon request). Homeodomain-deletion and swaps were done by fusion PCR.

Due to lethality issues with expressing *Hox* trangenes using attB integration sites, traditional P-element integration techniques were used, such that transgenes were inserted randomly into the *Drosophila* genome (Brand and Perrimon, 1993). For each construct, 3-7 independent lines were established that were homozygous viable on the second chromosome. Males homozygous for NE-Ftz constructs (NE-X), carrying *Dr/TM3SbUbx-lacZ* on chromosome III, were crossed with ftz^{9H34}, *ftz*K/ TM3Sb, *Ubx-lacZ* virgin females and males and females carrying one copy of NE-Ftz and ftz^{9H34}, *ftz*K/ *TM3SbUbx-lacZ* were selected and self-crossed (Figure 4-3C). Rescue efficiency was measured by calculating the percentage of embryos homozygous for ftz^{9H34} (β-galactosidase negative embryos) that showed Eve antibody staining in any number of RP2 neurons in stage 10-12 embryos. Rescue percentages from several independent transgene lines were averaged together. To confirm that the *ftz* cis-regulatory elements present in the transgene were sufficient to drive transgene expression in the Ftz+ cells of the CNS a transgene in which GFP-coding sequence was placed downstream of the NE, *ftz* basal promoter, and first 169 amino acids of the Ftz coding region was generated. GFP was detected in an identical pattern to native Ftz protein, as visualized by double antibody staining of GFP and Ftz (Figure 4-3D).

<u>*Arthropod care and embryo collection*</u>

Artemia and *Callosobruchus* were obtained, reared, and maintained as previously described (Heffer et al., 2010). *Tribolium* were reared on whole-wheat flour with 5% yeast at 30°C.

Drosophila were maintained at 25°C, with 60% humidity on standard cornmeal/yeast food. One-week old *Artemia* nauplii were fixed according to (Heffer et al., 2010). *Callosobruchus* embryos that were 2 days old were collected by first soaking mung beans with eggs in a dilute bleach solution, scraping the eggs off the beans with a paintbrush, and then fixing according to standard *Drosophila* protocols. *Drosophila* embryos were collected over 2 hours, aged for 5-6 hours on apple juice plates at 25°C, and then fixed according to standard protocols.

Analysis of gene expression patterns

In situ hybridizations were performed according to established protocols in *Drosophila* (Tautz and Pfeifle, 1989; Kosman and Small, 1997), *Tribolium* (Schinko et al., 2009), and *Artemia* (Manzanares et al., 1993; Copf et al., 2004; Heffer et al., 2010). *Callosobruchus* embryos were first dissected from their thick vitelline membrane, and then stained according to *Drosophila* protocols. Digoxygenin-labeled probes were made with T7/T3 polymerase using embryonic cDNA and detected with a sheep anti-digoxigenin antibody (1:2000, Roche), and stained with NBT+BCIP according to the manufacturer's instructions.

Drosophila antibody stainings were performed according to established antibody protocols (Gutjahr et al., 1994). Primary antibodies used were: mouse anti-Ftz (1:1000; (Kellerman et al., 1990)), guinea-pig anti Eve (1:1000; M. Frausch), rabbit anti-GFP (1:1000; Invitrogen). Secondary antibodies used were: anti-mouse Alexa488 (1:500, Molecular Probes), anti-rabbit Alexa568 (1:500, Molecular Probes), biotinylated anti-guinea pig (1:1000, KPL). Embryos were mounted in Vectashield mounting solution with DAPI (Vector Laboratories), and scored for rescue and visualized by Leica DMRB microscopy.

Chapter 5

Investigating the roles of ftz and ftz-f1 in the short-germ beetle Tribolium castaneum
[Preliminary findings]

5.1 Introduction

Segmentation is a critical developmental process that all insects undergo during embryogenesis to pattern the body plan. While modes of forming segments may differ, all insects have a clearly segmented body with distinct head, thoracic, and abdominal regions. Insects use one of three modes of germ development to pattern the early embryo, which mainly differ as to when segments are formed during embryogenesis (reviewed in Davis and Patel, 2002). In short-germ insects, the head segments are specified at the blastoderm stage of development, while thoracic and abdominal segments are added sequentially from a posterior proliferation region as development progresses. In long-germ insects, all head, thoracic, and abdominal segments are specified simultaneously at the blastoderm stage. Intermediate-germ insects develop somewhere in between, specifying the head and some thoracic segments at the blastoderm stage, and then adding the remaining segments from the growth zone.

Many studies have elucidated the pathways and genes involved in segmentation in the long-germ insect *Drosophila* (Nusslein-Volhard and Wieschaus, 1980; Akam, 1987; Clyde et al., 2003; Jaeger et al., 2004; Schroeder et al., 2004). Here, a hierarchy of segmentation genes is turned on by maternally deposited transcription factors. The gap genes divide the embryo into broad regions, and activate the pair-rule genes, which are expressed in seven stripes and pattern parasegments. Pair-rule genes then activate segment polarity genes, which give anterior-posterior identity to each segment, ultimately producing an embryo made up of small metameric regions. Mutants in each class of segmentation genes lack the regions which they pattern

(Nusslein-Volhard and Wieschaus, 1980). For example, *ftz* is a pair-rule gene expressed in seven stripes during the blastoderm stage of *Drosophila* development, and *ftz* mutants are missing half of their segments, specifically the regions that *ftz* patterns (Kuroiwa et al., 1984).

In recent years, there has been some effort to understand the gene networks underlying the other modes of segmentation. In *Tribolium*, the pair-rule genes have received the most attention, and the expression and function of all pair-rule orthologs have been reported (Choe et al., 2006; Choe and Brown, 2007). While pair-rule expression patterns are conserved during *Tribolium* embryogenesis, functional studies revealed two classes of pair-rule genes. "Primary" pair-rule genes, which include *eve*, *odd*, and *run*, produced severely truncated embryos when the gene was knocked-down, which suggested these genes are important for both segmentation and elongation. "Secondary pair-rule genes", such as *prd* and *slp*, gave pair-rule cuticle phenotypes in which half of the body segments were missing. Other pair-rule orthologs had pair-rule expression patterns, but produced no phenotype at all, such as *h, ftz, opa,* and *Ten-m*.

Through studies like this, it has become apparent that there are differences in the gene networks underlying segmentation in these insects. For example, the transcription factor *even-skipped* (*eve*) has both pair-rule expression and function in *Drosophila*, but no segmentation expression pattern in the short-germ grasshopper *Schistocerca* (Patel et al., 1992), suggesting *eve* doesn't have a function in the segmentation process in this insect. In both the short-germ beetle *Tribolium* and intermediate-germ milkweed bug *Oncopeltus*, *eve* is expressed in every segment and loss-of eve function produces truncated, or gap-like, phenotypes (Brown et al., 1997; Liu and Kaufman, 2005; Choe et al., 2006).

Here, we report that the nuclear receptor *ftz-f1* is expressed as a pair-rule gene in *Tribolium* and has a role in segmentation in short-germ development. We find that *ftz-f1* stripes overlap with *ftz* stripes, an interaction that is critical for proper segmentation in *Drosophila*. We also find that Ftz-F1 has two roles in *Tribolium* embryogenesis: first, an early role in segmentation, and second a role later in proper cuticle formation. Preliminary results also suggest that Ftz and Ftz-F1 are partially redundant early in development. Together, these results suggest that *ftz-f1* may have had an ancestral role as a pair-rule gene and had a role in segmentation before *ftz*.

5.2 Results

5.2.1 *ftz* and *ftz-f1* have pair-rule expression in *Tribolium*

In the long-germ dipteran *Drosophila melanogaster*, *ftz* is expressed in seven stripes during the blastoderm stage of development (Hafen et al., 1984), where it functions to pattern these regions, as they are missing in *ftz* mutants (Kuroiwa et al., 1984). In *Tribolium*, *ftz* (*Tc-ftz*) was first detected in a single stripe during the blastoderm stage of development (Figure 5-1A; (Brown et al., 1994)), and later in three pair-rule stripes, which overlap with every other Engrailed stripe (Figure 5-1B-D; (Brown et al., 1994)). *Drosophila ftz-f1* (*Dm-ftz-f1*) is a maternal gene, expressed ubiquitously throughout the embryo at the blastoderm stage of development (Yu et al., 1997). However, *Dm-ftz-f1* functions only in the regions that overlap with *ftz*, as *ftz-f1* mutant embryos are identical to *ftz* mutants (Florence et al., 1997; Guichet et al., 1997; Yu et al., 1997; Schwartz et al., 2001; Yussa et al., 2001). We examined *ftz-f1* expression in the short-germ beetle *Tribolium castaneum* (*Tc-ftz-f1*) and found that *ftz-f1* was also expressed in a single stripe during the blastoderm stage of embryogenesis (Figure 5-1E), and

then in 4 pair-rule stripes that overlap with every other Engrailed stripe (Figure 5-1F-H). Moreover, *Tc-ftz-f1* and *Tc-ftz* striped expression overlap, as simultaneous in situ hybridization did not produce any new stripes (data not shown). Together, these results show that both *ftz* and *ftz-f1* are expressed in pair-rule patterns during *Tribolium* development, suggesting they might play a role in segmentation in this organism.

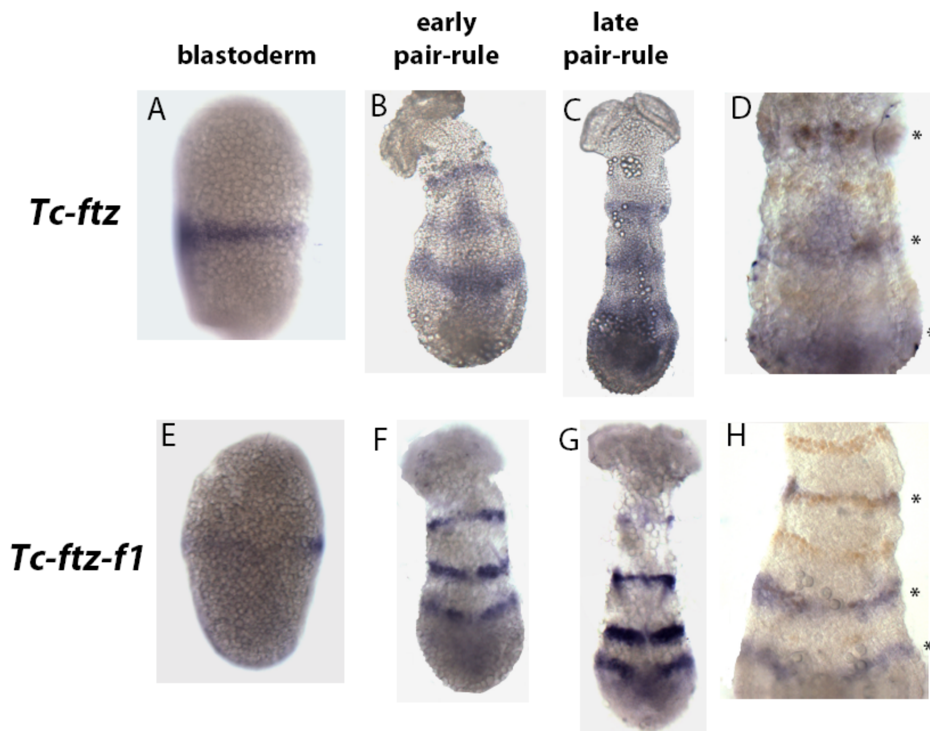


FIGURE 5-1. *ftz* and *ftz-f1* are expressed in pair-rule patterns in the short-germ beetle *Tribolium*. (A-D) *Tribolium ftz* expression. *ftz* was expressed in a single stripe during the blastoderm stage (A), which then resolved into pair-rule stripes as the germ band elongated (B,C). These stripes are pair-rule, as they overlap with alternating Engrailed stripes (D, stars). (E-H) *Tribolium ftz-f1* expression. *ftz-f1* was also expressed in a single stripe early in development (E) which resolved into stripes (F,G) that overlap with every other Engrailed stripe (H, stars). For all images, RNA was detected by *in situ* hybridization using *dig* probes (blue), and Engrailed protein was detected by antibody staining using the 4D9 anti-En antibody (Developmental Hybridoma Bank) and visualized using DAB staining (brown).

5.2.2 *ftz* and *ftz-f1* embryonic RNAi effects on cuticle formation

Parental RNAi has been used to elucidate the function of many embryonic genes in *Tribolium* (e.g., (Bucher et al., 2002; Choe et al., 2006; Farzana and Brown, 2008)). However, *Tc-ftz-f1* was shown to be necessary for oogenesis (Xu et al., 2010) and *Tc-ftz* was previously reported to have no function in *Tribolium* using parental RNAi and in analysis of a large genomic deletion (Stuart et al., 1991; Choe et al., 2006). Since parental RNAi either cannot be used or did not produce a phenotype, we performed embryonic RNAi in *Tribolium* (Posnien et al., 2009) to see if either of these genes are required for early development. Injection of *Tc-ftz-f1* double-stranded RNA (dsRNA) produced a range of cuticle phenotypes, from truncated legs (Figure 5-2B) to missing abdominal segments (Figure 5-2C), to a complete absence of cuticle formation ("strong phenotype"; Figure 5-2D). At high dsRNA concentrations, a majority of injected embryos displayed this strong phenotype, with eyespots visible through the developing eggshell, but no solid structure within (Figures 5-2D, 5-3). Because *ftz-f1* has a later role in embryogenesis in both flies (Ruaud et al., 2010) and nematodes (Asahina et al., 2000), we looked to see if the same might be true in *Tribolium*. While *ftz* transcripts could only be detected during the first two days of embryogenesis, *ftz-f1* transcripts could be detected throughout embryo development (Figure 5-4). These results suggest that the *ftz-f1* cuticle phenotype we see after knocking-down gene expression is likely due to a later role of *ftz-f1* in embryogenesis.

Injection of *Tc-ftz* dsRNA had no apparent effect in early development, even at high concentrations, as a similar percentage of embryos hatched compared to controls (Figure 5-3), and larval cuticles all appeared wild-type (data not shown). Experiments are in progress to test whether *ftz* RNA was effectively knocked-down by the dsRNA. Together, these results suggest that *Tc-ftz* does not function during embryogenesis, although it is expressed in stripes. *Tc-ftz-f1*

is important for proper cuticle formation late in embryogenesis, making it difficult to discern potential roles in segmentation.

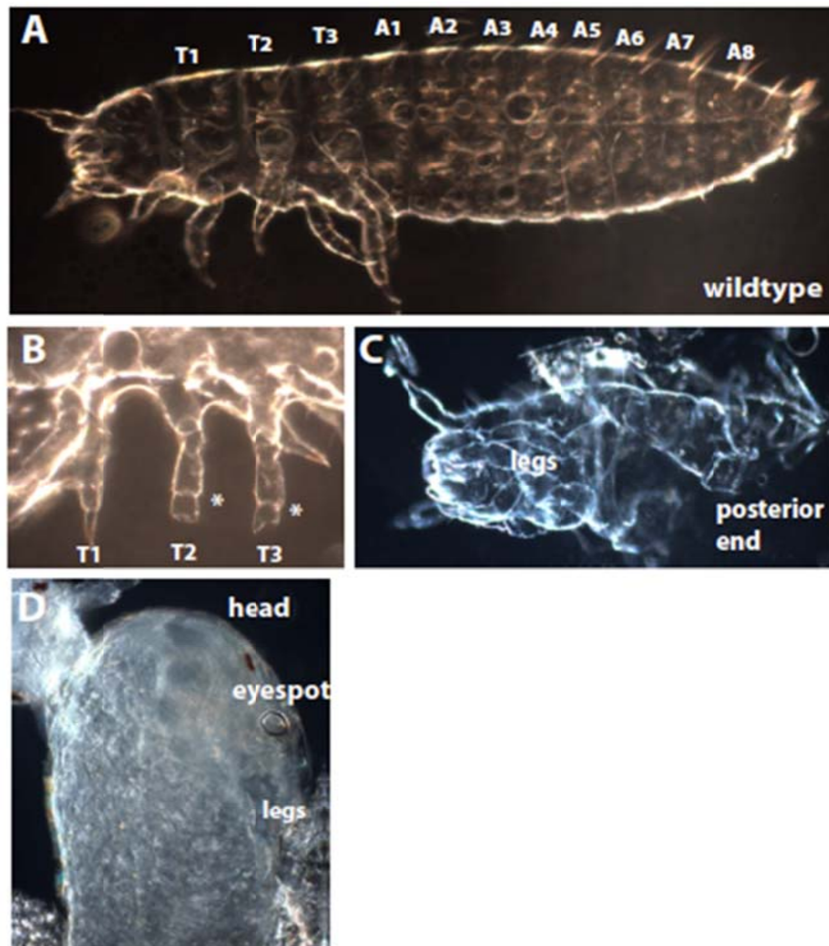


FIGURE 5-2. The effects of *Tc-ftz-f1* embryonic RNAi on cuticle formation. A) Wild-type cuticle, with a head, three thoracic segments, eight abdominal segments, and a telson. B) A weak *Tc-ftz-f1* RNAi phenotype, in which the T2 leg was truncated (asterisk). C) A stronger *Tc-ftz-f1* RNAi phenotype in which all abdominal segments were missing, but the thoracic segments and telson were present. D) A strong *Tc-ftz-f1* RNAi phenotype in which eyespots and an apparent head and thoracic region were visible through the vitelline membrane, but no solid structure could be dissected.

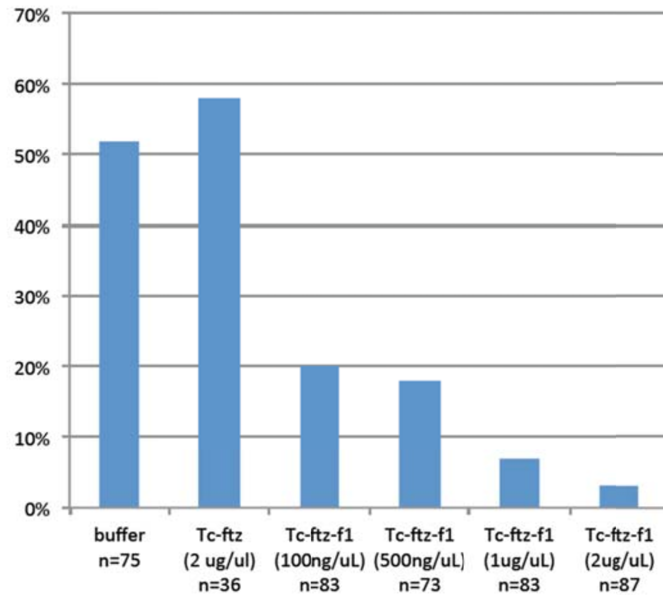


FIGURE 5-3. Hatching percentages of injected *Tribolium* embryos. Embryos were injected with varying concentrations of dsRNA before the blastoderm stage of development and allowed to develop for 4 days at 30°C. At this time, the number of larva that had hatched was counted. Injection of buffer was used as a control.

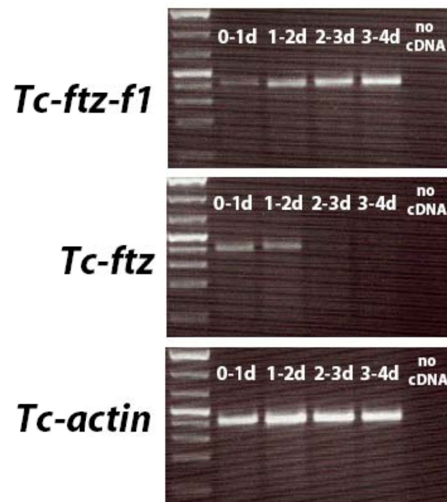


FIGURE 5-4. *ftz* and *ftz-f1* expression during embryogenesis as determined by RT-PCR. RNA was extracted from embryos were collected over a 24-hour period, and cDNA was made. PCR was performed using this cDNA as a template. *Tc-actin* was used as a positive control for the amount of template used in each reaction.

5.2.3 Investigating the roles of *Tribolium ftz* and *ftz-f1* in segmentation

To examine the possible roles *ftz* and *ftz-f1* may have in early embryogenesis, embryos were injected with dsRNA and stained with the segmentation marker Engrailed. Engrailed (En) is expressed in the posterior region of each segment during embryogenesis, and is used as a marker for each developing segment in many insects (Patel et al., 1989). While injection of *Tc-ftz* showed no effect on En expression, injection of *Tc-ftz-f1* showed a decrease in En expression in every other segment (Figure 5-5, arrows, n=2/5 embryos), suggesting a role in segmentation. Further experiments are being done to examine this apparent role in segmentation.

To investigate whether *ftz* or *ftz-f1* showed any delay in the timing of early embryogenesis, we examined embryos injected with dsRNA 24-hours post-injection and determined what percentage of embryos had developed a fully extended germband by DAPI staining (Figure 5-6). Injection of buffer or *Tc-ftz* dsRNA had no apparent effect on embryo growth, as a majority of the embryos had fully-extended germbands (71% and 82% embryos, respectively). Injection of two different *Tc-ftz-f1* dsRNAs showed a decreased number of embryos with a fully extended germband (50-60% embryos), with many embryos at earlier developmental stages (~30% embryos). Interestingly, injection of both *ftz* and *ftz-f1* dsRNAs revealed an even smaller percentage of embryos with an extended germband (32%), and many in either the blastoderm stage (26%) or early germband stage of embryogenesis (20%). Together, these results suggest that *ftz* and *ftz-f1* may have partially redundant roles in early development in *Tribolium*.

