Out on a *LIM*:

LIM homeobox genes in the ctenophore *Pleurobrachia bachei*.

Rachel Sanford[1,2], Leonid Moroz[1,2], Billie Swalla[1,3], Andrea Kohn[1,2]

Marine Genomics Research Apprenticeship 2012
Spring 2012

[1]Friday Harbor Laboratories, University of Washington, Friday Harbor, WA 98250
[2]Department of Neuroscience, University of Florida, Gainesville, FL 32611
[3]Department of Biology, University of Washington, Seattle, WA 98195

Contact Information:
Rachel Sanford
Department of Neuroscience
University of Florida
1149 South Newell Drive
Gainesville, FL 32611
rachelsusansanford@gmail.com

Keywords: LIM homeobox, *Pleurobrachia bachei*, neural development, evolution, ctenophore

Abstract

LIM homeobox (Lhx) gene family consists of six subfamilies of patterning genes which play a major role in the development of the nervous system and are found only in animals. We looked at their presence and expression in the ctenophore *Pleurobrachia bachei* to study their evolution and development. Ctenophores are among the most basal metazoans and one of the first animals to have a nervous system. We found four *Lhx* genes in *P. bachei*, *Lhx1/5, Lhx3/4, Islet* and *Lmx,* with our gene prediction models. *Lhx2/9* and *Lhx6/8,* the other two subfamilies in most metazoans, were absent. None of these genes would clone using an adult cDNA library but *Lhx3/4* and *Lmx* were successfully cloned using an embryonic cDNA library. Ctenophores and sponges are the only taxons that have four *Lhx* gene families instead of six. According to this genomic data, ctenophores and sponges are the most basally branched metazoans. Further research will be needed to understand the role that the Lhx gene family plays in the evolution and development of nerves in ctenophores.

Introduction

Ctenophores, also called comb jellies, are small bi-radial marine organisms made up of a jelly-like substance. They are one of the earliest branches of metazoans, their close relatives are poriferans, cnidarians, and placozoans. Their basal position in the Metazoa, along with the presence of a nervous system containing a nerve net and an apical organ (Figure 1) make ctenophores the ideal candidate for studies concerning the early evolution of the nervous system. Almost all ctenophores are predators and actively hunt by catching small marine organisms with the colloblasts, or sticky cells, on their tentacles. Movement is accomplished by eight rows of coordinated cilia or combs. This coordinated movement and predatory behavior may reflect the complexity of the neural system. We looked at the presence and expression of the LIM homeobox genes in *Pleurobrachia bachei* to study how the early nervous system develops.

The first LIM homeobox (*Lhx*) gene was discovered in *Caenorhabditis elegans* when the *Lhx* homolog *MEC-3* was isolated and shown to be required for neural development[3]. Later studies found the genes *LIN-11* and *Islet1*, which combined with *MEC-3*, comprise the LIM domain acronym[4,5]. The LIM domain contains highly conserved regions rich in cytosine and histidine which bind to zinc ions to create tandem zinc-finger domains. These domains function as a protein binding interface which mediates the action of the rest of the protein. LIM domains can be located by themselves or in combination with other domains such as a homeobox domain to make up a *Lhx* gene[6]. *Lhx* genes have a general structure of two zinc-finger LIM domains at the amino-terminal followed by a helix-turn-helix homeodomain at the carboxyl-terminal (Figure 2). The zinc-finger LIM domains interact with nuclear proteins while the homeodomain binds to DNA as a transcription factor[7]. *Lhx* proteins are involved in tissue specification particularly in the nervous system[2]. These proteins overlap to create a pattern which has been termed the "LIM

code"[7]. In vertebrates *Lhx* proteins are also involved in a diverse array of body patterning including the head, limbs, eyes, heart, and kidneys[8]. In basal metazoans such as *Nematostella* and *Mnemiopsis, Lhx* proteins are suggested to be involved neural specification and patterning[1,9].

The current literature suggests that *Lhx* genes are only present in metazoans. No *Lhx* genes have been found in choanoflagellates and other non-metazoans[9]. All metazoans genomes examined so far contain *Lhx* genes. This data suggest that these genes are a metazoan innovation (Figure 3). *Lhx* genes have been subdivided into six subfamilies: *Lhx1/5, Lhx2/9, Lhx3/4, Lhx6/7, Islet* and *Lmx*[2]. Twelve human *Lhx* genes have been found, six *Drosophila*, six *Nematostella*, six *Tricoplax*, four *Amphimedon*, and four *Mnemiopsis* [8,1].

For this study we examined *Lhx* genes in the ctenophore *Pleurobrachia bachei*. The recent sequencing and assembly of this genome provided us with a starting ground to study development of the ctenophore nervous system through *Lhx* genes and compare this to the rest of the metazoa.

