THE UNIVERSITY OF CHICAGO

THE MOLECULAR EMBRYOLOGY OF A CEPHALOPOD MOLLUSC, *OCTOPUS BIMACULOIDES*

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ORGANISMAL BIOLOGY AND ANATOMY

BY

CAROLINE BETH ALBERTIN

CHICAGO, ILLINOIS

AUGUST 2016

Copyright © 2016 Caroline B. Albertin

All rights reserved

TABLE OF CONTENTS

LIST OF FIGURES……...… iv

LIST OF TABLES……....……. ix

ACKNOWLEDGEMENTS……..……..…. x

ABSTRACT……...…... xii

CHAPTER 1. Introduction …….. 1

CHAPTER 2. Materials and methods …….…….. 22

CHAPTER 3. Developmental staging of the California two-spot octopus, *Octopus bimaculoides*…….. 36

CHAPTER 4. Cephalopod developmental control genes in the octopus genome……...…. 48

CHAPTER 5. Development and molecular patterning of the octopus brain……...… 105

CHAPTER 6. Gene expression in embryonic *Octopus bimaculoides* reveals conserved and novel mechanisms for establishing the body plan……. 142

CHAPTER 7. Conclusions and future directions……...……. 172

REFERENCES……...……. 185

LIST OF FIGURES

Chapter 1

1.1. Adult octopus anatomy ……...……. 2

1.2. Simplified phylogeny of cephalopods…….…. 3

1.3. Phylogenetic relationships of major molluscan groups ……...…… 5

1.4. Schematics of spiralian, octopus and zebrafish embryos at the 8-cell stage ……...……. 7

1.5. Bilaterian phylogeny ……...…….…… 9

1.6. Cephalopod axes……...…….. 12

Chapter 3

3.1. Lightfield photomicrographs of *O. bimaculoides* embryos

at early cleavage stages (one, four, and sixteen cells)……... 38

3.2. DAPI stained nuclei highlight the arrangement of cells in

early *O. bimaculoides* embryos…….. 39

3.3. *O. bimaculoides* embryos photomicrographed in lateral views

to demonstrate the progression of epiboly over stages 3-9…… 40

3.4. Nuclear staining at epibolic and organogenic stages

of *O. bimaculoides* development……... 41

3.5. SEM images of stages 3-10 *O. bimaculoides* embryos.……...…..…. 42

3.6. SEM of organ primordia at late stage 6……...……... 43

3.7. Later stages of octopus development…….. 45

Chapter 4

4.1. Octopus genome size relative to other bilaterians……. 51

4.2. Phylogenetic tree of eumetazoan hedgehogs……..……. 54

4.3. Metazoan FGF phylogenetic tree…… 57

4.4. Phylogenetic tree showing distribution of the 13 eumetazoan Wnts.……. 59

4.5. Phylogenetic tree showing distribution of eumetazoan TGFβ superfamily proteins.……. 61

4.6. Phylogenetic tree of the TBOX gene family in eumetazoans. ……... 63

4.7. Forkhead box transcription factors in metazoans. ……..…… 65

4.8. Phylogeny of metazoan HMG-box proteins.……..…… 67

4.9. Nuclear Hormone Receptors in selected bilaterians……... 70

4.10. Basic Helix-Loop-Helix transcription factors in metazoans……. 72

4.11. Eumetazoan homeodomain transcription factor phylogeny..……...… 75

4.12. Bilaterian Hox clusters.… ……..….. 77

4.13. Expression of *O. bimaculoides Hox* genes detected in development,

and in adult skin, suckers and neural tissues.……...……. 78

4.14. C2H2 genes in select eumetazoan species.……..……. 79

4.15. Eumetazoan C2H2 phylogenetic tree……...……. 80

4.16. Genomic organization of the largest octopus C2H2 cluster. ……..…. 81

4.17. Distribution of C2H2 Domains in octopus ZNFs. ……...… 83

4.18. Expression of 1,429 C2H2-containing genes in 12 octopus transcriptomes…… 85

4.19. Structure of a typical protocadherin.. ……... 86

4.20. Distribution of cadherins and protocadherins across metazoans. …… 88

4.21. Protocadherin expansion in octopus. ……...…… 89

4.22. Metazoan protocadherin expansions. ……..…. 90

4.23. Multiple sequence alignment of EC4 and EC5 of 14 protocadherins

located on Scaffold 30672. ……...…… 93

4.24. Maximum likelihood (ML) trees of EC1 and EC2 domains

of metazoan protocadherins……..…. 94

4.25. ML trees of EC3 and EC4 domains of metazoan protocadherins. ……...… 95

4.26. ML trees of EC5 and EC6 domains of metazoan protocadherins. ……...… 96

4.27. Tissue-specific expression of octopus-specific genes. ……..… 98

Chapter 5

