Title
Gene expression and evolution

Permalink
https://escholarship.org/uc/item/8vn2k5b0

Author
Lemons, Derek Scott

Publication Date
2010

Peer reviewed|Thesis/dissertation
Gene expression and evolution

A dissertation submitted in partial satisfaction of the requirements for
the degree Doctor of Philosophy

in

Biology

by

Derek Scott Lemons

Committee in charge:

Professor William McGinnis, Chair
Professor Xiang-Dong Fu
Professor Amy Pasquinelli
Professor Michael G. Rosenfeld
Professor Steven Wasserman

2010
The dissertation of Derek Scott Lemons is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

Chair

University of California, San Diego

2010
DEDICATION

To my father who showed me the wonder that is the natural world

To my mother who taught me to love and be loved
EPIGRAPH

_Somewhere, something incredible is waiting to be known._

Carl Sagan

_If we knew what it was we were doing it would not be called research._

Albert Einstein

_I'm sciencing as fast as I can!_

Professor Farnsworth
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Epigraph</td>
<td>v</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>x</td>
</tr>
<tr>
<td>Vita and Publications</td>
<td>xii</td>
</tr>
<tr>
<td>Abstract of the Dissertation</td>
<td>xiv</td>
</tr>
</tbody>
</table>

**Chapter I** Evolution, expression and function of the Hox complex miRNAs miR-10 and miR-10*  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Results</td>
<td>7</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>28</td>
</tr>
<tr>
<td>Figures</td>
<td>29</td>
</tr>
</tbody>
</table>

**Chapter II** Co-option of an anteroposterior head axis gene network for proximodistal patterning of appendages in early bilaterian evolution  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>51</td>
</tr>
<tr>
<td>Results</td>
<td>57</td>
</tr>
</tbody>
</table>
Chapter III Evolution of homeodomain proteins in early and pre-metazoans 65

Introduction 66

Results 69

Discussion 70

Materials and Methods 72

Figures 73

Appendix 77

References 98
LIST OF FIGURES

Chapter I

Figure 1. Alignment of available mir-10 hairpin sequences ........................................ 29
Figure 2. Schematic representations of Hox complexes and their miRNAs ............... 34
Figure 3. Alignment of hairpin sequences for miR-10 and related miRNAs .............. 35
Figure 4. Evolution of miR-10 and related miRNAs ................................................... 36
Figure 5. miR-10 precursor molecules and regulation ............................................... 37
Figure 6. pri-miR-10 in situ hybridization through embryogenesis ....................... 38
Figure 7. Conservation of mature miRNA sequences in miR-10 hairpins .................. 39
Figure 8. Both miR-10 and miR-10* are expressed as mature miRNAs during Drosophila embryogenesis ................................................................................................. 40
Figure 9. miR-10 has evolutionarily conserved target sites in the 3’UTRs of insect Sex combs reduced (Scr) orthologs ............................................................................. 42
Figure 10. miR-10* has evolutionarily conserved target sites in the 3’UTRs of insect Abdominal-B (Abd-B) orthologs ................................................................. 43
Figure 11. Scr protein and cytoplasmic transcript accumulation are downregulated in ventral T1 ........................................................................................................... 44
Figure 12. Mature miRNA is produced from transgenes and results in aberrant phenotypes ................................................................................................................. 45
Figure 13. pri-miR-10 is expressed in a pattern complementary to predicted targets . 46
Figure 14. Overexpression or lack of pri-miR-10 has no observable effect on putative target protein expression ................................................................. 48
Figure 15. Target site mutations for luciferase experiments ...................................... 49
Chapter II

Figure 16. Expression of core appendage network genes in anterior neurectoderm in Drosophila embryos ................................................................. 62

Figure 17. Expression of core network genes in Saccoglossus embryos ............... 63

Figure 18. Schematic diagram comparing expression of core network genes in limbs and anterior neurectoderm .......................................................... 64

Chapter III

Figure 19. Alignment of Homeodomains from various opisthokonts ....................... 73

Figure 20. Phylogenetic tree of opisthokont homeodomains .............................. 76
ACKNOWLEDGEMENTS

I am grateful to have found Bill McGinnis as a mentor. His outlook on science has shaped my development as a biologist, and has kept me from giving up even when I felt I could give no more. His fascination with all things evolutionary always piques my interest, and his willingness to allow his students to explore a project on their own helped me mature as a scientist.

I feel incredibly fortunate to have worked in a lab with such wonderful and supportive people. I would like to thank all the members of the McGinnis lab past and present. Special thanks to Dave Kosman and Cheryl Hsia for immeasurable assistance with materials and protocols without which I would have never accomplished any of this. Special thanks as well to Adam Pare, for all the beers at the pub, and all the non-scientific talks without which I fear I would no longer be sane.

My friends and family are the most important part of my life. Their support and encouragement throughout this lengthy process has kept me going through all the tough times and helped me to finish this manuscript.


anteroposterior head axis gene network for proximodistal patterning of appendages in early bilaterian evolution.

VITA

1999 Bachelor of Science, University of California, Santa Barbara
1999-2002 Research Technician, The Scripps Research Institute
2002-2009 Graduate Student, University of California, San Diego
2010 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS


Lemons, D., Sun, X., Zou, Z., Campagnolo, L., Xiong, J., Kuhnert, F., English, M. A., Licht, J. D., Stuhlmann, H. (submitted) Vascular Endothelial Zinc Finger 1 (VEZF1) is an endothelial transcriptional transactivator that displays sequence-specific DNA binding to the IL-3, FLK-1, and FLT-1 promoters.

Lemons, D., Pare, A., McGinnis, W. (in preparation) miRNAs from the *Drosophila* Hox complex have undetectable effects on the regulation of evolutionarily conserved Hox target genes.
The control of the spatial expression of gene products is one of the most important and complex processes contributing to the embryogenesis of metazoans. This control can be achieved through many levels of regulation. Transcription factors such as
homeodomain proteins are important regulators of the phenotypic output of cells in which they are expressed. The Hox genes are a genomic cluster of homeodomain containing master control proteins, each of which is expressed in, and controls the morphology of a specific anteroposterior domain. While transcription factors have long been known to be an important type of regulatory molecule, recent discoveries have shown that there is always more out there for biology to teach us. The discovery of microRNAs has led to a whole new paradigm, in which the expression of many genes is controlled by the activity of tiny RNAs that were once invisible to researchers.

The Hox complexes of bilaterian animals typically contain a few microRNAs, but nearly all of them have the mir-10 gene. This ancient regulatory molecule is conserved in its sequence, genomic position, and to some extent its spatial expression over evolutionary time scales. In Drosophila melanogaster, this gene produces two mature microRNA products, each of which has strong complementarity to conserved sequences in Hox genes. These microRNAs are also expressed in patterns that are complementary to these potential target genes. Despite these correlations, disruption of the normal expression of the mir-10 gene does not strongly effect the expression of these genes, but may instead function to canalize their expression.

Appendages are structures common to a wide variety of animal body plans. When compared between distantly related animals there are no obvious structural characteristics which suggest they are homologous, yet they are patterned by similar transcription factors. Since these structures are not likely ancestral, their seemingly conserved patterning mechanism may have been acquired through independent co-option of a pre-existing network of genes for a similar purpose. The core genes responsible for patterning
the proximodistal axis of appendages are also expressed in domains of the anterior neurectoderm that are conserved among distantly related animals and function to pattern the anteroposterior axis of the head. This ancient gene network was likely co-opted in order to pattern the axis of appendages during the early evolution of bilaterian animals.

The advent of multicellularity has provided our planet with a rich variety of organisms. It is unknown what was necessary to allow single celled organisms to give rise eventually to the people that would research them, but there are many proteins that are assumed to be important for this process. Multicellular organisms are composed of many different types of cells, and homeodomain transcription factors are often involved in the specification of these cell types. The nearest relatives to animals are choanoflagellates, which are simple single celled organisms that sometimes exist in colonies. Although most animals have a specific set of homeodomain containing genes, most these are either evolutionary novelties specific to animals or have been secondarily lost in choanoflagellates. For this reason it is unknown what role homeodomain proteins played in the process of multicellularization that gave rise to animals, but further studies may provide exciting answers.
Chapter I

Evolution, expression and function of the Hox complex miRNAs

miR-10 and miR-10*
INTRODUCTION

MicroRNAs (miRNAs) are a recently discovered class of regulatory molecule that has dramatically enhanced the standard view of gene regulation. These ~22-nucleotide RNAs are made from transcripts that are generated by RNA polymerase II in a two-step cleavage process that is catalyzed by the nuclear RNaseIII Drosha and the cytoplasmic RNaseIII Dicer (Bartel, 2004; Pasquinelli et al., 2005). After processing, the mature miRNA is packaged into a functional RNA-Induced Silencing Complex (RISC). The RISC can catalyze precise endonucleolytic cleavage of target RNAs if they have perfect or near perfect complementarity to RISC-bound miRNAs, or reduce protein production by unknown post-transcriptional mechanisms if there is imperfect complementarity.

Considering the large number of identified miRNAs, few validated in vivo targets are known at present. However, there are indications that Hox genes might be an important class of miRNA targets.

For some time, there has been evidence for post-transcriptional regulation of Hox gene expression. Mouse HOXB4 protein and Hoxb4 transcript patterns of accumulation were similar in somites 7–13 and in the posterior hindbrain of developing embryos, but transcripts were also detected in the posterior neural tube, whereas protein was not (Brend et al., 2003). In addition, there is evidence for post-transcriptional regulation of HOXC6 in the chick embryonic hindlimb (Nelson et al., 1996). There is also evidence that post-transcriptional regulation of the Sex combs reduced (Scr) ortholog in the crustacean Porcellio scaber might be involved in modification of the first thoracic segment to produce gnathal appendages as opposed to walking legs (Abzhanov and Kaufman, 1999). It is unknown whether these discrepancies between transcript and
protein expression patterns are due to miRNA regulation or other mechanisms such as localized protein instability.

Although a few Hox genes are post-transcriptionally regulated by unknown mechanisms, many Hox genes have been predicted \textit{in silico} to be direct targets of miRNAs in both vertebrates (Lewis et al., 2003) and invertebrates (Enright et al., 2003). On the basis of partial complementarity between miRNA sequence and 3’UTR sequences, \textit{Drosophila melanogaster} Scr, Antp, Ubx, abd-A and Abd-B transcripts are proposed targets of miRNA regulation (Enright et al., 2003). If true, this indicates that miRNA regulation of Hox protein expression, at least in flies, is the rule rather than the exception.

One validated Hox miRNA target is the protein-encoding Hox transcript from mouse \textit{Hoxb8}. The \textit{Hoxb8} transcript contains a near perfect miR-196 target site in its 3’UTR, and analysis of mouse embryonic RACE clones showed that some of the \textit{Hoxb8} transcripts are cleaved endonucleolytically at the tenth nucleotide of the miRNA target site (Mansfield et al., 2004; Yekta et al., 2004), which is typical for sites with perfect complementarity to miRNA regulators. This target site is conserved in vertebrate \textit{Hoxb8} genes, but it is not known what influence this miRNA-mediated regulation exerts on vertebrate axial morphology.

In addition to miR-196, other miRNAs that are encoded within Hox gene clusters might also be regulators of Hox protein expression. The \textit{mir-10} gene is phylogenetically conserved and is present in a conserved position in both insect and vertebrate Hox clusters, between the \textit{Dfd/Hox4} and \textit{Scr/Hox5} genes. In mammals, the \textit{mir-10a} gene maps adjacent to \textit{Hoxb4}, and the \textit{mir-10b} gene is adjacent to \textit{Hoxd4}. Additionally, at least one
other miRNA gene maps among the posterior Hox genes in both insect and vertebrate genomes, but the *Drosophila mir-iab-4* and vertebrate *mir-196* miRNA genes do not seem to be structural orthologs, as they have only fragmentary sequence identity. Although not as clear cut as the case for *Hoxb8* regulation, evidence is accumulating that the Hox miRNAs are regulators of Hox genes (Bender, 2008; McGlinn et al., 2009; Ronshaugen et al., 2005; Stark et al., 2008; Woltering and Durston, 2008).

The spatial expression patterns of Hox-cluster miRNAs suggest that they have roles in A–P axial patterning. One way to detect the expression domains of miRNAs is to use ubiquitously transcribed reporter transgenes with 3’ UTRs that contain sequences perfectly complementary to putative miRNA trans-regulators (miRNA sensors). The body region in which reporter protein expression is abolished is inferred to represent the expression pattern of the miRNA that is complementary to the target site (Brennecke et al., 2003). This strategy has allowed a rough mapping of spatial domains of miR-10a and miR-196 activity in mouse embryos (Mansfield et al., 2004). The miR-10 sensor indicates that miR-10 activity is present in regions of the posterior thorax and abdomen, but not in the head or tail bud. This is similar to the expression pattern of the adjacent gene, *Hoxb4*, which is consistent with the possibility that they are coordinately regulated (Mansfield et al., 2004). One of the limitations of current ‘sensor’ studies has been the difficulty of interpreting the ‘negative’ patterns at cellular resolution.

The primary transcripts of miRNAs are typically much longer than the mature miRNAs (Pasquinelli et al., 2005), and *in situ* hybridization conditions have been developed that allow the detection of these nascent transcripts at their chromosomal sites of transcription. This method was used to detect the *Drosophila pri-mir-10* expression
pattern in early embryos (Kosman et al., 2004). The pri-mir-10 transcripts are detected in a Hox-like pattern that spans the future thoracic and abdominal segments in blastoderm-stage embryos. Another method to detect miRNA expression patterns involves the use of LNA oligonucleotide probes (Valoczi et al., 2004). This allowed the in situ localization of a large subset of zebrafish mature miRNA transcripts (Wienholds et al., 2005), including those of miR-10a, miR-10b and miR-196a. All three of the Hox miRNAs are detected in zebrafish embryos in Hox-like patterns in the developing nerve cord and trunk, with the miR-10a and miR-10b expression borders being anterior to that of miR-196a. A comparison of the three studies that localized miR-10 expression in mouse, zebrafish and Drosophila embryos indicates that its Hox-like expression pattern in the posterior head, thoracic and abdominal primordia is conserved in many animal embryos (Kosman et al., 2004; Mansfield et al., 2004; Wienholds et al., 2005).

Why have mir-10, and perhaps mir-196/iab-4, genes been conserved in their specific genomic positions within Hox clusters since the deuterostome–protostome divergence? It might be that, like the protein-coding Hox-cluster genes, they rely on regulatory mechanisms such as shared enhancers and long-range Polycomb–Trithorax-mediated repression and activation (Duboule, 1998; Duboule and Deschamps, 2004; Ringrose and Paro, 2004). It has been suggested that miRNAs are derived from inverted duplications of the 3’ UTRs of future target genes (Allen et al., 2004). It is possible that this is the case for the Hox miRNAs, especially considering that they likely target adjacent Hox-protein-coding genes.

It remains to be seen what significance the mir-10, mir-196 or mir-iab-4 genes might have for regulating developmental patterning during bilaterian embryogenesis. It
will be exciting to discover the roles of these highly conserved miRNA genes that have been residents of animal Hox clusters for at least 550 million years. It is conceivable that the gain or loss of Hox miRNA target sites, or changes in spatio-temporal expression of Hox-targeting miRNAs could be an important mode of Hox gene regulation in animal development and evolution.
RESULTS

Evolution of miR-10

MicroRNAs have been found in multiple branches of the eukaryotic tree of life, but no individual miRNA sequence has been discovered to be shared between animals and plants. The proteins necessary for the function of miRNAs are found in the genomes of most eukaryotes. However, these proteins have other ancestral functions, and it is likely that miRNAs evolved separately in animals and plants by independent co-option of the common machinery present in the genomes of these organisms for processing double stranded RNA and utilization of small single stranded guide RNAs for post-transcriptional regulation (Cerutti and Casas-Mollano, 2006).