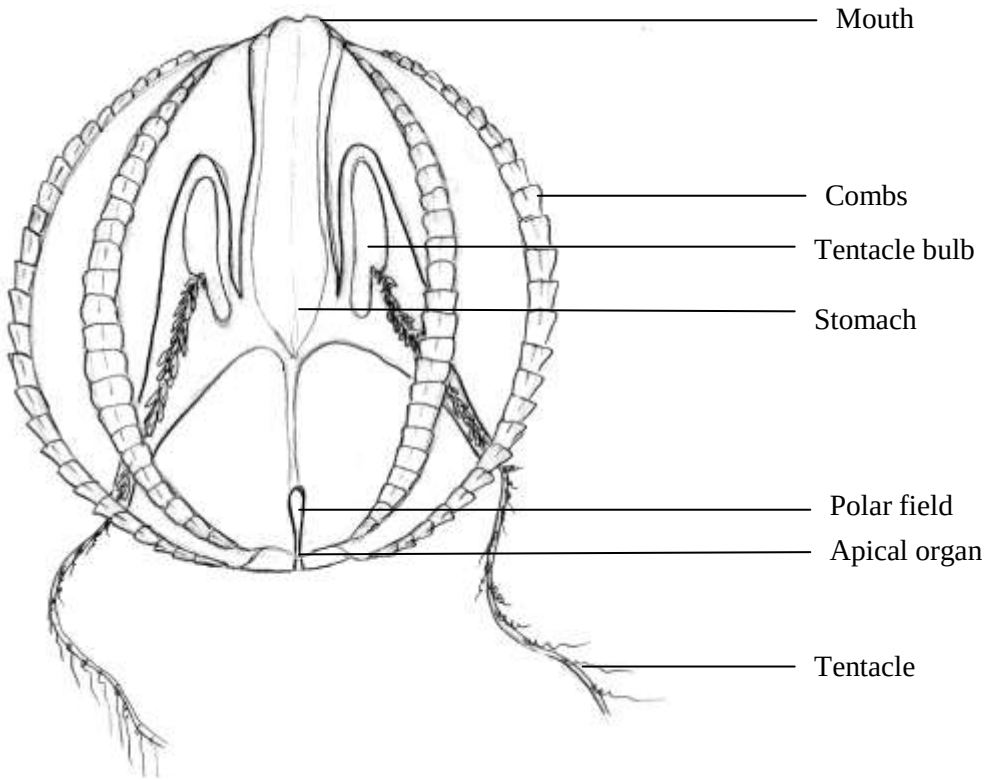


Figure 1. *Pleurobrachia bachei* anatomy. Ctenophores use their combs, which are made of cilia, for movement. The tentacle bulb controls the tentacle movement. The apical organ is the elementary brain.

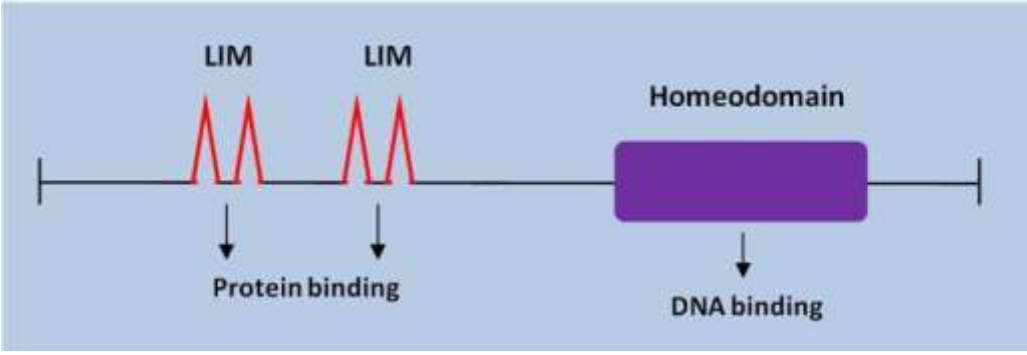
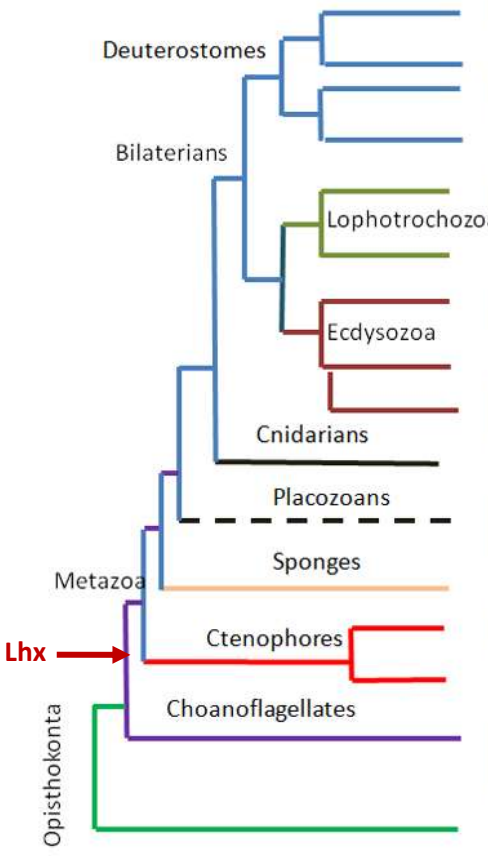


Figure 2. LIM homeobox general gene structure includes two LIM domains and one homeodomain LIM domains contain two zinc fingers. Adapted from Hobert et al. 2000[2]



	Lhx1/5	Lhx2	Lhx3/4	Lhx6/8	Islet	Lmx
Strongylocentrotus	1	1	1	1	1	-
Saccoglossus	1	1	1	1	1	1
Homo	1	1	2	4	2	2
Branchiostoma	1	1	1	1	1	1
Capitella	1	1	1	1	1	1
Lottia	-	1	1	1	1	1
Drosophila	1	1	1	1	1	-
Daphnia	1	1	1	1	1	1
Caenorhabditis	2	1	2	1	1	-
Nematostella	1	1	1	1	1	1
Trichoplax	1	1	1	1	1	1
Amphimedon	1	-	1	-	1	1
Pleurobrachia	1	-	1	-	1	1
Mnemiopsis	1	-	1	-	1	1
Monosiga	-	-	-	-	-	-
Capsaspora	-	-	-	-	-	-

Figure 3. Number of *Lhx* genes in each gene subfamily in selected taxa and estimate of when *Lhx* originated. This data came from NCBI BLAST searches and NCBI protein database. See supplement 1 for protein sequences for each of these genes.

Materials and Methods

Animals

Pleurobrachia bachei used in this study were collected on the dock of Friday Harbor Laboratories of the University of Washington during April and May of 2012. The animals were kept in sea tables with constant circulation of sea water at 10°C.

Predicted proteins and primers

Following genome sequencing by the Moroz lab and assembly and modeling by Mat Citeralla and David Girardo, we predicted these four *Lhx* genes in *Pleurobrachia bachei* using the gene models created by Mat and David: *Lhx1/5*, *Lhx3/4*, *Islet* and *Lmx*. Primers were then made using these predicted proteins by Dr. Andrea Kohn.