5.1.1. Relationships of extant molluscan clades and schematic of their adult nervous systems 107

5.1.2. Competing hypotheses of cephalopod brain development. ……..... 108

5.1.3. Phylogenetic analysis of *ELAV* and *ELAV-like* genes..…… 110

5.1.4. *ELAV* expression in *O. bimaculoides* hatchling brain..…… 111

5.1.5. *ELAV* gene expression in octopus embryos demonstrates a

cordal development to the nervous system..……...……. 112

5.1.6. Expression of neuronal differentiation genes at stage 8. ……..… 115

5.1.7. Apical-basal organization of octopus neurectoderm. …… 115

5.2.1. The octopus body plan in stage 8 embryos……..... 119

5.2.2. Expression of neural determination genes……...…….…. 120

5.2.3. Transcription factor expression in the clefts at organogenic stages. …… 121

5.2.4. Expression of octopus *WNT* genes at stage 8. ……...… 123

5.2.5. FGF signal in organogenic stages of *O. bimaculoides* development. ……...…… 124

5.2.6. Expression of *HH* and *DPP* during octopus development……...…….. 125

5.2.7. Schematic of expression of the anteroposterior patterning cassette

in mice, hemichordates, and octopus..……..…… 126

5.3.1. Distribution of nervous system types across the Eumetazoa..…...…… 130

5.3.2. Expression of neurosecretory markers in octopus embryos..…… 133

5.3.3. Expression of neurosecretory markers in octopus hatchling brains……...…. 133

5.3.4. Expression of markers of higher motor centers in the

central brain of stage 16 octopus embryos. ……..… 135

5.3.5. Markers of higher motor centers in octopus hatchling brain. ……...… 136

5.3.6. Expression of higher sensory associative markers in stage 16 octopus embryos…….. 137

5.3.7. The octopus hatchling frontal-vertical system

does not express *EMX*, *PAX6,* or *DACH*. ……...… 137

Chapter 6

6.1.1. Heat map of Hox gene expression across five developmental transcriptomes…….. 145

6.1.2. Heatmaps of expression of members of developmental transcription factor families... 147

6.1.3. Expression of the nuclear hormone receptors during octopus development……...….. 148

6.1.4. Expression of the C2H2 transcription factors during octopus development……...….. 149

6.1.5. Developmental expression of WNT genes in octopus……... 151

6.1.6. Expression of TGFβ and FGF genes across five stages of octopus development……. 152

6.1.7. Heatmaps of expression of stage specific genes…… 153

6.1.8. Proportion of stage-specific genes with and without matches

in animal and lophotrochozoan protein databases..……..… 154

6.2.1. *LABIAL* expression pattern in *O. bimaculoides*…..…… 163

6.2.2. *SCR* expression during octopus development…… 163

6.2.3. Expression of *LOX5* in *O. bimaculoides* embryos……. 164

6.2.4. *ANTP* is detected before and during organogenic stages in *O. bimaculoides*…..……. 165

6.2.5. *LOX4* expression in octopus embryos……...…. 165

6.2.6. *LOX2* expression in octopus embryos.……...… 166

6.2.7. Expression of *POST2* in octopus embryos……..… 167

6.2.8. Expression of *POST1* in octopus development…….. 167

6.2.9. Schematic of Hox expression in *O. bimaculoides* arm buds……. 169

6.2.10. Schematic of Hox gene expression and arrangement

in the genome in *O. bimaculoides*. ……...… 170

LIST OF TABLES

Chapter 2

2.1. *O. bimaculoides* transcriptome sequencing summary……..……..… 25

2.2. Summary of developmental transcriptome sequencing results……..……..… 25

2.3. *O. bimaculoides* transcriptome assembly summary for genome analysis……..…… 27

2.4. Non-cephalopod mollusc transcriptomes used to

identify candidate cephalopod-specific genes……... 29

2.5. PCR primers used for octopus cDNA isolation …….……..... 31

Chapter 4

4.1. Selected metazoan developmental control genes……..…….. 53

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Cliff Ragsdale, for his advice and support throughout my doctoral work. Thank you for providing a generous and open environment to study these fascinating, if difficult, animals. I am grateful for your help, insight, and guidance.