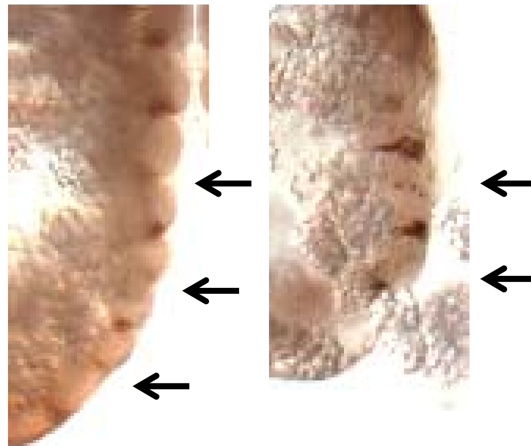


FIGURE 5-5. ***Tribolium ftz-f1* exhibits a role in segmentation.** Embryos injected with *ftz-f1* dsRNA show decreased Engrailed expression (brown) in every other segment. The two embryos shown here have fully extended germbands.

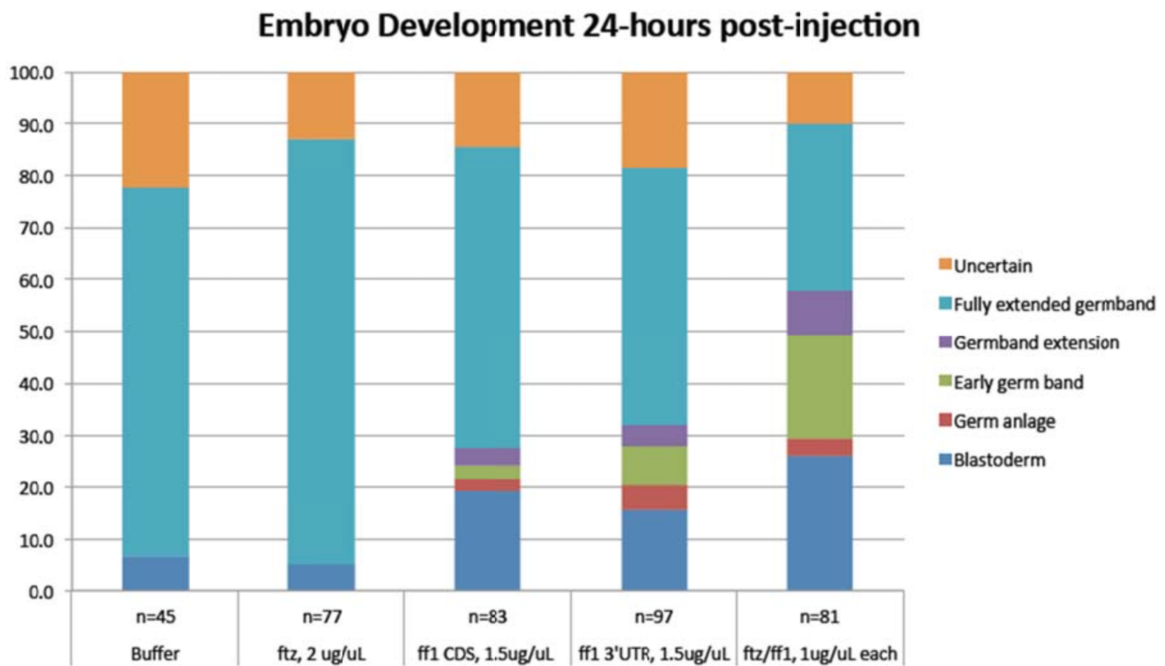


FIGURE 5-6. Effects of injecting *ftz* and *ftz-f1* dsRNA on developmental timing. Injected embryos were incubated at 30°C for 24 hours after injection and then visualized with DAPI staining. The percentage of embryos at each developmental stage is shown above.

5.3 Discussion and future experiments

Here we examined the expression and function of *ftz* and *ftz-f1* in *Tribolium*, two genes important for pair-rule segmentation function in the long-germ insect *Drosophila*. We found that both *ftz* and *ftz-f1* from this short-germ beetle are expressed in stripes (Figure 5-1); this differs from *Drosophila,* where *ftz-f1* exhibits ubiquitous expression early in development (Yu et al., 1997). This difference in *ftz-f1* expression can be explained by one of two evolutionary scenarios, both of which involve a change in a *ftz-f1* cis-regulatory element during insect evolution: 1) *ftz-f1* striped expression was the ancestral state and at some point(s) in holometabolous insects there was change in a cis-regulatory element such that the gene was expressed earlier in development and ubiquitously at the blastoderm stage, or 2) maternal *ftz-f1* expression was ancestral and a striped pattern was acquired in a lineage leading to *Tribolium*. In order to discriminate between these two scenarios, *ftz-f1* expression will have to be examined in other insects, both those that are holometabolous (such as honeybees, wasps, moths, or mosquitoes) and non-holometabolous insects (such as hemipterans and the firebrat).

We also report here that knock-down of *ftz-f1* expression revealed roles for this gene in segmentation and cuticle development during *Tribolium* embryogenesis (Figures 5-2,5-5,5-6). This later role in cuticle development is also seen in *Drosophila*, where another Ftz-F1 isoform – β-Ftz-F1 – was found to be important for maturation of the cuticle denticles (Ruaud et al., 2010). Whereas *Drosophila* has two Ftz-F1 isoforms (α and β), we were only able to identify one isoform in *Tribolium* at all stages of embryogenesis (data not shown). Similar to results reported by other, we found that knock-down of *ftz* expression showed no embryonic phenotype (Figure 5-6; (Choe et al., 2006)). Preliminary results suggest that *ftz* expression may not be effectively knocked-down in embryos, even at high concentrations (2ug/uL; data not shown). For this reason, we are currently testing higher dsRNA concentrations (up to 5ug/uL) and also knocking

down expression using morpholinos, an alternative approach to RNAi. Additionally, knock-down of both *ftz* and *ftz-f1* resulted in a delay in embryo development (Figure 5-6). While nearly all embryos injected with buffer and *ftz* dsRNAs had developed to the point of having a fully extended germband, a larger percentage of embryos injected with both dsRNAs (58%) were found with incomplete germbands when compared to injection of buffer (7%), *Tc-ftz* alone (5%), or *Tc-ftz-f1* alone (~30%). While this suggests that *ftz* and *ftz-f1* may exhibit some redundancy in function during *Tribolium* segmentation, further experiments need to be done to examine this.

Interestingly, another nuclear receptor – E75A – is expressed in pair-rule stripes at the blastoderm stage and during germ band development in the hemimetabolous insect *Oncopeltus* (Erezyilmaz et al., 2009). Additionally, E75A RNAi produced phenotypes that ranged from fusion of T2 and T3 legs, to complete absence of these legs, and often fewer abdominal segments. These results are reminiscent of what we found with *Tribolium ftz-f1* - another nuclear receptor – and may suggest that other nuclear receptors may have similar expression patterns and functions in insects.

Conclusions and Future Directions

The results presented above have demonstrated that *ftz* is much more labile during evolution than previously imagined, both in sequence and expression. Outside of the Ftz homeodomain, there are only two currently identified functional motifs that can be traced through evolutionary time: the LXXLL motif required for segmentation function in *Drosophila*, and the YPWM motif required for homeotic function. When we set out to map these two sequence changes through insect phylogeny, we expected to find that the LXXLL motif was acquired at one point during arthropod evolution, and that the YPWM motif degenerated at some later point. We were surprised to find that the YPWM motif had degenerated in many lineages independently and that they retain varying levels of homeotic function in the context of *Drosophila* Ftz. We were also surprised to find Ftz sequences lacking both of these functional motifs. If *ftz* genes continue to be isolated from arthropods, it will be interesting to see whether our current model continues to be supported or if we can find an LXXLL motif from a species that is closely related to the holometabolous insects, or if we can find a holometabolous Ftz sequence lacking an LXXLL motif. It would also be interesting to see if Ftz sequences from organisms in specific lineages (Orthoptera, for example) are similar and if any new functional motifs or domains may be identified.

Examining the expression pattern of *ftz* in the brine shrimp *Artemia* revealed that *ftz* had an additional pattern of expression, as a weak *Hox* gene. This was very interesting because it allowed us to propose a model for how organisms in nature can tolerate the large diversity we observe in Ftz sequence. While it would be very interesting to explore *ftz* function in *Artemia*, it is unlikely that we would see a phenotype, since the level of expression is low, and also because knock-down of other *Hox* genes doesn't have a high success rate (Will Sewell, personal

communication). Preliminary results expressing *Artemia*-Ftz in *Drosophila* revealed weak homeotic potential (A.H., unpublished data), suggesting there are other motifs important for homeotic function. It would be useful to identify these regions and look for conservation outside *Artemia*. In addition to identifying a new expression pattern of *ftz* in a crustacean, we are also able to say that the first change that led *ftz* towards a role in segmentation was a change in expression (cis-regulatory change), because *Thermobia ftz* is expressed in stripes but its coding region lacks an LXXLL motif. Currently, the testing of *ftz* function in *Thermobia* is underway. It would also be interesting to trace this acquisition of stripes more thoroughly by examining expression in the collembolan *Folsomia*.

Despite this flexibility we see in Ftz sequence and early *ftz* expression, we hypothesized that *ftz* has been maintained in all arthropod genomes examined to date because of its later conserved role in the developing central nervous system. We found that the highly conserved DNA-binding homeodomain is required for Ftz function in the CNS. Interestingly, the homeodomain from a neighboring Hox protein (Antp) was also able to provide Ftz CNS function, suggesting it is the homeodomain itself that is required for this function. Ongoing experiments are looking to see whether Ftz homeodomains from *Tribolium* and *Artemia* can also substitute for the *Drosophila* Ftz homeodomain. It would also be interesting to examine whether the entire coding sequence from Antp could be substituted for that of Ftz under the control of *ftz* regulatory sequences. This would allow us to see if it was only a cis-regulatory change that was required for Ftz CNS function, or whether other regions of the Ftz protein are important for this function. Also, it would be interesting to see if Ftz has a cofactor in CNS function or whether it functions alone or works with the basal transcription machinery.

The *Tribolium* results presented above show that *ftz-f1* expression has also changed during insect evolution. In order to resolve whether *Tribolium ftz-f1* was first expressed in stripes or as a maternal gene, *ftz-f1* expression will have to be examined in other insects. It will be interesting to see if *ftz-f1* was expressed as a pair-rule gene before *ftz* pair-rule expression was established. It would also be interesting to isolate the stripe cis-regulatory elements for *ftz* and *ftz-f1* in *Tribolium* to gain a better understanding of how these two genes are regulated

The rapid evolution of *ftz* in insects provides a unique opportunity to study transcription factor evolution over a large scale of evolutionary time, as one can comprehensively track protein changes important for function (i.e. the LXXLL and YPWM motifs) to better understand and predict the potential role of a gene in an organism. Insects are an excellent system to look at evolutionary questions like this, for there are many resources available for analytical and functional studies. Our studies of *ftz* evolution have contributed to the field of evo-devo by demonstrating that there is a great deal of flexibility in regulatory genes important for development, which was unexpected, since the precise timing and expression of early developmental genes is crucial for embryo development. While this flexibility has not yet been studied in as much detail in other regulatory genes, it would be of great interest to see how prevailent this plasticity is in nature.

APPENDIX I: Ftz and Ftz-F1 sequences

I. Ftz sequences

Artemia salina ftz sequence, obtained by 5’ and 3’ RACE of 0-2d nauplii

```
                                              M   N   P   Y   F   L
TATTGTCACAGTTTGCTGATATGTTCAGTAGTTAAT     ATG AAT CCT TAC TTT CTA
 P   S   Q   F   P   Q   S   P   F   F   G   T   Q   N   T   D
CCG TCA CAA TTT CCG CAA AGC CCT TTT TTT GGA ACA CAA AAT ACT GAC
 V   N   N   D   G   S   K   F   F   Q   A   C   F   Q   P   R
GTA AAT AAT GAT GGA TCT AAG TTT TTT CAA GCA TGT TTT CAA CCA AGG
 Q   I   N   V   A   C   D   F   K   S   D   Y   D   A   Q   K
CAA ATC AAC GTT GCC TGT GAC TTT AAA TCG GAC TAT GAT GCG CAA AAA
 D   L   T   H   N   T   Y   V   E   R   S   E   N   P   Q   H
GAC CTA ACT CAT AAT ACT TAT GTG GAG AGA TCA GAA AAT CCA CAA CAC
 C   M   R   S   G   Y   Y   P   T   N   F   V   Q   F   S   T
TGC ATG AGA AGT GGT TAC TAC CCA ACA AAC TTT GTG CAG TTT TCT ACT
 P   G   F   V   P   Y   H   Q   M   Q   M   S   N   S   S   I
CCT GGA TTT GTA CCT TAT CAT CAA ATG CAG ATG TCA AAT TCT TCT ATA
 A   P   L   Q   G   I   T   I   P   M   P   G   Q   K   R   T
GCT CCA CTG CAA GGA ATA ACT ATA CCA ATG CCT GGT CAA AAG CGA ACA
 R   Q   T   Y   T   K   Y   Q   T   L   E   L   E   K   E   F
CGT CAA ACA TAT ACG AAG TAC CAA ACA CTT GAA CTC GAA AAG GAA TTC
 L   Y   N   R   Y   L   T   R   V   R   R   M   D   I   S   S
TTG TAT AAT CGT TAC TTA ACC CGA GTT CGA CGA ATG GAT ATA TCG TCT
 K   L   Q   L   T   E   R   Q   I   K   I   W   F   Q   N   R
AAA TTG CAA TTG ACA GAA AGA CAA ATC AAG ATT TGG TTT CAA AAT CGA
 R   M   K   A   K   K   E   N   K   N   E   T   N   F   R   S
AGA ATG AAA GCG AAG AAG GAG AAT AAG AAT GAA ACA AAT TTT AGA AGT
 S   G   Q   S   C   D   A   S   D   E   M   V   S   T   S   S
TCT GGT CAA TCT TGT GAT GCA AGT GAT GAA ATG GTA TCC ACT AGC TCA
 M   T   Q   *
ATG ACC CAA TAA CTTTTACTAGAATTAATAATTTTGTACCTAAATAAATATGTAAATTG
TTCAA
```

Folsomia candida ftz sequence, obtained by 5' and 3' RACE of 0-4d eggs

```
TTTCGAAGGGGTGGACACCACGTGAATGAAATTTTAATGGACATGTGTCCATTATATAATTGA
                                             M   V   T   S   P
GGTAAAATATAATTGGATTTTTAGCCACATCTGCCCCCACG   ATG GTG ACT TCG CCC
 S   S   N   S   S   I   S   P   L   R   D   V   K   S   E   K
TCG TCC AAC TCG TCC ATC TCA CCG CTG AGA GAT GTC AAA TCG GAG AAG
 N   M   S   P   D   G   E   K   E   D   V   G   S   T   R   I
AAT ATG AGT CCG GAT GGA GAA AAG GAG GAT GTG GGC AGT ACT CGG ATT
 E   Y   P   P   W   L   K   R   G   S   Y   G   L   K   N   T
GAA TAT CCG CCA TGG CTA AAA CGC GGG TCT TAT GGT TTG AAA AAC ACC
 T   S   P   R   S   P   S   S   E   D   N   I   S   P   S   S
ACA TCG CCA CGT TCG CCC TCT TCG GAG GAT AAT ATT TCG CCC TCG TCG
 S   S   K   R   T   R   Q   T   Y   T   R   C   Q   T   L   E
TCA TCA AAA CGA ACC CGG CAA ACG TAC ACG CGC TGC CAA ACG CTC GAG
 L   E   K   E   F   H   F   N   K   Y   L   T   R   R   R   R
TTG GAG AAG GAA TTT CAC TTT AAT AAA TAC TTA ACG CGT CGG AGG AGA
 L   D   L   A   K   M   L   T   L   S   E   R   Q   I   K   I
CTT GAT TTG GCG AAA ATG TTA ACT CTG AGT GAA CGC CAG ATC AAA ATT
 W   F   Q   N   R   R   M   K   A   K   K   E   V   K   G   H
TGG TTT CAA AAT AGG CGG ATG AAG GCC AAG AAG GAG GTC AAA GGT CAC
 V   V   A   S   D   L   V   Q   R   H   G   N   T   N   S   E
GTG GTT GCC AGT GAT CTC GTC CAA CGA CAT GGT AAC ACT AAT TCC GAA
 S   N   S   C   Y   G   E   G   T   S   S   W   *
AGT AAT TCA TGT TAT GGA GAA GGA ACT TCT TCG TGG TAG GGAAAAGTGAG
AAAAAAAAAAAA
```

Thermobia domestica ftz sequence, obtained by 5' and 3' RACE of 0-9d eggs

```
ANAACCTCTTTCGGGGGCAGCTGACATTCAATGGGTGAAAGGAATGAGTAAAGGGGGTGGAGA
GAGGGACAGTCCAGAGAGGGGGTGTTTTTCGCGAAATGGCCGCGTGGTAACCCCTGGTTGATG
GTTAGACAGACAGACGCCGTGACACATATCGAAGTATATCTATACATATTCGGAAGGACACAA
ATCATTTTCACCCTGAACCAACTAGGCCATTTTCAGCTATAGAGCCAACCAGCACGCACATAA
                                                         M   S
AATAGCAATAGAATCAGGGGGTTTTATAGTGTTATCCAAGTGGTCTAACTACAGT ATG AGC
 A   A   P   Y   F   S   N   G   S   G   V   T   T   N   C   W
GCG GCT CCT TAT TTT AGC AAT GGA TCT GGT GTT ACA ACG AAC TGT TGG
 G   S   T   N   G   G   L   S   S   S   H   E   Q   N   P   Y
GGA TCC ACA AAC GGT GGA CTA TCT TCG AGC CAT GAG CAG AAT CCG TAT
 Q   P   F   Y   F   T   H   P   A   S   S   S   T   K   Y   G
CAA CCG TTT TAC TTC ACA CAT CCT GCA TCT TCA AGT ACT AAG TAT GGT
 L   V   S   S   T   Y   S   S   E   H   H   L   P   V   L   T
CTT GTG TCT TCC ACA TAT TCT AGT GAA CAT CAT TTG CCA GTG CTA ACG
 G   T   P   T   S   H   H   P   F   V   P   R   Y   S   T   S
GGT ACC CCT ACA TCT CAC CAC CCT TTC GTA CCC AGA TAC TCA ACT TCG
 P   P   S   A   V   T   A   T   N   P   T   F   H   P   N   P
CCT CCG TCT GCA GTG ACA GCG ACT AAT CCT ACA TTC CAT CCA AAT CCA
 L   T   R   G   L   K   P   D   S   E   P   T   T   T   T   T
CTT ACA CGG GGA CTA AAA CCG GAT TCT GAA CCC ACG ACT ACC ACA ACA
 E   S   S   P   P   I   T   S   T   T   P   V   S   V   A   T
GAA TCG TCA CCA CCA ATA ACA TCA ACA ACG CCA GTT AGT GTC GCT ACA
 T   A   N   N   V   N   N   N   L   Q   P   D   S   F   F   S
ACC GCG AAC AAT GTT AAC AAC AAT CTT CAA CCT GAT TCG TTC TTT TCA
 S   A   R   T   N   D   H   S   P   P   S   S   V   S   Q   L
TCG GCT CGA ACC AAC GAC CAT TCA CCA CCG TCT TCA GTG TCT CAG CTG
 F   M   D   S   G   R   D   L   I   A   N   G   C   K   V   S
TTC ATG GAC AGT GGA CGC GAC CTG ATT GCT AAT GGA TGT AAA GTG TCA
 S   F   C   A   P   N   N   I   G   V   P   D   S   S   S   L
TCG TTT TGT GCT CCA AAT AAT ATT GGT GTT CCA GAT TCG TCA TCG CTT
 M   V   Q   Q   G   F   D   V   T   R   P   L   D   C   L   Q
ATG GTT CAA CAG GGA TTT GAT GTC ACC AGG CCC TTG GAC TGC CTG CAG
 Q   P   F   V   G   K   G   P   A   N   Y   F   P   W   M   K
CAA CCT TTC GTT GGA AAA GGA CCG GCA AAC TAT TTC CCA TGG ATG AAG
 S   Y   T   D   T   G   H   G   P   K   R   T   R   Q   T   Y
TCG TAC ACA GAC ACT GGC CAT GGT CCA AAA CGA ACA CGA CAA ACC TAC
 T   R   F   Q   T   L   E   L   E   K   E   F   H   F   N   K
ACC CGT TTC CAA ACA CTA GAA CTG GAG AAA GAG TTT CAT TTC AAC AAA
 Y   L   T   R   R   R   R   I   E   I   A   H   S   L   G   L
TAT TTG ACC CGC AGG CGA CGG ATA GAG ATT GCT CAC TCA CTT GGC CTG
 S   E   R   Q   I   K   I   W   F   Q   N   R   R   M   K   A
TCC GAA CGA CAG ATC AAA ATC TGG TTC CAG AAC AGG CGG ATG AAG GCG
 K   K   E   I   K   M   Q   P   Q   P   V   S   N   G   T   E
AAG AAA GAG ATC AAA ATG CAG CCG CAG CCG GTG AGT AAT GGG ACA GAA
```

```
 D   D   I   L   E   K   G   M   A   T   T   P   P   D   A   Q
GAT GAC ATC TTG GAG AAA GGT ATG GCG ACG ACA CCT CCC GAT GCA CAG
 V   F   D   K   D   V   I   K   M   Q   N   Q   I   H   I   P
GTA TTT GAT AAA GAC GTT ATT AAA ATG CAG AAT CAG ATA CAC ATT CCC
 F   A   G   I   K   P   E   N   L   Q   F   P   T   I   K   E
TTC GCT GGC ATT AAG CCA GAA AAT CTG CAA TTT CCA ACG ATA AAG GAA
 E   F   Q   H   Q   T   D   I   S   I   C   S   S   D   T   *
GAA TTT CAA CAC CAA ACT GAT ATC AGC ATA TGT AGT AGT GAC ACG TGA
```