Cloning

PCR was done using the Invitrogen TOPO TA Cloning Kit. PCR was done in 33 cycles at 53°C. We used cDNA libraries made by Dr. Andrea Kohn with RNA from embryos collected by Gabrielle Winters, Emily Dabe, Caleb Bostwick and Dr. Billie Swalla. The PCR products were visualized and isolated on a 2% SizeSelect E-Gel. Products were then ligated into the TOPO vector and transformed into competent *E. coli* using the Invitrogen Cloning Kit. The transformed *E. coli* was then plated onto Luria broth and kanamycin agar plates and incubated at 37°C overnight. Two colonies were selected from the plates and grown up in Luria broth and kanamcyin bacterial cultures overnight. Minipreps were then performed on these cultures using the Qiagen QIAprep Spin Miniprep kit. The samples were then sent to SeqWright for sequencing.

Sequence analysis

Sequences obtained from SeqWright were BLASTed against the gene models to confirm the sequence of interest was present. Sequences were analyzed and converted into amino acids using the programs SeqMan.

Phylogenetic analysis

Sequences from a wide selection of other phyla were obtained by using NCBI protein BLAST and searching NCBI's protein database (See supplement 1). Some sequences were also acquired from the Simmons et al. 2012[1] supplement. All 74 sequences were aligned using the program MEGA5[10], which implemented the MUSCLE[11] application. A maximum parsimony and a maximum likelihood tree was then made using MEGA5[10]. Domain comparisons for these proteins were found using Pfam[12].

Probe Design

Based on the sequences obtained from SeqWright, RNA probes were made for these genes. These probes were manufactured by Yelena Bobkova at the Whitney Laboratory, University of Florida. Forward direction sequences were cut with *Pme3* and transcribed using T7 RNA polymerase while backward direction sequences were cut with *NotI* and transcribed using T3 polymerase.

***In-situ* Hybridization**

Adult

This protocol was adapted from Derelle and Manuel 2007[13]. Animals were fixed in 4% paraformadehyde in Filtered Sea Water (FSW) overnight at 4°C. The next day, animals were rinsed three times for 10 minutes each in PTW at room temperature then washed in a solution of 1:1 Methanol (MeOH)/ PTW for 10 minutes at room temp. Animals then stored in 100% MeOH at -20°C for 2 hours up to a week. When ready, animals were rehydrated 10 min each in

solutions of MeOH/PTW 3:1, 1:1, 1:3, 0:1 at room temp. Then animals were placed in a 1:1 solution hybridization buffer (HB) and PTW for 15 min. Animals were incubated in prehybridization solution for 1 hour at 60°C. We subsequently incubated animals in HB with DIG-RNA probe at 60°C overnight. The following day the animals were washed in HB for 30 min at 60°C, 1:1 HB/PTW for 30 min at 60 °C, and PTW for 30 min at room temp. We blocked the animals in 10% Goat Serum for 60 min at room temp and then incubated them in anti-DIG 1/2000 at 4°C overnight. On the day of development we washed the animals four times 30 min each with PBS at room temp. Detection buffer was made and aliquoted into wells of a 24 well plate. When ready to develop add 20 uL of NBT/BICP was added to the wells. Animals were placed into detection buffer, covered with tin foil and put on ice. Animals were checked periodically until staining occured. The reaction was stopped by placing the animals in 4% PFA in MeOH for 30 min. The animals were washed three times in EtOH at room temp and stored in 100% EtOH at 4°C until ready to be mounted onto slides. To mount, the animals were cut in half and flattened out manually. They were placed into methylsalicylate until they sunk, placed onto slide and cover slipped with Permount.

Embryo

Embryo *in-situ* hybridization protocol is very similar to the above with a few important changes. The embryos need to have their outer membranes removed after fixation and storage in MeOH. This was done by manually pipetting the embryos repeatedly. The embryos were then washed according to the above protocol above by removing most of the liquid from the embryos then adding 1 ml of the wash.

Embryo Collection

The day prior to collection three to four jars of *Pleurobrachia* were set up for spawning. Each jar, which sat in a 10°C sea water tank to keep the temperature down, contained six to eight animals and was covered with black plastic around 9 pm. The lights in the lab were turned off until collection time which fluctuated with the cycle of spawning the *Pleurobrachia* were on. Water samples from the bottom of the jars were removed with turkey basters and embryos manually collected under a dissecting microscope using P20 pipettes.

Results

Our gene models predicted four *Lhx* genes in Pleurobrachia: *Lhx1/5, Lhx3/4, Islet,* and *Lmx.* We successfully cloned *Lhx3/4* and *Lmx* from a combined stage embryonic cDNA library (See Figure 5). None of these showed up with an adult cDNA library. When PCR was done with individual stages of embryonic development *Lhx3* showed a band in the 4 cell, 8 cell, 16 cell, 32 cell, 64 cell, early gastrulation, very strongly in late gastrulation and 1 day embryos. Lmx showed a band in the 2 cell stage, the 32 cell, 64 cell and early gastrulation (Figure 6).

Lhx1/5

Currently we only have a partial sequence for this gene that just contains the homeodomain region. We have found one LIM domain on a different scaffold but the other LIM domain is still missing. We suspect it is located on a scaffold by itself and will not be possible to find at this time. We have evidence that *Lhx1/5* is present in *Pleurobrachia* but we will not be able to clone it or do any expression work within the time frame of the FHL apprenticeship.

Lmx3/4

We do not have a full length sequence for this gene. We cloned *Lhx3/4* using a combined stage cDNA library. We have also cloned *Lhx3/4* from separate cDNA libraries for each stage in

development. *Lhx3/4* was present in 4 cell embryos, 16 cell embryos, 32 cell embryos, 64 cell embryos, early gastrulation, very highly in late gastrulation and in 1 day embryos (Figure 6).

In-situ hybridization on adult *Pleurobrachia* showed cell specific staining under one side of the comb rows in the larger animal tested (See Figures 12 and 13). The smaller animal tested showed no staining here. We suspect that these are egg cells because the smaller animal might not have developed gonads yet while the larger one had. Also, ctenophores carry eggs under one side of the comb row and sperm on the other. *In situ* hybridization results from embryos showed specific expression in micromeres of cleaving embryos and localized in the tentacle bulbs of 1.5 day old embryos (Figure 16).