To the members of the Ragsdale lab, past and present, thank you. My thanks especially to Rahul Parnaik, who first brought octopuses to the Ragsdale lab and did the hard work of establishing a new system, and to Shuichi Shigeno, who has been a great resource and collaborator. I am, of course, grateful to my wonderful labmates, particularly Yan Wang and Judit Pungor, who are all the things, Stavroula Assimacopoulos, who has been a source of laboratory techniques and friendship, Joanna Rowell, Jennifer Dugas-Ford, Forrest Gulden, and Steve Briscoe. I was also fortunate to have Abby Point, Shuqi Kang, Holly Gui, and Magda Glotzer help me with everything from feeding many small beaks to cloning out genes.

My thanks go out to my fantastic collaborators, especially Dan Rokhsar and Oleg Simakov, with whom I enjoyed working on the genome project. I am grateful to the Marine Biological Laboratory in Woods Hole, MA, particularly Roger Hanlon, Steve Senft, Louis Kerr, and Dave Remsen for giving me the opportunity to see cephalopods in a whole new light, and for the UChicago Graduate Student Research Award that made it possible. I would also like to thank Dr. Chuck Winkler, who has been my source of octopuses over the years.

I am grateful to the members of my committee: Ed Munro, Chip Ferguson, and Robert Ho. Thank you all for your suggestions, support, and encouragement. I feel very fortunate to have had access to the incredible resources at the University of Chicago, especially the genomics and microscopy core facilities. I am grateful to Vytas Bindokas and Christine Labno for their

help with imaging, and to Bill Buikema, Chris Hall, and Pieter Faber for sequencing. My thanks also to Titus Brown, who gave me a spot in his crash-course on bioinformatics, which changed the direction of my work.

I am indebted to my undergraduate advisor Stan Rachootin, whose enthusiasm for all creatures weird and wonderful is infectious, and who instilled in me a passion for basic research. I would also like to thank Dirk Hentschel, who trained me in molecular biology, embryology, and animal husbandry, Yoshiyuki Yamamoto and Una Ren for their advice, and Patrick Callaerts, who gave me my first experience working with molluscs.

I feel incredibly fortunate to have shared this experience with dear friends in the Darwinian cluster. In particular, Laura Merwin, Alice MacQueen, Kristen Jenkins, Tom Stewart, Aaron Olsen, Justin Lemberg, Colin Kyle, Chris Schell, Ben Rubin, Ben Winger, Ben Krinsky, Kacy Gordon, Qiyan Mao, Kate Criswell, Haley Stinnett, Darcy Ross, grad school would not have been the same, or as sane, without you.

I would also like to thank my parents, JoAnn and Tom Albertin, for encouraging me to keep tadpoles, go tidepooling, and to always be curious about the world.

And thank you to Justin for always being there, for the late-night runs in to lab to feed the octopuses, for your enthusiasm, encouragement, and, well, for everything.

ABSTRACT

Cephalopods have a highly derived body plan and a suite of innovations with no obvious correlates in other animals. One of the most striking novelties in cephalopods is their embryogenesis, which lacks any trace of the spiral cleavage pattern found in non-cephalopod molluscs and other spiralians. Instead, cephalopod embryos undergo bilateral, meroblastic cleavage on top of a large yolk, greatly resembling early embryogenesis in fish and constituting a striking convergence between these two distantly related groups. In this thesis, I explore cephalopod development from a molecular perspective, leveraging genomics, transcriptomics, and gene expression studies to shed light on the development of these remarkable animals.

To identify the gene networks important in this highly derived developmental program, we analyzed the genome and transcriptomes of the California two-spot octopus, *Octopus bimaculoides*. The core developmental gene repertoire of the octopus is broadly similar to that found in typical invertebrate bilaterians, except for massive expansions in the protocadherins and the C2H2 zinc finger transcription factor gene families. I identified hundreds of cephalopod-specific genes, many of which showed elevated expression levels in such specialized cephalopod structures as the skin, the suckers, and the nervous system. Comparative genome structure analysis suggests that substantial expansion of a handful of gene families, along with extensive remodeling of genome linkages, drove the evolution of cephalopod morphological novelties.