Although no pan-eukaryotic miRNAs exist there are many miRNAs which are shared between the deep branches of animal phylogeny (Wheeler et al., 2009). One of the well conserved animal miRNAs is miR-10 which has been found to be produced in both deuterostome and protostome model organisms (Lagos-Quintana et al., 2001; Lagos-Quintana et al., 2003; Lim et al., 2003). Other than being conserved broadly in animals, this miRNA is interesting because it was found to be present in a homologous position between Hox4 and Hox5 orthologs in the Hox complex in both *Drosophila* and vertebrates.

A survey of publicly available genome data revealed that miR-10 sequences are present in the genomes of nearly all bilaterian animals that have had their genomes sequenced (Fig. 1). A single miR-10 sequence is found in all arthropods that have available genome sequence including a crustacean (*Daphnia pulex*), two chelicerates (*Tetranychus urticae*, and *Ixodes scapularis*) and many insects (Fig. 1 - section 1) and the
21 base pair sequence (ACCCTGTAGATCCGAATTTGT) is conserved in all of these animals. Sequences that correspond to miR-10 miRNAs are additionally found in the genomes of many lophotrochozoans (Fig. 1 - section 2) and these animals often have multiple miR-10-related miRNAs with multiple nucleotide changes or insertions in the mature miRNA sequence (Fig. 1 - section 8). Non-vertebrate deuterostomes contain a single miR-10 sequence (Fig. 1 - bottom of section 3) although no miR-10 ortholog is found in available urochordate genome sequences. All vertebrates have at least two miR-10 sequences in their genomes with miR-10a (most similar to arthropod miR-10 of the deuterostome miRNAs) present in the HoxB complex (HoxBb in fish), and miR-10b located in the HoxD complex. The sequence of the mature miR-10a is identical to arthropod miR-10 (Fig. 1 - section 3), although the 5’ and 3’ ends are shifted in mature sequences of some animals (Griffiths-Jones et al., 2006). The miR-10b sequences are identical to that of miR-10a except for a single nucleotide change in the center of the mature miRNA (Fig. 1 - section 4). The ray finned fish (Actinopterygii) have additional miR-10 sequences due to their secondary genome duplication following the divergence with the lineage leading to tetrapods. The HoxBb clusters of most fish retain the ancestral miR-10a sequence, while the HoxBa clusters typically contain miR-10c (Fig. 1 - section 6) which has diverged with a single nucleotide change that is shared between all fish and must have occurred shortly after the genome duplication. Fish HoxDa clusters retain the ancestral miR-10b (miR-10b-1) sequence (Fig. 1 - section 4), while HoxDb clusters contain miR-10d sequences (Fig. 1 - section 5) which differ by a single nucleotide from miR-10b. Fish HoxC clusters additionally contain miR-10b (miR-10b-2) sequences (Fig. 1 - section 4). The salmon (Salmo salar) genome has undergone an additional recent
genome duplication and contains at least 7 miR-10 sequences in its various Hox clusters (Mungpakdee et al., 2008). Coelacanth (Latimeria menadoensis), frog (Xenopus tropicalis), and lizard (Anolis carolinensis) genomes also have miR-10 sequences in their HoxC clusters which was lost in birds and mammals. This sequence in frog and lizard differs from that in coelacanth by one nucleotide (Fig. 1 - section 7), a change which likely occurred prior to the first appearance of terrestrial vertebrates. Some animal genomes contain additional miR-10 sequences (Fig. 1 - section 8) that are not direct orthologs of any of the aforementioned sequences but are derived duplicate miRNAs.

The miR-10 sequences present in assembled genomes seem to be present in the Hox complex and almost invariably between the Hox4 and Hox5 orthologous genes (data not shown). Genome sequence assemblies for many species from multiple major insect clades all showed a similar genomic location of miR-10 and it is likely that this should be case for most insects. A number of vertebrate genomes have been assembled and a similar case is found in all of these animals. Available genome sequence from a few ascidians suggests that they have lost the miR-10 miRNA. These animals have either completely, or at least partially, disrupted Hox complexes, and so it is possible that the loss of a coherent Hox complex may coincide with a lack of necessity for regulation by and subsequent loss of miR-10. In order to test this possibility, the genomic location of miR-10 sequences was examined in animals for which extensive whole genome shotgun data was publicly available that had not yet been assembled. Prior to this, no Lophotrochozoan genome or non-insect Arthropod Hox genomes had been assembled.

The sequences for all of the orthologous Hox homeodomains were collected following BLAST searches of the Lottia gigantea (limpet), Capitella (polychaete worm),
and *Daphnia pulex* (crustacean) whole genome shotgun data from the NCBI trace archives (data not shown). These sequences were then used as primers for the Tracinator program (Rebeiz and Posakony, unpublished) which iteratively BLASTs the sequence ends against the genomic trace database, collects the best hits, and assembles them into a larger contig. The contigs that were output from this analysis were then combined with some additional alignment by hand to create local assemblies of the Hox complex (if present) of each animal.

In these three cases complete Hox complexes were found with preserved co-linear gene order and no intervening genes (Fig. 2) similar to the presumed ancestral state of the Hox complex. The *Capitella* sequence data did not allow for a full assembly connecting the continuous sequences between the Hox8 and Hox9 orthologs. However, the ends of each half of the complex contained no additional non-Hox genes and further sequencing would presumably allow for a full assembly of this complex. Additional non-insect protostome Hox complex assemblies were attempted, but high levels of repetitive sequence or low coverage prevented these from being successful (data not shown).

In each of these three animals, miR-10 sequences were found between the Hox4 and Hox5 orthologs and in the same transcriptional direction (Fig. 2A-C). The *Daphnia* Hox complex contains sequences for *mir-iab-4* and *mir-iab-4-as* hairpins in positions similar to that of insect Hox complexes (Fig. 2A), but these miRNA genes may be restricted to Arthropod lineages as they were not found in either *Lottia* or *Capitella* Hox complexes or genome sequences. In both *Daphnia* and *Capitella* a second miR-10-related hairpin sequence was found between the Hox3 and Hox4 orthologs in the opposite orientation to Hox gene transcription (Fig. 2A-B), which has apparently been lost in
Lottia. The 5’ arms of these hairpins are very similar to miR-10, but their 3’ arms contain sequences orthologous to insect miR-993 (data not shown). This miRNA hairpin is found in many Arthropod genomes, and although there is extensive similarity to miR-10, this miRNA mainly produces mature miRNA from the 3’ arm of the hairpin in developing Drosophila (Ruby et al., 2007). However, it is unknown if this is the case in all protostomes and it may produce a miR-10-related mature miRNA from the 5’ of the hairpin or possibly miRNAs from both arms in some animals. There is a conserved Lophotrochozoan miR-10-related hairpin found between Hox8 and Hox9 orthologs in both Lottia and Capitella Hox complexes (Fig. 2B-C). The miRNA sequence is not identical between the two animals (data not shown), but due to shared location and conserved sequence differences from the ancestral miR-10 they are assumed to be orthologous. In addition to these sequences, the Capitella Hox complex contains two additional miR-10-related hairpins. One of these is near the 3’ end of the Hox7 ortholog, and the other is in the same intergenic interval as miR-10 but transcribed in the opposite direction, likely a result of a local inverted duplication (Fig. 2B).

The data from Lophotrochozoan Hox complexes, as well as genome data from numerous other bilaterian animals, indicate that the presence of a coherent Hox cluster always corresponds with the presence of a miR-10 gene between the Hox4 and Hox5 orthologs. However, there are a number of genomes which contain sequences for miR-10 hairpins which do not have coherent co-linear Hox complexes. The purple sea urchin (Strongylocentrotus purpuratus) genome has a Hox complex in which the gene order has been dramatically altered from the ancestral state and has lost the Hox4 ortholog (Cameron et al., 2006). However, this complex does contain a miR-10 gene near the
Hox3 ortholog, which may be ancestral despite the reorganization of the complex and loss of Hox4. The parasitic platyhelminth worm *Schistosoma mansoni* has a miR-10 gene which has multiple nucleotide changes in the mature sequence (Fig. 1 - section 2). While *Schistosoma* does not have a full complement of Hox genes, some of the remaining genes are found in small clusters in its genome (Koziol et al., 2009). The miR-10 sequence is found in a region containing Hox4 and Hox2 orthologs, to the 5’ of the *Schistosoma* Hox4 ortholog. Although *Schistosoma* does not have a Hox5 ortholog (Koziol et al., 2009) this is a similar genome location to the ancestral linkage. One of the two miR-10 genes in the platyhelminth flatworm *Schmidtea mediterranea* (Fig. 1 - section 2) is found closely linked to a derived homeodomain gene that is not clearly orthologous to any typical animal homeodomain gene (data not shown) and may either be residual ancestral Hox linkage or just a coincidence. The leech *Helobdella robusta* has a derived complement of Hox genes including duplicated and absent genes, which exist in five separate mini-clusters (data not shown). The miR-10 gene in this genome resides in one of these mini-clusters, which contains Hox5 and Hox7 orthologs. Its position near the 3’ end of the Hox7 ortholog and sequence are similar to the *Capitella* miR-10-related gene in this genomic position and likely is a direct orthologs of this miRNA hairpin. This means that this miR-10 duplicate must have arisen sometime prior to evolution of the major annelid classes and that the ancestral bilaterian miR-10 ortholog was lost in the lineage leading to *Helobdella*. Together the genomic positions of miR-10 in animals which do not have coherent Hox complexes, suggest that while is not necessary to maintain a coherent cluster in order to retain the miR-10 gene, linkage to the Hox genes
is typically retained, although none of these animals (other than possibly *Schmidtea*) have completely lost Hox gene clustering.

The miR-10 miRNA seems to be common to all bilaterian animals (ignoring secondary loss in a few lineages), but the origin of this miRNA is unknown. Although it has been claimed that miR-10 is found in cnidarians (Prochnik et al., 2007; Sempere et al., 2006) no conclusive evidence has shown this to be the case and definitive miR-10 hairpin sequences are not found in available cnidarian genomes. There are no miRNAs shared between sponges and higher animals, but miR-100 is conserved in cnidarians and bilaterians. This miRNA sequence is similar in sequence to miR-10 (Fig. 3) and it is likely that miR-10 was derived from miR-100, and as cnidarians do not have the full complement of Hox genes found in bilaterian genomes or true Hox complexes it is likely that miR-10 appeared prior to or during the formation of the first true Hox complex.

Other miR-10 related miRNAs are found in the genomes of many animals. Although nematodes genomes do not have miR-10 sequences, they do have a miRNA which is likely derived from miR-10. The 5’ half of the nematode miR-57 mature sequence is identical to miR-10 and additional sequence similarity is found in sequences throughout the hairpins of these two miRNAs (Fig. 3). miR-99 is a vertebrate miRNA which is derived from and very similar to miR-100 (Fig. 3). miR-125 is a bilaterian miRNA, which likely also derived from a miR-100 duplicate as it is found to be clustered with miR-100 in many animals (Sokol et al., 2008). miR-993 is a miRNA found in the Hox complex of all arthropods (data not shown) and at least one lophotrochozoan (*Capitella*) (Fig. 2B) between the Hox3 and Hox4 orthologs and is transcribed in the opposite direction of Hox gene and miR-10 transcription. The 5’ arm of this hairpin
contains sequences that are very similar to miR-10 (Fig. 3) and likely was created during and inverted duplication event in an ancient Hox complex prior to the split between ecdysozoans and lophotrochozoans.

The most likely path of sequence evolution starts with the novel appearance of miR-100 prior to the evolution of bilaterian animals. This miRNA then gave rise to miR-10 and miR-125 through duplication and subsequent sequence alterations early in bilaterian evolution. Subsequently, in protostomes a miR-10 a local inverted duplication the Hox complex gave rise to miR-993. In the nematode lineage, one of these became miR-57 through sequence changes while the other was lost. It is not possible to make the determination of which scenario is more likely because nematodes do not have a coherent Hox complex or full complement of Hox genes, and so linkage to the Hox3-Hox4 interval or Hox4-Hox5 interval can be determined. It is possible that additional sequencing of cnidarians may show that miR-10 was present in early animal evolution, in which case it would be possible for miR-10 to be the progenitor animal miRNA, and not miR-100. The data from this section is summarized in Figure 4.

Expression of Drosophila mir-10 transcripts

miRNAs are initially transcribed as PolII primary transcripts which contain a hairpin structure that is excised by the nuclear RNaseIII enzyme Drosha, exported to the cytoplasm by Exportin-5, and then the internal loop is cleaved by the cytoplasmic RNaseIII enzyme Dicer1 (Ambros, 2003). In order to determine the primary transcript that produces miR-10 (pri-miR-10), 5’ and 3’ RACE were performed on a cDNA library created from polyA RNA from a 0-24 hr embryo collection. The 3’ RACE products
reveal the polyA cleavage site to be about 700 bp downstream of the hairpin sequence (Fig. 5A). 5’ RACE gave multiple products, the longest of which revealed exon-intron boundaries and a putative transcriptional start site about 6.8 kb upstream of the hairpin (Fig. 5A). The sequence just upstream of the 5’ end of the RACE products contains a likely basal promoter sequence containing INR, DPE, and MTE elements (Butler and Kadonaga, 2002; Lim et al., 2004). These motifs are common to many promoters and are typically found at discrete intervals relative to each other. These motifs in the upstream sequence of miR-10 are found to conform to the sequence consensus and spacing requirements of a proper PolII promoter. This basal promoter structure is also conserved in amongst Drosophilids (Fig. 5C), further suggesting its likelihood as the miR-10 promoter. Together these results indicate that at least one pri-miR-10 transcript in Drosophila melanogaster spans an approximately 7.5 kb region of the chromosome between Scr and Dfd with a single approximately 5.5 kb intron (Fig. 5A), although secondary transcripts are possible.