GTAAACCGTACACAGTATTGTTCCTGAATTGTAACTTATGATCAAAATATTGTGGTAGCTATT
ACTTAATACAGCTTTAAAAAAAGAAAAAAAAAAAAAAAA

Pedetontus saltator sequence, obtained by RIGHT

```
ACCTAAGAATTTGGACAATAATTGTGTACCATGCACAGCCTAAGTATTTATTTTTTATTATAAT
ATAAAAATCGATGATATTCTGGAAGATTTGGGGCGTCTGGTAAACATTGCGTATATCCTANGTG
ATCAGAAAGAACTTATGCAATGCGATCTAACGGGAGAAAAAATACAGCAAGACCACAGTAATTA
TGGACATATATAGGGATCTATGCGAAACAGTTTGCAATTCATTGATCTAACACTCCTGCTCGAT
TTAATTAATAGACACAAACATAGGTAGTTAGTTTTTAGCAACTTCCTATTTCTCATTTGAATTC
ATCATTTTTTGTTTTGATCTGGCAATACCCTCATTCAAAAATTTAAAATTCCCTCGTTTTTGTT
TTGTTTTTGTGTTGCGTAGACGTTATTTATTCCCCTCTGGCCGGTTGTGTTAAAAAAAATGAAA
                                                      M   K   S
TCACGTGATCACACAGGCTGCCAACCTTACACTGGCTTGGCGCGCCTGTGCT ATG AAA AGC
 Q   A   V   G   R   H   I   Y   I   P   Q   M   S   V   F
CAA GCG GTC GGC CGC CAT ATA TAC ATT CCA CAA ATG TCC GTC TTT T GT
TTTTCTTAGAATTTGTACTCTAGAGATCGAAACTGAACTGTTTTAGAAGAAGAAGAGTGTCAAA
ACGAATTGTGAGTAAGTGTGTTTGTTTGTTTCCGTTTCTTTTTGGTCTTGTGCAACGAAAAT
              Y   S   S   P   G   P   K   R   T   R   Q   T   Y
GACGACAAAACAG AT TCC AGC CCT GGG CCA AAG AGA ACT CGC CAA ACG TAC
 T   R   V   Q   T   L   E   L   E   K   E   F   H   F   N   R
ACC CGT GTG CAG ACA CTG GAA TTG GAG AAG GAA TTT CAT TTC AAT CGA
 Y   L   T   R   R   R   R   I   E   I   A   H   A   L   G   L
TAT CTG ACG AGA AGA CGA CGA ATT GAA ATC GCT CAT GCC CTG GGC CTC
 T   E   R   Q   I   K   I   W   F   Q   N   R   R   M   K   A
ACA GAA CGG CAA ATC AAA ATA TGG TTC CAA AAT CGA AGA ATG AAA GCC
 K   K   E   S   K   L   Q   E   V   R   E   H   E   Y   V   G
AAA AAG GAA AGC AAA CTC CAG GAA GTT CGA GAA CAT GAA TAT GTG GGG
 Q   D   S   T   V   S   E   T   P   V   S   A   T   P   S   T
CAA GAC AGT ACA GTA AGT GAA ACA CCC GTT TCT GCA ACA CCT TCG ACT
 E   S   I   K   I   L   E   H   S   I   P   P   I   K   V   E
GAA TCT ATT AAG ATA CTA GAA CAT AGT ATT CCT CCT ATT AAG GTA GAA
 A   G   M   L   I   N   S   *
GCA GGT ATG TTA ATT AAC AGT TAG TATTATGGCAAATGCATATTAGTATTAAGAAC
TGAAATCAATAAAATCGAGAAGCACCGATTTTAACTTCCTATAAAATATAATTACGCATTCAAA
GCCAAGCTCAAAATATTTTGACAAATGGAAACTGTTGTGACAGTGGTATATATATTTTCTTTTC
AAATGTATTTCTTACCTGCCATACATTCCTGTTGTGAAAGCCAACACTTTTAGTGTATTGTTAG
TTTATGATTTGCTCAGGTGGGCTAAGAAAGGGTTTAACATTTTATGGCTCTCGTCTACCTGACT
TACAAAATAACAGAAAATATTTTTGCTGGGAATT
```

Forficula aricularia sequence, obtained by RIGHT

```
GAATTCTGTGTACTCCTGTGTACTTCCATGTTACCATTAATTAAACAGATAATAAAAATAGGCA
TAAATCGAAGAATACTGTCTTCTGATATATATCCATATTTTGTTTATTTCAGATTACAATTTTC
CATCTAGGATTAGTGAGCTAAGCATCCCTCACAGGAGAAAATGGCTAATTAGTTGATGAAAAAG
TAGTTTTAGTAAATGATCAATATTATAAAAACATTTAAAATTAGCCAAAAATAAATCTTATTAC
TAAATTTTCTAAATTAAAATAAACGCGGTGTATCCTTGCTTTGGCTGAATAACTCCTTCAACCC
ACTACAAATTTACCACAGNCACGATCTGTGATCTTCAACGAATAATAGTATAAAGCTCATTTTA
ACTGCGTAATATTTGTAATTTTAATAGCGAAAAATTACCGAAAATCTTCATCAGGGCACCGAGA
TATACAGGGTAGGGGTATAGTAATTTTATGTGAATAGCCATATTCTCTTATTTCTGATAGTTAC
                                        M   K   N   S   G   K
CATAGTATTACTGAATTAATAGACACGATAACTCAAAA ATG AAA AAT AGT GGA AAG
 F   I   L   I   A   Y   L   M   F   A   I   S   I   P   Q   R
TTC ATT TTA ATT GCG TAC CTA ATG TTT GCA ATT TCA ATA CCT CAA CGG
 K
AAA A GTGATTGGATTACCAAAAATCCGAGACACCGGGCACCGAGATATGCTTCAACCCTAGA
CTAAAACAAATATTTCACTTGAATATTTAATAATTAAATATACATACACCGAGGAGTGAACCGC
GACCTAATATTGTGTACACCTACTTTCACTTTCATTTCCAAGTAGAAGGATCTACAAATGTAGG
GACTTTATATAATGAATTCCACATAATTCCAAAGGCGTGAACACCTAATTCCAAAAGCGGTTCA
CCCCTCGATATACACAGACACACACA
                                                             K
TGGTAAAGAGTATTTTTATGATTATGTTTTTATGTTTTTATTTCTAAATTGATTTTCAG AA
 T   T   S   G   S   K   R   S   R   Q   T   Y   S   R   Y   Q
ACG ACA TCA GGC TCG AAA CGC AGC CGG CAA ACC TAC TCC CGT TAC CAA
 T   L   E   L   E   K   E   F   H   F   N   K   Y   L   T   R
ACA CTG GAA CTC GAG AAA GAG TTT CAC TTC AAC AAG TAC TTG ACC AGG
 R   R   R   I   E   I   A   N   A   L   H   L   T   E   R   Q
CGT CGG CGA ATT GAA ATC GCT AAT GCA TTA CAC TTA ACA GAA CGC CAA
 I   K   I   W   F   Q   N   R   R   M   K   E   K   K   T   R
ATC AAA ATC TGG TTC CAA AAT CGA CGA ATG AAG GAG AAA AAG ACG CGA
 S   T   E   A   D   M   N   S
TCC ACA GAG GCC GAT ATG AAT TC
```

Callosobruchus maculatus ftz sequence, obtained by degenerate PCR (gDNA), followed by 5' and 3' RACE of 0-2d eggs

```
                                       M   S   A   S   A   Q
AGTGTGCAATTTTTCCACGCGTCGAACGGTGACGAAGT  ATG AGT GCC TCC GCG CAA
 F   G   S   C   E   Y   Y   N   Q   Y   G   Y   N   F   Y   G
TTC GGG TCT TGC GAG TAC TAC AAT CAG TAC GGG TAC AAT TTC TAT GGA
 D   A   Q   R   M   P   G   Y   G   N   M   G   Y   Q   Y   H
GAC GCG CAA AGG ATG CCT GGG TAC GGG AAT ATG GGA TAC CAG TAT CAT
 N   A   Y   P   Y   G   G   Y   T   E   K   R   D   A   F   A
AAT GCC TAC CCC TAC GGT GGG TAC ACC GAG AAA AGG GAT GCG TTC GCT
 E   A   Y   G   D   V   K   E   E   P   S   A   C   R   F   D
GAA GCG TAC GGC GAC GTG AAA GAA GAA CCG TCC GCG TGC AGG TTC GAC
 A   H   A   N   Q   G   Y   S   N   P   V   C   E   P   D   D
GCG CAC GCA AAT CAA GGG TAT TCA AAT CCG GTG TGC GAG CCG GAC GAT
 S   I   S   R   R   P   V   I   N   Q   A   Y   Q   P   T   G
TCT ATT AGT CGC CGG CCG GTC ATC AAT CAG GCT TAC CAA CCG ACC GGC
 Y   G   S   L   A   T   S   L   S   P   P   R   A   A   N   E
TAC GGT TCA TTA GCG ACC AGT TTG TCT CCA CCT AGA GCT GCT AAC GAG
 D   D   S   T   T   G   S   S   S   A   E   K   T   G   K   M
GAT GAT TCG ACG ACG GGG TCT TCT TCG GCG GAA AAA ACT GGA AAG ATG
 E   E   D   S   S   A   L   R   A   L   L   S   K   P   G   G
GAG GAG GAC TCG TCC GCG CTG AGA GCG CTG CTC AGC AAG CCC GGA GGC
 E   K   I   T   Y   D   Y   T   E   L   R   K   T   H   S   P
GAG AAG ATT ACC TAC GAT TAC ACG GAG CTG AGG AAG ACA CAC TCG CCT
 A   D   Y   E   V   H   A   S   N   M   S   L   D   C   D   E
GCG GAT TAC GAG GTG CAC GCG AGT AAT ATG AGT TTG GAC TGC GAT GAA
 D   L   S   S   C   G   K   E   K   T   S   E   A   A   E   D
GAT TTA TCT TCT TGC GGC AAG GAG AAG ACG TCC GAA GCA GCT GAA GAT
 A   L   T   A   A   Q   N   N   F   Y   P   W   M   R   S   S
GCC CTG ACG GCG GCA CAG AAC AAC TTC TAC CCG TGG ATG AGG AGC TCG
 N   D   H   T   A   K   G   N   K   R   T   R   Q   T   Y   T
AAC GAT CAT ACA GCA AAG GGC AAT AAA AGG ACC CGT CAG ACC TAC ACC
 R   Y   Q   T   L   E   L   E   K   E   F   H   F   N   K   Y
AGA TAC CAG ACA CTC GAG CTA GAA AAG GAG TTT CAT TTT AAC AAG TAT
 L   T   R   R   R   R   I   E   I   A   H   T   L   C   L   T
CTC ACC AGA AGG AGA AGG ATA GAG ATT GCG CAT ACT CTC TGC CTC ACT
 E   R   Q   I   K   I   W   F   Q   N   R   R   M   K   A   K
GAA CGT CAA ATC AAG ATC TGG TTC CAG AAC CGA AGG ATG AAG GCT AAA
 K   G   D   K   L   A   V   P   A   Q   Q   V   D   F   P   T
AAG GGT GAC AAA CTG GCA GTC CCC GCT CAA CAA GTA GAC TTC CCC ACC
 I   Q   D   V   N   M   N   Q   H   L   Y   P   A   M   S   P
ATT CAG GAC GTG AAC ATG AAC CAG CAC CTA TAC CCC GCA ATG AGT CCT
 A   T   P   A   S   Y   Y   S   C   G   S   S   E   G   L   P
GCA ACT CCG GCC AGC TAC TAC TCC TGC GGG AGT TCT GAA GGA CTC CCG
 E   T   S   T   L   D   A   N   R   N   S   F   D   A   N   R
GAG ACT TCA ACG CTG GAT GCG AAC AGA AAC TCC TTT GAC GCT AAC AGA
```

```
 N   S   F   D   A   E   V   A   R   P   L   T   A   L   K   N
AAC TCC TTT GAT GCT GAA GTT GCG AGG CCT CTA ACA GCC CTC AAA AAT
 I   P   G   P   P   L   S   P   P   Q   *
ATT CCA GGA CCT CCT CTA TCG CCA CCT CAG TGA CAAATTCGGTGAGTTGTGA
TTTGTCCAAAGAGCGCATATCAAAACTGTTGATTGTATATAGTTTGAGAGTGTGGTTTTACCT
GGGAGTGTTGGTATTGCAAAGTGTGGAGGAATTATTGGAAAATTAGTTAATTAGATCTATCAG
ACAGTATCTAAAGGTTTTTGTCCATATCTGTTTAAAGATATCTGAAAGGATGTTACTTGAAAT
GGGACAAAAACTTCTCCAATCGAAGGTTCAGGTAAAACTAATGCTTATTATGTATTTACAAGA
ATTTATGAAACTCTTGAAATGATATACCAAATTAATGACATTTTACAAATTTTGACGTTCATT
TTATCTTCTTGTTACAATATGAGTCCAAGTACAAAACACTGTTTTCATAGGCTTTGAAAAACT
GCTTTGGAGATAACTTAAACGACCTTATCAACTTCTACCAGATATTTCATCTATCTAATTTCA
TTTCATCATCATCATTTTGTTTGAGTATTTTATAGCTTCATTTCTTCCACACCATTCATGTTT
ACAATTTGACAGATACAAAAGTAACTTAAGGAAACGATGTTGGTATATCACCCATACGAAAAC
TGGAATCGCCAAAGAAACAAAATTTATCATTTAGATGTTTTCACCTGGTAAAAATACTTAAAT
TAGATAACATTGAACTATATAAATAAAATCGTTAAATTAGGATCCACCCGCAACGGTACTTTT
TGATTCATACATTTAACGTATTGTCTTCTTAACCTCTTATAAATCTTGTCAGTATATAAGGTG
TTATAATCAGGATGACCAATTTTAGTATCGCCATATCATTAGACAGTTATCTTATCTAAAAAC
GAAAAACATTTACTTAAGTTTTGATTATAGCACTCCCAGGTATCGTTTAGTGCCAAACTTAGA
AATTTAGTTACTACTAATCAATAAATATATCTATAAATACCTATTTTTATATAACTTTATAGT
CATTAGTTACTAAGCTTAAAAATAAAGTACTATTTATTTACCTAAATCCTATATTTCCTAACT
ATTCAAGAATAAAGCTTGTATATAGACAAAAAAAAAAAA
```

Dermestes maculatus ftz sequence, obtained by 5’ and 3’ RACE of 0-3d eggs, RIGHT

```
ACACCGATTTTAGTTACACTTCATATTTGACAGTGACACAGGCGACTGCAAATTTTTAACAGT
                     M   S   A   S   T   G   Y   N   Y   D   Y
GTGCTAGAGTGACAAA    ATG AGT GCC TCA ACT GGA TAT AAT TAT GAC TAC
 W   S   Q   H   P   T   T   Y   Q   Q   Y   R   S   N   I   P
TGG AGT CAA CAT CCA ACT ACA TAT CAA CAG TAT CGT AGC AAT ATA CCG
 L   S   S   S   E   R   P   L   T   N   Y   N   V   S   P   A
TTA TCT TCA TCC GAA AGA CCG CTT ACT AAT TAC AAC GTA TCA CCA GCA
 S   L   N   Y   N   N   E   I   D   N   Y   R   S   V   T   N
AGT CTG AAT TAC AAC AAC GAA ATA GAT AAC TAC CGC AGT GTA ACA AAT
 L   N   G   F   N   P   Y   G   Y   M   N   E   G   L   L   K
TTG AAT GGT TTT AAT CCC TAT GGA TAT ATG AAT GAA GGA CTA TTA AAA
 T   V   N   N   F   D   K   L   R   S   T   V   N   D   Y   G
ACA GTG AAT AAT TTT GAT AAA CTA AGA AGT ACA GTG AAT GAT TAC GGT
 I   S   A   D   I   I   S   N   N   E   P   I   I   N   P   T
ATT AGT GCA GAT ATC ATA AGT AAT AAT GAA CCT ATT ATA AAC CCT ACT
 H   S   E   N   Y   N   I   Q   N   T   F   S   P   N   F   Q
CAT TCC GAA AAC TAT AAC ATT CAA AAT ACC TTT TCT CCT AAC TTT CAA
 V   H   N   P   S   G   G   L   N   D   A   T   N   I   S   P
GTT CAC AAT CCG AGT GGA GGT TTA AAT GAT GCC ACT AAT ATA TCA CCG
 K   M   S   T   D   T   T   I   S   P   K   K   E   I   E   D
AAA ATG TCA ACT GAT ACT ACT ATA TCC CCA AAG AAA GAG ATT GAA GAT
 D   S   P   A   L   R   A   L   L   T   K   P   H   I   R   K
GAT TCA CCT GCA CTT AGA GCG TTA TTA ACT AAA CCA CAC ATA AGA AAA
 P   Y   D   F   Y   E   T   N   K   P   I   D   Y   Q   N   Q
CCT TAT GAT TTC TAT GAA ACA AAC AAA CCT ATT GAT TAT CAA AAC CAA
 F   Y   S   H   V   N   E   F   A   C   N   K   N   I   K   T
TTT TAT TCA CAT GTG AAT GAA TTT GCA TGT AAC AAG AAT ATT AAA ACT
 T   P   T   P   A   V   I   P   Q   D   E   I   N   S   S   E
ACA CCA ACA CCG GCA GTT ATA CCT CAA GAT GAA ATA AAT TCC TCC GAA
 N   I   S   N   T   N   S   V   T   P   T   N   N   I   Y   P
AAT ATA TCT AAT ACG AAT AGT GTG ACA CCA ACT AAT AAT ATA TAT CCT
 W   M   K   A   N
TGG ATG AAA GCA AAT G GTAGTAAGGAAGTATTAAATAAAATTTGTAATTGTATTTTA
                     A   E   A   T   N   H   G   G   K   R   T
ACATTTTTATTTTTTACAG  CC GAA GCA ACA AAT CAT GGT GGT AAA AGG ACA
 R   Q   T   Y   T   R   Y   Q   T   L   E   L   E   K   E   F
AGA CAA ACT TAT ACC AGA TAC CAA ACT CTA GAA CTA GAG AAA GAA TTC
 H   F   N   K   Y   L   T   R   R   R   R   I   E   I   A   H
CAT TTC AAT AAA TAT TTA ACT CGT CGG AGA AGG ATA GAG ATT GCA CAC
 A   L   C   L   S   E   R   Q   I   K   I   W   F   Q   N   R
GCG CTG TGT TTA TCA GAA CGC CAA ATA AAA ATA TGG TTT CAA AAT AGA
 R   M   K   A   K   K   D   N   K   F   T   L   Q   E   F   T
AGA ATG AAA GCA AAA AAA GAT AAC AAA TTC ACA TTA CAA GAA TTC ACT

 E   D   I   N   M   N   Q   N   Q   L   I   A   N   S   P   C
```

GAA GAC ATC AAC ATG AAT CAG AAT CAG TTA ATT GCT AAT TCG CCC TGT
 A   N   N   A   L   Y   M   S   N   V   S   P   Q   E   T   S
GCA AAC AAT GCC TTA TAT ATG AGC AAT GTA TCA CCA CAG GAA ACA TCA
 T   G   G   Q   E   P   N   A   L   N   E   G   I   V   E   A
ACA GGA GGT CAA GAA CCA AAC GCG CTC AAC GAA GGC ATC GTC GAA GCA
 L   T   Q   F   R   N   I   S   G   P   P   C   I   S   *
TTA ACG CAG TTC AGA AAT ATA TCA GGA CCA CCT TGC ATA TCT TAA TTC
AATTTTACATTTAACTTATTTTTACGCGAAGAGCTCT