Islet

Our *Islet* gene model is full length but has not yet been successfully cloned. Phylogenetic tree analyses shows this gene clustering with all the other *Islets* tested and most closely related to *Mneniopsis.* In this section of the tree all of the basal metazoans (cnidarians, poriferians, ctenophores, and placazoans) branch earlier than the other taxa but their relationship to each other cannot be discerned with this tree (See supplement 2).

Lmx

We have a full length *Lmx* sequence which has a deletion of one of the LIM domains. When compared with the *Mnenopsis* sequence our sequence matches very well except for a large deletion of one of the LIM domains (Figure 7). Lmx has successfully been cloned in the combined stage embryonic cDNA library. For the individual stage PCR, Lmx was present in 2 cell, 32 cell, 64 cell, and early gastrulation (Figure 6).

In-situ hybridization showed there was staining present in the tentacles (Figure 14 and 15). A long structure down the main tentacle was stained. There was no significant staining in

the remainder of the adult animal. The embryo *in situ* hybridization showed specific expression in the micromeres during cleavage (Figure 17).

Lhx gene tree

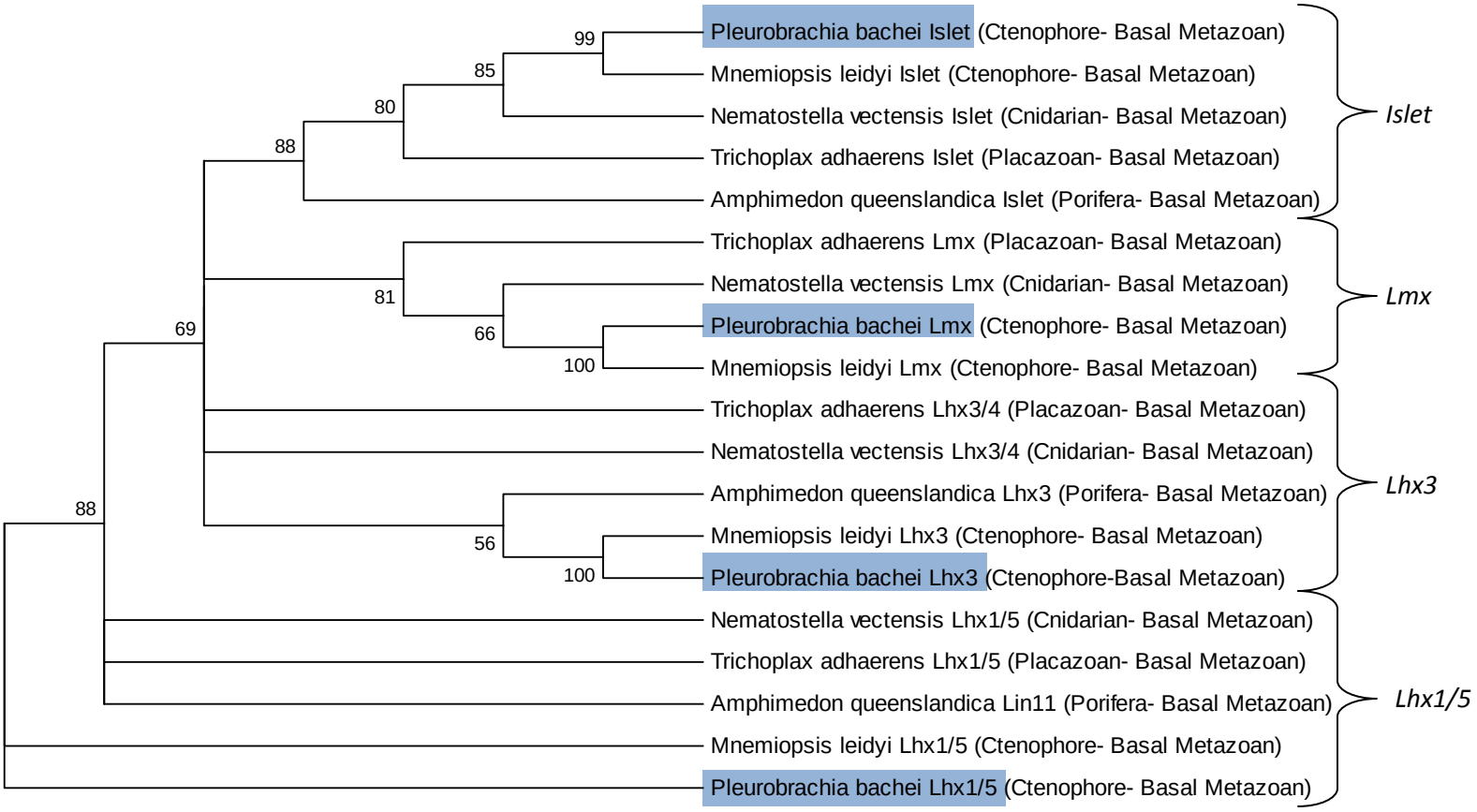


Figure 4. Maximum parsimony tree for the four *Lhx* genes present in ctenophores. These sequences were found on NCBI using BLAST or found in the supplement of Simmons et al. 2012[1]. See Supplement 2 for additional trees.
The evolutionary history was inferred using the Maximum Parsimony method. Tree #1 out of 3 most parsimonious trees (length = 2245) is shown. The consistency index is (0.770370), the retention index is (0.559081), and the composite index is 0.458720 (0.430699) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 1 in which the initial trees were obtained with the random addition of sequences (10 replicates). The analysis involved 19 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 579 positions in the final dataset. Evolutionary analyses were conducted in MEGA5

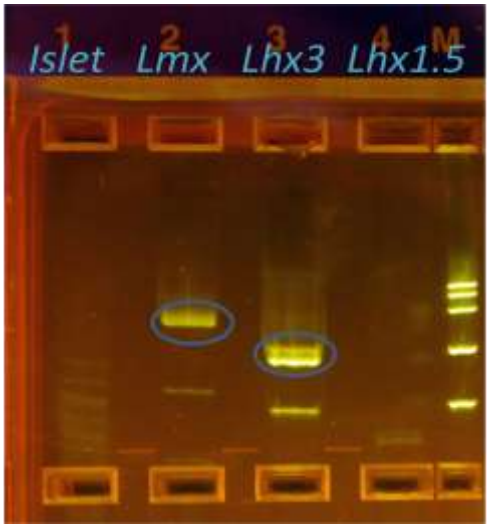


Figure 5. Successfully cloned *Lmx* and *Lhx3/4* using a combined embryonic cDNA library.