A prominent cephalopod innovation is their large, complex nervous system. The morphological structure of the cephalopod embryonic brain was studied with molecular markers of neuronal development. Expression of pan-neuronal genes indicated that early neurogenic territories in octopus are arranged as concentric cords rather than pairs of ganglia, an arrangement hypothesized to be characteristic of the ancestral molluscan nervous system. The

expression of a highly conserved developmental transcription factor cassette that is characteristic of the vertebrate midbrain-hindbrain boundary identified the major division in the cephalopod brain, that between the supraesophageal and subesophageal masses. Similar to the vertebrate midbrain-hindbrain boundary, this territory is a signaling center, although the signaling ligands detected in octopus are greatly expanded from those described in other animals. Gene expression study later in development indicated a shared transcription factor "fingerprint" between cephalopods and other animals of neurosecretory and higher motor centers. In contrast, the cephalopod frontal-vertical system, which is a higher integrative center implicated in learning, memory and decision-making, proved to have a different molecular signature from that of the analogous structures in vertebrates and annelids.

Finally, I examined the deployment of highly conserved developmental toolkit genes using both bioinformatics and gene expression. Analysis of Hox gene expression indicated that, despite the absence of a Hox genomic cluster, these genes are expressed in a canonical pattern of anterior-posterior nested domains, one modified to reflect the radial morphology of the cephalopod embryo. Bioinformatic analyses detected the Hox genes and other highly conserved developmental toolkit genes primarily during stages that coincide with the emergence of the body plan, but not before. Notably, many of the genes differentially expressed in the early transcriptomes are taxonomically restricted. Our results support a mid-developmental period of highly conserved toolkit gene expression preceded by the deployment of taxonomically restricted genes. These results suggest the surprising conclusion that it is cephalopod-specific genes that underlie the "fish-like" embryogenesis of cephalopods.

CHAPTER 1

Introduction

Coleoid cephalopods – octopuses, cuttlefish, and squid - are exceptional molluscs. They have dexterous arms lined with hundreds of suckers, which function as specialized tactile and chemosensory organs. An elaborate adaptive coloration system, under direct neural control, allows them to rapidly change the color and texture of their skin (Hanlon and Messenger, 1996). These are visual animals, with large, camera eyes and a lens that can focus, which, in some species, has visual acuity comparable to that of amniotes (Sweeney et al., 2007), and which serve as a classic example of convergent evolution. Further examples of convergence between cephalopods and vertebrates, particularly fish, include the presence of a closed circulatory system, a blood-brain barrier, giant axons, and a lateral line system (Packard, 1972).

Coleoid cephalopods have elaborate brains, the largest of the invertebrates, rivaling those of vertebrates. The central nervous system of the common octopus, *Octopus vulgaris*, for example, contains half a billion neurons, which is six times the number in a mouse brain (Herculano-Houzel et al., 2006; Young, 1971). This central nervous system generates sophisticated behaviors, including observational learning and complex problem solving (Grasso and Basil, 2009). The cephalopod brain, while comparable to that of vertebrates in both size and complexity, is distinctly invertebrate in its arrangement, with a circumesophageal central brain, and lobes arranged with a rind of neuronal cell bodies surrounding a central neuropil rather than the vertebrate arrangement of a supraesophageal brain with neuronal cell bodies embedded within the neuropil.

While cephalopods exhibit a number of convergent features with vertebrates, they are also distinctly molluscan. Like other molluscs, they possess a mantle, a radula, and a foot, but

these and other features are greatly modified (Figure 1.1). The molluscan foot evolved into dexterous, sucker-lined arms. Free of the external shell, the muscular mantle enables the jet-propelled lifestyle of coleoid cephalopods. Perhaps most striking are the innovations that set cephalopods apart from other animals: among them, the ink sac, the chemosensory suckers lining the arms, and the adaptive coloration system that permits extraordinary displays of camouflage. The exceptional combination of convergence and innovation in the cephalopod clade represents a deeply attractive opportunity to study the evolution of novelty.

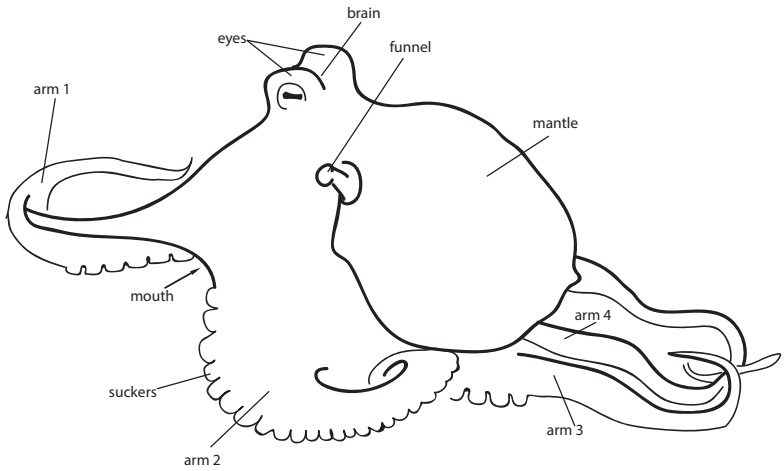


Figure 1.1. Adult octopus anatomy.
The mouth is located at the center of the arm crown. The central brain is located between the eyes, with the esophagus passing between the supraesophageal and subesophageal masses.