Ftz-F1 sequences

Artemia salina, obtained by degenerate PCR and RACE using RNA from 0-2d nauplii

```
                      M   S   Q   T   S   E   H   R   K   Y   F
TATTTCATGGACCATAACA ATG AGT CAA ACT TCT GAG CAT AGG AAA TAT TTT
 N   E   E   I   F   E   K   D   L   V   L   D   L   S   S   E
AAT GAA GAG ATT TTT GAA AAA GAC CTT GTT TTG GAT TTG TCA TCA GAA
 N   S   K   K   V   F   N   L   I   S   L   E   L   Q   S   S
AAT TCC AAA AAA GTT TTC AAC TTG ATA TCT CTT GAG CTG CAG AGC AGT
 E   Q   D   F   Q   T   G   L   S   E   G   S   L   E   F   E
GAG CAG GAT TTT CAG ACT GGA TTA TCG GAA GGT TCC CTG GAG TTC GAA
 A   T   P   E   T   A   T   R   P   M   S   V   D   S   G   G
GCA ACA CCG GAA ACT GCA ACC AGA CCT ATG AGC GTT GAC AGT GGC GGG
 D   L   R   T   A   D   P   P   D   I   K   E   G   I   Q   E
GAT TTA CGT ACT GCT GAT CCC CCA GAT ATC AAG GAA GGA ATT CAA GAG
 L   C   P   V   C   G   D   K   V   S   G   Y   H   Y   G   L
TTA TGC CCA GTT TGT GGT GAT AAA GTT TCT GGA TAT CAC TAT GGT CTC
 L   T   C   E   S   C   K   G   F   F   K   R   T   V   Q   N
TTG ACA TGT GAA TCG TGC AAG GGA TTT TTC AAG CGA ACA GTG CAG AAC
 K   K   V   Y   T   C   V   A   D   R   S   C   H   I   D   K
AAA AAG GTA TAT ACG TGT GTG GCC GAT AGA AGT TGT CAT ATA GAT AAA
 S   Q   R   K   R   C   P   F   C   R   F   Q   K   C   L   E
AGC CAG AGA AAG CGC TGT CCT TTT TGC CGG TTC CAA AAG TGC CTG GAA
 V   G   M   K   L   E   A   V   R   A   D   R   M   R   G   G
GTT GGG ATG AAA TTA GAA GCC GTT CGC GCC GAT AGA ATG CGA GGT GGA
 R   N   K   F   G   P   M   Y   K   R   D   R   A   R   K   M
AGG AAC AAA TTC GGC CCC ATG TAC AAG AGA GAC CGT GCT AGA AAA ATG
 Q   I   V   R   E   R   Q   F   C   S   P   G   E   T   P   T
CAA ATT GTC CGA GAG AGA CAG TTT TGT TCA CCT GGT GAA ACC CCG ACA
 P   P   A   N   G   V   I   Y   P   G   G   H   Q   I   T   G
CCC CCT GCG AAC GGT GTA ATT TAT CCA GGG GGT CAT CAA ATC ACC GGT
 A   E   I   A   L   T   Y   S   A   A   T   G   T   F   Q   E
GCT GAA ATT GCA TTA ACA TAT TCT GCA GCG ACA GGA ACG TTT CAA GAA
 T   V   K   H   D   I   Q   I   Q   Q   V   S   S   L   T   S
ACA GTG AAA CAT GAC ATT CAA ATC CAA CAG GTT TCT AGC TTA ACT TCC
 S   P   D   S   S   P   N   S   Q   I   N   T   S   L   G   F
TCC CCT GAT TCA AGT CCT AAC TCT CAG ATT AAC ACA TCT CTT GGT TTT
 G   N   L   Q   S   E   K   L   W   T   I   S   S   N   G   V
GGA AAC TTA CAG TCT GAA AAA TTA TGG ACA ATC TCC TCT AAT GGA GTA
 S   I   P   Q   A   M   S   P   K   A   Y   Q   F   E   S   L
AGT ATC CCC CAA GCA ATG TCA CCG AAG GCA TAC CAG TTT GAA TCC TTG
 L   N   S   E   S   S   S   L   N   N   T   V   S   S   G   K
TTA AAC AGT GAA TCG TCG TCT CTC AAT AAT ACA GTG TCC AGT GGT AAA
 M   P   P   I   L   S   D   L   V   Q   S   L   D   D   K   E
ATG CCA CCA ATA TTA AGT GAT CTA GTT CAA AGC TTG GAT GAT AAA GAG
```

```
 W   Q   S   A   L   F   G   L   L   Q   N   Q   T   Y   N   Q
TGG CAG TCA GCA CTC TTT GGA TTA CTT CAA AAT CAA ACA TAT AAT CAA
 C   E   V   D   L   F   E   L   L   C   K   V   L   D   Q   S
TGT GAG GTC GAT TTA TTT GAG CTT CTG TGT AAA GTT TTA GAT CAA AGT
 L   F   T   Q   V   D   W   A   R   N   S   A   F   F   K   E
CTG TTT ACC CAA GTC GAC TGG GCC CGT AAC TCT GCT TTC TTT AAA GAA
 L   K   V   D   D   Q   M   K   L   L   Q   N   S   W   S   D
CTA AAG GTA GAT GAT CAA ATG AAG CTA CTA CAA AAC TCC TGG TCT GAC
 L   L   V   L   D   H   I   H   Q   R   M   H   N   N   L   P
CTT CTG GTT CTA GAC CAT ATT CAT CAG AGG ATG CAC AAT AAT TTA CCA
 E   E   T   Q   L   A   N   G   Q   K   F   D   L   L   S   L
GAA GAA ACT CAA CTA GCC AAT GGA CAA AAG TTC GAT CTT CTT TCG CTG
 A   I   L   G   S   Q   S   L   A   E   P   L   F   A   V   I
GCA ATA CTA GGA TCA CAG TCC CTT GCA GAG CCG CTT TTT GCT GTG ATA
 S   K   L   N   D   L   R   F   D   L   H   D   Y   V   C   I
TCC AAG TTG AAC GAC CTC CGG TTC GAC TTG CAT GAT TAT GTC TGT ATT
 K   F   L   I   L   L   N   P   D   V   R   G   I   V   N   R
AAG TTC TTG ATC CTT TTA AAT CCA GAT GTG CGT GGT ATT GTG AAT CGA
 R   L   V   S   D   A   H   D   Q   I   R   Q   A   L   F   D
CGA CTT GTC TCA GAT GCG CAT GAT CAA ATT CGT CAA GCG TTG TTT GAT
 F   C   V   N   C   H   S   N   T   V   D   K   F   S   K   L
TTT TGT GTT AAT TGT CAC TCA AAC ACA GTG GAC AAA TTC AGT AAA CTA
 L   G   L   I   P   D   L   R   A   I   S   S   R   G   E   D
CTG GGC CTG ATA CCT GAT TTA AGG GCA ATA TCG TCT AGA GGA GAG GAT
 F   L   Y   L   K   H   L   N   G   C   A   P   T   Q   T   L
TTT TTG TAT TTG AAG CAT CTA AAC GGT TGC GCC CCT ACT CAG ACT CTG
 L   M   E   M   L   H   A   K   R   R   *
CTA ATG GAA ATG TTG CAT GCT AAA AGA AGA TGA   AGAAGAATAAGTTCTAAG
GTATTGTAAGCGTTTTCATTTCTTTTTCTTACTCTAAGAGTTATTCAACGTCTATTGAATTTTT
AATCACCGATCTAATACTTTTCTTTTCCTTTATTTGCATGCTTATATTAAAAAAAAAAAA
```

Folsomia candida, obtained by degenerate PCR and RACE using RNA from 0-4d embryos

```
TATTCTCCGACCTTGACACACATTTATATTTAGAAGTTGAAAAGTATTGCCAAATCACATAAAA
CATTGGCAAATCTAAAAATCCTACATCAGGAACATAGTTTACTTACTACTACGAAAAAATCAAT
CACTGGTTGGAACTAGTAATTACCTTCCCTACTTGCGGTGGTCGGGTTGGGTGATCCTTGACTA
GATATGCATAAGCTCATTGGTCTATAGAACTGTATGCAGTTGAATACTTAAAAACGAATAATTT
CAACCAACGAGTACCGCATAACACGACAACTAAATACCTGAGACCAATTTCTAATTACAATAAA
                         M   H   E   E   D   E   A   S   T   T
TTAAAAAATAAGCGTTCGTCAAA ATG CAC GAG GAG GAC GAA GCC AGT ACG ACA
 S   V   E   K   V   V   I   I   E   I   G   P   E   Q   A   E
TCG GTG GAG AAA GTT GTT ATC ATT GAA ATT GGC CCA GAA CAA GCC GAG
 G   S   S   T   S   E   S   H   H   D   N   L   D   N   S   N
GGT AGC TCC ACC TCA GAA TCT CAT CAT GAC AAT TTG GAC AAC TCC AAC
 S   T   T   A   E   S   P   P   F   T   G   G   N   Q   S   S
AGC ACG ACG GCC GAG TCT CCT CCC TTC ACG GGG GGT AAC CAG TCA TCG
 G   A   T   P   S   G   L   E   Y   T   T   A   I   C   Q   D
GGA GCC ACC CCC AGT GGG CTG GAG TAC ACG ACG GCT ATT TGT CAG GAC
 Q   P   D   T   K   E   G   I   E   E   L   C   P   V   C   G
CAG CCG GAT ACA AAG GAG GGG ATC GAA GAG TTG TGT CCC GTC TGT GGG
 D   K   V   S   G   Y   H   Y   G   L   L   T   C   E   S   C
GAC AAA GTG TCC GGC TAC CAC TAC GGC TTG CTC ACG TGC GAG TCC TGC
 K   G   F   F   K   R   T   V   Q   N   K   K   V   Y   T   C
AAG GGT TTT TTT AAA CGC ACT GTC CAG AAC AAG AAA GTC TAC ACC TGC
 V   A   D   R   N   C   H   I   D   K   T   Q   R   K   R   C
GTC GCC GAC AGA AAC TGC CAC ATT GAC AAA ACA CAA CGA AAG AGG TGT
 P   Y   C   R   F   Q   K   T   L   A   V   G   M   K   L   E
CCA TAT TGC AGA TTC CAG AAA ACT TTG GCT GTT GGT ATG AAA CTG GAA
 A   V   R   A   D   R   M   R   G   G   R   N   K   F   G   P
GCC GTC CGA GCT GAC CGA ATG CGG GGT GGA CGA AAC AAA TTC GGA CCG
 M   Y   K   R   D   R   A   R   K   L   Q   M   M   R   Q   R
ATG TAC AAG AGG GAC AGA GCC AGA AAA TTA CAA ATG ATG CGT CAA CGA
 Q   L   A   I   Q   Q   A   R   Q   Q   G   L   A   I   V   S
CAA TTA GCT ATC CAA CAA GCT CGT CAA CAA GGT TTA GCC ATC GTA TCT
 D   T   L   P   L   S   Y   S   N   G   S   P   Y   G   Q   G
GAC ACC CTC CCG TTG TCC TAC AGC AAT GGA TCT CCT TAC GGA CAA GGC
 V   T   I   K   Q   E   I   Q   I   P   Q   V   S   S   L   T
GTC ACG ATT AAG CAA GAA ATT CAA ATA CCT CAG GTG TCT TCA TTG ACC
 S   S   P   D   S   S   P   S   P   L   A   T   L   G   M   V
TCA TCC CCC GAC TCC TCG CCA TCT CCT CTC GCC ACC CTG GGC ATG GTC
 N   G   S   G   Q   S   S   I   N   L   A   D   P   N   S   G
AAC GGC TCG GGA CAA TCA AGT ATA AAC CTG GCA GAT CCC AAC TCT GGG
 P   L   P   P   R   P   Q   N   S   S   S   T   K   H   F   V
CCA CTT CCT CCA CGT CCC CAA AAT TCG TCA TCT ACA AAG CAT TTT GTG
 Y   D   S   N   N   P   S   A   N   N   S   S   I   G   N   Q
TAT GAT TCC AAT AAC CCG TCT GCC AAT AAC TCA TCA ATT GGA AAT CAA
 H   A   G   N   D   P   S   Q   G   H   I   V   S   V   D   H
CAT GCT GGA AAT GAC CCA TCC CAG GGC CAC ATT GTG AGC GTT GAC CAC
```

N S S G Q N S T S S S S V G S K
AAT TCT TCG GGA CAA AAT TCA ACA TCT AGT TCT AGC GTC GGA AGC AAA
M P P L I R D L L S S L D D K E
ATG CCA CCC CTG ATT AGA GAC CTT TTA TCG TCA TTA GAT GAC AAG GAG
W Q H S L F N L L Q N Q T Y N Q
TGG CAA CAT TCC TTG TTC AAT CTT TTG CAG AAT CAA ACT TAT AAC CAA
C E V D L F E L L C K V L D Q N
TGC GAG GTC GAT TTG TTT GAG CTG CTC TGT AAA GTT CTA GAT CAA AAC
L F A Q V D W A R N S I F F K D
CTC TTT GCA CAG GTG GAC TGG GCA AGA AAC TCC ATC TTC TTC AAG GAC
L K V D D Q M K L L Q Y S W S D
CTC AAG GTG GAC GAT CAA ATG AAG CTA CTT CAA TAT TCT TGG TCG GAT
M L V L D H I H H R V H N H L P
ATG CTC GTT CTC GAC CAC ATT CAC CAC CGT GTA CAC AAT CAT TTA CCA
D D A P L P N G Q K F D L L S L
GAC GAT GCA CCC CTT CCC AAC GGA CAA AAG TTT GAT CTG CTG TCC CTC
A L L G V P A S I D R F N E V T
GCC CTA TTA GGC GTC CCT GCC TCC ATC GAT CGC TTC AAT GAG GTC ACC
L K L Q E I K F D Q A D Y I C L
TTA AAG CTC CAG GAA ATT AAG TTT GAT CAA GCG GAT TAC ATT TGC TTA
K F L L L L N P D V K S L M S R
AAG TTC CTC CTG CTT CTT AAT CCA GAC GTC AAG TCG TTG ATG AGT AGG
K H V Q E T H D H V Q Q S L L T
AAG CAT GTT CAG GAA ACT CAT GAC CAT GTG CAG CAG TCG CTT CTC ACC
Y C I N C Y P Q V Q E K F T K L
TAT TGC ATA AAC TGC TAC CCA CAA GTT CAA GAA AAG TTC ACC AAG CTT
L T L L P D I R Q V A S R G E D
CTT ACT CTA CTG CCA GAT ATA CGG CAA GTA GCA TCA AGG GGT GAA GAT
Y L Y F K H I N G G A P T Q T L
TAC CTG TAC TTT AAG CAC ATT AAC GGA GGA GCT CCA ACC CAA ACC CTT
L M E M L H A K R K *
CTT ATG GAA ATG CTA CAC GCG AAA AGA AAA TAA GCTTCGTCACATAGTCTC
CAAAAAGAAAGCTACTACTTTTATTTGCTGACTTATGTAGTTGTATTGGATAATAAAGAAGCTC
TAGAGATACTCACTATATATGTAACTATCGATGACCTGTAAATGAATGATGAAACTGTAAATGG
TGATGAAATGAAGAAAAAAAAAAAA

Thermobia domestica, obtained by degenerate PCR and RACE using RNA from 0-9d embryos

```
CATTTTCAGCATCCAATTTTTTTTTTCACTACGTAACGAAGAAGATTTTAGTGAAACCTTGGCA
TAGACGAAAAGTTCAGGGCTTTCTGAAGCTCTGGGACTCCGGGGGTGCCTCAACATTATTGTTC
                  M   H   E   E   A   T   S   M   S   V   P   N
GATAACGCAAAAAAC ATG CAT GAA GAG GCC ACA AGC ATG AGC GTT CCA AAC
 T   A   A   A   T   C   T   T   T   Q   P   T   D   T   E   L
ACA GCT GCA GCA ACT TGC ACC ACT ACA CAG CCC ACG GAC ACA GAA CTA
 Q   V   S   Y   S   S   G   T   A   G   S   S   G   M   E   Y
CAA GTG TCA TAC TCG TCT GGC ACT GCA GGG AGT TCC GGG ATG GAG TAT
 T   G   G   L   P   S   Q   D   L   P   D   T   K   E   G   I
ACC GGA GGC CTG CCG TCC CAG GAC CTT CCA GAC ACC AAG GAA GGC ATC
 E   E   L   C   P   V   C   G   D   K   V   S   G   Y   H   Y
GAA GAA CTG TGT CCA GTG TGT GGA GAC AAA GTG TCC GGA TAT CAC TAC
 G   L   L   T   C   E   S   C   K   G   F   F   K   R   T   V
GGT CTC CTG ACG TGT GAA TCT TGC AAA GGA TTC TTT AAA CGG ACT GTG
 Q   N   K   K   V   Y   T   C   V   A   D   R   S   C   H   I
CAA AAT AAA AAA GTT TAT ACG TGT GTG GCG GAT AGG AGC TGT CAT ATC
 D   K   T   Q   R   K   R   C   P   Y   C   R   F   Q   K   C
GAC AAA ACG CAA AGA AAA AGG TGT CCG TAC TGC AGA TTC CAG AAG TGC
 L   E   V   G   M   K   L   E   A   V   R   A   D   R   M   R
CTC GAA GTT GGA ATG AAA TTG GAA GCC GTC CGA GCG GAC CGA ATG CGG
 G   G   R   N   K   F   G   P   M   Y   K   R   D   R   A   R
GGA GGG AGG AAT AAA TTC GGA CCC ATG TAC AAA AGA GAC CGA GCT CGA
 K   L   Q   L   M   R   Q   R   Q   L   S   A   Q   R   P   G
AAA CTA CAA TTG ATG AGA CAG CGG CAG CTC TCA GCC CAG AGG CCG GGT
 M   T   V   P   G   I   N   E   A   V   T   L   T   Y   S   T
ATG ACG GTA CCC GGT ATT AAC GAA GCG GTT ACC CTC ACG TAT AGC ACA
 P   A   G   T   G   H   F   A   A   A   P   G   G   S   N   L
CCT GCT GGA ACG GGT CAT TTC GCT GCA GCG CCT GGT GGT TCG AAT CTT
 H   I   K   Q   E   I   Q   I   P   Q   V   S   S   L   T   S
CAC ATC AAA CAG GAG ATC CAA ATC CCT CAA GTT TCG TCG CTT ACG TCG
 S   P   D   S   S   P   S   P   I   N   Q   S   L   A   P   L
TCC CCT GAC TCG TCG CCC AGT CCC ATC AAC CAG TCC CTC GCC CCG CTC
 G   V   G   T   T   T   A   T   T   A   T   T   N   N   H   V
GGC GTT GGG ACG ACC ACG GCC ACC ACC GCG ACG ACC AAC AAC CAC GTG
 A   A   N   G   P   A   I   L   G   A   D   H   K   L   W   A
GCA GCC AAC GGC CCA GCG ATC CTC GGA GCG GAT CAT AAA CTG TGG GCG
 S   P   N   S   T   T   P   S   P   L   S   L   S   P   K   T
AGT CCA AAC TCT ACA ACC CCG TCA CCG CTT TCG CTC TCG CCC AAG ACC
 F   Q   Y   D   G   A   V   P   N   T   V   K   I   S   P   L
TTC CAA TAC GAC GGG GCA GTT CCC AAC ACG GTG AAG ATC TCG CCG CTC
 I   R   D   F   V   Q   A   V   D   D   R   E   W   Q   N   S
ATC AGG GAC TTT GTG CAA GCA GTG GAC GAT AGA GAG TGG CAA AAC TCG
 L   F   G   L   L   Q   N   Q   T   Y   N   Q   C   E   V   D
TTA TTC GGT CTC CTT CAG AAT CAG ACC TAC AAC CAG TGT GAA GTG GAT
```

```
 L   F   E   L   M   C   K   V   L   D   Q   N   L   F   S   Q
CTT TTC GAA CTT ATG TGC AAA GTG CTG GAT CAA AAC CTC TTC TCT CAA
 V   D   W   A   R   N   S   I   F   F   K   D   L   K   V   D
GTC GAT TGG GCG CGG AAC TCG ATA TTC TTC AAG GAC CTG AAG GTG GAT
 D   Q   M   K   L   L   Q   H   S   W   S   D   M   L   V   L
GAC CAA ATG AAG CTT CTG CAG CAT TCG TGG TCG GAT ATG TTG GTG TTG
 D   H   M   H   Q   R   M   H   N   N   L   P   D   E   T   T
GAC CAT ATG CAC CAA CGA ATG CAT AAT AAT CTT CCG GAC GAG ACT ACA
 L   P   N   G   Q   K   F   D   L   L   C   L   G   L   L   G
TTA CCC AAC GGA CAG AAG TTC GAT CTC TTG TGC CTG GGG CTG TTG GGC
 V   P   T   L   A   D   Q   F   N   D   L   A   A   K   L   H
GTC CCC ACG CTT GCC GAC CAA TTT AAT GAC CTC GCT GCC AAG CTC CAT
 E   L   K   F   D   I   S   D   Y   I   C   I   K   F   L   L
GAG CTC AAA TTC GAC ATC AGC GAC TAT ATC TGT ATC AAG TTC CTC CTT
 L   L   N   P   E   V   R   G   L   M   N   K   K   H   V   Q
CTA CTC AAC CCT GAG GTT CGA GGG TTA ATG AAT AAG AAA CAC GTC CAG
 D   G   H   D   Q   V   Q   Q   A   L   L   D   Y   T   V   N
GAT GGT CAT GAC CAA GTA CAA CAG GCA CTA CTC GAC TAT ACG GTG AAT
 C   Y   P   Q   I   Q   D   K   F   T   K   L   M   M   L   L
TGT TAT CCA CAA ATT CAG GAC AAG TTC ACG AAG CTT ATG ATG CTG TTA
 P   E   I   H   Q   L   A   T   R   G   E   E   H   L   Y   H
CCG GAG ATT CAT CAG CTG GCC ACT CGA GGG GAA GAG CAT TTG TAC CAC
 K   H   C   S   G   G   A   P   T   Q   T   L   L   M   E   M
AAA CAT TGC AGC GGA GGA GCG CCC ACT CAG ACG TTG CTA ATG GAA ATG
 L   H   A   K   R   K   *
CTA CAC GCG AAG AGA AAA TAG   TAGATCCACTTCAGCTTCAGAATTTGAAAATTTT
TCTCCAGATATTTCATGGATTGCAAGAAAGCTGTGAATCAAATGCACAGAGGCCGGCCCCTCTG
TTTACCTATGTCCAGTGACTCTTTGCATTATGAAATTATGAGTCTCAGAGCTCCATACACCTCC
TCAGGAAGCAAAAAAAAAAA
```