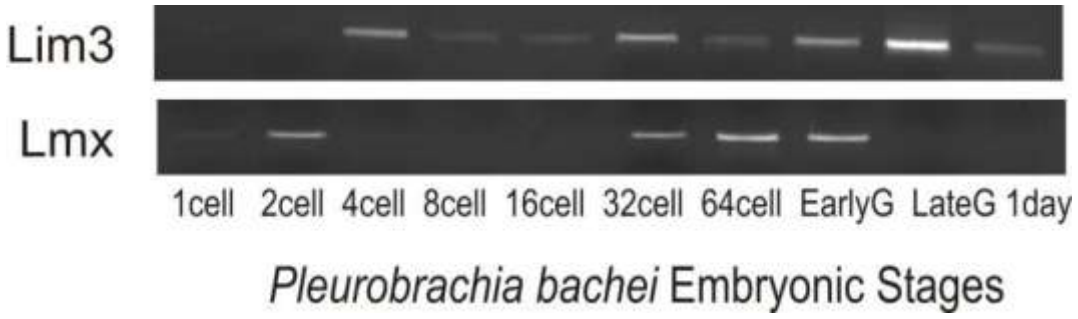


Figure 6. Temporal expression of *Lim3* and *Lmx* differs throughout development.

```
>lcl|60613 MlLmx
Length=428

 Score =  388 bits (996),  Expect = 2e-136, Method: Compositional matrix adjust.
 Identities = 215/379 (57%), Positives = 248/379 (65%), Gaps = 87/379 (23%)

Pb  Query  1    MVESGIVHSTVKEEPVTVS----------EATSIPSSTTTQP----------------K  33
            MVE+ IVHSIVKEE   V+          E  S P    ++P                    
Ml  Sbjct  1    MVETSIVHSIVKEETDIVTRPLDHVSRPLEHISRPLEQVSRPLVCKSSDLVMCNVPNNTS  60

Query  34   KSPICGACRLPIEETLLMKVNDVYWHEQCLRCCLCNVQLSESCFFDKQQRLYCKRDHYQR  93
            KSP+C ACRLPIEE+LLMKVNDV+WHEQCLRCCLCN+ LS+SCFFDK QRLYCK DHYQR
Sbjct  61   KSPVCIACRLPIEESLLMKVNDVFWHEQCLRCCLCNVPLSDSCFFDKHQRLYCKADHYQR  120

Query  94   VLDSQCC----------------------------------------DQLMHGNGDLVCKI  116
            +LD+ CC                                        D+Y MHGNGDLVCK 
Sbjct  121  ILDASCCICNGSINANDLVYRVHGSLYHIRCFTCIRCSSSLRKGDFYMHGNGDLVCKT  180

Query  117  CSDDTLREQYFLQNRLCRRDRREVKRPRTVLSSVQRK----VFKEAFDMIPRPCRKERE  172
            CSDD LR+ ++LG +P CRRDRREVKRPRTVLSSVQRK    VFKEAFD IPRPCRKERE
Sbjct  181  CSDDGLRDHFYLGVDPSCRRDRREVKRPRTVLSSVQRKEHLSVFKEAFDRIPRPCRKERE  240

Query  173  KLSSQTGLSVRVVQVWFQNQRAKVKKLARRKMNKQAEGGGATVQGAHSDSDMDDRCLMDK  232
            KLSSQTGLSVRVVQVWFQNQRAKVKKLARRK+NKQA          +SDSD+D+RCL DK
Sbjct  241  KLSSQTGLSVRVVQVWFQNQRAKVKKLARRKLNKQA----------VNSDSDLDERCLHDK  291

Query  233  TGEAKVTSSVEGYPDSVSPVSCALGDQNYTQLASPSLKKHFGDFATSHCSYADL--QVGQ  290
            TGE+KVTSSVEGYP++ SP SC+L DQ YT L+SP   K F +     C Y D+  Q+ +
Sbjct  292  TGESKVTSSVEGYPEATSPASCSL-DQTYTNLSSPGGMKPFQN-----CPYVDIHSQISR  345

Query  291  PDL--GGHTLAPHPPPDFL  307
            P++       L PH P + +
Sbjct  346  PEIPQPDSYLLPHSPDNLM  364
```

Figure 7. *Lmx* gene comparison between *Mnemiopsis leidyi* (Ml) and *Pleurobrachia bachei* (Pb) showing a large deletion of a LIM domain starting at amino acid 103.

Domain organization

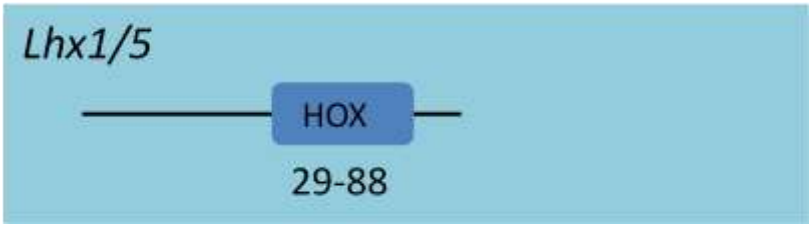


Figure 8. *P. bachei Lhx1/5* domain organization. Not full length. Homeobox domain from amino acid 30 to 84. Domain organization from Pfam[12].

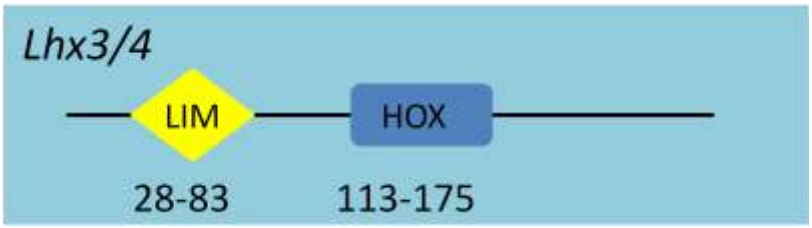


Figure 9. *P.bachei Lmx3/4* domain organization. Not full length. LIM domain from amino acid 29 to 88. Domain organization from Pfam[12].