To help determine the function of any gene it is important to know in which tissues it is expressed through development. In order to determine the expression pattern of miR-10, in situ hybridizations were performed using a probe antisense to pri-miR-10 sequence and analyzed its expression throughout embryogenesis. miR-10 primary transcripts are first expressed during the blastoderm stage of embryogenesis. At this point pri-miR-10 is expressed in a broad band which corresponds to the “trunk” of the embryo and is reminiscent of the types of patterns seen with gap gene or Hox gene expression (Fig. 6A). The expression pattern at this stage is largely complementary to that of Hunchback, suggestive of a possible negative regulatory interaction. However,
expression of pri-miR-10 was unaltered when examined in embryos mutant for the Hunchback gene (data not shown). It is likely that the pattern of pri-miR-10 in early embryos is regulated by some combination of maternally deposited morphogens and early gap genes, although no other possibilities were tested. The miR-10 primary transcript is also expressed in early embryos in the cells of the yolk (data not shown). During late blastoderm development and the beginning of gastrulation the expression of pri-miR-10 becomes downregulated in a subset of cells, taking on a striped appearance, which is similar to, but not as refined as that of the pair rule genes (Fig. 6B). The stripes of pri-miR-10 at these stages are in approximate register with that of the Hox complex gene fushi tarazu (data not shown), which in Drosophila has taken on a patterning role dissimilar to its ancestral Hox gene function. Shortly after the beginning of gastrulation, the transcription of pri-miR-10 appears to shut off and then re-initiates during the early stages of germband elongation in an entirely different pattern. As germband elongation proceeds, pri-miR-10 transcription is initiated in the anal pad and large intestine primordia and in ventral neurectoderm of the trunk segments (Fig. 6C). At this stage pri-miR-10 is expressed in developing neuroblasts as visualized in embryos stained for miR-10 and hunchback protein (Fig. 6D). After germband retraction pri-miR-10 is transcribed in the anal pads and large intestine as well as a band in the central midgut (Fig. 6E). In the large intestine, transcription is restricted to the endoderm as seen by triple staining for pri-miR-10, en, and Dpp, which are expressed in endodermal cells but not mesodermal tissue (Fig. 6F). In the central midgut pri-miR-10 is also endoderm specific as seen in embryos doubly stained for miR-10 and Dpp, which is exclusively in the mesoderm (Fig. 6G). In late embryogenesis pri-miR-10 staining can be seen in Malpighian tubules (Fig.
6H), as well as in the ventral nerve cord of the central nervous system (CNS) (Fig. 6I). The anteroposterior extent of pri-miR-10 expression in the CNS is reminiscent of its blastoderm expression pattern. The most anterior expression seen just posterior to the sub-esophageal ganglion, and extends posteriorly to near the terminal end of the CNS (Fig. 6J). It is possible that this pattern of expression in the CNS may be conserved in all bilaterians as a similar expression pattern is seen in zebrafish embryos (Wienholds et al., 2005).

**Expression of primary transcript products**

After a pre-miRNA is cleaved by Dicer, one strand of the resulting ~22nt dsRNA is packaged into an RNA Induced Silencing Complex (RISC). As research on miRNAs progressed it became clear that, for a subset of miRNAs, either strand of this dsRNA duplex can enter into RISC. The strand which becomes active mature miRNA is determined by the relative stability of the ends of the RNA duplex, where the strand with the 5’ end that has a lower binding energy should be packaged more often into RISC (Schwarz et al., 2003). This type of analysis predicts that there is no biochemical reason why miR-10 should be preferentially packaged into RISC over the opposite strand (miR-10*), and in fact suggest that miR-10* may actually be packaged into RISC preferentially over miR-10.

Both miR-10 (Aravin et al., 2003 Aug; Lagos-Quintana et al., 2001) and miR-10* (Lagos-Quintana et al., 2001) were cloned in early efforts to determine the miR-10 complement of the *Drosophila* genome. However, the data from these studies was not enough to statistically determine the relative abundance of either strand for most miRNA
duplexes. Due to the presence of miR-10 sequences in most other animals the possibility of a second miRNA from this locus was largely ignored. Although the miR-10 sequence from the 5’ arm of the hairpin is well conserved in animals the sequence from the 3’ arm of the hairpin is not particularly well conserved. The conservation that does exist among disparate animal clades (Fig. 1) may be due mostly to the necessity for complementary bases which allow formation of a hairpin capable of being processed by the miRNA machinery. The situation is very different if only arthropods are examined, where both miR-10 and miR-10* sequences are conserved (Fig. 7A). In deuterostomes only the miR-10 sequence is completely conserved and the 3’ arms of these hairpins have limited conservation (Fig. 7B). Although the putative “seed” sequence of the potential deuterostome miR-10* is conserved, free energy calculations suggest that it is unlikely to be packaged into RISC at appreciable levels (data not shown).

In order to confirm the prediction that both strands of the miR-10 duplex are present as mature miRNAs in Drosophila, northern blot analysis on RNA from multiple embryonic stages was performed using antisense probes complementary to each arm of the hairpin. Both miR-10 and miR-10* are present in embryos, with levels increasing throughout embryogenesis and miR-10* is found at significantly higher levels at all stages (Fig. 8A). In addition to Northern analysis of miRNAs, it has been demonstrated that mature miRNAs can be detected by in situ hybridization by using modified oligos composed of LNA nucleotides (Wienholds et al., 2005). Both miR-10 (Fig. 8B) and miR-10* (Fig. 8C) can be detected during embryogenesis utilizing this technique and the expression pattern of each is found to be largely similar to that of the primary transcript (compare Fig. 8B to Fig. 5E) but is found exclusively in the cytoplasm (Fig. 8D).
**Identification of putative targets**

One way to elucidate the function of miRNAs in the tissues where they are expressed is through the identification of their *in vivo* target genes. Due to the conserved location of the *miR-10* gene in animal Hox complexes, and the potential for miRNAs to have been spawned by and target nearby genes (Allen et al., 2004), it seemed likely that miR-10 and miR-10* might be regulators of Hox gene expression. Simple sequence motif searches of the 3’ UTR sequences of *Drosophila* Hox genes resulted in identification of potential target sites of varying strength.

The best putative target site for miR-10 was found in the 3’UTR of the *Sex combs reduced* gene (*Scr*). This sequence would likely form a duplex with miR-10 primed RISC. Although pairing with the 5’ “seed” sequence of miR-10 is not as extensive as a prototypical miRNA-mRNA pairing (a continuous helix of base pairs 3-7 as opposed to the prototypical “seed” sequence of base pairs 2-8), there is potential for a long continuous helix formed with the 3’ end of the miRNA (bases 11-22) (Fig. 9). Analysis of the putative 3’UTR sequences of other insects indicates that this sequence is found to be conserved among nearly all species examined (Fig. 9). This site is likely to be orthologous and have been maintained since the evolutionary split between the major insect groups, as the regions of potential pairing are maintained. In most of the non-*Drosophilid* insects there are additional sequences in the “seed” region which would provide added strength to the potential interaction. Additionally, the sequence surrounding the putative target site in these UTRs is not well conserved even among the relatively closely related *Drosophilids* suggesting this conservation has been selected for
(data not shown). The development of algorithms which attempt to predict miRNA target sites have been developed that have had varying success at prediction of true positive miRNA-mRNA interactions (Alexiou et al., 2009). Computational prediction of genome wide target sites in the 3’UTRs of all Drosophila genes (data not shown) using the RNAHybrid program (Rehmsmeier et al., 2004), along with the conservation among arthropods, suggests that the sequence identified in the 3’UTR of Scr is likely the best putative target of miR-10.

Computational prediction of genome wide targets (data not shown) identified a putative target site for miR-10* in the 3’UTR of the Hox gene Abdominal-B (Abd-B) (Fig. 10). Similar to the putative target site for miR-10 in the Scr 3’UTR, the putative miR-10* target site in Abd-B contains long potential continuous helix pairings. Additionally, potential target sites are found in the putative 3’UTR sequences of most arthropod Abd-B orthologs (Fig. 10), although these are not preserved in their pairing structure as well as the miR-10 target sites in Scr orthologs. These target sites are also conserved within a block of non-conserved UTR sequence (data not shown), similar to what was observed in Scr. This is one of the top potential target sites for miR-10* in the Drosophila genome and the only one which has such a high level of evolutionary conservation.

**Functional analysis of miR-10**

There are no published examples of post-transcriptional regulation of Scr during Drosophila embryogenesis. However, there is apparent post-transcriptional regulation of the Porcelio scaber Scr ortholog in the maxilliped segment during embryogenesis
miRNAs are may function in a number of ways to downregulate their target genes. Most initial studies suggested that miRNA targeted genes were downregulated by inhibition of protein production without affecting the levels of cytoplasmic mRNA. The expression of mRNA and protein products from the Scr gene overlap, for the most part, in all tissues during embryogenesis. However, in germband extended embryos, there is field of cells in the ventral T1 segment which transcribes Scr and contains low level accumulation of cytoplasmic mRNA, but that never exhibits any detectable protein accumulation (Fig. 11). During this stage pri-miR-10 is not actively transcribed in all of these cells, but it was possible that miR-10 produced from early transcripts would be present in these cells. LNA antisense oligo staining for mature miR-10 did not detect appreciable levels of miRNA in these cells at this stage. Additionally, in situ analysis of the expression of sequences 5’ of the miR-10 hairpin compared to sequences to the 3’ suggests that full length pri-miR-10 transcripts are not made in early blastoderm embryo ectoderm (data not shown) and that this early expression is likely from an alternate transcript from this genomic region. Taken together these data suggest that, although Scr is likely post-transcriptionally regulated in these cells, it is not through the activity of miR-10.

Further research on the function of miRNAs suggested that, RISC packaged miRNAs sequester target RNAs to cytoplasmic P-bodies (Liu et al., 2005) and that target RNAs may be degraded without having perfect complementarity to the target RNA (Lai et al., 1998). If this is the mode of activity of regulation of Scr by miR-10, then transcript pattern and levels would be altered in embryos ectopically expression or lacking miR-10. Available chromosomal deletions which lack miR-10 sequences are also deficient for Scr,
precluding the possibility of ascertaining the change in Scr expression in embryos lacking miR-10.

In order to determine the potential change in expression in embryos ectopically expressing miR-10, UAS transgene expression constructs were created. These transgenic flies allow expression of pri-miR-10 under the control of a variety of spatially and temporally specific GAL4 drivers. This transgene can produce significant amounts of mature miRNA, as visualized by antisense LNA oligo staining in embryos containing both the UAS-miR-10 transgene as well as a prd-GAL4 transgene (Fig. 12A).

Ectopic expression of miR-10 by nearly ubiquitous GAL4 drivers results in death during early larval development and noticeable cuticle phenotypes in the anterior of the early larvae. The transgenic larvae die typically during the first instar and occasionally do not hatch, but none survive to adulthood. These larvae have malformations of their anterior cuticle which appear as cuticular breaks near the mouthhooks (compare Fig. 12C-D with Fig. 12B). Although these cuticle phenotypes are reproducible, they do not correlate with any known Scr mutant phenotype. Typical cuticle phenotypes seen in Scr mutants include reduction of the T1 beard structure, but this is not seen in animals ectopically expressing pri-miR-10 (data not shown).

Scr and pri-miR-10 are expressed in largely complementary patterns in both early embryos (Fig. 13D-F) and late stage CNS (Fig. 13A-C). This correlation would make sense if miR-10 is negatively regulating Scr, but ectopic expression of pri-miR-10 does not corroborate this interaction. The en-GAL4 driver expresses in a small number of cells which overlap the Scr expression pattern. When pri-miR-10 is expressed under the control of en-GAL4, these cells have no detectable decrease in SCR protein levels when
compared to wildtype controls (compare Fig. 13 C-D with A-B). Additionally, embryos ectopically expressing pri-miR-10 from a variety of other GAL4 drivers do not exhibit detectable changes in either pattern or levels of Scr transcript or protein accumulation (data not shown).

In order to test the functionality of the putative miR-10 target site in Scr, constructs were created attaching the Scr 3’UTR sequence to a luciferase gene. The sequence of the putative miR-10 target site was then mutated in order to disrupt the potential pairing between miR-10 and Scr 3’UTR sequences (Fig. 15). These constructs were transfected into S2 cells and the expression levels of luciferase were analyzed either with or without co-transfection of a second transgene expressing pri-miR-10. These experiments showed either no significant change in expression of luciferase production or were inconclusive in their results.

**Functional analysis of miR-10**

The putative target site for miR-10* in the 3’UTR of Abd-B is predicted to form an extremely stable duplex and is comparable in structure and extent of binding of some of the best characterized miRNA-mRNA interactions. In combination with the conservation of target sites in other insect Abd-B orthologs, this is highly suggestive that this interaction is likely to be biologically important.

During embryogenesis miR-10 and Abd-B are expressed in mostly non-overlapping regions. While Abd-B is expressed in the majority of the posterior ectoderm, it is completely excluded from the developing anal pads, where miR-10* is strongly expressed (Fig. 13A-C and G-I). Abd-B is also expressed throughout later embryogenesis
in the visceral mesoderm surrounding the large intestine of the hindgut, whereas miR-10* is found in the endodermal cells of this region (Fig. 13J). This complementarity is suggestive of either unidirectional or mutual negative regulatory interactions. Embryos which are mutant for *Abd-B* gene products do not show altered expression of *pri-miR-10* (data not shown). Conversely embryos which are deficient for the *mir-10* gene, and thus lacking miR-10*, do not exhibit ectopic expression of *Abd-B* mRNA or protein in either the large intestine endoderm or in the anal pads, suggesting that miR-10* does not regulate *Abd-B* in these tissues (data not shown).

Although transcription is largely complementary in most tissues, in the ventral nerve cord of later stage embryos, both *pri-miR-10* and *Abd-B* are transcribed in parasegments 10-14 (Fig. 13A-C and G-I). The *Abd-B* gene produces multiple transcript variants which code for two different proteins. Of these, the r-type protein (shorter protein with regulatory function) is produced in the CNS only in the most posterior regions (PS14) whereas the m-type protein (longer protein with morphogenetic function) is expressed in more anterior regions of the CNS (PS13-10). *Abd-B* may utilize either of two polyA signals to terminate transcripts, the longer of which contains the putative miR-10* target site. The majority of *Abd-B* expression domains only show expression of the shorter 3’UTR, but the CNS displays expression of at least the longer version of 3’UTR if not a combination of both in parasegments 10-14. Since the domain of *pri-miR-10* transcription extends into the posterior of the CNS but appears to decrease toward the extreme posterior (Fig. 5I, J), it is possible that miR-10* may be downregulating the expression of r-type ABD-B in more anterior parasegments of the CNS, functioning as a protein type switch. To test this possibility, *Abd-B*D14 (an allele mutant specifically for
ABD-B m-type protein) was recombined with Df(3R)CP1. This strain can produce embryos which are deficient for both miR-10* and ABD-B m-type protein. Analysis of ABD-B protein expression shows no ectopic expression of ABD-B r-type protein in this mutant background (Fig. 14L). This suggests that the longer 3’UTRs found in parasegments 10-13 belong exclusively to transcripts encoding ABD-B m-type protein.

To test the possibility that miR-10* is regulating Abd-B expression in the CNS in general, transheterozygote Df(3R)CP1: Df(3R)LIN embryos deficient for miR-10 were analyzed for alterations in Abd-B transcript and protein levels in parasegments 10-14 of the CNS. The staging of the transheterozygote embryos is difficult due to altered morphology caused by the lack of genes other than miR-10. However, no significant differences were seen in the expression levels or anterior extent of expression of either Abd-B transcripts or protein (Fig. 14K) when compared to wildtype embryos (Fig. 14I). Ectopic expression of pri-miR-10 in the CNS either through ubiquitous overexpression (Fig. 14J) or in a subset of cells (Fig. 14G-H) failed to result in noticeable changes in levels of ABD-B protein compared to wildtype (Compare to Fig. 14E, F, and I).
DISCUSSION

The miR-10 gene is one of the most ancient animal miRNA genes. It is likely to have developed from an even more ancient miRNA, early in bilaterian evolution. Subsequently it has been duplicated, deleted, and altered many times in many lineages, resulting in a large family of related miRNAs. Most of these are found in the Hox complexes of bilaterian animals but have unknown functions. The *Drosophila mir-10* gene is expressed in the form of a large primary transcript from the Hox complex, and its mature miRNA products have strong evolutionarily conserved putative target sequences in nearby Hox genes. Expression of the primary transcript and its products occurs in multiple tissues and in a mostly complementary pattern to that of the putative target genes *Abd-B* and *Scr*. However, embryos lacking the *mir-10* gene or ectopically expressing the primary transcript do not show observable alterations in the expression of these Hox genes. Despite this, ubiquitous overexpression of *pri-mir-10* results in cuticle malformations and early larval death, but the root cause of this is still unknown. The *mir-10* gene miRNAs may have many targets which are unidentified that may be downregulated resulting in this phenotype.