Callosobruchus maculatus, obtained by degenerate PCR and RACE using RNA from 0-2d embryos

```
ACAAAGTGTAAGCGAAGTGGCCTAGAGACCTACATTTATAGAACATTCGATTTTTACGGCGCTA
AAAGAAATAGATATTTGTGAAACCTTGACGCGAAATTGAAGGTCGAAATTTCTCGGGACCTTGG
                                                              M   H
GGCGGGTAAATTCGGTTATCATTGCGGGTTAGAGTGCGTATTTGGGTAAACA         ATG CAT
 E   E   A   T   S   M   S   V   P   G   S   V   A   V   S   T
GAA GAA GCG ACA AGT ATG AGC GTT CCT GGC TCT GTT GCC GTA TCT ACG
 G   Q   T   E   N   G   T   S   T   I   S   V   E   N   A   D
GGA CAA ACA GAA AAC GGG ACA TCG ACA ATT TCA GTG GAA AAT GCC GAT
 L   D   L   E   V   T   T   D   D   D   N   V   E   L   Q   M
TTA GAC CTG GAA GTA ACC ACA GAC GAT GAC AAC GTT GAG CTG CAG ATG
 S   F   S   S   N   S   G   G   G   A   A   G   G   L   P   G
TCT TTT TCT TCA AAT TCA GGA GGT GGC GCC GCG GGC GGC CTC CCC GGT
 L   E   M   G   G   P   G   G   G   Y   S   M   P   G   I   A
CTG GAA ATG GGC GGC CCG GGC GGC GGA TAC TCC ATG CCG GGC ATC GCG
 P   M   T   P   A   G   D   M   P   D   T   K   D   G   I   E
CCT ATG ACG CCC GCC GGA GAC ATG CCC GAC ACC AAG GAC GGC ATC GAG
 E   L   C   P   V   C   G   D   K   V   S   G   Y   H   Y   G
GAA CTG TGC CCC GTC TGC GGG GAC AAA GTG TCC GGG TAC CAC TAC GGA
 L   L   T   C   E   S   C   K   G   F   F   K   R   T   V   Q
CTG CTC ACG TGC GAG TCG TGC AAG GGC TTC TTC AAG CGC ACC GTT CAG
 N   K   N   V   Y   T   C   V   A   E   R   S   C   H   I   D
AAC AAG AAT GTG TAC ACG TGC GTC GCC GAA CGG AGC TGT CAC ATC GAC
 K   T   Q   R   K   R   C   P   F   C   R   F   Q   K   C   L
AAG ACG CAG CGG AAA CGG TGT CCC TTC TGC AGG TTC CAG AAG TGC CTC
 D   V   G   M   K   L   E   A   V   R   A   D   R   M   R   G
GAC GTC GGA ATG AAG CTA GAA GCC GTC CGA GCG GAC AGG ATG CGC GGC
 G   R   N   K   F   G   P   M   Y   K   R   D   R   A   R   K
GGC AGG AAC AAA TTC GGC CCG ATG TAC AAA CGG GAC AGG GCG CGG AAG
 L   Q   T   M   L   L   G   A   A   G   T   A   G   D   T   K
CTG CAG ACG ATG CTG CTC GGG GCC GCG GGC ACC GCT GGA GAC ACC AAG
 L   W   A   A   N   S   T   T   T   S   P   H   S   L   S   P
CTC TGG GCG GCC AAC TCT ACC ACC ACC TCG CCG CAC TCG CTC AGT CCT
 K   V   F   Q   F   D   S   G   V   P   A   P   T   A   S   S
AAG GTG TTC CAG TTC GAC TCA GGT GTG CCA GCC CCT ACG GCG TCT AGC
 V   G   A   K   L   S   P   L   I   R   D   F   I   Q   G   I
GTA GGA GCG AAG CTG TCG CCC CTG ATC AGG GAC TTC ATC CAG GGT ATA
 D   D   R   E   W   Q   N   S   L   Y   G   L   L   Q   N   Q
GAC GAC CGG GAG TGG CAG AAC TCC CTC TAT GGC CTG CTC CAG AAC CAG
 T   Y   N   Q   C   E   V   D   L   F   E   L   M   C   K   V
ACG TAT AAT CAG TGC GAG GTG GAC TTA TTC GAA CTT ATG TGT AAA GTG
 L   D   Q   N   L   F   S   Q   V   D   W   A   R   N   S   I
TTA GAC CAG AAC CTA TTC TCG CAA GTC GAC TGG GCG AGG AAC TCT ATA
 F   F   K   D   L   K   V   D   D   Q   M   K   L   L   Q   N
TTC TTT AAG GAT CTC AAG GTG GAC GAC CAG ATG AAG TTG CTG CAG AAC
```

```
 S   W   S   D   M   L   V   L   D   H   L   H   Q   R   M   H
TCT TGG TCG GAC ATG TTG GTG TTG GAC CAC CTT CAC CAA CGG ATG CAC
 N   S   L   P   D   E   T   T   L   H   N   G   Q   K   F   D
AAC AGT CTG CCA GAC GAG ACG ACG TTG CAC AAC GGC CAG AAG TTC GAT
 L   L   S   L   G   L   L   G   V   P   A   L   A   E   H   F
CTG CTC AGT CTG GGA CTC CTC GGT GTT CCG GCG CTC GCG GAA CAC TTC
 N   D   I   T   A   K   L   Q   E   L   K   F   D   I   S   D
AAC GAC ATC ACC GCC AAG TTG CAA GAA TTG AAA TTC GAT ATC AGC GAC
 Y   I   C   I   K   F   M   L   L   L   N   P   D   V   R   G
TAT ATC TGC ATC AAA TTC ATG CTG CTT CTT AAT CCA GAT GTT CGA GGC
 I   T   N   R   K   H   V   E   E   G   F   E   Q   V   Q   Q
ATC ACA AAT AGG AAA CAT GTA GAG GAA GGC TTT GAG CAA GTC CAA CAG
 A   L   L   E   Y   T   V   T   C   Y   P   Q   I   Q   D   K
GCA TTA TTA GAA TAT ACG GTG ACA TGT TAC CCA CAA ATT CAG GAC AAA
 F   H   K   M   Q   Q   L   L   S   E   I   H   D   I   A   V
TTC CAC AAG ATG CAG CAA CTG CTG TCG GAG ATC CAC GAT ATC GCC GTA
 R   G   E   E   H   L   Y   H   K   H   C   S   G   G   A   P
AGG GGG GAG GAG CAC CTG TAC CAC AAG CAT TGC AGC GGC GGA GCG CCG
 T   Q   T   L   L   M   E   M   L   H   A   K   R   R   *
ACG CAG ACG CTG CTC ATG GAG ATG CTG CAC GCG AAG CGG AGA TAA CCT
TCCGCGCGTCCACAGCTTCAGCCGCAGCTTCAGCTGCAGCAAAG
```

Dermestes maculatus, obtained by degenerate PCR and RACE using RNA from mixed embryos

```
 M   H   E   E   A   A   S   M   S   N   L   D   A   S   Y   L
ATG CAT GAA GAG GCG GCA AGC ATG AGT AAT CTG GAC GCG TCG TAT TTG
 F   S   P   G   G   G   G   G   G   V   L   P   G   G   G   V
TTT TCC CCA GGG GGT GGA GGC GGC GGC GTC CTT CCC GGA GGC GGC GTC
 D   M   G   P   S   Y   Q   L   T   G   P   A   T   S   L   T
GAC ATG GGC CCT AGC TAT CAG CTG ACC GGT CCG GCG ACT TCG CTT ACG
 T   G   D   L   P   D   T   K   D   G   I   E   E   L   C   P
ACC GGC GAT CTG CCC GAC ACC AAG GAC GGA ATC GAG GAG TTG TGT CCA
 V   C   G   D   K   V   S   G   Y   H   Y   G   L   L   T   C
GTG TGC GGC GAC AAA GTG TCC GGA TAT CAC TAC GGT CTT CTC ACG TGT
 E   S   C   K   G   F   F   K   R   T   V   Q   N   K   K   V
GAA TCA TGC AAG GGT TTC TTC AAG AGG ACC GTA CAA AAC AAG AAG GTG
 Y   T   C   V   A   E   R   S   C   H   I   D   K   T   Q   R
TAT ACG TGC GTC GCC GAG AGG AGT TGT CAC ATA GAC AAA ACG CAA AGG
 K   R   C   P   F   C   R   F   Q   K   C   L   E   V   G   M
AAA AGG TGT CCG TTC TGT CGT TTT CAA AAG TGC CTG GAA GTT GGC ATG
 K   L   E   A   V   R   A   D   R   M   R   G   G   R   N   K
AAG CTA GAA GCT GTA CGA GCA GAT CGG ATG AGA GGA GGA AGG AAT AAA
 F   G   P   M   Y   K   R   D   R   A   R   K   L   Q   M   M
TTT GGT CCG ATG TAC AAA AGA GAT AGA GCC CGG AAA TTG CAA ATG ATG
 R   Q   R   Q   L   A   A   Q   T   L   R   G   S   L   G   D
AGG CAA AGG CAG TTG GCG GCG CAG ACG TTG CGG GGG TCG CTG GGC GAT
 S   S   M   Y   S   S   Q   P   G   T   S   P   F   A   N   I
AGC AGC ATG TAT AGC AGC CAG CCA GGC ACG TCG CCC TTC GCA AAC ATC
 H   I   K   Q   E   I   Q   I   P   Q   V   S   S   L   T   S
CAC ATC AAG CAG GAG ATC CAA ATC CCG CAG GTA TCG TCG CTG ACG TCC
 S   P   D   S   S   P   S   P   I   A   V   A   L   G   Q   V
TCC CCG GAT TCA TCA CCA AGC CCC ATC GCT GTC GCT CTG GGT CAA GTG
 N   S   Q   L   A   Q   P   A   S   S   Q   Q   P   T   L   Q
AAT TCT CAA TTA GCC CAA CCC GCC TCT AGC CAA CAG CCG ACC CTG CAA
 I   V   G   V   P   G   G   G   G   P   H   T   S   M   I   M
ATA GTC GGG GTG CCG GGC GGC GGC GGC CCA CAC ACC TCC ATG ATC ATG
 G   P   E   N   K   L   W   G   S   A   N   S   A   T   T   S
GGC CCC GAG AAC AAA CTC TGG GGT TCC GCC AAC TCG GCC ACG ACG TCA
 P   H   S   L   S   P   K   A   F   Q   F   D   T   V   V   P
CCT CAT TCC CTG AGC CCG AAG GCG TTC CAG TTC GAC ACG GTG GTG CCC
 G   G   S   A   P   P   S   S   K   M   S   P   L   I   R   D
GGC GGC AGC GCG CCC CCG TCT TCT AAA ATG TCG CCC CTT ATC AGG GAC
 F   V   Q   A   I   D   D   R   E   W   Q   N   L   L   Y   G
TTT GTG CAG GCG ATT GAC GAT CGC GAG TGG CAG AAC TTA CTT TAT GGA
 L   L   Q   S   Q   T   Y   N   Q   C   E   V   D   L   F   E
CTC CTA CAG AGC CAA ACT TAT AAT CAA TGT GAA GTT GAC TTG TTT GAA
 L   M   C   K   V   L   D   Q   N   L   F   S   Q   V   D   W
CTT ATG TGT AAA GTG TTG GAC CAA AAC CTC TTC TCG CAG GTC GAC TGG
```

```
 A   R   N   S   I   F   F   K   D   L   K   V   D   D   Q   M
GCG CGA AAT TCC ATC TTC TTT AAG GAT CTC AAG GTT GAT GAC CAA ATG
 K   L   L   Q   H   S   W   S   D   M   L   V   L   D   H   I
AAA CTC TTG CAA CAT TCG TGG TCG GAT ATG TTA GTG TTG GAC CAC ATA
 H   Q   R   M   H   N   N   L   P   D   E   T   T   L   H   N
CAT CAG CGT ATG CAC AAT AAT TTG CCG GAT GAA ACC ACC CTC CAT AAT
 G   Q   K   F   D   L   L   N   L   G   L   L   G   V   P   T
GGG CAG AAG TTC GAT TTG CTC AAT TTG GGT TTG TTA GGG GTG CCC ACG
 L   A   D   H   F   N   D   I   T   A   K   L   Q   E   L   K
CTG GCG GAT CAC TTC AAT GAC ATC ACA GCC AAG CTG CAG GAG CTC AAG
 F   D   I   S   D   Y   I   C   V   K   F   L   L   L   L   N
TTT GAT ATA AGC GAC TAT ATA TGC GTG AAA TTC TTG CTA CTC CTC AAT
 P   A   F   L   D   V   R   G   I   T   N   R   K   H   V   Q
CCA GCT TTT CTA GAT GTA CGC GGC ATC ACC AAT CGG AAA CAC GTC CAA
 E   G   Y   E   Q   V   Q   Q   A   L   L   Q   Y   T   I   S
GAG GGC TAT GAG CAA GTG CAA CAG GCT CTA TTG CAG TAT ACC ATT TCA
 C   Y   P   E   V   Q   D   K   F   N   K   M   L   Q   L   L
TGT TAC CCA GAA GTT CAG GAT AAG TTC AAC AAG ATG TTG CAG CTG CTG
 P   E   I   H   S   L   A   A   R   G   E   E   H   L   Y   H
CCA GAG ATC CAC TCG TTG GCA GCA CGC GGA GAG GAG CAC CTA TAC CAC
 K   H   C   N   G   S   A   P   T   Q   T   L   L   M   E   M
AAG CAC TGC AAC GGC AGC GCC CCC ACC CAA ACA CTA TTA ATG GAA ATG
 L   H   A   K   K   K   *
CTA CAT GCG AAA AAA AAA TAA
```

Appendix II: Detailed Materials and Methods

I. (General) Molecular biology and cloning

i. *Polymerase Chain Reaction (PCR)*
DNA templates (genomic DNA, cDNA, plasmid) were diluted to 10ng-100ug per reaction. Primers were stored at -20°C as 100uM stocks, and diluted to 10pol/uL (10uM) for PCR. Buffer supplied from the manufacturer containing Mg^{2+} was diluted to 1X (from 5X-10X stock), and a mixture of dNTPs (dATP, dTTP, dCTP, dGTP) was diluted to a final concentration of 0.2-0.4mM. *Taq* or *Vent* polymerase (0.2-0.5uL) was added immediately before being put in the tube in the thermocycler. A standard thermocycler program included a 3-5 minute initial incubation at 95 ^{o}C, and 30 cycles of denaturing (30 seconds at 95°C), annealing (30 seconds at appropriate temperature), and extending DNA products (72°C at 1kb/minute), and one final longer extention (5 minutes at 72°C). PCR products were checked by running 5-50uL of the reaction on a 1-2% agarose gel.

ii. *Site-directed mutagenesis*
Site-directed mutagenesis was performed using 10-50ng plasmid DNA as a template, 5uM each primer containing the mutated nucleotides, and 0.2mM dNTPs in 1X reaction buffer. *Vent* polymerase was used to minimize the number of errors incorporated into the DNA. A thermocycler program that included an initial incubation at 95°C (5 minutes) and 18 cycles of denaturing (30 seconds at 95°C), annealing (1 minute at an appropriate temperature), and extending DNA products (72°C at 1kb/minute) was executed. Immediately upon completion, 20U *DpnI* was added to each reaction to degrade DNA template, and the reaction was incubated at 37°C. After 2 hours, 1-5uL of the reaction was transformed into DH5α cells and grown overnight on LB+amp (100ug/mL) at 37°C. Several colonies were picked, and mini-preps of the DNA were made and sequenced to identify which contained the mutation.

iii. *Restriction-enzyme digests*
All restriction enzyme digests were done in the buffer recommended by the supplier (NEB). When the two enzymes being used were not compatible in the same buffer, two consecutive digests were done. Vector DNA (5-10ug) or insert DNA (up to 5ug) were digested in parallel in the recommended buffer and BSA (if required) for 2-4 hours in a 37°C waterbath. Halfway through the digestion, 1uL Calf Intestine Phosphatase (CIP) was added to the vector digest to prevent re-ligation of vector fragments. Reactions were cleaned up with a PCR-purification kit (Qiagen) or the correct-sized fragment was purified from an agarose gel using a Qiagen Gel Extraction Kit. Complete digestion and correct sized fragments were checked on an agarose gel by gel electrophoresis.

iv. *Ligation of DNA fragments*
A known amount of vector and insert DNA were run alongside each other on an agarose gel to compare concentration levels. Vector and insert DNA were incubated in T4 Ligase buffer with Ligase at room temperature for several hours or overnight at 14°C.

v. *Transformation into bacteria*

After ligation, the entire reaction (20uL) was used for transformation into DH5α or HB101 bacterial cells. Cells (50-100uL) frozen at -80°C were thawed on ice for 15 minutes, the ligation reaction was added to the cells and incubated on ice for an additional 15 minutes. Tubes were placed in a 37°C waterbath for 2 minutes, and 800uL LB (no antibiotic) was added and allowed to incubate at 37°C for an additional 45 minutes. Cells were spun down, resuspended in ~150uL supernatant, plated on LB + ampicillin (50ug/uL?) plates, and incubated at 37°C overnight.

vi. *Screening for positive transformation clones by PCR*

The number of clones selected for screening was determined by comparing the number of clones on the vector+insert plate with those on the vector only plate. Single colonies were picked with a toothpick, streaked on a fresh LB plate with antibiotic, and then dipped in an epindorff tube with 20uL water. Tubes were boiled for 5 minutes and centrifuged at maximum speed for 2 minutes. The supernatant (5uL) was used as DNA template with PCR primers that would positively identify the presence of an insert. PCR products were run on an agarose gel, and those clones that contained the insert were grown as larger preps (see below).

vii. *Small-scale plasmid preps (mini-preps)*

Small-scale DNA preps were made for sequencing DNA that had been obtained from PCRs of genomic DNA or cDNA. A single colony (or cells grown from the streaked plate described above) was grown overnight at 37°C in LB+antibiotic. After 12-18 hours, half the culture was spun down and cells were resuspended in ~100uL of supernatant. Cells were lysed in 300uL Lysis Buffer (give composition) and incubated for 5 minutes at room temperature. DNA was neutralized by adding 150uL Neutralization Buffer (give composition) and mixed. Tubes were centrifuged at 13,000 at 4°C for 8 minutes. The supernatant was transferred to a new tube and DNA was precipitated by addition of 600uL isopropanol. DNA was pelleted by centrifugation (13,000rpm, 4°C, 10 minutes), washed with 70% ethanol, and the dried pellet was resuspended in TE+RNaseA (10ug/mL). Plasmid quality and concentration were checked by running DNA on a 0.8% agarose gel next to a DNA ladder of known concentration.

viii. *Large-scale plasmid preps (midi-preps)*

Once the DNA sequence of the clone was found to be correct, a larger-scale DNA prep (midi-prep) was made. Cells (100uL) used to make the mini-prep were grown in a 100mL LB+antibiotic culture overnight and plasmid DNA was extracted using Qiagen's Midi-Prep Kit according to the manufacturer's instructions.