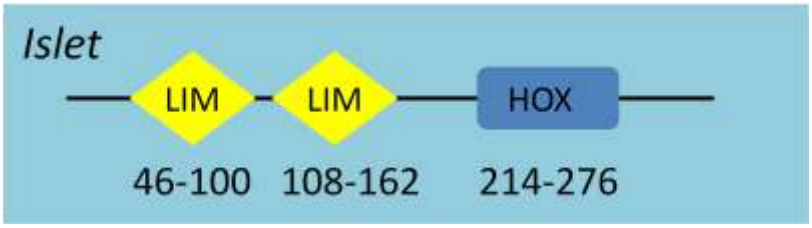


Figure 10. *P. bachei Islet* domain organization. Full length. LIM domain from amino acid 21 to 79 and 83 to 141. Homeobox domain from 189 to 245. Domain organization from Pfam[12].

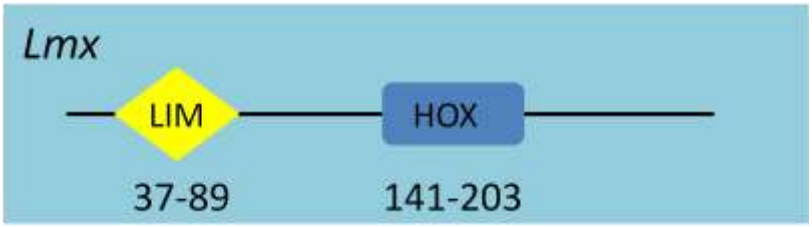


Figure 11. *P. bachei Lmx* domain organization. Full length. LIM domain from amino acid 38 to 94. Homeobox domain from amino acid 142 to 198. Domain organization from Pfam[12].

***In-situ* Hybridization Pictures**

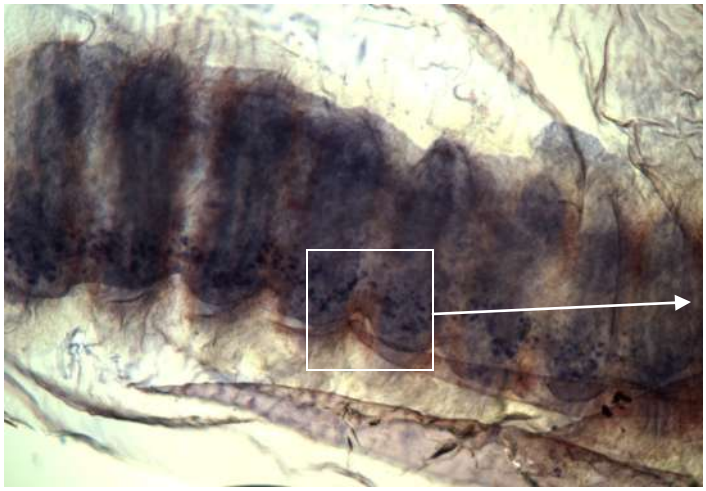


Figure 12. *In situ* hybridization using *Lhx3/4* probe in adult *Pleurbrachia bachei*. Cell specific is expression localized to one side of comb row. These are likely to be egg cells. Magnification 4x.

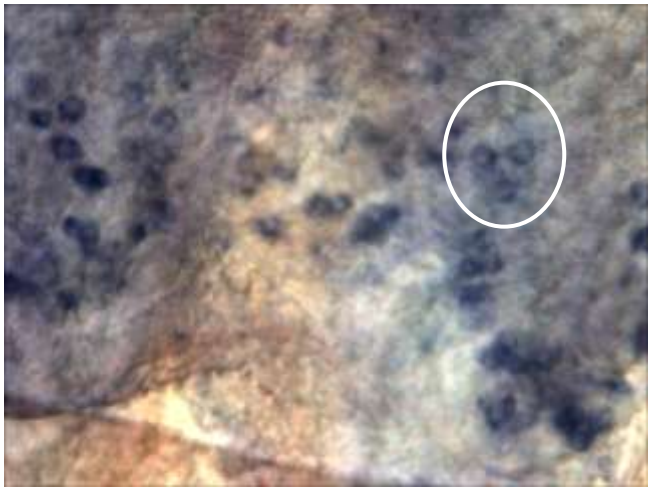


Figure 13. *Lhx3/4* localization. Close up of specific cells localization from figure 7. You can see from the cells under the white circle that the mRNA is cytoplasmic. For larger picture, see supplement 3. Magnification 40x.

Figure 14. *Lmx* localization in a *Pleurobrachia bachei* tentacle. Magnification 4x.

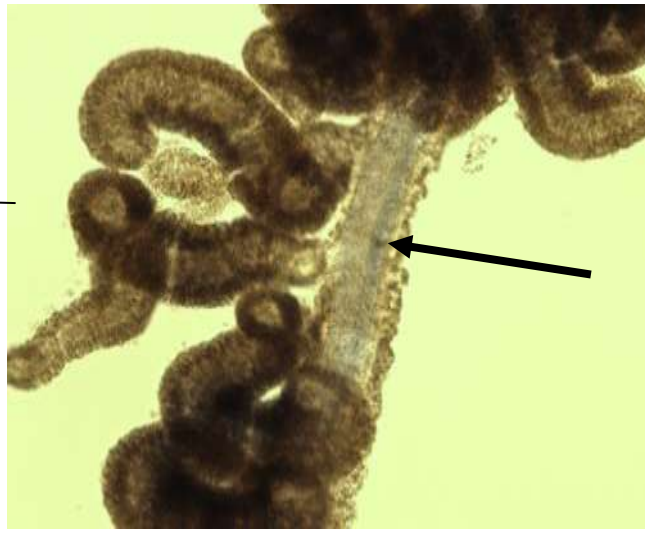


Figure 15. *Lhx* localization. Close up from figure 9. Arrow is pointing to expression a particular structure down the tentacle. For larger picture, see supplement 3. Magnification 40x.