It is possible that the evolutionarily conserved function of Hox miRNAs is to downregulate Hox proteins in sub-regions of the nerve chord. This possibility is raised by the conserved expression of miR-10 in the CNS of *Drosophila* and zebrafish (Wienholds et al., 2005). Members of the other vertebrate Hox miRNA family, miR-196, are also expressed in the CNS of zebrafish (Wienholds et al., 2005), and it is possible that *mir-iab-4* (not orthologous by sequence, but orthologous by chromosomal position in the Hox complex) also has expression in the CNS. It may be that the main function of the *mir-10*
locus miRNAs is to refine the expression of Hox proteins in the CNS to provide specific identities to the regions where this downregulation is occurring. It is possible that the combinatorial activity of multiple miRNAs acting on Hox mRNAs might be required for proper downregulation and this may explain the lack of regulation seen when *mir-10* gene products are absent or overexpressed. It is also possible that these miRNAs serve to canalize expression of their target genes in order to achieve robust, reproducible, and precise levels and spatial limits of expression in order to maintain a reproducible phenotypic output (Wu et al., 2009).

MATERIALS AND METHODS

MicroRNA sequences were obtained using BLAST searches of NCBI sequence databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequence alignments were created using BioEdit software (Hall, 1999). Hox complex assemblies were created using the Tracinator program (Rebeiz and Posakony, unpublished). MicroRNA precursor folding diagrams were created with Mfold (Zuker, 2003). Cuticle preparations were performed as in Bergson and McGinnis, 1990. Northern blots were performed as in Pasquinelli et al., 2003. Standard in situ hybridizations and immunohistochemistry were performed as in Kosman et al., 2004. SCR and ABD-B antibodies were obtained from the Developmental Studies Hybridoma Bank, and the EN antibody was a gift from Patrick O’Farrell. LNA in situ hybridizations were performed as in Kloosterman et al., 2006.
Figure 1. Alignment of available mir-10 hairpin sequences. Sequence alignment starting at nucleotide 1 of vertebrate miR-10 mature miRNA (Drosophila miR-10 starts at nucleotide 2 of alignment). Highlighted bases are 90% conserved among all sequences. Section divisions in far right column divide the alignment into 1 - Arthropod miR-10, 2 - Lophotrochozoan miR-10, 3 - vertebrate miR-10a (miR-10 in non-vertebrate deuterostomes), 4 - vertebrate miR-10b, 5 - fish miR-10d, 6 - fish miR-10c, 7 - miR-10e, 8 - duplicated orphan miR-10. Special indications: Acal* - nucleotides 13-14 (TA insertion) removed from this sequence in order to fit alignment; // - nucleotides have been removed starting at these locations to fit alignment. Organism abbreviations are as follows: Dmel - Drosophila melanogaster, Dsec - Drosophila sechellia, Dsim - Drosophila simulans, Dyak - Drosophila yakuba, Dere - Drosophila erecta, Dana - Drosophila ananassae, Dpse - Drosophila pseudoobscura, Dper - Drosophila persimilis, Dmir - Drosophila miranda, Dvir - Drosophila virilis, Dmoj - Drosophila mojavensis, Dwil - Drosophila willistoni, Dgri - Drosophila grimshawi, Gnor - Glossina morsitans, Mdes - Mayetiola destructor, Agam - Anocephalus gambiae, Aeag - Aedes aegypti, Cpip - Culex pipiens quinquefasciatus, Bnor - Bombyx mori, Tcas - Tribolium castaneum, Amel - Apis mellifera, Nvit - Nasonia vitripennis, Nlon - Nasonia longicornis, Rpro - Rhodnius prolixus, Nlug - Nilaparvata lugens, Pham - Pediculus Humanus, Apis - Acyrthosiphum pism, Lmig - Locusta migratoria, Dpul - Daphnia pulex, Isca - Isodes scapularis, Turt - Tetranuchus urticae, Vdes - Varroa destructor, Acal - Aplysia californica, Lgig - Lottia gigantea, Cap - Capitella, Hrob - Helobdella robusta, Smed - Schmidtea mediterranea, Sman - Schistosoma mansoni, Emil - Echinococcus multilocularis, Hsap - Homo sapiens, Ppan - Pan paniscus, Ptro - Pan troglodytes, Ggor - Gorilla gorilla, Ppyg - Pongo pygmaeus, Fabe - Pongo abelii, Pham - Papio hamadryas, Mmul - Macaca mulatta, Nleu - Nomascus leucogenys, Slab - Saginus labiatus, Ageo - Ateles geoffroyi, Cjac - Callithrix jacchus, Tsyr - Tarsius syrichta, Ogar - Otolemur garnettii, Mmur - Microcebus murinus, Tbel - Tupaia belangeri, Mmus - Mus musculus, Mmol - Mus musculus molossinus, Dord - Dipodomys ordii, Rnor - Rattus norvegicus, Cpor - Cavia porcellus, Opir - Ochotona princeps, Stri - Spermophilus tridecemlineatus, Sara - Sorex araneus, Ocum - Orictolagus cuniculus, Eeur - Erinaceus europaeus, Pvm - Pteropus vampyrus, Mluc - Myotis lucifugus, Cper - Carollia perspicillata, Cfam - Canis familiaris, Aliu - Ailuropoda melanoleuca, Trtu - Tursiops truncatus, Fcat - Felis catus, Btau - Bos Taurus, Bbub - Bubalus bubalis, Ecab - Equus caballus, Sscr - Sus scrofa, Dnov - Dasypus novemcinctus, Laf - Lexodonta africana, Emax - Elephas maximus, Etel - Echinops telfairi, Mdom - Monodelphis domestica, Meug - Macropus eugenii, Oana - Ornithorhynchus anatinus, Tacu - Tachyglossus aculeatus, Tgtu - Taeniopygia guttata, Ggal - Gallus gallus, Acar - Anolis carolinensis, Xtro - Xenopus tropicalis, Lmen - Latimeria menadoensis, Amia - Amia calva, Pbic - Polypterus bichir, Drer - Danio rerio, Cmil - Callorhinus milii, Ccon - Copadichromis conophorus, Abur - Astatotilapia burtoni, Mcon - Mchenga conophoros, Tnig - Tetraodon nigroviridis, Trub - Takifugu rubripes, Snep - Sphoeroides niphela, Blan - Branchiostoma lanceolatum, Bflo - Branchiostoma floridae, Skow - Saccoglossus kowalevskii, Spur - Strongylocentrotus purpuratus, Afru - Alloccentrotus fragilis, Mnem - Macaca nemestrina, Rfer - Rhinolophus ferrumequinum, Chof - Chloepus hoffmannii, Pman - Peromyscus maniculatus, Vic - Vicugna vicugna, Lpac - Lama pacos, Pcap - Procavia capensis, Olat - Oryzias latipes, Ssal - Salmo salar, Hfra - Heterodontus francisci, Gacu - Gasterosteus aculeatus, Onyk - Oncorhynchus mykiss, Mamb - Megalobrama amblycephala, Ipun - Ictalurus punctatus, Onil - Oreochromis niloticus.
Figure 1. continued
Figure 1. continued
Figure 2. Schematic representations of Hox complexes and their miRNAs. Transcriptional direction of Hox genes is from right to left in all cases. Hairpins on top are transcribed from right to left. Hairpins on bottom are transcribed from left to right. (A) *Daphnia pulex* Hox complex and locations of miR-10, miR-993, miR-iab-4, and miR-iab-4-as hairpins. (B) *Capitella* Hox complex and locations of miR-10, miR-993, and 3 miR-10-related (miR-10-r) hairpins. // - indicates lack of sequence data to definitively determine linkage. (C) *Lottia gigantea* Hox complex and locations of miR-10 and a miR-10-related (miR-10-r) hairpin.
Figure 3. Alignment of hairpin sequences for miR-10 and related miRNAs. Sequence alignment starting at nucleotide 1 of vertebrate miR-10 mature miRNA (*Drosophila* miR-10 starts at nucleotide 2 of alignment). Highlighted bases are 60% conserved among all sequences. Sequences are as follows: *Dm* - *Drosophila melanogaster*, *Mm* - *Mus musculus*, *Dr* - *Danio rerio*, *Ac* - *Anolis carolinensis*, *Tc* - *Tribolium castaneum*, *Ce* - *Caenorhabditis elegans*, *Nv* - *Nematostella vectensis*.
Figure 4. Evolution of miR-10 and related miRNAs. Phylogenetic tree showing gain and loss of miR-10 and related miRNAs. Colored squares indicate genome duplication events and consequently duplicated Hox complexes and miR-10 genes. Colored circles indicate gains of miR-10 and related miRNAs. Colored diamonds indicate loss of specific miRNAs.
**Figure 5. miR-10 precursor molecules and regulation.** (A) Diagram of *pri-miR-10* genomic structure determined by 5’ and 3’ RACE. Indicated as red and blue boxes are the terminal exons of *Scr* and *Dfd* respectively, and the black bar denotes the position of the pre-miRNA hairpin. (B) *pre-miR-10* hairpin which is excised from the primary transcript. (C) Alignment of basal promoter regions of *pri-miR-10* from six *Drosophila* species. Pink shaded residues are not conserved with *Dmel*. Highlighted in red, green and blue are promoter motifs Initiator (Inr), Motif Ten Element (MTE), and Downstream Promoter Element (DPE) respectively. Consensus sequences for motifs are given above alignment and lighter shaded nucleotides indicate lack of conformity to consensus. Organism abbreviations are as in Figure 1.
Figure 6. *pri-miR-10* in situ hybridization through embryogenesis. *pri-miR-10* antisense probe shows spatially localized staining at all stages after the beginning of cellularization. *pri-miR-10* is expressed in an early “gap” pattern (A). In later blastoderm stages *pri-miR-10* develops a “pair rule” pattern (B). In the later stages of embryogenesis *pri-miR-10* transcription occurs in anal pad and large intestine primordia (C,E) as well as ventral ectoderm (C,D) and regions of midgut (E). Large intestine and midgut expression is endoderm specific (F, and G respectively). Staining is also seen in late stage embryos in Malpighian tubules (H) and in the CNS (I). CNS staining has an anterior boundary near the sub-esophageal ganglion and extends to near the posterior end of the ventral nerve chord (J). *pri-miR-10* staining is in green, *Dpp* is in red (F,G), blue is either *en* (F) or DAPI (G,J).
Figure 7. Conservation of mature miRNA sequences in miR-10 hairpins. Highlighted and outlined sequences are 90% conserved in each alignment. (A) Both arms of Arthropod miR-10 hairpins which contribute to the Dicer cleavage product are conserved. (B) In deuterostomes only the miR-10 sequence is conserved.
Figure 8. Both miR-10 and miR-10* are expressed as mature miRNAs during Drosophila embryogenesis. (A) Northern blots of total RNA purified from Drosophila embryos at various developmental stages using $^{32}$P end labeled antisense oligo probes against miR-10 and miR-10*. Time course shows both miR-10 and miR-10* are expressed at very low levels in blastoderm embryos and during early germ band extension (0-4 hrs) and progressively increase in abundance throughout embryogenesis. miR-10* is found at much higher levels throughout embryogenesis. (B) Dorsal view of a stage 14 Drosophila embryo stained for mature miR-10 by LNA in situ hybridization. (C) Lateral view of a stage 14 Drosophila embryo stained for mature miR-10* by LNA in situ hybridization. Both miRNAs show identical patterns and their expression at this stage matches that of pri-miR-10 transcripts (compare to Fig. 5E). (D) Close up of anal pad cells showing cytoplasmic localization of miR-10* staining.
Figure 9. miR-10 has evolutionarily conserved target sites in the 3'UTRs of insect Sex combs reduced (Scr) orthologs. Alignment of conserved sequences found in the 3’UTRs (or 3’ of the stop codon in putative UTR sequence) of Scr genes in insects and complementarity to mature miR-10 sequence. All species have at least 12 nucleotides of complementarity to the 3’ end of miR-10 and 5 or more nucleotides of complementarity in the “seed” region near the 5’ end. Highlighted in blue is the conserved region which can pair with miR-10. Dark blue indicates G-U base pairing. Green highlights additional regions of pairing in some species. Bases that are not highlighted would not contribute to pairing between miRNA and mRNA. Species abbreviations are as in Figure 1, except Ngir - Nasonia giraulti, and Ppap - Phlebotomus papatasi.
<table>
<thead>
<tr>
<th>Species</th>
<th>3'UTR Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dmel</td>
<td>5' CGACCUCUCAC-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dsim</td>
<td>5' CGACCUCUCAC-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dsec</td>
<td>5' CGACCUCUCAC-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dyak</td>
<td>5' CGACCUCUCAC-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dere</td>
<td>5' CGACCUCUCAC-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dana</td>
<td>5' CGACCUCUCAC-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dpse</td>
<td>5' CAACCUCUCAG-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dper</td>
<td>5' CAACCUCUCAG-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dwil</td>
<td>5' CGACCUCUCAG-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dmoj</td>
<td>5' CGACCUCUCAG-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dvir</td>
<td>5' CGACCUCUCAG-UAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Dgri</td>
<td>5' CGACCUCUCAG-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Aeg</td>
<td>5' AAACCUCUC---GAAGAUAUUA 3'</td>
</tr>
<tr>
<td>Cpip</td>
<td>5' AAACCUCUCUCGAG-CAUCGAAUUUA 3'</td>
</tr>
<tr>
<td>Agam</td>
<td>5' AAACCUCUC---AAACGAAUUUA 3'</td>
</tr>
<tr>
<td>miR-10*</td>
<td>3' UUUGGAGAGAUCUUGGGCUUAAAC 5'</td>
</tr>
<tr>
<td>Bmor</td>
<td>5' CGACCUCUCAGUAUCCGAAUUUC 3'</td>
</tr>
<tr>
<td>Tcas</td>
<td>5' AAACUUCUCUAGAACAUAAGGAAUUA 3'</td>
</tr>
<tr>
<td>Amel</td>
<td>5' AAACUUUUUAAACAAUAUGGAAUUA 3'</td>
</tr>
<tr>
<td>Nvit</td>
<td>5' GUAUUUGUCUCUGG-UAUGAAUUGG 3'</td>
</tr>
</tbody>
</table>