II. Isolation of *ftz* homologs

i. *Isolation of genomic DNA*

Genomic DNA was isolated from arthropods that were collected in the field and stored in 70% ethanol at -20°C, or collected in the lab and frozen at -80°C before extracting DNA. The organism was homogenized in DNA Extraction Buffer (100mM Tris pH 7.5, 100mM EDTA, 100mM NaCl, 0.5% SDS) and incubated at 65°C for 30 minutes. Neutralization buffer (1.4M KCl, 4.3M LiCl) was added, and the mixture was incubated on ice for 30 minutes. Debris was

removed by centrifugation, and DNA was precipitated with isopropanol and centrifugation. After washing with 70% ethanol, DNA pellets were dried and resuspended in TE+RNaseA (10ug/mL). Intact genomic DNA was checked by gel electrophoresis on a 0.8% agarose gel in TAE.

ii. *Isolation of embryonic RNA*
Embryos were collected and frozen at -80°C in TRIzol (Invitrogen) before RNA extraction. Tubes were thawed on ice, and embryos were homogenized in TRIzol using a pestle until the solution became cloudy. Debris was removed by centrifugation, and the supernatant was mixed with chloroform (0.2 v/v) by vigorous shaking. After centrifugation (15 minutes, 13,000rpm, 4°C), the aqueous phase was transferred to a new tube and mixed gently with 1 volume 70% ethanol. An RNA Extraction Kit (Qiagen) was then used to separate RNA from other nucleic acids, which was then precipitated overnight with 3M NaOAc (volume?) and cold 100% ethanol. RNA was spun down, washed, and resuspended in nuclease-free water. RNA concentration and purity was measured by spectrophotometry. All RNA was stored at -80°C until use.

iii. *Rapid Amplification of cDNA Ends (RACE)*
5' and 3' RACE was performed as described by the manufacturer's instructions (Ambion). Two rounds of nested PCR was enough to visualize products when the entire PCR product was run on a 1-2% agarose gel.

iv. *Rapid Isolation of Gene Homologs across Taxa (RIGHT)* --see text (Chapter 2) or supplemental data in publication for more detailed protocol

III. Plasmid construction

i. *'degen-YPWM' constructs*
All 'degen-YPWM' constructs were made in an FtzLRAAA background to eliminate interaction with Ftz-F1. The "FNWS" motif in *Drosophila* Ftz (Dm-Ftz) was changed by site-directed mutagenesis to degenerate motifs found in *Artemia*-Ftz (YHQM) and *Folsomia*-Ftz (YPPWLK) in pKS and then cloned into pUAST-myc (5' myc-tag sequence: MGTEQKLISEEDLNEF) using restriction enzyme cloning and the EcoRI site in-frame with the myc tag. As a negative control, the FNWS was changed to AAAA, which should completely abolish interaction with Exd. All cloning was confirmed by restriction digests and changes to the FNWS motif were confirmed by sequencing. Primer sequences for the site-directed mutagenesis were: *Artemia*: AAT GGA GCC GGC GAT **TAC CAC CAG ATG** CAC ATC GAG GAG ACT, and the reverse-complement primer; *Folsomia*: AAT GGA GCC GGC GAT **TAC CCT CCT TGG CTG AAG** CAC ATC GAG GAG ACT, and the reverse-complement primer; *negative control*: TGG AGC CGG CGA T**GC CGC TGC CGC C**CA CAT CGA G, and the reverse-complement primer.

ii. *CNS rescue constructs*
The PFK4 vector, which includes the 10kb *ftz* regulatory region, promoter, *ftz* gene, and 3' region, was used as a PCR template to generate fragments for cloning. The *ftz* promoter (40 base pairs upstream of the translation start site) and coding region up to the SalI site in *ftz* was isolated

and cloned into KS-Dm*ftz* using NotI and SalI restriction ends. An additional 200 bases of sequence was added to the 3'UTR of KS-Dm*ftz* using the endogenous EcoRI site and adding an XbaI site at the 3' end. This entire ~2kb *ftz* fragment was cloned into pCasper4 using the NotI and XbaI sites on the ends. The 2.2kb neurogenic element, defined by XbaI and BalI sites in PFK4, was amplified by PCR from PFK4 with primers that had KpnI and NotI ends. This region was then cloned into pCasper4 along with the *ftz* coding region (NE-Ftz). All fragments were verified by restriction digest and sequenced.

Alteration to the *ftz* coding region resulted in constructs that allowed for testing of the requirement of different motifs and domains in the Ftz protein for CNS function. Site-directed mutagenesis changed the LRALL to LRAAA (NE-FtzLRAAA) and FNWS to AAAA (NE-FtzAAAA). The N-terminal arm of the homeodomain was changed from that of Ftz (SKRTRQTY) to that of Antp (<u>R</u>KR<u>G</u>RQTY) by site-directed mutagenesis (NE-FtzNTAntp). The entire Ftz homeodomain was swapped for that of Antp (NE-FtzAntpHD) using PCR and ligation of the 5' Ftz end, Antp homeodomain, and 3' Ftz end. The construct lacking the homeodomain (NE-FtzΔHD) was created by Uli Lohr. As a positive control for *ftz* CNS expression and negative control for CNS rescue, GFP was cloned in frame with the 5' end of Ftz at the SalI site just downstream of the LRALL motif. Primer sequences used for site-directed mutagenesis were: *LRAAA*: GCA CAC TGA GGG CT**G CAG CC**A CCA ATC CC, and the reverse-complement primer; *AAAA*: TGG AGC CGG CGA T**GC CGC TGC CGC C**CA CAT CGA G, and the reverse-complement primer; *NTAntp*: GAT TGC AAA GAC **CGC AAA CGC GGA** CGT CAG ACG TAC, and the reverse-complement primer; *AntpHD*: forward, 5' CGC AAA CGC GGA AGG CAG and reverse, 5' CTT GTT CTC CTT CTT CCA CTT C.

iii. *<u>Constructs for in situs</u>*

All sequences used to make probes for *in situs* were cloned into pKS using restriction enzymes. located in the pKS multiple cloning site. *Artemia Antp* (NCBI accession: AF435786) was cloned into pKS with XbaI, and *Artemia cad* (NCBI accession: AJ567452), *ftz* (see Appendix I), and *ftz-f1* (see Appendix I) were cloned into pKS with XhoI and BamHI.

<u>IV. Work with *Drosophila melanogaster*</u>

i. *<u>Maintenance</u>*

Drosophila melanogaster were kept in plugged vials at 20°C or 25°C and fed standard medium (molasses, cornmeal, bakers yeast and agar). Stock vials were kept at 20°C and flipped once a month.

ii. *<u>Transgenic flies</u>*

Independent transgenic fly lines were established from DNA injected into w- embryos by Rainbow Transgenics Inc (California). Surviving larva were received in vials, and each adult that eclosed was crossed to 2-3 w- flies of the opposite sex. Typically, 100 flies were screened. Male w+ flies recovered from this cross was crossed to 3 female w-. All lines that had one

insertion (equal number of male and female w- and w+ flies) were crossed to second (Cyo) and third (Sb) chromosome balancers to figure out where the transgene had been inserted.

iii. *Embryo collection and fixation*

Embryos were collected on apple juice plates (2.5% sugar (w:v); 2.5% agar (w:v) and 25% apple juice (v:v)) and staged for the desired time. Embryos were rinsed with water and removed from the plate with a brush, and transferred to a mesh trap. They were dechorionated in a 3% bleach solution for 3 minutes, and transferred to an epindorff tube and fixed in embryo fixation solution (4% PFA in PBS, 50% heptane (v:v)) for 20 minutes while moving quickly on the rotating wheel. The lower (aqueous) phase was removed, and 800uL methanol was added. Embryos were devitillinized by manual shaking for 15 seconds and allowed to sink to the bottom of the tube. After 2-3 washes in methanol, embryos were stored in methanol at -20°C until use.

iv. *Antibody stainings*

Embryos were rehydrated through 3 washes in in PBST (1X PBS with 0.05% Tween20) at room temperature. The primary antibody was diluted to an appropriate concentration in PBST and incubated with the embryos overnight at 4°C on a rotating wheel. Embryos were rinsed three time (invert tube several time, allow to settle), and then washed 3 times (on rotating platform). The embryos were incubated in the secondary antibody (diluted to an appropriate concentration) for 1.5-2 hours at room temperature. After 3 rinses and 3 washes, embryos were incubated in ABC for 1 hour at room temperature. After washing (either several times or overnight), DAB was added and the color reaction was monitored in concave glass dishes. Once stained appropriately, embryos were rinsed several times in PBST, and then incubated in 90% glycerol. Microscopy was performed using a Leica DMRB microscopy.

v. *in situ hybridizations*

Embryos were collected and fixed as described above. Digoxigenin-UTP RNA probes were made using 1ug linearized plasmid and either the T7/T3 promoter for reverse transcription. Probe reactions (1ug linearized DNA, 1X transcription buffer, 1X dig U-NTP mix, 5mM DTT, 50U RNase inhibitor, T3/T7 polymerase) were incubated at 37°C for 2hrs. Probes were fragmented in Carbonate Buffer (120 mM Na_2CO_3, 80 mM $NaHCO_3$, pH 10.2) at 65°C. Fragmentation was stopped with 0.2 M NaAc (pH 6.0) and RNA was precipitated (4M LiCl, tRNA, EtOH), spun down, washed, and dissolved in Hybidization Solution (50% Formamide, 5X SSC, 100ug/mL heparin, 100ug/mL denatured salmon sperm, 0.1% Tween 20) and stored at -20°C.

Embryos stored in EtOH were rinsed twice in methanol, washed in MeOH and PBT/formaldehyde (1:1) for 5 minutes and incubated in PBT/formaldehyde for 25 minutes. After 3 washes in PBT, embryos were incubated at 95°C for 5 minutes and rinsed with PBT/Hybridization Solution (1:1). Embryos were prehybridized in Hybridization Solution at 55°C for 2 hours. Probes (2uL in 100uL Hybridization Solution) were heated at 95°C for 5 minutes, cooled on ice, added to the embryos, and incubated overnight at 55°C. Embryos were washed several times in Hybridization Solution at 55°C, followed by washes in Hyb.Sol/PBT

(1:1), and PBT at room temperature, and incubated in anti-Digoxigenin FEB-fragment AP (source) at 1:2000 for one hour. After several washes in PBT, embryos were washed in Staining Buffer (100mM NaCl, 50mM $MgCl_2$, 100mM Tris pH 9.5, 0.1% Tween 20), and the NBT+BCIP (4.5uL NBT at 100mg/ml + 3.5uL 50mg/ml BCIP per 1mL staining buffer) reaction was carried out until a clear staining pattern was seen.. Staining was stopped by several rinses in PBT, and then embryos were carried through a series of rinses: 1X PBT/MeOH (1:1), 2X MeOH, 1X EtOH, 2X MeOH, 1X MeOH/PBT, 2X PBT. Embryos were incubated in 90% glycerol overnight and transferred to microscope slides were viewing and photography.

V. Work with other arthropods

i. *Rearing and maintenance*

Tribolium were reared in glass Mason jars at room temperature on whole-wheat flour with 5% dry yeast. *Callosobruchus* were reared on a layer of mung beans in plastic containers with holes in the container lid. *Dermestes* were reared in cages that had a layer of wood chip shavings on the bottom, a sponge or porous surface for egg-laying, and occasionally fed protein (bacon, meat). *Thermobia* were maintained at 30°C in a container with folded paper surfaces and fed oatmeal and hermit crab food. *Folsomia* were reared on petri dishes with a moist charcoal/plaster of paris surface and fed dry yeast pellets. *Artemia* cysts were rehydrated in 3% aerated salt water in and fed a dilute yeast solution.

ii. *Embryo collections/fixation*

All embryos were staged at the appropriate temperature until the desired stage of development, and fixed embryos were stored at -20°C until use.

Artemia collected by concentrating them in one area with a light source. They were fixed in 4% PFA for 2 hours at room temperature and then taken through a series of methonal/PBT washes: 25%, 50%, 75%, and rinsed 4 times in 100% MeOH before storing at -20°C.

Tribolium adults were placed on all-purpose white flour and allowed to lay eggs for the desired amount of time at 30°C. Eggs were collected by sifting the flour through a 350mm sieve and transferred to a collection basket (same as *Drosophila*). Flour was removed by several rinses with water, and embryos were dechorionated and fixed using the *Drosophila* protocol.

iii. *in situ hybridizations*

Artemia: Nauplii were taken through a series of washes to get them into PBT (75%, 50%, 25% MeOH in PBT) and rinsed 3X in PBT. To increase reagent penetration, *Artemia* were sonicated for 10sec, power setting 20 (Fisher 50 Sonic Dismembrator) and washed several times in PBT. Nauplii were treated with ProteinaseK in PBT (10ug/mL) for 5 mins at room temperature, washed in PBT, and fixed in 4% PFA in PBT for 1hr at room temperature. Following fixation,

Artemia were washed several times for 15 minutes each: 3X PBT, 1X 1:1 PBT/Hyb. Soln, 1X Hyb. Soln. Nauplii were incubated in Hybridization Solution for 2hrs at 60°C, inverting the tubes a few times during the incubation. Probes were prepared at described for *Drosophila* and incubated with the *Artemia* nauplii overnight at 60°C. *Artemia* were washed several time at 60°C with a series of Hyb Soln:PBT solutions (4:1, 3:2, 2:3, 1:4) for 15 mins each, followed by 2 washes in PBT. Nauplii were incubated in a pre-absorbed α-DIG-frag-AP antibody (1:2000) for 2 hours at room temperature or overnight at 4°C. *Artemia* were washed 3X in PBT for 30mins, 1X 10 minutes in PBT/Staining Buffer (1:1) and 2X for 10mins in Staining Buffer. The NBT/BCIP reaction and subsequent washes were carried out as described for *Drosophila*.

Tribolium: *Tribolium* embryos were rinsed once in MeOH and then incubated in 50% Xylenes (50% MeOH) and vortexed for 30 seconds every 10 minutes for 30 minutes to completely remove the vitelline membrane and separate young embryos from the yolk. Embryos were rinsed twice in MeOH, once in MeOH/PBT (1:1), and washed three times (5 minutes each) in PBT. After a 15 minute incubation in 4%PFA, embryos were washed three times in PBT, and incubated at 92°C for 5 minutes. Embryos were prehybridized in Hybridization Buffer at 60°C for at least 2 hours before incubating overnight in an appropriate concentration of dig-labeled probe. The remaining staining steps are the same as that used for *Drosophila*.

iv. *Tribolium embryonic RNAi*

Tribolium ftz and *ftz-f1* sequences used for RNA probes were also used to for functional studies. Plasmid DNA was linearized with a restriction site at the 5' and 3' end of the gene, and single-stranded RNA was made using 1ug of each digest as a template and T7/T3 MegaScript Kit (Ambion) according to the manufacturer's instructions. RNA yields from each reaction were measured on a spectrophotometer and equal concentrations of single-stranded RNA were mixed an annealed in a PCR tube with the following program: 98°C for 2', -2°C/1' down to 4°C. Double-stranded RNA (dsRNA) concentration was measured on a spectophotomoter and quality was checked on an agarose gel before diluting to an appropriate concentration for injection.

Embryo preparation and injection was performed following the protocol of Posnien et. al, 2009. Embryos were collected on white all-purpose flour for 2 hours at room temperature, and then the adults were collected by passing the flour through a 700mm sieve. The eggs were allowed to incubate for 1 hour more in the flour and then were collected in a 300um sieve. Once washed with water, the eggs were transferred to a piece of moist black filter paper and then placed in a line along the side of a microscope slide (orientation not important). Double-stranded RNA (with green dye added for visualization) was loaded into an injection needle and embryos were injected under the microscope. After injection, embryos were placed in a petri dish with a paper towel soaked in a 15% NaCl solution, parafilmed closed, and incubated at 30°C for 4 days. After this time, larval cuticles were examined for defects.

v. *Tribolium cuticle preps*

Larva (hatched or dissected) were placed on a microscope slide, covered with a few drops of 95% lactic acid/5% EtOH and slide cover, and incubated at 55°C for 1-2 days. Cuticles were visualized by dark-field microscopy.

References

Abzhanov, A. and Kaufman, T. C. (2000) 'Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk', *Development* 127(11): 2239-49.

Abzhanov, A., Popadic, A. and Kaufman, T. C. (1999) 'Chelicerate Hox genes and the homology of arthropod segments', *Evol Dev* 1(2): 77-89.

Akam, M. (1987) 'The molecular basis for metameric pattern in the Drosophila embryo', *Development* 101(1): 1-22.

Akam, M. (1995) 'Hox genes and the evolution of diverse body plans', *Philos Trans R Soc Lond B Biol Sci* 349(1329): 313-9.

Alonso, C. R., Maxton-Kuechenmeister, J. and Akam, M. (2001) 'Evolution of Ftz protein function in insects', *Curr Biol* 11(18): 1473-8.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L. et al. (1998) 'Zebrafish hox clusters and vertebrate genome evolution', *Science* 282(5394): 1711-4.

Angelini, D. R. and Kaufman, T. C. (2005) 'Comparative developmental genetics and the evolution of arthropod body plans', *Annual review of genetics* 39: 95-119.

Asahina, M., Ishihara, T., Jindra, M., Kohara, Y., Katsura, I. and Hirose, S. (2000) 'The conserved nuclear receptor Ftz-F1 is required for embryogenesis, moulting and reproduction in Caenorhabditis elegans', *Genes Cells* 5(9): 711-23.

Averof, M. and Akam, M. (1993) 'HOM/Hox genes of Artemia: implications for the origin of insect and crustacean body plans', *Curr Biol* 3: 73-78.

Averof, M. and Akam, M. (1995) 'Hox genes and the diversification of insect and crustacean body plans', *Nature* 376(6539): 420-3.

Bateson, W. (1894) *Materials for the Study of Variation Treated with Especial Regards to Discontinuity in the Origin of Species.*: Macmillan.

Beeman, R. W. and Stauth, D. M. (1997) 'Rapid cloning of insect transposon insertion junctions using 'universal' PCR', *Insect Mol Biol* 6(1): 83-8.

Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B. and Hogness, D. S. (1983) 'Molecular Genetics of the Bithorax Complex in Drosophila melanogaster', *Science* 221(4605): 23-9.

Bender, W., Weiffenbach, B., Karch, F. and Peifer, M. (1985) Domains of *cis*-Interaction in the Bithorax Complex *Cold Spring Harbor Symposia on Quantitative Biology, Molecular Biology of Development*, vol. L.

Berry, M. and Gehring, W. (2000) 'Phosphorylation status of the SCR homeodomain determines its functional activity: essential role for protein phsophatase 2A,B'.', *EMBO J.* 19: 2946-2957.

Biedler, J., Qi, Y., Holligan, D., della Torre, A., Wessler, S. and Tu, Z. (2003) 'Transposable element (TE) display and rapid detection of TE insertion polymorphism in the Anopheles gambiae species complex', *Insect Mol Biol* 12(3): 211-6.

Bowler, T., Kosman, D., Licht, J. D. and Pick, L. (2006) 'Computational Identification of Ftz/Ftz-F1 target genes', *Developmental Biology* 299: 78-90.

Brand, A. H. and Perrimon, N. (1993) 'Targeted gene expression as a means of altering cell fates and generating dominant phenotypes', *Development* 118(2): 401-15.

Brayer, K. J., Lynch, V. J. and Wagner, G. P. (2011) 'Evolution of a derived protein-protein interaction between HoxA11 and Foxo1a in mammals caused by changes in

intramolecular regulation', *Proceedings of the National Academy of Sciences of the United States of America* 108(32): E414-20.
Brennan, R. G. and Matthews, B. W. (1989) 'The helix-turn-helix DNA binding motif', *J Biol Chem* 264(4): 1903-6.
Britten, R. J. and Davidson, E. H. (1969) 'Gene regulation for higher cells: a theory', *Science* 165(891): 349-57.
Brown, S. J., Fellers, J. P., Shippy, T. D., Richardson, E. A., Maxwell, M., Stuart, J. J. and Denell, R. E. (2002) 'Sequence of the Tribolium castaneum homeotic complex: the region corresponding to the Drosophila melanogaster antennapedia complex', *Genetics* 160(3): 1067-74.
Brown, S. J., Hilgenfeld, R. B. and Denell, R. E. (1994) 'The beetle Tribolium castaneum has a fushi tarazu homolog expressed in stripes during segmentation', *Proc Natl Acad Sci U S A* 91(26): 12922-6.
Brown, S. J., Parrish, J. K., Beeman, R. W. and Denell, R. E. (1997) 'Molecular characterization and embryonic expression of the even-skipped ortholog of Tribolium castaneum', *Mech Dev* 61(1-2): 165-73.
Brunetti, C. R., Selegue, J. E., Monteiro, A., French, V., Brakefield, P. M. and Carroll, S. B. (2001) 'The generation and diversification of butterfly eyespot color patterns', *Current biology : CB* 11(20): 1578-85.
Bucher, G., Scholten, J. and Klingler, M. (2002) 'Parental RNAi in Tribolium (Coleoptera)', *Curr Biol* 12(3): R85-6.
Burglin, T. R. (1997) 'Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals', *Nucleic Acids Res* 25(21): 4173-80.
Cameron, R. A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J. P., Berney, K., Arenas-Mena, C., Martinez, P., Lucas, S., Richardson, P. M. et al. (2006) 'Unusual gene order and organization of the sea urchin hox cluster', *J Exp Zool B Mol Dev Evol* 306(1): 45-58.
Carroll, S. B. (2008) 'Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution', *Cell* 134(1): 25-36.
Carroll, S. B., Grenier, J. K. and Weatherbee, S. D. (2005) *From DNA to diversity: molecular genetics and the evolution of animal design. Second Edition.*: Blackwell Science Ltd.
Carroll, S. B. and Scott, M. P. (1985) 'Localization of the *fushi tarazu* protein during *Drosophila* embryogenesis', *Cell* 43: 47-57.
Carroll, S. B., Weatherbee, S. D. and Langeland, J. A. (1995) 'Homeotic genes and the regulation and evolution of insect wing number', *Nature* 375(6526): 58-61.
Casa, A. M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S. and Wessler, S. R. (2000) 'Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize', *Proc Natl Acad Sci U S A* 97(18): 10083-9.
Chai, C. L., Zhang, Z., Huang, F. F., Wang, X. Y., Yu, Q. Y., Liu, B. B., Tian, T., Xia, Q. Y., Lu, C. and Xiang, Z. H. (2008) 'A genomewide survey of homeobox genes and identification of novel structure of the Hox cluster in the silkworm, Bombyx mori', *Insect Biochem Mol Biol* 38(12): 1111-20.
Chang, C.-P., Shen, W.-F., Rozenfeld, S., Lawrence, H. J., Largman, C. and Cleary, M. L. (1995) 'Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins', *Genes & Dev.* 9: 663-674.