Lhx3

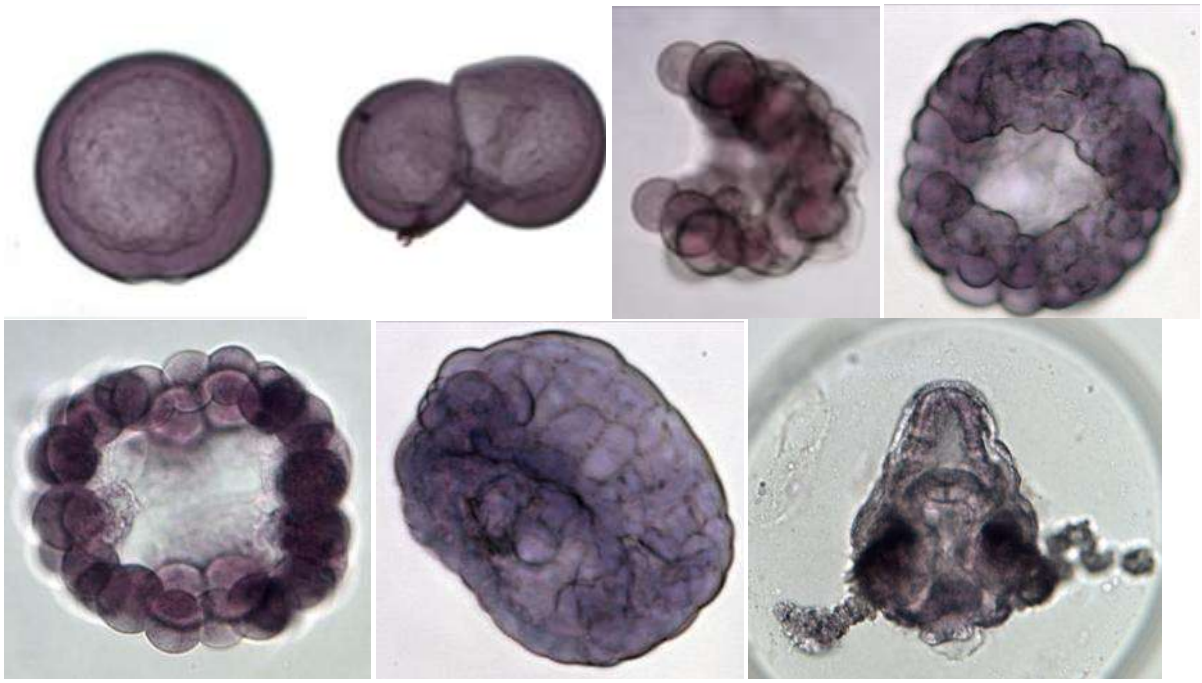


Figure 16. Embryo expression shows localization of *P. bachei Lhx3* in the micromeres. A 1.5 day cyddipid larvae show localization in the tentacle bulbs.

Lmx

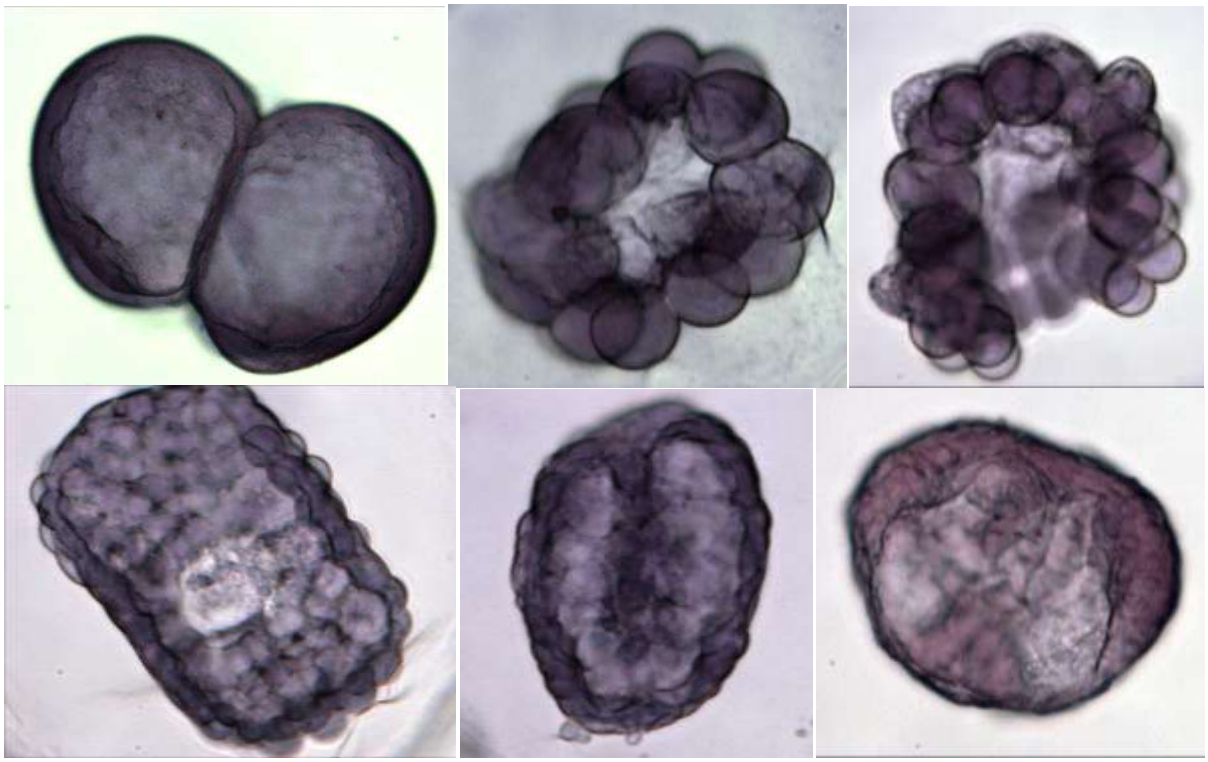


Figure 17. Embryo expression shows localization of *P. bachei Lmx* in the micromeres.