**Figure 10.** miR-10* has evolutionarily conserved target sites in the 3'UTRs of insect *Abdominal-B (Abd-B)* orthologs. Alignment of conserved sequences found in the 3'UTRs (or 3' of the stop codon in putative UTR sequence) of *Abd-B* genes in insects and complementarity to mature miR-10* sequence. Potential target-miRNA pairings are not equivalent in structure, but all species have large blocks of complementarity to both 5' and 3' ends of the miR-10* sequence. Highlighted in blue is the conserved region which can pair with miR-10. Dark blue indicates G-U base pairing. Green highlights additional regions of pairing in some species. Bases that are not highlighted would not contribute to pairing between miRNA and mRNA. Species abbreviations are as in Figure 1.
Figure 11. Scr protein and cytoplasmic transcript accumulation are downregulated in ventral T1. The pattern of Scr cytoplasmic transcript accumulation in the ectoderm includes all of the labial segment as well as lateral regions of the first thoracic segment (A). The cells which are transcribing Scr (seen with intron probe – green in B) include all of the ventral and lateral domains of both the labial segment and T1. Scr protein accumulation does not occur in all cells which are transcribing Scr (C) but coincides completely with those cells that show cytoplasmic transcript accumulation (D).
Figure 12. Mature miRNA is produced from transgenes and results in aberrant phenotypes. When expressed from a **prd**-GAL4 driver, miR-10* is detectable in the **prd** pattern at high levels in a germband extended embryo (A). When expressed from an Actin-GAL4 (C) or da-GAL4 driver (D) malformations are noticeable in the head cuticle when compared to wild type (B).
Figure 13. *pri-miR-10* is expressed in a pattern complementary to predicted targets. Images of *in situ* hybridizations of *pri-miR-10* (green) *Scr* (red) and *Abd-B* (blue) at various stages of development. (A) *pri-miR-10* at stage 15. (B) *Scr* and *Abd-B* in same embryo as (A). (C) Overlay of (A) and (B). (D) *pri-miR-10* and (E) *Scr* in a blastoderm stage embryo. (F) Overlay of (D) and (E). (G-I) *pri-miR-10* and *Abd-B* expression in the posterior of the extended germ band of a stage 11 embryo. Note overlapping expression in posterior neuroectoderm. (J) Field of cells from the large intestine of the hindgut of a stage 15 embryo. *pri-miR-10* is expressed in the endodermal cells, while *Abd-B* is expressed in the visceral mesoderm surrounding the gut.
Figure 14. Overexpression or lack of pri-miR-10 has no observable effect on putative target protein expression. Embryos expressing UAS-pri-miR-10 under the control of en-GAL4 (C,D,G,H) exhibit no significant decrease in SCR protein (A-D) or ABD-B protein (E-H) in cells expressing ectopic miRNA (marked with anti-en in green) versus control cells in wild type (W1118) embryos (A,B,E,F). No significant change in ABD-B protein (I-L) levels is seen in the CNS of W1118 (I) compared to embryos ubiquitously expressing UAS-pri-miR-10 under the control of da-GAL4 (J), or embryos deficient for the miR-10 locus ((Df3R)CP1). Embryos mutant for the Abd-B-m specific transcripts (Abd-B^{D14}) and deficient for miR-10 do not exhibit ectopic r-type ABD-B protein (L).
**Figure 15. Target site mutations for luciferase experiments.** Alignment of *Scr* and *Abd-B* 3’UTR sequences with miR-10 or miR-10* respectively. Vertical bars indicate typical base pairs and colons indicate G-U base pairing. Highlighted in red are point mutations which disrupt continuous helix pairing between mRNA and miRNA particularly in “seed” regions.
Chapter II

Co-option of an ancestral anterior head patterning gene network for proximodistal patterning of appendages in early bilaterian evolution
INTRODUCTION

The enormous diversity of extant animal forms is a testament to the power of evolution, and much of this diversity has been achieved through the emergence of novel morphological traits. The origin of novel morphological traits is an extremely important issue in biology, and a frequent source of this novelty is co-option of pre-existing gene networks for new purposes (Carroll et al., 2008). Appendages, such as limbs, fins and antennae, are structures common to many animal body plans which must have arisen at least once, and probably multiple times, in lineages which lacked appendages. We provide evidence that appendage proximodistal patterning genes are expressed in similar registers in the anterior embryonic neurectoderm of *Drosophila melanogaster* and *Saccoglossus kowalevskii* (a hemichordate). These results, in concert with existing expression data from a variety of other animals suggest that a pre-existing gene network for anterior head patterning was co-opted for patterning of the proximodistal axis of appendages of bilaterian animals.

Since the advent of molecular biology, many morphological traits that are shared between disparate animal clades have been found to be controlled by conserved underlying gene networks (Bier, 1997; Holley et al., 1995; McGinnis and Krumlauf, 1992; Olson, 2006; Silver and Rebay, 2005). Morphological novelty, on the other hand, involves the evolution of new traits that are often patterned by co-opted genes or gene networks that originally performed other developmental functions. For example, eye spots on butterfly wings are patterned through the re-deployment, in small foci corresponding to the eye spots, of a gene regulatory network that also controls the growth and pattern of the entire insect wing (Keys et al., 1999). Another example is seen in the
redeployment of a few Hox genes to pattern the paired appendages of vertebrates; these genes having been co-opted from an ancestral role in patterning posterior structures on the main body axis of chordates (McGinnis and Krumlauf, 1992; Zakany and Duboule, 2007). In this study we wished to explore the origins of the proximodistal appendage patterning network.

The patterning of *Drosophila* appendages is a well studied system of proximodistal axis specification. Although numerous genes participate in patterning of *Drosophila* appendages, there is a core network which is responsible for establishing the gross morphological divisions. The gene pair *buttonhead* (*btd*):*D-Sp1*, and the genes *Distal-less* (*Dll*), *dachshund* (*dac*), and *homothorax* (*hth*), exist in an interacting network that regulates the growth and boundaries of the distal, medial, and proximal appendage domains (Kojima, 2004). We refer to these genes as the core proximodistal appendage patterning network.

In early *Drosophila* embryos, *hth* and *btd* are both expressed in the appendage primordia. *btd*:*D-Sp1* (a recently duplicated and diverged gene pair of *Sp8* orthologs in *Drosophila*), are required for the activation of *Dll* in embryonic thoracic appendage primordia (Estella et al., 2003). As the domains of *Dll* expressing cells expand, *hth* becomes excluded from a subset of these cells in response to repression by *Dll* (Bolinger and Boekhoff-Falk, 2005). Cells from these primordia go on to form the larval Keilin’s organs and imaginal discs.

Early in leg imaginal disc development, as in the embryonic appendage primordia, cells are divided into two major domains by a central cluster of *Dll* expressing cells surrounded by *hth* expressing cells (Abu-Shaar and Mann, 1998; Wu and Cohen,
1999). As development progresses, \textit{dac} expression comes on in a medial region of leg discs (Abu-Shaar and Mann, 1998; Wu and Cohen, 1999). Small regions of overlapping expression of \textit{Dll}, \textit{dac}, and \textit{hth} exist in later leg disc development, although the genes also exhibit mutually repressive interactions in some cells (Abu-Shaar and Mann, 1998; Dong et al., 2001; Wu and Cohen, 1999). Although the regulatory relationships between \textit{btd}:\textit{D-Sp1} and other core appendage patterning genes in the imaginal discs are unknown, the \textit{btd} gene is expressed in a disc domain that overlaps with \textit{hth}, \textit{dac} and \textit{Dll}, but is excluded from the most distal and most proximal cells (Estella et al., 2003). The developing \textit{Drosophila} antenna has a slightly different imaginal disc expression profile than that of the leg, with the medial \textit{dac} domain being smaller and \textit{hth} expression overlapping that of both \textit{dac} and \textit{Dll} (Dong et al., 2001). However, expression domains of these genes still roughly correspond to the same proximodistal fates in both developing leg and antennae (Dong et al., 2001), and similarly ordered and overlapping expression domains of the core genes are conserved in many developing arthropod appendages (Angelini and Kaufman, 2005; Beermann et al., 2004; Schaeper et al., 2009).

Investigation of genes underlying proximodistal development of vertebrate appendages has revealed that, despite structural dissimilarity to arthropod appendages, they develop under the control of genetic networks that include orthologs of \textit{Drosophila} \textit{btd}:\textit{D-Sp1}, \textit{Dll}, \textit{dac}, and \textit{hth} genes (Pueyo and Couso, 2005). Vertebrate \textit{Sp8} genes are expressed in evolutionarily conserved patterns in distal ectoderm of limb buds, and knockdown of \textit{Sp8} function in chick results in defects of limb outgrowth and patterning (Kawakami et al., 2004). \textit{Dlx} family genes (\textit{Dlx1}, 2, 5 & 6; \textit{Dll} orthologs) are also expressed in distal ectoderm of mouse limb buds, and \textit{Dlx5}:\textit{Dlx6} double mutants have
distal limb defects (Kraus and Lufkin, 2006; Panganiban and Rubenstein, 2002). *Dach1* (a *dac* ortholog) is expressed in a complex pattern in developing mouse limb buds, with a transient stage when expression is limited to anterior-medial limb bud cells (Davis et al., 1999; Hammond et al., 1998). *Meis1* (a vertebrate *hth* ortholog) is expressed in the proximal regions of vertebrate limb buds, and required for the normal development of the proximal domain of chick appendages (Mercader et al., 1999).

Available fossil data from the Pre-Cambrian does not allow us to be sure of the body plan of the last common ancestor of vertebrates and arthropods (Valentine, 2004). However, a synthesis of comparative morphology suggests that it either existed with rudimentary appendages or lacked them entirely (Shubin et al., 1997). The appendages of disparate extant bilaterian groups almost certainly evolved independently in multiple lineages subsequent to their divergence from a common ancestor which lacked appendages (Shubin et al., 1997). If animal appendages are not derived from a common ancestral appendage, the involvement of a common gene network in proximodistal patterning could be due to random convergence of the same set of genes to pattern non-homologous appendages, or independent co-option of the same gene network which functioned to pattern an ancestral structure shared by both vertebrates and arthropods (Davidson and Erwin, 2006; Panganiban et al., 1997; Tabin et al., 1999). Involvement of a gene network in essential developmental roles (e.g. insect wing patterning) may make the regulatory interactions within the network resistant to change (Davidson and Erwin, 2006). This does not, however, preclude redeployment of such a patterning network using different genetic inputs and outputs, which could then contribute to novel morphological
structures (Davidson and Erwin, 2006), such as butterfly wing eye spots (Keys et al., 1999).

It has been previously proposed that lateral appendages might have originated through the co-option of a pre-existing group of genes, including Dll, that controlled a rudimentary appendage-like outgrowth in the ancestor of vertebrates and arthropods (Tabin et al., 1999). It has also been proposed that the appendages of vertebrates and arthropods might be modified duplicates of the entire anteroposterior body axis (Minelli, 2000). This proposal is based in part on an ancestral role of Hox genes in patterning the main body axis, and the involvement of a subset of Hox genes in patterning the proximal-distal axis of vertebrate appendages (Zakany and Duboule, 2007). Hox genes do not have similar expression patterns in vertebrate and arthropod appendages, so for this and other reasons the model that the entire anteroposterior body axis patterning system is redeployed in most animal appendages (Minelli, 2000) is not well supported in our opinion. Our proposition relates an ancient conserved network for patterning the anterior neurectoderm of animals to the proximodistal patterning of bilateral animal appendages.

A survey of previous research provides data from a few different animal groups on expression patterns of the core proximodistal appendage patterning genes in various tissues. We noticed that these genes, as well as other genes that are part of the proximodistal appendage patterning network in *Drosophila*, such as *aristaless (al)*, *apterous (ap)*, and *BarH1*, are expressed in discrete domains in the anterior embryonic neurectoderm of many chordates and arthropods (Appendix). We considered the hypothesis that a shared anteroposterior expression regimen of these genes in head neurectoderm might be common in bilateral animals. We wished to evaluate this
hypothesis by testing the relative expression patterns of core appendage patterning genes in the anterior neurectoderm of *Drosophila* embryos, as well as in embryos of *Saccoglossus*, a basal deuterostome that lacks bilateral appendages.
RESULTS

Genes of the core proximodistal appendage patterning network are expressed in a spatially and temporally complex manner during *Drosophila* development. However, in the anterior neurectoderm of *Drosophila* embryos, these genes are expressed in a clear anteroposterior order. We determined the relative expression patterns of *Dll, dac, hth*, and *btd* using combinatorial in situ hybridizations, and analyzed their relative expression patterns in germband extended (stage 11) embryos (Fig. 16b-e). At this stage the procephalic neurectoderm has been divided into anterior, central, and posterior protocerebral areas, and a more posterior deuterocerebral area, with the most anterior cells being those flanking the dorsal midline of the procephalon and posterior cells located more ventrolaterally (Younossi-Hartenstein et al., 1996) (Fig. 16a). *btd:D-Sp1* and *Dll* are transcribed in overlapping patches covering most of the anterior protocerebral neurectoderm (Fig. 16b-c). The domain of *dac* transcription is mainly in central and posterior protocerebral neurectoderm, with small regions of overlap with *Dll* and *btd:D-Sp1* (Fig. 16d). *hth* transcripts are largely absent in anterior neurectoderm (Fig. 16e), and are completely excluded from domains which transcribe *Dll* and *btd:D-Sp1* (compare to Fig. 16c). The posterior protocerebral region contains cells which transcribe both *hth* and *dac*, but the majority of *hth* transcription is found in the deuterocerebrum and more posterior neurectoderm (compare Figs. 16d and 16e). Taken together, these data reveal an expression order of *btd:D-Sp1* and *Dll* in the most anterior neurectodermal cells, *dac* in medial cells, and *hth* in posterior cells, with small zones of overlap at the borders of the three major domains (Fig. 16f).
Using in situ hybridization, we also tested the expression patterns of the orthologous genes in *Saccoglossus* embryos ranging from gastrula to early gill slit stages. In post gastrulae, Sp8 (the *btd/Sp1* ortholog) is transcribed at high levels in the anterior third of the embryonic neurectoderm, and to a lesser degree in multiple medial to posterior stripes (Fig. 17a). In embryos at the one gill slit stage, expression is similar with the major expression confined to the proboscis (Fig. 17b). The *Saccoglossus dac* ortholog (*Dach*) is transcribed at high levels in a neurectodermal stripe just anterior to the collar, and at low levels throughout most of the rest of the embryo, both in post gastrula (Fig. 17c) and one gill slit stage embryos (Fig. 17d). *Meis*, the ortholog of *hth*, is transcribed at high levels in a broad band in the trunk neurectoderm, as well as in two dorsal patches - one just anterior to the collar; the other in the posterior trunk (Fig. 17e-f). The *Saccoglossus Lhx2/9* gene, orthologous to *apterous*, is transcribed throughout the proboscis neurectoderm in late gastrulae (Fig. 17g) and then becomes restricted mainly to a strong stripe just anterior to the collar (Fig. 17h). Along with previously documented expression patterns for the *Saccoglossus BarH1* and *Dlx* orthologs (Lowe et al., 2003) we provide an expression model (Fig. 17i) summarizing the transcription domains of all of these genes.
DISCUSSION

Based on the above data and previously published work (Appendix), we estimated the ancestral expression domains of the anteroposterior head gene network, and compared them to the approximate domains of the proximodistal appendage gene network in the developing *Drosophila* leg (Fig 18). Our proposal posits that a “head-appendage” network, consisting of the btd/Sp8, Dll/Dlx, dac/Dach, hth/Meis genes (and likely other genes, some of which are shown in Figure 18), was present in a bilaterian ancestor that lacked appendages, where the network functioned to pattern the anteroposterior head axis. The evolution of this network may even have contributed to the process of cephalization in early animals. Subsequently, this network was co-opted to pattern the proximodistal axis of bilateral appendages through modification of input and output connections.

After a network for anteroposterior head patterning had been co-opted for proximodistal appendage patterning, it could be used specify and diversify the pattern of many body wall outgrowths (e.g. sensory structures, locomotory appendages, external genitalia, feeding appendages, etc.) through changes in the network. These could include variations in the regulatory relationships and expression patterns of the core appendage patterning genes, as well as further modifications of input and output connections (Dong et al., 2001). Consistent with this theory, inputs into this network during *Drosophila* appendage formation, such as *Dpp* and *wg*, are not conserved in this role among insects (Angelini and Kaufman, 2005). At least part of the same network has apparently been co-opted for the development of beetle horns (Moczek et al., 2006; Moczek and Rose, 2009), an appendage-like body wall outgrowth, long after the evolutionary advent of bilateral
appendages. It will be interesting to study the expression pattern of the head-appendage patterning genes in other branches of the evolutionary tree, especially in cnidarians, acoel flatworms, and lophotrochozoans.