Chesebro, J., Hrycaj, S., Mahfooz, N. and Popadic, A. (2009) 'Diverging functions of Scr between embryonic and post-embryonic development in a hemimetabolous insect, Oncopeltus fasciatus', *Dev Biol* 329(1): 142-51.
Choe, C. P. and 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(1): 281-94.
Choe, C. P., Miller, S. C. and 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(17): 6560-4.
Cifuentes, F. J. and Garcia-Bellido, A. (1997) 'Proximo-distal specification in the wing disc of Drosophila by the nubbin gene', *Proceedings of the National Academy of Sciences of the United States of America* 94(21): 11405-10.
Clyde, D. E., Corado, M. S., Wu, X., Pare, A., Papatsenko, D. and Small, S. (2003) 'A self-organizing system of repressor gradients establishes segmental complexity in Drosophila', *Nature* 426(6968): 849-53.
Cohn, M. J. and Tickle, C. (1999) 'Developmental basis of limblessness and axial patterning in snakes', *Nature* 399(6735): 474-9.
Cook, C. E., Smith, M. L., Telford, M. J., Bastianello, A. and Akam, M. (2001) 'Hox genes and the phylogeny of the arthropods', *Curr Biol* 11(10): 759-63.
Copf, T., Schroder, R. and Averof, M. (2004) 'Ancestral role of caudal genes in axis elongation and segmentation', *Proc Natl Acad Sci U S A* 101(51): 17711-5.
Cracraft, J. (2005) 'Phylogeny and evo-devo: characters, homology, and the historical analysis of the evolution of development', *Zoology (Jena)* 108(4): 345-56.
Crow, K. D., Stadler, P. F., Lynch, V. J., Amemiya, C. and Wagner, G. P. (2006) 'The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts', *Mol Biol Evol* 23(1): 121-36.
Curtis, C. D., Brisson, J. A., DeCamillis, M. A., Shippy, T. D., Brown, S. J. and Denell, R. E. (2001) 'Molecular characterization of Cephalothorax, the Tribolium ortholog of Sex combs reduced', *Genesis* 30(1): 12-20.
Damen, W. G., Janssen, R. and Prpic, N. M. (2005) 'Pair rule gene orthologs in spider segmentation', *Evol Dev* 7(6): 618-28.
Damen, W. G. and Tautz, D. (1998) 'A Hox class 3 orthologue from the spider Cupiennius salei is expressed in a Hox-gene-like fashion', *Dev Genes Evol* 208(10): 586-90.
Davidson, E. H. (2006) *The regulatory genome : gene regulatory networks in development and evolution*, Burlington, MA ; San Diego: Academic.
Davidson, E. H. and Erwin, D. H. (2006) 'Gene regulatory networks and the evolution of animal body plans', *Science* 311(5762): 796-800.
Davis, G. K. and Patel, N. H. (2002) 'Short, long, and beyond: molecular and embryological approaches to insect segmentation', *Annual review of entomology* 47: 669-99.
Dawes, R., Dawson, I., Falciani, F., Tear, G. and Akam, M. (1994) 'Dax, a locust Hox gene related to fushi-tarazu but showing no pair-rule expression', *Development* 120(6): 1561-72.
de Rosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., Akam, M., Carroll, S. B. and Balavoine, G. (1999) 'Hox genes in brachiopods and priapulids and protostome evolution', *Nature* 399(6738): 772-6.

Dearden, P., Grbic, M., Falciani, F. and Akam, M. (2000) 'Maternal expression and early zygotic regulation of the Hox3/zen gene in the grasshopper Schistocerca gregaria', *Evol Dev* 2(5): 261-70.

Dearden, P. K., Wilson, M. J., Sablan, L., Osborne, P. W., Havler, M., McNaughton, E., Kimura, K., Milshina, N. V., Hasselmann, M., Gempe, T. et al. (2006) 'Patterns of conservation and change in honey bee developmental genes', *Genome Res* 16(11): 1376-84.

DeCamillis, M. A., Lewis, D. L., Brown, S. J., Beeman, R. W. and Denell, R. E. (2001) 'Interactions of the Tribolium Sex combs reduced and proboscipedia orthologs in embryonic labial development', *Genetics* 159(4): 1643-8.

Denell, R. E. (1973) 'Homoeosis in Drosophila. I. Complementation studies with revertants of Nasobemia', *Genetics* 75(2): 279-97.

Denell, R. E., Hummels, K. R., Wakimoto, B. T. and Kaufman, T. C. (1981) 'Developmental studies of lethality associated with the Antennapedia gene complex in *Drosophila melanogaster*', *Dev. Biol.* 81: 43-50.

Devenport, M. P., Blass, C. and Eggleston, P. (2000) 'Characterization of the Hox gene cluster in the malaria vector mosquito, Anopheles gambiae', *Evol Dev* 2(6): 326-39.

Doe, C. Q., Hiromi, Y., Gehring, W. J. and Goodman, C. S. (1988) 'Expression and function of the segmentation gene fushi tarazu during Drosophila neurogenesis', *Science* 239(4836): 170-5.

Duboule, D. (1994) *Guidebook to the Homeobox Genes*, Oxford: Oxford University Press.

Duncan, I. (1987) 'The bithorax complex', *Annual review of genetics* 21: 285-319.

Duncan, I. W. and Kaufman, T. C. (1975) 'Cytogenic analysis of chromosome 3 in Drosophila melanogaster: mapping of the proximal portion of the right arm', *Genetics* 80(4): 733-52.

Ebner, A., Cabernard, C., Affolter, M. and Merabet, S. (2005) 'Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex', *Development* 132(7): 1591-600.

Ekker, S. C., Young, K. E., vonKessler, D. P. and Beachy, P. A. (1991) 'Optimal DNA sequence recognition by the *Ultrabithorax* homeodomain of *Drosophila*', *EMBO J.* 10: 1179-1186.

Erezyilmaz, D. F., Kelstrup, H. C. and 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(1): 300-10.

Falciani, F., Hausdorf, B., Schroder, R., Akam, M., Tautz, D., Denell, R. and Brown, S. (1996) 'Class 3 Hox genes in insects and the origin of zen', *Proc Natl Acad Sci U S A* 93(16): 8479-84.

Farzana, L. and Brown, S. J. (2008) 'Hedgehog signaling pathway function conserved in Tribolium segmentation', *Development genes and evolution* 218(3-4): 181-92.

Ferrier, D. E. and Akam, M. (1996) 'Organization of the Hox gene cluster in the grasshopper, Schistocerca gregaria', *Proc Natl Acad Sci U S A* 93(23): 13024-9.

Florence, B., Guichet, A., Ephrussi, A. and Laughon, A. (1997) 'Ftz-F1 is a cofactor in Ftz activation of the Drosophila engrailed gene', *Development* 124(4): 839-47.

Force, A., Cresko, W. A., Pickett, F. B., Proulx, S. R., Amemiya, C. and Lynch, M. (2005) 'The origin of subfunctions and modular gene regulation', *Genetics* 170(1): 433-46.

Frank, L. H. and Rushlow, C. (1996) 'A group of genes required for maintenance of the amnioserosa tissue in Drosophila', *Development* 122(5): 1343-52.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. (1986) 'Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network', *Cell* 47(5): 735-46.

Frobius, A. C., Matus, D. Q. and Seaver, E. C. (2008) 'Genomic organization and expression demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan Capitella sp. I', *PLoS One* 3(12): e4004.
Furukubo-Tokunaga, K., Flister, S. and Gehring, W. J. (1993) 'Functional specificity of the Antennapedia homeodomain', *Proc Natl Acad Sci U S A* 90(13): 6360-4.
Galant, R. and Carroll, S. B. (2002) 'Evolution of a transcriptional repression domain in an insect Hox protein', *Nature* 415: 910-913.
Galant, R., Walsh, C. M. and Carroll, S. B. (2002) 'Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites', *Development* 129(13): 3115-26.
Garcia-Bellido, A. (1975) 'Genetic control of wing disk development in *Drosophila*', *Ciba Found. Symp.* 29: Cell Patterning: pp. 161-182.
Gehring, W. J. (1966) 'Bildung eines vollstandigen Mittlebeines mit Sternopleura in der Antennenregion ei der Mutante *Nasobemia* (*Ns*) von *Drosophila melanogaster*.', *Arch. Jul. Klaus Stift. Vererbungsforsch.* 41: 44-54.
Gehring, W. J. (1985) 'The homeo box: a key to the understanding of development?', *Cell* 40(1): 3-5.
Gehring, W. J. and Hiromi, Y. (1986) 'Homeotic genes and the homeobox', *Ann. Rev. Genet.* 20: 147-173.
Gehring, W. J., Kloter, U. and Suga, H. (2009) 'Evolution of the Hox gene complex from an evolutionary ground state', *Current topics in developmental biology* 88: 35-61.
Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G. and Wutrich, K. (1994) 'Homeodomain-DNA recognition', *Cell* 78: 211-233.
Gibson, G. (2000) 'Evolution: hox genes and the cellared wine principle', *Curr Biol* 10(12): R452-5.
Gibson, G., Schier, A., LeMotte, P. and Gehring, W. J. (1990) 'The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain', *Cell* 62(6): 1087-103.
Goldshmidt, R. (1940) *The Material Basis of Evolution*, New Haven: Yale University Press.
Goltsev, Y., Hsiong, W., Lanzaro, G. and Levine, M. (2004) 'Different combinations of gap repressors for common stripes in Anopheles and Drosophila embryos', *Dev Biol* 275(2): 435-46.
Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. and Carroll, S. B. (2005) 'Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila', *Nature* 433(7025): 481-7.
Gonzalez-Gaitan, M. A., Micol, J. L. and Garcia-Bellido, A. (1990) 'Developmental genetic analysis of Contrabithorax mutations in Drosophila melanogaster', *Genetics* 126(1): 139-55.
Greer, J. M., Puetz, J., Thomas, K. R. and Capecchi, M. R. (2000) 'Maintenance of functional equivalence during paralogous *Hox* gene evolution.', *Nature* 367: 83-87.
Grenier, J. K. and Carroll, S. B. (2000) 'Functional evolution of the Ultrabithorax protein', *Proc Natl Acad Sci U S A* 97(2): 704-9.
Grenier, J. K., Garber, T. L., Warren, R., Whitington, P. M. and Carroll, S. (1997) 'Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade', *Curr Biol* 7(8): 547-53.

Guichet, A., Copeland, J. W., Erdelyi, M., Hlousek, D., Zavorszky, P., Ho, J., Brown, S., Percival-Smith, A., Krause, H. M. and Ephrussi, A. (1997) 'The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors', *Nature* 385(6616): 548-52.

Guo, T., Mandai, K., Condie, B. G., Wickramasinghe, S. R., Capecchi, M. R. and Ginty, D. D. (2011) 'An evolving NGF-Hoxd1 signaling pathway mediates development of divergent neural circuits in vertebrates', *Nature neuroscience* 14(1): 31-6.

Gutjahr, T., Vanario Alonso, C. E., Pick, L. and Noll, M. (1994) 'Multiple regulatory elements direct the complex expression patterns of the *Drosophila* segmentation gene *paired*', *Mech. Dev.* 48: 119-128.

Hafen, E., Kuroiwa, A. and Gehring, W. J. (1984) 'Spatial distribution of transcripts from the segmentation gene fushi tarazu during Drosophila embryonic development', *Cell* 37(3): 833-41.

Hall, B. K. (2003) 'Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution', *Biol Rev Camb Philos Soc* 78(3): 409-33.

Hanes, S. D. and Brent, R. (1989) 'DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9', *Cell* 57(7): 1275-83.

Hawthorne, D. J. (2001) 'AFLP-based genetic linkage map of the Colorado potato beetle Leptinotarsa decemlineata: sex chromosomes and a pyrethroid-resistance candidate gene', *Genetics* 158(2): 695-700.

Hayashi, S. and Scott, M. P. (1990) 'What determines the specificity of action of Drosophila homeodomain proteins?', *Cell* 63(5): 883-94.

Heffer, A., Lohr, U. and Pick, L. (2011) 'ftz evolution: findings, hypotheses and speculations (response to DOI 10.1002/bies.201100019)', *BioEssays : news and reviews in molecular, cellular and developmental biology* 33(12): 910-8.

Heffer, A., Shultz, J. and Pick, L. (2010) 'Surprising flexibility in a conserved Hox transcription factor over 550 million years of evolution', *Proc. Natl. Acad. Sci. USA* 107: 18040-18045.

Hiromi, Y. and Gehring, W. J. (1987) 'Regulation and function of the Drosophila segmentation gene fushi tarazu', *Cell* 50(6): 963-74.

Hiromi, Y., Kuroiwa, A. and Gehring, W. J. (1985) 'Control elements of the *Drosophila* segmentation gene *fushi tarazu*', *Cell* 43: 603-613.

Hoekstra, H. E. and Coyne, J. A. (2007) 'The locus of evolution: evo devo and the genetics of adaptation', *Evolution* 61(5): 995-1016.

Hogness, D. S., Lipshitz, H. D., Beachy, P. A., Peattie, R. B., Saint, R. B., Goldschmidt-Clermont, M., Harte, P. J., Gavis, E. R. and Helfand, S. L. (1985) Regulation and Products of the *Ubx* Domain of the Bithorax Complex, vol. Cold Spring Harbor Symposia on Quantitative Biology, Molecular Biology of Development.

Hou, H. Y., Heffer, A., Anderson, W. R., Liu, J., Bowler, T. and Pick, L. (2009) 'Stripy Ftz target genes are coordinately regulated by Ftz-F1', *Dev Biol* 335(2): 442-53.

Howard, M. L. and Davidson, E. H. (2004) 'cis-Regulatory control circuits in development', *Developmental Biology* 271(1): 109-18.

Hrycaj, S., Chesebro, J. and Popadic, A. (2010) 'Functional analysis of Scr during embryonic and post-embryonic development in the cockroach, Periplaneta americana', *Developmental Biology* 341(1): 324-34.

Hsia, C. C. and McGinnis, W. (2003) 'Evolution of transcription factor function', *Curr Opin Genet Dev.* 13: 199-206.
Hsia, C. C., Pare, A. C., Hannon, M., Ronshaugen, M. and McGinnis, W. (2010) 'Silencing of an abdominal Hox gene during early development is correlated with limb development in a crustacean trunk', *Evol Dev* 12(2): 131-43.
Hughes, C. L. and 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(17): 3683-94.
Hughes, C. L. and Kaufman, T. C. (2002a) 'Exploring myriapod segmentation: the expression patterns of even-skipped, engrailed, and wingless in a centipede', *Dev Biol* 247(1): 47-61.
Hughes, C. L. and Kaufman, T. C. (2002b) 'Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede', *Development* 129(5): 1225-38.
Hughes, C. L., Liu, P. Z. and 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(6): 393-401.
Irvine, S. Q., Warinner, S. A., Hunter, J. D. and Martindale, M. Q. (1997) 'A survey of homeobox genes in Chaetopterus variopedatus and analysis of polychaete homeodomains', *Mol Phylogenet Evol* 7(3): 331-45.
Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Vanario-Alonso, C. E., Samsonova, M. et al. (2004) 'Dynamic control of positional information in the early Drosophila embryo', *Nature* 430(6997): 368-71.
Janssen, R. and Budd, G. E. (2010) 'Gene expression suggests conserved aspects of Hox gene regulation in arthropods and provides additional support for monophyletic Myriapoda', *Evodevo* 1(1): 4.
Janssen, R. and Damen, W. G. (2006) 'The ten Hox genes of the millipede Glomeris marginata', *Dev Genes Evol* 216(7-8): 451-65.
Johnson, F. B., Parker, E. and 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', *Proc Natl Acad Sci U S A* 92(3): 739-43.
Kaufman, T. C., Lewis, R. and Wakimoto, B. (1980) 'Cytogenetic Analysis of Chromosome 3 in DROSOPHILA MELANOGASTER: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B', *Genetics* 94(1): 115-33.
Kellerman, K. A., Mattson, D. M. and Duncan, I. (1990) 'Mutation affecting the stability of the *fushi tarazu* protein of *Drosophila*', *Genes & Development* 4: 1936-1950.
Khila, A., Abouheif, E. and Rowe, L. (2009) 'Evolution of a novel appendage ground plan in water striders is driven by changes in the Hox gene Ultrabithorax', *PLoS genetics* 5(7): e1000583.
Kosman, D. and Small, S. (1997) 'Concentration-dependent patterning by an ectopic expression domain of the Drosophila gap gene knirps', *Development* 124(7): 1343-54.
Kourakis, M. J., Master, V. A., Lokhorst, D. K., Nardelli-Haefliger, D., Wedeen, C. J., Matindale, M. Q. and Shankland, M. (1997) 'Conserved anterior boundaries of Hox gene expression in the central nervous system of the leech *Helodella*', *Dev. Biol.* 190: 284-300.
Kuroiwa, A., Hafen, E. and Gehring, W. J. (1984) 'Cloning and transcriptional analysis of the segmentation gene fushi tarazu of Drosophila', *Cell* 37(3): 825-31.
Lamka, M. L. and Lipshitz, H. D. (1999) 'Role of the amnioserosa in germ band retraction of the Drosophila melanogaster embryo', *Dev Biol* 214(1): 102-12.

Lawrence, P. A. (1992) *The Making of a Fly: The genetics of animal design*, Oxford: Blackwell Scientific Publications.

Lemke, S., Busch, S. E., Antonopoulos, D. A., Meyer, F., Domanus, M. H. and Schmidt-Ott, U. (2010) 'Maternal activation of gap genes in the hover fly Episyrphus', *Development* 137(10): 1709-19.

Lemke, S., Stauber, M., Shaw, P. J., Rafiqi, A. M., Prell, A. and Schmidt-Ott, U. (2008) 'Bicoid occurrence and Bicoid-dependent hunchback regulation in lower cyclorrhaphan flies', *Evolution & development* 10(4): 413-20.

LeMotte, P. K., Kuroiwa, A., Fessler, L. I. and Gehring, W. J. (1989) 'The homeotic gene *Sex combs reduced* of *Drosophila*: gene structure and embryonic expression', *EMBO J.* 8: 219-227.

Levine, M. (2010) 'Transcriptional enhancers in animal development and evolution', *Current biology : CB* 20(17): R754-63.

Levine, M. and Davidson, E. H. (2005) 'Gene regulatory networks for development', *Proceedings of the National Academy of Sciences of the United States of America* 102(14): 4936-42.

Lewis, E. B. (1963) 'Genes and developmental pathways', *Am. Zool.* 3: 33-56.

Lewis, E. B. (1978) 'A gene complex controlling segmentation in Drosophila', *Nature* 276(5688): 565-70.

Lewis, E. B. (1998) 'The bithorax complex: the first fifty years', *The International journal of developmental biology* 42(3): 403-15.

Lewis, R. A., Wakimoto, B. T., Denell, R. E. and 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.

Lin, L. and McGinnis, W. (1992) 'Mapping functional specificity in the Dfd and Ubx homeo domains', *Genes Dev* 6(6): 1071-81.

Liu, P. Z. and 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(9): 2081-92.

Lohr, U. and Pick, L. (2005) 'Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of Drosophila melanogaster', *Curr Biol* 15(7): 643-9.

Lohr, U., Yussa, M. and Pick, L. (2001) 'Drosophila fushi tarazu. a gene on the border of homeotic function', *Curr Biol* 11(18): 1403-12.

Lynch, J. and Desplan, C. (2003) 'Evolution of development: beyond bicoid', *Curr Biol* 13(14): R557-9.

Lynch, M. and Force, A. (2000) 'The probability of duplicate gene preservation by subfunctionalization', *Genetics* 154(1): 459-73.

Lynch, V. J., Brayer, K., Gellersen, B. and Wagner, G. P. (2009) 'HoxA-11 and FOXO1A cooperate to regulate decidual prolactin expression: towards inferring the core transcriptional regulators of decidual genes', *PLoS One* 4(9): e6845.

Lynch, V. J., Roth, J. J., Takahashi, K., Dunn, C. W., Nonaka, D. F., Stopper, G. F. and Wagner, G. P. (2004) 'Adaptive evolution of HoxA-11 and HoxA-13 at the origin of the uterus in mammals', *Proceedings. Biological sciences / The Royal Society* 271(1554): 2201-7.

Lynch, V. J., Tanzer, A., Wang, Y., Leung, F. C., Gellersen, B., Emera, D. and Wagner, G. P. (2008) 'Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals', *Proc Natl Acad Sci U S A* 105(39): 14928-33.

Lynch, V. J. and Wagner, G. P. (2008) 'Resurrecting the role of transcription factor change in developmental evolution', *Evolution* 62(9): 2131-54.

Mahaffey, J. W. and Kaufman, T. C. (1987) 'Distribution of the *Sex combs reduced* gene products in *Drosophila melanogaster*', *Genetics* 117: 51-60.