Discussion

The presence of four *Lhx* genes in *Pleurobrachia* indicates that basal metazoans may share the same pattern of early neuronal development as the rest of the metazoa. The lack of the subfamilies *Lhx2/5* and *Lhx6/8* indicate that ctenophores branched off earlier than cnidarians and placazoans, who do have these subfamilies in their genomes. Because the poriferan *Amphimedon* also possesses the same four *Lhx* families we cannot speculate as to the phylogenic relationship between the sponges and ctenophores, other than to say that they are closely related.

In situ expression

The cell specific staining of one side of the comb rows with *Lhx3/4* indicates that this gene is present in the egg cells and might be a maternal factor. The *Lmx* staining of the structure in the tentacle might indicate that *Lmx* is active in the neural innervations of the tentacle. In embryos, the specific staining in the micromeres suggests these *Lhx* genes are playing a role in development.

Lmx deletion

The *Lmx* gene in *Pleurobrachia* contains only one *LIM* domain and a homeobox domain, as opposed to most *Lhx* genes which contain two *LIM* domains followed by a homeobox domain[9]. Comparison of a closely related species gene, *Mnemiopsis*, made it clear that there was a deletion in *Pleurobrachia's* genome which caused a deletion of one of the *LIM* domains in this gene. This gene still seems to be functional as shown by the expression seen in the tentacle.

Phylogenetic Analysis

The gene tree that was produced from the sequences we acquired lacks good bootstrap support at many nodes and does not lend to drawing many phylogenic conclusions. All of the *Lhx* families cluster together and the *Pleurobrachia* sequences fit in there nicely as well.

Pleurobrachia also clusters closely with *Mnemiopsis* and the rest of the basal metazoans. The *Pleurobrachia* and *Mnemiopsis Lmx* is shown as a sister group to the rest of the taxa, with good support, which might indicate that ctenophores are some of the most basal organisms.

Conclusion

Out of the six subfamilies of *Lhx* genes present in metazoans, four present in the ctenophore *Pleurobrachia bachei*. The presence of these genes suggest that ctenophores share the same developmental neural networks as the rest of the metazoa. Additionally, along with sponges, ctenophores are the only taxon that does not have representatives from the other two subfamilies. This suggests that sponges and ctenophores are the earliest branches on the metazoan tree.

Future Directions

A full length sequence for *Lhx1/5* needs to be found and cloned in embryonic cDNA libraries with subsequent adult and embryo *in-situ* hybridizations. A full length sequence needs to be found for *Lhx3/4*. We need to try the *Islet* primer with later stage embryonic cDNA libraries. Our embryonic cDNA library was concentrated more on early stages, before 24 hours following fertilization. Once we have cloned *Islet*, we need to do adult and embryo *in situ* hybridizations with this gene. Repetitions of the embryo *in situ* hybridizations in *Lhx3/4* and *Lmx* needs to be completed for all genes.

Acknowledgments

I want to thank Dr. Leonid Moroz for convincing me to join him at Friday Harbor Labs and for helping me shape the progress of this project. I want to thank Dr. Andrea Kohn for teaching me so much amazing molecular biology in such a short time and helping me push my

project forward. I want to thank Dr. Billie Swalla for her wonderful knowledge of embryos which contributed greatly to the success of this project. I also want to thank Mat Citarella for his crucial bioinformatic work which made my whole project possible. Great thanks go to Gabrielle Winters who helped me organize the experiments and embryo collections. I also want to thank Emily Dabe and Caleb Bostwick for hours of help in the lab, some very early in the morning. Without everyone's amazingly hard work this project would have never have happened so I thank everyone very much!

References

1 Simmons, D. K., Pang, K. & Martindale, M. Q. Lim homeobox genes in the Ctenophore Mnemiopsis leidyi: the evolution of neural cell type specification. *EvoDevo* **3**, 2, doi:10.1186/2041-9139-3-2 (2012).

2 Hobert, O. & Westphal, H. Functions of LIM-homeobox genes. *Trends in genetics : TIG* **16**, 75-83 (2000).

3 Way, J. C. & Chalfie, M. The mec-3 gene of Caenorhabditis elegans requires its own product for maintained expression and is expressed in three neuronal cell types. *Genes & development* **3**, 1823-1833 (1989).

4 Freyd, G., Kim, S. K. & Horvitz, H. R. Novel cysteine-rich motif and homeodomain in the product of the Caenorhabditis elegans cell lineage gene lin-11. *Nature* **344**, 876-879, doi:10.1038/344876a0 (1990).

5 Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. & Edlund, T. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* **344**, 879-882, doi:10.1038/344879a0 (1990).

6 Kadrmas, J. L. & Beckerle, M. C. The LIM domain: from the cytoskeleton to the nucleus. *Nature reviews. Molecular cell biology* **5**, 920-931, doi:10.1038/nrm1499 (2004).

7 Shirasaki, R. & Pfaff, S. L. Transcriptional codes and the control of neuronal identity. *Annual review of neuroscience* **25**, 251-281, doi:10.1146/annurev.neuro.25.112701.142916 (2002).

8 Koch, B. J., Ryan, J. F. & Baxevanis, A. D. The diversification of the LIM superclass at the base of the metazoa increased subcellular complexity and promoted multicellular specialization. *PloS one* **7**, e33261, doi:10.1371/journal.pone.0033261 (2012).

9 Srivastava, M. *et al.* Early evolution of the LIM homeobox gene family. *BMC biology* **8**, 4, doi:10.1186/1741-7007-8-4 (2010).

10 Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731-2739, doi:10.1093/molbev/msr121 (2011).

11 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

12 Punta, M. *et al.* The Pfam protein families database. *Nucleic acids research* **40**, D290-301, doi:10.1093/nar/gkr1065 (2012).

13 Derelle, R. & Manuel, M. Ancient connection between NKL genes and the mesoderm? Insights from Tlx expression in a ctenophore. *Development genes and evolution* **217**, 253-261, doi:10.1007/s00427-007-0131-x (2007).