MATERIALS AND METHODS

*Drosophila in situ* hybridizations were performed as in Kosman et al., 2004. *btd* antisense probes were made from a 2.6 kb genomic fragment starting 49 bp 5’ of the coding region. *Dll* antisense probes were made from a 1.4 kb EcoRI cDNA fragment (Cohen et al., 1989). *dac* antisense probes were made from a genomic PCR fragment cloned into pCRII (Invitrogen), the primers for the *dac* fragment were: 5’ AAGCAAAGTATAGAACGGATTAGCA 3’; 5’ TCCAACGAATCTTTCACTTCG 3’.

*Saccoglossus in situ* hybridizations were performed as in Lowe et al., 2004. Antisense probes for *Sp8*, *Lhx2/9*, *Dach*, and *Meis* were made from *Saccoglossus kowalevskii* cDNAs (Freeman et al., 2008), accession numbers NM_001168189, NM_001164971, NM_001164944, and GU384871 respectively.
Figure 16. Expression of core appendage network genes in anterior neurectoderm in *Drosophila* embryos. Data are presented as maximum projections of confocal sections through embryonic procephalic neurectoderm of a stage 11 *Drosophila* embryo. (A) Schematic of the procephalic region of a stage 11 *Drosophila* embryo displaying subdivisions of the neurectoderm. Indicated are anterior, central, and posterior protocerebral (Pa, Pc, and Pp respectively) and deuterocerebral regions (adapted from Younossi-Hartenstein et al., 1996). (B) *btd* is transcribed mainly in the anterior protocerebral neurectoderm with a small region of expression in central regions. Also seen are antennal and maxillary segment expression outside the neurectoderm. The D-Sp1 transcription pattern is the same as *btd* in this region (data not shown). (C) *Dll* is transcribed in nearly the same neurectodermal pattern as *btd* at this stage. (D) *dac* is transcribed mainly in central and posterior protocerebral neurectodermal cells and overlaps with the posterior expression of *Dll* and *btd*. (E) *hth* is expressed in posterior protocerebral cells and deuterocerebral cells, overlapping with *dac* expression in its posterior expression domain. (F) Diagram indicating relative expression domains of *Dll*, *btd*, *dac*, and *hth* in the procephalic neurectoderm.
Figure 17. Expression of core network genes in Saccoglossus embryos by in situ in post gastrula (A,C,E, and G) and one gill slit stages (B,D,F, and H). Data are presented as sagittal optical sections with anterior to the upper right of each panel. (A,B) High levels of Sp8 transcripts could be detected in two broad lateral patches in the proboscis. Expression is absent from both the dorsal and ventral midlines and the most anterior region of the proboscis. Sp8 transcripts are additionally detected at low levels in two broad lateral stripes in the trunk ectoderm with both dorsal and ventral midlines and ciliated band free of expression. (C,D) Dach is expressed at high levels in an ectodermal stripe just anterior to the collar. Additional low level ectodermal expression is detected throughout much of the embryo. (E,F) Meis is expressed strongly in the trunk ectoderm excluding the ciliated band and at the base of the proboscis at early developmental stages, which then subsequently refines to a strong dorsal domain. (G,H) At post gastrula stage Lhx2/9 is expressed throughout the proboscis ectoderm. By one gill slit stage expression becomes restricted mainly to a strong stripe at the base of the proboscis. (I) A schematic combining these data with previously published expression data for Dlx and BarH1 (Lowe et al., 2003) indicating the relative levels and anteroposterior extents of neurectodermal expression of the appendage network genes. Dashed lines indicate expression in a subset of cells for the indicated anteroposterior domain.
Figure 18. Schematic diagram comparing expression of core network genes in limbs and anterior neurectoderm. (A) Estimated ancestral expression patterns in embryonic anterior neurectoderm (based on conserved domains of expression) is displayed on a generalized diagram of a chordate brain. (B) Expression of proximodistal appendage patterning genes is displayed on a diagram of an adult *Drosophila* leg (adapted from Kojima, 2004).
Chapter III

Evolution of homeodomain proteins in early and pre-metazoans
Introduction

Metazoan animals comprise a monophyletic group of multicellular organisms which arose from a unicellular ancestor long before the appearance of the typical animal body forms seen in Cambrian fossil deposits. The transition from unicellularity to multicellularity has occurred multiple times in the ancient past in the ancestors of plants, fungi and animals. Development of a multicellular mode of existence likely allowed for greater adaptability and utilization of resources providing advantages for newly developed phyla to exploit underutilized niches. The multicellularization that gave rise to the metazoans was an extremely important event, which provided the basis for a group of organisms which have evolved to live in nearly every ecosystem. However, little is known about how this multicellularization occurred or what molecules were involved in the process.

There are a number of classes of proteins that would likely be essential to the process of multicellularization. Cell adhesion proteins and extracellular matrix components would allow for the construction of rudimentary multicellular structure. However, an indistinct mass of cells does not make a functioning multicellular organism. True multicellular organisms (as opposed to colonial forms of unicellular organisms) have multiple cell types which perform different functions. In order to accomplish this there must be a mechanism for cells to communicate with each other and respond through changes in cell type or cell function. Signal transduction pathways allow for cell to cell communication, and transcription factors provide proper responses to signals by altering the expression of batteries of downstream effector genes.
Transcription factors are ubiquitous proteins found in all organisms which provide control over when genes activated or repressed. However, homeodomain transcription factors comprise an ancient family restricted to eukaryotes. The homeodomain protein family is distinctive in part due to its size and diversity in multicellular organisms, but principally because most homeodomain proteins assign developmental fates to individual cells, or to groups of adjacent cells in developing multicellular organisms. Most eukaryotes encode one or more genes of the non-TALE homeodomain superclass, with 60 amino acid homeodomains; as well as one or more genes in the TALE superclass, which usually encode 63 amino acid homeodomains (Burglin, 1997; Derelle et al., 2007). In multicellular animals, the non-TALE and TALE superclasses have both duplicated and diverged into many classes and subclasses, often with hundreds of homeodomain genes per genome.

In order to determine what the genetic toolkit of the first animals was comprised of, it is necessary to compare both a variety of animal genomes as well as allied outgroups such as fungi, which together with animals form the Opisthokonta. While complete genomes for many fungal and eumetazoan animals exist, the addition of genomic sequence from intermediate species will make elucidation of the process of multicellularization more feasible. The need for these data, in part, has led to the whole genome sequencing of the choanoflagellate Monosiga brevicollis and the sponge Amphimedon queenslandica (formerly known as Reniera). Sponges are the earliest branching metazoan animals and have extremely simple bodies and a low number of cell types when compared to higher metazoans. Choanoflagellates are the nearest outgroup to metazoans and exist either as colonies or single cells which, interestingly, closely
resemble one type of cell found in sponges. It is thought that a unicellular organism similar to a choanoflagellate, or perhaps a particular species of choanoflagellate, was the ancestor of all animals. While it is formally possible that the common ancestor of both fungi and animals was a multicellular organism, this is not thought to be likely and as such the choanoflagellates occupy a unique phylogenetic position which may provide many answers about our simple origins. As part of the genome project for the choanoflagellate *Monosiga brevicollis*, the homeodomain encoding genes in its genome were analyzed as well as those from the recently sequenced species of sponge, *Amphimedon queenslandica*. 
RESULTS

Analysis of the demosponge *Amphimedon queenslandica* genomic sequence indicates that it contains at least 31 homeodomain genes. Sponges have simple body plans when compared to other animals and do not exhibit the complex anteroposterior diversification seen in many of the higher animals. As such, it is not surprising that the sponge genome does not contain Hox genes. However, a number of sequences for animal specific types of homeodomain proteins of the ANTP superclass (a diverse group of non-TALE class homeodomain genes) including one each of the BARH, TLX, MSH, BSH, and HEX class genes as well as four NK class genes (data not shown). Additionally, *Amphimedon* has sequences for seven PRD class homeodomains, three LIM class homeodomains, and four POU class homeodomains, all of which are non-TALE class type homeodomain subclasses (Figs. 19). TALE class homeodomains found in the sponge include one each of MEIS, PBX, and SIX class genes as well as five IRO class genes.

Only two homeodomain genes are detected in the *Monosiga brevicollis* genome, both of which group with the MEIS subclass of TALE homeodomains as determined by Bayesian phylogenetic analysis and characteristic C-terminal amino acid residues (Figs. 19 and 20). Although the genome of the last common ancestor of opisthokonts (fungi, choanozoans, and animals) encoded both non-TALE and TALE homeodomain genes (Derelle et al., 2007), the *Monosiga* genome only encodes TALE class homeodomain genes.
DISCUSSION

The analysis of homeodomains from choanoflagellates and sponges, in comparison with human and various fungal homeodomains, indicates that the last common ancestor of the opisthokont lineage had a definitive MEIS class gene (Fig. 20 - red branch). However, none of the other TALE class homeodomains and non-TALE class homeodomains of fungi definitively group with any of the specific sub-classes found in animals (Fig. 20). This would suggest that, while the opisthokont ancestor had both TALE and non-TALE type homeodomain genes that, most of these genes have diverged significantly since the division between animals and fungi, with each group developing its own characteristic sub-classes.

Sponges have at least one gene of many of the major sub-classes of animal homeodomain genes, indicating that these classes are common to all animals and arose either early in animal diversification or prior to the evolution of the first metazoan. Further amplification and diversification of homeodomain genes occurred in the lineage leading to bilaterian animals, and most of the homeodomain types found in triploblast animals, including the HOX subclass, must have been present in the last common ancestor of cnidarians and bilaterians (Ryan et al., 2006).

The lineage leading to Monosiga apparently lost all non-TALE homeodomain genes. It is unexpected that none of the other animal specific classes of homeodomain genes are found in this genome, but it is possible that the non-colony forming choanoflagellate Monosiga has evolved a simplified genome in an adaptation to its exclusively unicellular existence. It will be interesting to see if this genomic state is common to all choanoflagellates or if this is a derived state found in non-colony forming
species. Ongoing sequencing of a colony forming species of choanoflagellate 
(*Proterospongia*) may provide further insights into the evolution of homeodomain genes
in the lineage leading to animals, as well as the evolution of multicellularity in general.

Chapter III contains material previously published in King, N., Westbrook, M. J.,
Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y.,
Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K. J., Zuzow, R., Dirks,
W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J. B., Morris, A., Nichols, S.,
Richter, D. J., Salamov, A., Sequencing, J. G., Bork, P., Lim, W. A., Manning, G.,
Miller, W. T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I. V., and Rokhsar, D.
(2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of
MATERIALS AND METHODS

Phylogenetic analysis was performed on aligned sequences with Mr. Bayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) run with mixed amino acid and inverse gamma settings for 3 million iterations with a burnin of 75000. The unrooted dendrogram was made using FigTree (Andrew Rambaut, http://tree.bio.ed.ac.uk/). *Monosiga brevicollis* sequences were found by BLAST of the assembled genome at JGI. *Homo sapiens* sequences obtained from the NCBI homeodomain resource. Sponge sequences were found by BLAST of the *Reniera* (*Amphimedon queenslandica*) trace data from the NCBI trace archives. Fungal sequences were obtained from the Broad Institute (*Aspergillus nidulans*, *Coprinus cinerea*, *Cryptococcus neoformans*, *Neurospora crassa*, *Rhizopus oryzae*) and JGI (*Aspergillus niger*, *Laccaria bicolor*, *Phanerochaete chrysosporium*, *Phycomyces blakesleeanus*).
Figure 19. Alignment of Homeodomains from various opisthokonts. Highlighted bases are 30% conserved amongst all sequences. Amino acid positions given at the top of the alignment are based on sequence numbering in TALE superclass homeodomains. Homo sapiens sequences are labeled with HUMAN and include gene name labels. Sponge (Amphimedon queenslandica - formerly known as Reniera) sequences are labeled with Ren and include labels of homeodomain family grouping based on phylogenetic analysis in Figure 2. Fungal sequences are labeled with species abbreviations (A.nid - Aspergillus nidulans; C.cin - Coprinus cinerea; C.neo - Cryptococcus neoformans; N.cra - Neurospora crassa; R.ory - Rhizopus oryzae; A.nig - Aspergillus niger; L.bic - Laccaria bicolor; P.chr - Phanerochaete chrysosporium; P.bla - Phycomyces blakesleeanus).
<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HESX_HUMAN</td>
<td>PAX6_HUMAN</td>
<td>LHX2_HUMAN</td>
</tr>
<tr>
<td>PMX1_HUMAN</td>
<td>PMX2_HUMAN</td>
<td>Lh61_HUMAN</td>
</tr>
<tr>
<td>OTX1_HUMAN</td>
<td>ISL1_HUMAN</td>
<td>Lmxb_HUMAN</td>
</tr>
<tr>
<td>CRT1_HUMAN</td>
<td>Ren_LIM1</td>
<td>Ren_LIM2</td>
</tr>
<tr>
<td>FRII1_HUMAN</td>
<td>R.ory1</td>
<td>R.ory2</td>
</tr>
<tr>
<td>PIX1_HUMAN</td>
<td>P.chr3</td>
<td>P.chr4</td>
</tr>
<tr>
<td>GSC_HUMAN</td>
<td>Ren_POU1</td>
<td>Ren_POU2</td>
</tr>
<tr>
<td>FAX6_HUMAN</td>
<td>Ren_prd1</td>
<td>Ren_prd2</td>
</tr>
<tr>
<td></td>
<td>Ren_prd3</td>
<td>Ren_prd4</td>
</tr>
<tr>
<td></td>
<td>Ren_prd6</td>
<td>Ren_prd7</td>
</tr>
<tr>
<td></td>
<td>LhX2_HUMAN</td>
<td>Lh61_HUMAN</td>
</tr>
<tr>
<td></td>
<td>Lmxb_HUMAN</td>
<td>Lmxb_HUMAN</td>
</tr>
<tr>
<td></td>
<td>Ren_LIM1</td>
<td>Ren_LIM2</td>
</tr>
<tr>
<td></td>
<td>R.ory1</td>
<td>R.ory2</td>
</tr>
<tr>
<td></td>
<td>P.chr3</td>
<td>P.chr4</td>
</tr>
<tr>
<td></td>
<td>Ren_POU1</td>
<td>Ren_POU2</td>
</tr>
<tr>
<td></td>
<td>Ren_prd1</td>
<td>Ren_prd2</td>
</tr>
<tr>
<td></td>
<td>Ren_prd3</td>
<td>Ren_prd4</td>
</tr>
<tr>
<td></td>
<td>Ren_prd6</td>
<td>Ren_prd7</td>
</tr>
</tbody>
</table>

**Alignment Details**: 74 identities with a total of 5467 columns.
Figure 20. Phylogenetic tree of opisthokont homeodomains. An unrooted dendrogram of selected opisthokont homeodomains determined by Bayesian phylogenetic analysis. Nodes are labeled with posterior probabilities. Fungal gene labels are in light blue and those from Monosiga are labeled in red. Branches leading to MEIS class genes are highlighted in red, IRO class in dark blue, SIX class in purple, POU class in green, and LIM class in orange.
APPENDIX

Developmental expression and functional data for genes of the proximodistal appendage patterning network in appendages and anterior neurectoderm

Dll/Dlx

Arthropods

Fly (Drosophila melanogaster)

Appendage:
Expression of Dll in distal domains of leg and antennal imaginal discs with a small proximal domain in developing legs. Distal leg and antennal truncations and defects in mutants (Abu-Shaar and Mann, 1998; Dong et al., 2001; Panganiban and Rubenstein, 2002; Wu and Cohen, 1999).