Mahfooz, N., Turchyn, N., Mihajlovic, M., Hrycaj, S. and Popadic, A. (2007) 'Ubx regulates differential enlargement and diversification of insect hind legs', *PLoS One* 2(9): e866.

Mahfooz, N. S., Li, H. and Popadic, A. (2004) 'Differential expression patterns of the hox gene are associated with differential growth of insect hind legs', *Proceedings of the National Academy of Sciences of the United States of America* 101(14): 4877-82.

Maier, D., Preiss, A. and Powell, J. R. (1990) 'Regulation of the segmentation gene fushi tarazu has been functionally conserved in Drosophila', *EMBO J* 9(12): 3957-66.

Malicki, J., Bogarad, L. D., Martin, M. M., Ruddle, F. H. and McGinnis, W. (1993) 'Functional analysis of the mouse homeobox gene *HoxB9* in *Drosophila* development', *Mech. Dev.* 42: 139-50.

Malicki, J., Schughart, K. and McGinnis, W. (1990) 'Mouse *Hox-2.2* specifies thoracic segmental identity in *Drosophila* embryos and larvae', *Cell* 63: 961-967.

Mann, R. and Carroll, S. (2002) 'Molecular mechanisms of selector gene function and evolution', *Curr Opin Genet Dev* 12(5): 592.

Mann, R. S. (1995) 'The specificity of homeotic gene function', *Bioessays* 17(10): 855-63.

Mann, R. S. and Chan, S.-K. (1996) 'Extra specificity from *extradenticle*: the partnership between HOX and PBX/EXD homeodomain proteins', *TIGS* 12: 258-262.

Mann, R. S., Lelli, K. M. and Joshi, R. (2009) 'Hox specificity unique roles for cofactors and collaborators', *Current topics in developmental biology* 88: 63-101.

Manzanares, M., Marco, R. and Garesse, R. (1993) 'Genomic organization and developmental pattern of expression of the engrailed gene from the brine shrimp Artemia', *Development* 118(4): 1209-19.

Marshall, C. R., Raff, E. C. and Raff, R. A. (1994) 'Dollo's law and the death and resurrection of genes', *Proc Natl Acad Sci U S A* 91(25): 12283-7.

McDonald, J. A., Fujioka, M., Odden, J. P., Jaynes, J. B. and Doe, C. Q. (2003) 'Specification of motoneuron fate in Drosophila: integration of positive and negative transcription factor inputs by a minimal eve enhancer', *Journal of neurobiology* 57(2): 193-203.

McGinnis, N., Kuziora, M. A. and McGinnis, W. (1990) 'Human *Hox-4.2* and *Drosophila Deformed* encode similar regulatory specificities in *Drosophila* embryos and larvae', *Cell* 63: 969-976.

McGinnis, W. (1994) 'A century of homeosis, a decade of homeoboxes', *Genetics* 137(3): 607-11.

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. and Gehring, W. J. (1984a) 'A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans', *Cell* 37(2): 403-8.

McGinnis, W. and Krumlauf, R. (1992) 'Homeobox genes and axial patterning', *Cell* 68(2): 283-302.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J. (1984b) 'A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes', *Nature* 308(5958): 428-33.

McGregor, A. P. (2005) 'How to get ahead: the origin, evolution and function of bicoid', *Bioessays* 27(9): 904-13.

Miko, I., Friedrich, F., Yoder, M. J., Hines, H. M., Deitz, L. L., Bertone, M. A., Seltmann, K. C., Wallace, M. S. and Deans, A. R. (2012) 'On dorsal prothoracic appendages in treehoppers (Hemiptera: Membracidae) and the nature of morphological evidence', *PLoS One* 7(1): e30137.

Moczek, A. P. (2008) 'On the origins of novelty in development and evolution', *BioEssays : news and reviews in molecular, cellular and developmental biology* 30(5): 432-47.

Moczek, A. P. (2010) 'Phenotypic plasticity and diversity in insects', *Philos Trans R Soc Lond B Biol Sci* 365(1540): 593-603.

Moczek, A. P., Andrews, J., Kijimoto, T., Yerushalmi, Y. and Rose, D. J. (2007) 'Emerging model systems in evo-devo: horned beetles and the origins of diversity', *Evolution & development* 9(4): 323-8.

Mouchel-Vielh, E., Blin, M., Rigolot, C. and Deutsch, J. S. (2002) 'Expression of a homologue of the fushi tarazu (ftz) gene in a cirripede crustacean', *Evol Dev* 4(2): 76-85.

Neuteboom, S. T., Peltenburg, L. T., van Dijk, M. A. and Murre, C. (1995) 'The hexapeptide LFPWMR in Hoxb-8 is required for cooperative DNA binding with Pbx1 and Pbx2 proteins', *Proc Natl Acad Sci U S A* 92(20): 9166-70.

Nusslein-Volhard, C. and Wieschaus, E. (1980) 'Mutations affecting segment number and polarity in Drosophila', *Nature* 287(5785): 795-801.

Ohno, S. (1970) *Evolution by gene duplication*, Berlin: Springer-Verlag.

Panfilio, K. A. (2009) 'Late extraembryonic morphogenesis and its zen(RNAi)-induced failure in the milkweed bug Oncopeltus fasciatus', *Dev Biol* 333(2): 297-311.

Panfilio, K. A. and Akam, M. (2007) 'A comparison of Hox3 and Zen protein coding sequences in taxa that span the Hox3/zen divergence', *Dev Genes Evol* 217(4): 323-9.

Panfilio, K. A., Liu, P. Z., Akam, M. and Kaufman, T. C. (2006) 'Oncopeltus fasciatus zen is essential for serosal tissue function in katatrepsis', *Dev Biol* 292(1): 226-43.

Panfilio, K. A. and Roth, S. (2010) 'Epithelial reorganization events during late extraembryonic development in a hemimetabolous insect', *Dev Biol* 340(1): 100-15.

Papillon, D. and Telford, M. J. (2007) 'Evolution of Hox3 and ftz in arthropods: insights from the crustacean Daphnia pulex', *Dev Genes Evol* 217(4): 315-22.

Passalacqua, K. D., Hrycaj, S., Mahfooz, N. and Popadic, A. (2010) 'Evolving expression patterns of the homeotic gene Scr in insects', *The International journal of developmental biology* 54(5): 897-904.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S. and Aggarwal, A. K. (1999) 'Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex', *Nature* 397(6721): 714-9.

Patel, N. H., Ball, E. E. and Goodman, C. S. (1992) 'Changing role of even-skipped during the evolution of insect pattern formation', *Nature* 357(6376): 339-42.

Patel, N. H., Kornberg, T. B. and Goodman, C. S. (1989) 'Expression of engrailed during segmentation in grasshopper and crayfish', *Development* 107(2): 201-12.

Patraquim, P., Warnefors, M. and Alonso, C. R. (2011) 'Evolution of Hox Post-Transcriptional Regulation by Alternative Polyadenylation and MicroRNA Modulation Within 12 Drosophila Genomes', *Molecular Biology and Evolution* 28(9): 2453-60.

Pearson, J. C., Lemons, D. and McGinnis, W. (2005) 'Modulating Hox gene functions during animal body patterning', *Nat Rev Genet* 6(12): 893-904.

Petruk, S., Sedkov, Y., Riley, K. M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J. B., Brock, H. W. and Mazo, A. (2006) 'Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference', *Cell* 127(6): 1209-21.

Phelan, M. L., Rambaldi, I. and Featherstone, M. S. (1995) 'Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif', *Mol Cell Biol* 15(8): 3989-97.

Pick, L. and Heffer, A. (2012) 'Hox gene evolution: multiple mechanisms contributing to evolutionary novelties', *Annals of the New York Academy of Sciences*.

Popadic, A., Abzhanov, A., Rusch, D. and Kaufman, T. C. (1998) 'Understanding the genetic basis of morphological evolution: the role of homeotic genes in the diversification of the arthropod bauplan', *The International journal of developmental biology* 42(3): 453-61.

Posnien, N., Schinko, J., Grossmann, D., Shippy, T. D., Konopova, B. and Bucher, G. (2009) 'RNAi in the red flour beetle (Tribolium)', *Cold Spring Harbor protocols* 2009(8): pdb prot5256.

Postlethwait, J. H. and Schneiderman, H. A. (1971) 'Pattern formation and determination in the antenna of the homoeotic mutant Antennapedia of Drosophila melanogaster', *Developmental Biology* 25(4): 606-40.

Powers, T. P., Hogan, J., Ke, Z., Dymbrowski, K., Wang, X., Collins, F. H. and Kaufman, T. C. (2000) 'Characterization of the Hox cluster from the mosquito Anopheles gambiae (Diptera: Culicidae)', *Evol Dev* 2(6): 311-25.

Project, B. D. G. (2012).

Prud'homme, B., Gompel, N. and Carroll, S. B. (2007) 'Emerging principles of regulatory evolution', *Proc Natl Acad Sci U S A* 104 Suppl 1: 8605-12.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A., Williams, T. M., Yeh, S. D., True, J. R. and Carroll, S. B. (2006) 'Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene', *Nature* 440(7087): 1050-3.

Prud'homme, B., Minervino, C., Hocine, M., Cande, J. D., Aouane, A., Dufour, H. D., Kassner, V. A. and Gompel, N. (2011) 'Body plan innovation in treehoppers through the evolution of an extra wing-like appendage', *Nature* 473(7345): 83-6.

Rafiqi, A. M., Lemke, S., Ferguson, S., Stauber, M. and Schmidt-Ott, U. (2008) 'Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of zen', *Proc Natl Acad Sci U S A* 105(1): 234-9.

Regier, J. C., Shultz, J. W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J. W. and Cunningham, C. W. (2010) 'Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences', *Nature* 463: 1079-1083.

Riley, P. D., Carroll, S. B. and Scott, M. P. (1987) 'The expression and regulation of Sex combs reduced protein in *Drosophila* embryos', *Genes & Dev.* 1: 716-730.

Rogers, B. T. and Kaufman, T. C. (1996) 'Structure of the insect head as revealed by the EN protein pattern in developing embryos', *Development* 122(11): 3419-32.

Rogers, B. T., Peterson, M. D. and Kaufman, T. C. (1997) 'Evolution of the insect body plan as revealed by the Sex combs reduced expression pattern', *Development* 124(1): 149-57.

Ronshaugen, M., McGinnis, N. and McGinnis, W. (2002) 'Hox protein mutation and macroevolution of the insect body plan', *Nature* 415(6874): 914-7.
Ruaud, A. F., Lam, G. and Thummel, C. S. (2010) 'The Drosophila nuclear receptors DHR3 and betaFTZ-F1 control overlapping developmental responses in late embryos', *Development* 137(1): 123-31.
Rushlow, C., Doyle, H., Hoey, T. and Levine, M. (1987a) 'Molecular characterization of the zerknullt region of the Antennapedia gene complex in Drosophila', *Genes Dev* 1(10): 1268-79.
Rushlow, C., Frasch, M., Doyle, H. and Levine, M. (1987b) 'Maternal regulation of zerknullt: a homoeobox gene controlling differentiation of dorsal tissues in Drosophila', *Nature* 330(6148): 583-6.
Saenko, S. V., Marialva, M. S. and Beldade, P. (2011) 'Involvement of the conserved Hox gene Antennapedia in the development and evolution of a novel trait', *Evodevo* 2: 9.
Schinko, J., Posnien, N., Kittelmann, S., Koniszewski, N. and Bucher, G. (2009) 'Single and double whole-mount in situ hybridization in red flour beetle (Tribolium) embryos', *Cold Spring Harbor protocols* 2009(8): pdb prot5258.
Schlosser, G. and Wagner, G. P. (2004) *Modularity in Development and Evolution*, Chicago: University of Chicago Press.
Schmidt-Ott, U. (2005) 'Insect serosa: a head line in comparative developmental genetics', *Curr Biol* 15(7): R245-7.
Schmidt-Ott, U. and Wimmer, E. A. (2004) Starting the segmentation gene cascade in insects. in G. P. Wagner (ed.) *Modularity in Development and Evolution*. Chicago: The University of Chicago Press.
Schneuwly, S., Klemenz, R. and Gehring, W. J. (1987a) 'Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia', *Nature* 325(6107): 816-8.
Schneuwly, S., Kuroiwa, A., Baumgartner, P. and Gehring, W. J. (1986) 'Structural organization and sequence of the homeotic gene Antennapedia of Drosophila melanogaster', *The EMBO journal* 5(4): 733-9.
Schneuwly, S., Kuroiwa, A. and Gehring, W. J. (1987b) 'Molecular analysis of the dominant homeotic Antennapedia phenotype', *The EMBO journal* 6(1): 201-6.
Schroeder, M. D., Pearce, M., Fak, J., Fan, H., Unnerstall, U., Emberly, E., Rajewsky, N., Siggia, E. D. and Gaul, U. (2004) 'Transcriptional control in the segmentation gene network of Drosophila', *PLoS biology* 2(9): E271.
Schwartz, C. J., Sampson, H. M., Hlousek, D., Percival-Smith, A., Copeland, J. W., Simmonds, A. J. and Krause, H. M. (2001) 'FTZ-Factor1 and Fushi tarazu interact via conserved nuclear receptor and coactivator motifs', *The EMBO journal* 20(3): 510-9.
Scott, M. P. (1992) 'Vertebrate homeobox gene nomenclature', *Cell* 71: 551-553.
Scott, M. P. (1993) 'A rational nomenclature for vertebrate homeobox (HOX) genes', *Nucleic Acids Res* 21(8): 1687-8.
Scott, M. P. and Weiner, A. J. (1984) 'Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila', *Proc Natl Acad Sci U S A* 81(13): 4115-9.
Shippy, T. D., Rogers, C. D., Beeman, R. W., Brown, S. J. and Denell, R. E. (2006) 'The Tribolium castaneum ortholog of Sex combs reduced controls dorsal ridge development', *Genetics* 174(1): 297-307.

Shubin, N., Tabin, C. and Carroll, S. (2009) 'Deep homology and the origins of evolutionary novelty', *Nature* 457(7231): 818-23.
Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H. J. et al. (2011) 'Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins', *Cell* 147(6): 1270-82.
Sprules, T., Green, N., Featherstone, M. and Gehring, K. (2003) 'Lock and key binding of the HOX YPWM peptide to the PBX homeodomain', *J Biol Chem* 278(2): 1053-8.
Stern, D. L. and Orgogozo, V. (2008) 'The loci of evolution: how predictable is genetic evolution?', *Evolution* 62(9): 2155-77.
Stern, D. L. and Orgogozo, V. (2009) 'Is genetic evolution predictable?', *Science* 323(5915): 746-51.
Struhl, g. (1982) 'Genes controlling segmental specification in the *Drosophila* thorax', *Proc. Natl. Acad. Sci. USA* 79: 7380-7384.
Struhl, G. (1984) 'A universal genetic key to body plan?', *Nature* 310(5972): 10-1.
Struhl, G. (1985) 'Near-reciprocal phenotypes caused by inactivation or indiscriminate expression of the *Drosophila* segmentation gene *ftz*', 318: 677-680.
Stuart, J. J., Brown, S. J., Beeman, R. W. and Denell, R. E. (1991) 'A deficiency of the homeotic complex of the beetle Tribolium', *Nature* 350(6313): 72-4.
Subramanian, R. A., Arensburger, P., Atkinson, P. W. and O'Brochta, D. A. (2007) 'Transposable element dynamics of the hAT element Herves in the human malaria vector Anopheles gambiae s.s', *Genetics* 176(4): 2477-87.
Suzuki, T., Kawasaki, H., Yu, R. T., Ueda, H. and Umesono, K. (2001) 'Segmentation gene product Fushi tarazu is an LXXLL motif-dependent coactivator for orphan receptor FTZ-F1', *Proceedings of the National Academy of Sciences of the United States of America* 98(22): 12403-8.
Suzuki, T., Kawasaki, H., Yu , R. T., Ueda, H. and 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.
Tautz, D. and Pfeifle, C. (1989) *Chromosoma* 98: 81-85.
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(6): 349-52.
Telford, M. J. and Thomas, R. H. (1998) 'Of mites and zen: expression studies in a chelicerate arthropod confirm zen is a divergent Hox gene', *Dev Genes Evol* 208(10): 591-4.
Thomsen, S., Azzam, G., Kaschula, R., Williams, L. S. and Alonso, C. R. (2010) 'Developmental RNA processing of 3'UTRs in Hox mRNAs as a context-dependent mechanism modulating visibility to microRNAs', *Development* 137(17): 2951-60.
Tomoyasu, Y., Wheeler, S. R. and Denell, R. E. (2005) 'Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum', *Nature* 433(7026): 643-7.
Tour, E., Hittinger, C. T. and McGinnis, W. (2005) 'Evolutionarily conserved domains required for activation and repression functions of the Drosophila Hox protein Ultrabithorax', *Development* 132(23): 5271-81.
Tschopp, P., Fraudeau, N., Bena, F. and Duboule, D. (2011) 'Reshuffling genomic landscapes to study the regulatory evolution of Hox gene clusters', *Proceedings of the National Academy of Sciences of the United States of America* 108(26): 10632-7.

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

Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M. E., Botas, J. and Cohen, S. M. (1992) 'Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less', *Cell* 71(3): 437-50.

van der Zee, M., Berns, N. and Roth, S. (2005) 'Distinct functions of the Tribolium zerknullt genes in serosa specification and dorsal closure', *Curr Biol* 15(7): 624-36.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. et al. (1995) 'AFLP: a new technique for DNA fingerprinting', *Nucleic Acids Res* 23(21): 4407-14.

Wagner, G. P., Amemiya, C. and Ruddle, F. (2003) 'Hox cluster duplications and the opportunity for evolutionary novelties', *Proc Natl Acad Sci U S A* 100(25): 14603-6.

Wagner, G. P. and Lynch, V. J. (2008) 'The gene regulatory logic of transcription factor evolution', *Trends Ecol Evol* 23(7): 377-85.

Wakimoto, B. T. and Kaufman, T. C. (1981) 'Analysis of larval segmentation genotypes associated with the *Antennapedia* gene complex in *Drosophila melanogaster*', *Dev. Biol.* 81: 51-64.

Wakimoto, B. T., Turner, F. R. and Kaufman, T. C. (1984) 'Defects in embryogenesis in mutants associated with the antennapedia gene complex of Drosophila melanogaster', *Dev Biol* 102(1): 147-72.

Walldorf, U., Fleig, R. and Gehring, W. J. (1989) 'Comparison of homeobox-containing genes of the honeybee and Drosophila', *Proceedings of the National Academy of Sciences of the United States of America* 86(24): 9971-5.

Warren, R. W., Nagy, L., Selegue, J., Gates, J. and Carroll, S. (1994) 'Evolution of homeotic gene regulation and function in flies and butterflies', *Nature* 372(6505): 458-61.

Wasik, B. R., Rose, D. J. and Moczek, A. P. (2010) 'Beetle horns are regulated by the Hox gene, Sex combs reduced, in a species- and sex-specific manner', *Evolution & development* 12(4): 353-62.

Weatherbee, S. D., Halder, G., Kim, J., Hudson, A. and Carroll, S. (1998) 'Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere', *Genes & Development* 12(10): 1474-82.

Weatherbee, S. D., Nijhout, H. F., Grunert, L. W., Halder, G., Galant, R., Selegue, J. and Carroll, S. (1999) 'Ultrabithorax function in butterfly wings and the evolution of insect wing patterns', *Current biology : CB* 9(3): 109-15.

Wittkopp, P. J. (2010) 'Variable transcription factor binding: a mechanism of evolutionary change', *PLoS biology* 8(3): e1000342.

Wittkopp, P. J., Haerum, B. K. and Clark, A. G. (2008) 'Independent effects of cis- and trans-regulatory variation on gene expression in Drosophila melanogaster', *Genetics* 178(3): 1831-5.

Wray, G. A. (2007) 'The evolutionary significance of cis-regulatory mutations', *Nat Rev Genet* 8(3): 206-16.

Wu, S., Ying, G., Wu, Q. and Capecchi, M. R. (2007) 'Toward simpler and faster genome-wide mutagenesis in mice', *Nature Genetics* 39(7): 922-30.

Xu, J., Tan, A. and 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(10): 1471-80.

Yasukochi, Y., Ashakumary, L. A., Wu, C., Yoshido, A., Nohata, J., Mita, K. and Sahara, K. (2004) 'Organization of the Hox gene cluster of the silkworm, Bombyx mori: a split of the Hox cluster in a non-Drosophila insect', *Dev Genes Evol* 214(12): 606-14.

Yu, Y., Li, W., Su, K., Yussa, M., Han, W., Perrimon, N. and Pick, L. (1997) 'The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz', *Nature* 385(6616): 552-5.

Yussa, M., Lohr, U., Su, K. and Pick, L. (2001) 'The nuclear receptor Ftz-F1 and homeodomain protein Ftz interact through evolutionarily conserved protein domains', *Mech. Dev.* 107(1-2): 39-53.

Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A. and Scott, M. P. (1993) 'Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity', *Development* 118(2): 339-52.

Zhao, J. J., Lazzarini, R. A. and Pick, L. (1993) 'The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila Sex combs reduced* gene', *Genes & Dev.* 7: 343-354.

Zhao, J. J., Lazzarini, R. A. and Pick, L. (1996) 'Functional dissection of the mouse Hox-a5 gene', *The EMBO journal* 15(6): 1313-22.