Anterior neurectoderm:
Dll expression (using a Dll-lacZ reporter line) in a complex pattern of the larval brain in the optic lobe (Kaphingst and Kunes, 1994). Dll has also been found to be expressed in anterior parts of the late embryonic brain, including glial cells in the protocerebrum, as well as other clusters of brain cells in the protocerebrum and deuterocerebrum, but not in tritocerebrum and posterior deuterocerebrum (Panganiban and Rubenstein, 2002; Sprecher et al., 2007).

Other:
Unknown

Beetle (Tribolium castaneum)

Appendage:
Expression of Dll starts prior to appendage budding and is in medial and distal domains of early leg buds and is downregulated later in development in medial domains as dac expression increases. This results in similar pattern to Drosophila where Dll in expressed in distal domains with secondary proximomedial stripe and partial overlap with dac. Fused segments and other distal appendage defects in mutants (Beermann et al., 2004; Beermann et al., 2001; Prpic et al., 2001).
Anterior neurectoderm:

*Dll* expression in the embryonic procephalon (Beermann et al., 2001).

Other:

*Dll* expression in small groups of PNS cells (Beermann et al., 2001).

**Milkweed bug** (*Oncopeltus fasciatus*)

**Appendage:**

Expression of *Dll* starts before appendage budding begins. As the buds develop expression becomes restricted to distal regions with a small stripe more proximally developing later in thoracic limbs reminiscent of the *Drosophila* leg imaginal disc expression patterns. Distal expansion of *dac* and *hth* expression domains with absent or defective distal appendage structures in RNAi embryos (Angelini and Kaufman, 2004).

Anterior neurectoderm:

*Dll* expression in the procephalon during embryogenesis (Rogers et al., 2002).

Other:

*Dll* expression in pleuropodial glands (Angelini and Kaufman, 2004).

**Cricket** (*Gryllus bimaculatus*)

**Appendage:**

Expression of *Dll* in distal domains of developing legs. Early expression mostly complementary to *hth* comprising medial and distal domains. After robust *dac* expression appears in medial regions *Dll* is restricted to distal regions and a secondary domain at the border of *hth* and *dac* expression. Partially overlapping expression with *dac* and *hth* reminiscent of expression in *Drosophila* legs (Inoue et al., 2002).

Anterior neurectoderm:

*Dll* expression in the procephalon during embryogenesis (Rogers et al., 2002).

Other:

Unknown

**Millipede** (*Glomeris marginata*)

**Appendage:**
Expression of *Dll* in the center of appendage domains before appendage budding and in distal domains of developing appendages. (Prpic and Tautz, 2003).

**Anterior neurectoderm:**
*Dll* expression in embryonic brain (Prpic and Tautz, 2003).

**Other:**
*Dll* expression in anal valves (Prpic and Tautz, 2003).

**Vertebrates**

**Mouse (Mus musculus)**

**Appendage:**
Expression of *Dlx2*, *Dlx5*, and *Dlx6* in distal limb domains including AER cells with altered limb patterning and distal limb defects in *Dlx5/6* mutant mice (Panganiban and Rubenstein, 2002; Robledo et al., 2002).

**Anterior neurectoderm:**
*Dlx1*, *Dlx2*, *Dlx5*, and *Dlx6* are expressed in domains in the telencephalon and diencephalon of developing embryos. *Dlx5/6* mutants have anterior head defects including exencephaly (Panganiban and Rubenstein, 2002; Robledo et al., 2002).

**Other:**
Expression of *Dlx* genes in neural crest, branchial arches, olfactory placode, otic placode, optic cup, and retina (Panganiban and Rubenstein, 2002).

**Chick (Gallus gallus)**

**Appendage:**
Expression of *Dlx3* and *Dlx5* starts early in limb bud development in distal cells including the AER (Ferrari et al., 1995; Pera and Kessel, 1999).

**Anterior neurectoderm:**
*Dlx* orthologs expressed in various domains in developing telencephalon and diencephalon (Borghjid and Siddiqui, 2000; Fernandez et al., 1998; Pera and Kessel, 1999; Puelles et al., 2000; Zhu and Bendall, 2006).

**Other:**
Expression of *Dlx3* in olfactory placode, otic placode, branchial arches and posterior tailbud (Pera and Kessel, 1999).
Frog (*Xenopus laevis*)

**Appendage:**
Unknown

**Anterior neurectoderm:**
Expression of *X-dll3* and *X-dll4* in telencephalon and parts of the diencephalon during development (Papalopulu and Kintner, 1993).

**Other:**

Zebrafish (*Danio rerio*)

**Appendage:**
Expression of all 8 *Dlx* genes in apical ectodermal cells of pectoral fin buds as well as in the medial fin fold (Akimenko et al., 1994; Ellies et al., 1997; Zerucha and Ekker, 2000).

**Anterior neurectoderm:**
Expression of 5 *Dlx* orthologs in telencephalic and diencephalic domains during development (Akimenko et al., 1994; Ellies et al., 1997; Zerucha and Ekker, 2000).

**Other:**
Many, but not all, of the *Dlx* orthologs are expressed in olfactory placode, otic vesicle, and visceral arches (Akimenko et al., 1994; Ellies et al., 1997; Zerucha and Ekker, 2000).

Others

Tunicate (*Ciona intestinalis*)

**Appendage:**
N/A

**Anterior neurectoderm:**
Expression of *Dlx* in anterior neurectoderm in tailbud stage embryos (Harafuji et al., 2002).
Amphioxus (Branchiostoma floridae)

Appendage: N/A

Anterior neurectoderm:
*Dlx* expression in neural plate and anterior ectoderm in neurula stage embryos in cells that will become part of the anterior cerebral vesicle in later stage embryos (Holland et al., 1996).

Other:
*Dlx* expression is fairly broad in the dorsal ectoderm of early embryos (Holland et al., 1996).

btd/Sp8

Arthropods

Fly (Drosophila melanogaster)

Appendage:
Expression of *btd:Sp1* (*Drosophila Sp8* class genes) throughout entire appendage primordia in early embryogenesis and remaining throughout embryogenesis (Wimmer et al., 1996). During imaginal disc development expression is seen in most cells except the most proximal and most distal cells (Estella et al., 2003). Keilin’s organs are lost or defective, and *Dll* expression is lost in antennal and thoracic appendage primordia in mutants (Estella et al., 2003).

Anterior neurectoderm:
*Btd* and *Sp1* expression in anterior procephalic neurectoderm starting in early embryogenesis (Wimmer et al., 1996).

Other:
Strong stripe of *btd* expression in intercalary/mandibular region of blastoderm embryos, where it functions as a gap gene (Cohen and Jurgens, 1990). Later expression in subsets of PNS and ventral nerve cord cells (Wimmer et al., 1996).

Beetle (Tribolium castaneum)

Appendage:
Expression of *Sp8* starts prior to formation of appendage buds, and in early limb buds is fairly uniform. As limb bud development progresses expression becomes
refined into multiple medial and distal stripes with the most proximal stripe being just proximal to the Dll expression domain. Appendage defects including missing or fused segments in RNAi embryos (Beermann et al., 2004).

**Anterior neurectoderm:**
Sp8 expression seen in head lobes (Beermann et al., 2004).

**Other:**
Unknown

**Milkweed bug (Oncopeltus fasciatus)**

**Appendage:**
Expression of Sp8/9 is broad during earliest stages of appendage bud development encompassing the entirety of the buds. In later stages expression is restricted to two separate distal domains in the antennae and multiple stripes from proximal to distal in the legs. Fused segments and other appendage malformations seen in RNAi embryos with all appendages being shorter (Schaeper et al., 2009).

**Anterior neurectoderm:**
Sp8 is expressed in anterior neurectoderm in cells of the protocerebrum and cells that will give rise to the optic center (Schaeper et al., 2009).

**Other:**
Sp8 is expressed in a subset of cells in the ventral nerve cord (Schaeper et al., 2009).

**Vertebrates**

**Mouse (Mus musculus)**

**Appendage:**
Expression of Sp8 in limb buds and high levels in AER cells. Expression in the telencephalon. Loss of distal limb structures in mutants (Bell et al., 2003; Treichel et al., 2003).

**Anterior neurectoderm:**
Expression of Sp8 in telencephalon, olfactory placode, and midbrain-hindbrain border. Forebrain developmental defects, exencephaly, and patterning defects in mutants (Bell et al., 2003; Treichel et al., 2003; Zembrzycki et al., 2007).

**Other:**
Expression of Sp8 in neural tube, genital tubercle, nasal processes, and spinal cord (Bell et al., 2003; Treichel et al., 2003).

**Chick (Gallus gallus)**

**Appendage:**
Expression of Sp8 in early limb field becoming confined to distal regions later in development with strong expression in the AER. Missing or malformed distal limb structures when dominant negative Sp8 is expressed (Kawakami et al., 2004).

**Anterior neurectoderm:**
Expression of Sp8 in anterior telencephalon and at the midbrain-hindbrain boundary (Kawakami et al., 2004).

**Other:**
Expression of Sp8 in the neural tube (Kawakami et al., 2004).

**Zebrafish (Danio rerio)**

**Appendage:**
Expression of Sp8 in distal regions of developing pectoral fins (apical fold). Reduced pectoral fin outgrowth in morpholino injected embryos (Kawakami et al., 2004).

**Anterior neurectoderm:**
Expression of Sp8 in forebrain and at the midbrain-hindbrain boundary (Kawakami et al., 2004).

**Other:**
Expression of Sp8 in the neural tube, neural keel, and tail bud (Kawakami et al., 2004).

**Others**

**Tunicate (Ciona intestinalis)**

**Appendage:**
N/A

**Anterior neurectoderm:**
Expression of Sp8 in the anteriormost neurectoderm in tailbud stage embryos, likely corresponding to palps (Imai et al., 2004).

**dac/Dach**

**Arthropods**

**Fly (Drosophila melanogaster)**

**Appendage:**
Expression of *dac* in medial cells of developing leg and antennal imaginal discs (Abu-Shaar and Mann, 1998; Dong et al., 2001; Lecuit and Cohen, 1997; Wu and Cohen, 1999). Shortened and fused medial segments in mutant legs (Mardon et al., 1994).

**Anterior neurectoderm:**
Expression of *dac* starts around stage 9 in medial protocerebral neurectoderm in the embryonic procephalon with expression expanding more posteriorly by stage 10. These cells will give rise to parts of the mushroom body and nearby brain (Noveen et al., 2000).

**Other:**
Expression of *dac* in the developing eye, and in many cells of the ventral nerve cord in late embryonic stages (Mardon et al., 1994).

**Beetle (Tribolium castaneum)**

**Appendage:**
Strong expression of *dac* begins proximally in early appendage budding which as the appendage buds grow corresponds to medial domains. Later in embryonic appendage development minor proximal domains appear (Beermann et al., 2004; Prpic et al., 2001). The major domains of *dac* expression overlap with the proximal region of *Dll* expression (Prpic et al., 2001).

**Anterior neurectoderm:**
Expression of *dac* starting in early embryonic development in multiple anterior head ectoderm domains (possibly future mushroom body) (Prpic et al., 2001).

**Other:**
Expression of *dac* in embryonic dorsal mesoderm, Malpighian tubules, posterior gut, a subset of cells in the ventral nerve cord, and eye primordia (Prpic et al., 2001).
Milkweed bug (*Oncopeltus fasciatus*)

**Appendage:**
Expression of *dac* in medial segments of developing legs and antennae with more broad expression in mandibular and maxillary appendages. Medial expansion of *Dll* expression domain in thoracic limbs and absent or defective medial appendage structures in *dac* RNAi embryos (Angelini and Kaufman, 2004).

**Anterior neurectoderm:**
Expression of *dac* seen in CNS (extent difficult to determine) (Angelini and Kaufman, 2004).

**Other:**

Cricket (*Gryllus bimaculatus*)

**Appendage:**
Strong expression of *dac* in medial domain starting in early leg development. In later embryonic stages expression corresponds to distal femur, tibia, and proximal tarsus. Partially overlapping expression with *dll* and *hth* reminiscent of expression in *Drosophila* legs (Inoue et al., 2002).

**Anterior neurectoderm:**
Expression of *dac* in early head neurectoderm and persisting during development in a subset of cells in the protocerebrum and deuterocerebrum. This expression likely corresponds to expression domain seen later in adult brains (Inoue et al., 2004).

**Other:**
Expression of *dac* in embryonic eye primordia starting in early development (Inoue et al., 2004).

Millipede (*Glomeris marginata*)

**Appendage:**
Expression of *dac* in rings around the center of appendage domains before appendage budding and in medial domains of developing appendages with lower level proximal expression and distal PNS cell expression starting in later limb development. (Prpic and Tautz, 2003).

**Anterior neurectoderm:**
Expression of *dac* in subset of protocerebrum and deuterocerebrum (Prpic and Tautz, 2003).

**Other:**
Expression of *dac* in anal valves, heart primordium, and a subset of cells in the ventral nerve cord (Prpic and Tautz, 2003).

**Vertebrates**

**Mouse** (*Mus musculus*)

**Appendage:**
Expression of *Dach1* in limb buds begins around E11 in anterior and posterior medial domains, and by E14.5 is found in distal and medial regions of the limb bud (Caubit et al., 1999; Davis et al., 1999; Hammond et al., 1998). Expression of *Dach2* is largely similar to that of *Dach1* in limb buds (Davis et al., 2001).

**Anterior neurectoderm:**
Expression is dynamic in the developing CNS with strongest expression in the telencephalon and near the midbrain/hindbrain boundary (Caubit et al., 1999; Davis et al., 1999). Expression of *Dach2* is largely similar to that of *Dach1* except that *Dach2* is expressed in a subset of the telencephalon pattern of *Dach1* (Davis et al., 2001).

**Other:**
Expression in neural tube, eye, otic pit, trunk mesoderm, branchial arches, and mammary gland primordia (Caubit et al., 1999; Davis et al., 1999; Davis et al., 2001; Hammond et al., 1998).

**Chick** (*Gallus gallus*)

**Appendage:**
Expression in the developing limb buds, mainly in proximal and medial regions and strongest in the anterior, but also expressed in the AER (Heanue et al., 2002). Derepression of *Meis2* in limb buds when dominant negative *Dach* is expressed (Kida et al., 2004).

**Anterior neurectoderm:**
Expression of *Dach1* is dynamically expressed in many tissues of the developing brain with highest levels in the telencephalon and midbrain (Heanue et al., 2002).

**Other:**
Expression of Dach1 in branchial arches, neural tube, somites, nasal placode, eye/retina, and large intestine (Heanue et al., 2002).

Zebrfish (Danio rerio)

Appendage:
Expression of all three dach genes in pectoral fin buds with dachB and dachC displaying somewhat distal expression and dachA expressed in anterior and posterior limb bud regions similar to mouse Dach (Hammond et al., 2002).

Anterior neurectoderm:
The three Dach orthologs display dynamic expression in the brain during development with expression seen in fore, mid and hindbrain at various points in development (Hammond et al., 2002).

Other:
Expression of dachA in the eye/retina and somites, dachB in the lateral line system, dachC in the pronephros and neural crest, and all three genes in the otic vesicle (Hammond et al., 2002).

Others

Tunicate (Ciona intestinalis)

Appendage:
N/A

Anterior neurectoderm:
Dach is expressed in the nerve cord just posterior to the head-tail border in tailbud stage embryos (Mazet et al., 2005).

Other:
Dach expression at the tailbud stage is detected in the buccal cavity, in head mesenchyme. At larval stages, Dach is expressed in mesenchyme associated with the palps as well as in posterior mesenchyme (Mazet et al., 2005).

Amphioxus (Branchiostoma floridae)

Appendage:
N/A

Anterior neurectoderm:
Expression of *Dach* in posterior cells of the cerebral vesicle as well as multiple more posterior neural domains (Candiani et al., 2003).

**Other:**
Expression of *Dach* in paraxial mesoderm in gastrula stage embryos corresponding to somites and pharynx endoderm in later stages. Expression also in the developing frontal eye and ventrolateral neurons in the nerve cord (Candiani et al., 2003).

**hth/Meis**

**Arthropods**

**Fly (Drosophila melanogaster)**

**Appendage:**
Expression of *hth* starts early in embryogenesis and covers most of the blastoderm stage embryo except the anterior and posterior termini. In later embryonic stages *hth* expression disappears from many of the cells of the appendage primordia (Kurant et al., 1998; Riechhof et al., 1997). In leg imaginal discs, *hth* is expressed in proximal cells, and *hth* mutants exhibit proximal leg defects (Abu-Shaar and Mann, 1998; Wu and Cohen, 1999).

**Anterior neurectoderm:**
Expression of *hth* in posterior head neurectoderm but largely excluded from anterior procephalon (Kurant et al., 1998).

**Other:**
Expression of *hth* is very broad and includes most of trunk ectoderm, gastric caeca, midgut endoderm, eye imaginal disc, and ventral nerve cord (Kurant et al., 1998; Riechhof et al., 1997).

**Beetle (Tribolium castaneum)**

**Appendage:**
Expression of *hth* in proximal cells of early limb buds and, as development progresses, in an additional small stripe of medial cells (Prpic et al., 2003).

**Anterior neurectoderm:**
Extent of *hth* expression not clear, but appears to be absent in anteriormost cells (Prpic et al., 2003).
Other:
Expression of hth in a large proportion of the body except for the distal limbs and anterior head (Prpic et al., 2003).

**Milkweed bug (Oncopeltus fasciatus)**

Appendage:
Expression of hth orthologs in body wall and proximal regions of developing appendages of Oncopeltus embryos. Absent or fused proximal appendage structures in hth RNAi embryos (Angelini and Kaufman, 2004).

Anterior neurectoderm:
Expression of hth seen in most of the developing head except for the anteriormost region (Angelini and Kaufman, 2004).

Other:
Expression of hth throughout most of the embryo except distal limbs and anterior head (Angelini and Kaufman, 2004).

**Cricket (Gryllus bimaculatus)**

Appendage:
Expression of hth in proximal regions of limb buds. Partially overlapping expression with dll and dac reminiscent of expression in Drosophila legs (Inoue et al., 2002).

Anterior neurectoderm:
Unknown

Other:
Unknown

**Millipede (Glomeris marginata)**

Appendage:
Expression of hth in developing appendages is proximal and medial in legs and relatively uniform in others, but exd is only expressed proximally. (Prpic and Tautz, 2003).

Anterior neurectoderm:
Expression of hth in subset of protocerebrum and deuterocerebrum (Prpic and Tautz, 2003).
Other:
Expression of *hth* in lateral plates and many other tissues (Prpic and Tautz, 2003).

Vertebrates

**Mouse (Mus musculus)**

**Appendage:**
Expression of *Meis1* and *Meis2* in body wall and proximal regions of developing mouse limb buds (Mercader et al., 1999).

**Anterior neurectoderm:**
Expression of *Meis1* and *Meis2* in overlapping but distinct domains throughout much of the hindbrain and midbrain (Zhang et al., 2002).

**Other:**
Expression of *Meis1* and *Meis2* in the trunk mesoderm, developing eyes, and nasal placode (Mercader et al., 1999).

**Chick (Gallus gallus)**

**Appendage:**
Expression of *Meis1* and *Meis2* in body wall and proximal regions of developing chick limb buds and proximalization of limb structures in *Meis* overexpression embryos (Capdevila et al., 1999; Mercader et al., 1999).

**Anterior neurectoderm:**
Strong expression of *Meis1* and *Meis2* in the midbrain with *Meis1* expressed in subdomains of the diencephalon (hindbrain not analyzed) (Ferran et al., 2007).

**Other:**
Expression of *Meis1* and *Meis2* in the trunk mesoderm (Mercader et al., 1999).

**Frog (Xenopus laevis)**

**Appendage:**
Unknown

**Anterior neurectoderm:**
Expression of *Meis1* early in the midbrain primordia with low level expression in the neural plate. Later in development expression is in separate domains in the
midbrain and in the hindbrain (Maeda et al., 2001). Expression of *Meis3* in hindbrain in developing *Xenopus* embryos. When *Meis* is ectopically expressed, there are missing anterior structures, inhibited anterior neural marker expression and ectopic posterior neural marker expression (Salzberg et al., 1999).

Other:
Expression of *Meis1* in the optic cup, branchial arches, otic vesicle, olfactory placode, and spinal cord (Maeda et al., 2001).

**Zebrafish (Danio rerio)**

Appendage:
Expression of *meis1.1* in proximal cells of pectoral fin buds in later developmental stages (Waskiewicz et al., 2001).

Anterior neurectoderm:
Expression of *meis2* during early embryogenesis in the hindbrain primordia. Strong hindbrain expression persists throughout development and small domains in the telencephalon and midbrain appear during somitogenesis (Biemar et al., 2001; Zerucha and Prince, 2001). Expression of *meis1.1* and *meis2.2* in similar but distinct domains in forebrain, midbrain, and hindbrain over many developmental stages, with *meis3.1* restricted to the hindbrain (Waskiewicz et al., 2001).

Other:
Expression of *meis2* in spinal cord (stronger more anteriorly), somites, lateral mesoderm, heart tube, pharyngeal arches, and retina (Biemar et al., 2001; Zerucha and Prince, 2001). Expression of *meis1.1* in retina and *meis1.1*, *meis2.2*, and *meis3.1* expression in the spinal cord (Waskiewicz et al., 2001).

Others

**Tunicate (Ciona intestinalis)**

Appendage:
N/A

Anterior neurectoderm:
Expression of *Meis* comprises the posterior half of the ectoderm and sensory vesicle primordia during neurula stages. During tailbud stages expression is in the ventral anterior sensory vesicle and posterior head ectoderm (Imai et al., 2004).

Other:
Expression of *Meis* in anterior tail ectoderm during tailbud stages (Imai et al., 2004).

*al/Alx/Arx*

**Arthropods**

**Fly (Drosophila melanogaster)**

**Appendage:**
Expression of *al* in embryonic appendage primordia and in distal most cells of leg and antennal discs as well as in cells which contribute to body wall/proximal most appendage (Campbell et al., 1993; Schneitz et al., 1993). Loss of distal most structures in legs and antennae of *al* mutants (Schneitz et al., 1993). Loss of *al* expression in Dll mutant leg discs (Campbell and Tomlinson, 1998). Reciprocal repression between *al* and *BarH1* in leg discs (Kojima et al., 2000).

**Anterior neurectoderm:**
Expression of *al* in anterior neurectoderm not observed in early embryos (Schneitz et al., 1993).

**Other:**
Expression of *al* in histoblasts, dorsal ectoderm, stomodaeum, and cells lining the amnioserosa in stage 11 embryos (Schneitz et al., 1993).

**Beetle (Tribolium castaneum)**

**Appendage:**
Expression of *al* at the distal tips of developing appendages in *Tribolium* with secondary proximal domain near the body wall and later in development an additional domain in a medial position. Missing or aberrant distal most structures of antennae and legs in *al* RNAi treated embryos (Beermann and Schroder, 2004).

**Anterior neurectoderm:**
No *al* expression detected

**Other:**
Expression of *al* in ventral histoblasts and dorsal ectoderm (Beermann and Schroder, 2004).

**Cricket (Gryllus bimaculatus)**
Appendage:
Expression of *al* is very similar to that of *Tribolium*. Expression starts in distal tips of early appendage buds as well as a secondary proximal domain (coxopodite) in many appendages, with additional medial domains developing later in embryogenesis in leg buds (Miyawaki et al., 2002).

Anterior neurectoderm:
Expression of *al* not detected

Other:
Expression of *al* in embryonic hindgut (Miyawaki et al., 2002).

**Vertebrates**

**Mouse (*Mus musculus*)**

Appendage:
Expression of *Alx4* in anterior of early (E10.5) embryonic limb buds but by E15.5 expression is only in the distal tips of the forming digits (Qu et al., 1997).

Anterior neurectoderm:
Expression of *Arx* throughout much of the telencephalon and parts of the diencephalon (Miura et al., 1997).

Other:
Expression of *Alx-4* in embryos in anterior craniofacial mesenchyme and first branchial arch (Qu et al., 1997). Expression of *Arx* in optic areas, somites, and floor plate (Miura et al., 1997).

**Chick (*Gallus gallus*)**

Appendage:
Expression of *Alx4* broadly in early limb buds, becoming confined to mainly anterior proximal cells as development progresses (Takahashi et al., 1998).

Anterior neurectoderm:
Expression of *Arx* in dynamic domains of the telencephalon and diencephalon and reduced expression in *Dlx1/2* mutant embryos (Cobos et al., 2005).

Other:
Unknown
**Frog (Xenopus laevis)**

**Appendage:**
Unknown

**Anterior neurectoderm:**
Expression of \( Arx \) in dynamic domains of the telencephalon and diencephalon and forebrain defects in morpholino injected embryos (Seufert et al., 2005).

**Other:**
Expression of \( Arx \) in the floor plate (Seufert et al., 2005).

**Zebrafish (Danio rerio)**

**Appendage:**
Unknown

**Anterior neurectoderm:**
Expression of \( Arx \) in dynamic domains of the telencephalon and diencephalon (Miura et al., 1997).

**Other:**
Expression of \( Arx \) in somites and floor plate (Miura et al., 1997).

**ap/Lhx2/9**

**Arthropods**

**Fly (Drosophila melanogaster)**

**Appendage:**
Expression of \( ap \) in fourth tarsal segment during pupal leg development (Cohen et al., 1992; Pueyo et al., 2000), and \( ap \) mutant flies are missing the fourth tarsal segment (Pueyo et al., 2000).

**Anterior neurectoderm:**
Expression of \( ap \) in the embryonic anterior neurectoderm/brain starting at stage 11 (Cohen et al., 1992).

**Other:**
Expression of \( ap \) in body wall muscle, pharynx, optic lobe, and a subset of cells in the ventral nerve cord (Rincon-Limas et al., 1999).
Vertebrates

Mouse (*Mus musculus*)

**Appendage:**
Expression of *Lhx2* and *Lhx9* in distal regions of developing limb buds starting at E10 except for the AER (Retaux et al., 1999; Rincon-Limas et al., 1999). Severe malformations and truncations of distal limb structures in *Lhx2/Lhx9* double mutants and altered gene expression of limb patterning genes (e.g. FGF10) (Tzchori et al., 2009).

**Anterior neurectoderm:**
Expression of *Lhx2* and *Lhx9* in the developing brain starting at E10.5 in distinct but overlapping patterns in the telencephalon and diencephalon with additional expression seen in both midbrain and hindbrain (Retaux et al., 1999; Rincon-Limas et al., 1999). Forebrain malformations in *Lhx2* mutants (Porter et al., 1997). Patterning defects in telencephalon of *Lhx2* mutants (Bulchand et al., 2001).

**Other:**

Chick (*Gallus gallus*)

**Appendage:**
Expression of *Lhx2* in distal regions of developing chick limb buds (but not the AER) starting at the beginning of limb bud development and continuing throughout limb development. Expression of dominant negative *Lhx2* results in truncated limbs and misexpression of limb patterning genes (Rodriguez-Esteban et al., 1998).

**Anterior neurectoderm:**
Expression of *Lhx2* in anterior CNS (Rodriguez-Esteban et al., 1998).

**Other:**
Expression of *Lhx2* in eye and nasal epithelia and spinal cord (Rodriguez-Esteban et al., 1998).

Zebrafish (*Danio rerio*)
Appendage:
Expression of Lhx2 in developing pectoral fin buds, widespread early (26hpf) and later restricted to posterior regions (48hpf) (Seth et al., 2006).

Anterior neurectoderm:
Expression of Lhx2 and Lhx9 in forebrain of early embryos. Strong expression of Lhx2 throughout development in the telencephalon and diencephalon with small segmentally repeated domains in the mid and hindbrain. Lhx9 is expressed in a very similar pattern but is only in a subset of the telencephalon cells that Lhx2 is found in (Ando et al., 2005). Forebrain axon guidance defects and altered forebrain patterning gene expression (e.g. Dlx2) seen in Lhx2 mutants (Seth et al., 2006).

Other:
Expression of Lhx2 in the optic vesicle (Seth et al., 2006).

**B-H1/BarH**

**Arthropods**

**Fly (Drosophila melanogaster)**

Appendage:
Expression of BarH1 (B-H1) in cells corresponding to tarsal segment 4 in early 3rd instar leg discs and expression in tarsal segments 4 and 5 during late 3rd instar leg disc and pupal leg development. Fusion of tarsal segments in mutants. Expression in distal region (but not the most distal cells) of antennal discs and loss of distal structures in mutants. Reciprocal repression between BarH1 and dac and between BarH1 and al. Activation of BarH1 by Dll. Activation of ap by BarH1 (Kojima et al., 2000).

Anterior neurectoderm:
Expression of BarH1 in anterior and central procephalic neurectoderm starting around stage 12 and continued expression in the brain in late stage embryos (Higashijima et al., 1992).

Other:
Expression in segment specific patterns in the PNS and posterior CNS and in developing eyes (Higashijima et al., 1992).

**Vertebrates**
**Mouse (Mus musculus)**

**Appendage:**
No expression detected for *Barhl1* and unknown for *Barhl2*.

**Anterior neurectoderm:**
Expression of *Barhl1* and *Barhl2* in the diencephalon starting at E9.5, and at later stages in the midbrain and hindbrain (Bulfone et al., 2000; Mo et al., 2004).
*Barhl2* is expressed in the telencephalon starting at E14.5 (Mo et al., 2004).

**Other:**
Expression in spinal cord (both), inner ear (*Barhl1*), and retina (*Barhl2*) (Bulfone et al., 2000; Mo et al., 2004).

**Frog (Xenopus laevis)**

**Appendage:**
Unknown

**Anterior neurectoderm:**
Expression of two BarH genes (*XBH1* and *XBH2*) in separate domains of the diencephalon of postneurula stage embryos with lower level domains in the midbrain and hindbrain appearing later in development (Patterson et al., 2000).

**Other:**
Expression of *XBH1* in the developing retina (Patterson et al., 2000).

**Zebrafish (Danio rerio)**

**Appendage:**
Unknown

**Anterior neurectoderm:**
Expression of *barhl2* in diencephalon primordia of early embryos. Expression of the three BarH genes (*barhl1.1, barhl1.2* and *barhl2*) in partially overlapping domains of the diencephalon after 14 somite stage. All three are found in largely similar midbrain and hindbrain domains by 48hpf, with *barhl2* also being expressed in the telencephalon during middle embryonic stages (Colombo et al., 2006).

**Other:**
Expression of all three BarH genes in developing retina (Colombo et al., 2006).
REFERENCES