Cephalopods are an ancient clade comprising approximately 700 extant species organized into two groups: the nautiloids and the coleoids (Kroger et al., 2011; Nixon and Young, 2003). The nautiloids have an external, chambered shell, pinhole eyes, and two pairs of gills. They also have dozens of digital tentacles that lack suckers, which are quite different from the arms characteristic of coleoids.

The great diversity of extant cephalopods is found in the soft-bodied coleoids (Figure 1.2), which consist of the Decapodiformes (squids and cuttlefish) and the Octopodiformes

(*Vampyroteuthis* and octopods). These groups have eight (Octopodiformes) or ten (Decapodiformes) sucker-lined appendages and one pair of gills, and have internalized or lost the shell. The adaptive coloration system is unique to coleoids. The decapodiforms include the largest and the smallest cephalopods: the giant and colossal squids, *Architeuthis* and *Mesonychoteuthis,* and the pygmy squid, *Idiosepius*, as well as spirulids, sepiids (cuttlefish), sepoilids, and teuthids (squid). The octopodiforms include the vampire squid *Vampyroteuthis*, and cirrate and incirrate octopuses. Cirrate octopuses have fins on their mantle and cirri, or muscular papillae, lining the arm, both of which are absent in incirrate octopuses. Current molecular diversification estimates suggest that nautiloids and coleoids split roughly 415 million years ago (mya), while the divergence between decapodiforms and octopodiforms was around 276 mya (Kroger et al., 2011). While these clades appear phylogenetically robust, the relationships within them are not well resolved (Lindgren et al., 2012; Strugnell et al., 2005).

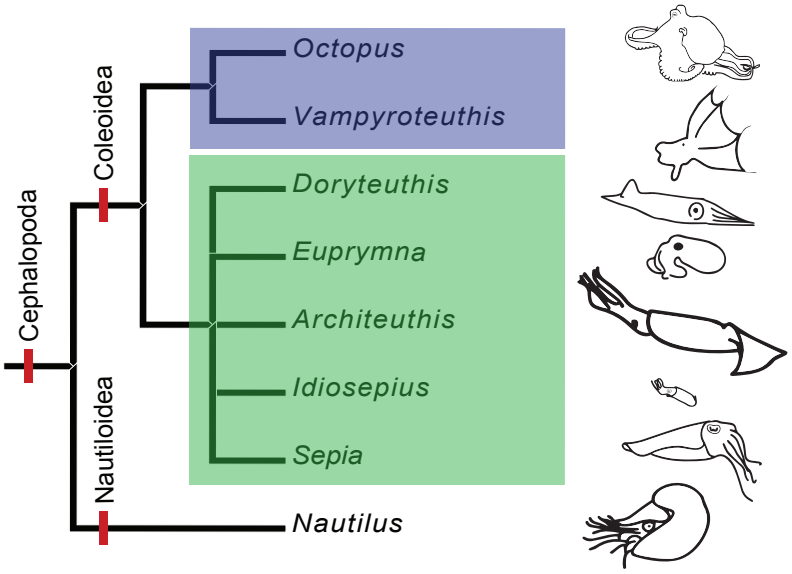


Figure 1.2. Simplified phylogeny of cephalopods.
Coleoid cephalopods are found in two major groups: the Decapodiformes (green) and the Octopodiformes (blue). Representative genera are at the tips, along with a cartoon indicating overall morphology. *Doryteuthis* is a teuthid squid, *Euprymna* is a sepiolid, *Sepia* is a sepiid, and *Octopus* is an incirrate octopus genus. Phylogeny based on Lindgren et al. (2012).

The Mollusca

Cephalopods are molluscs, the second largest clade in the Metazoa with more than 100,000 described species (Brusca and Brusca, 2003). There are seven major molluscan clades: the aplacophorans, polyplacophorans, monoplacophorans, cephalopods, scaphopods, bivalves and gastropods. The molluscs are united by a number of shared features, including a dorsal mantle that secretes a calcareous shell or spicules, a muscular ventral foot, and a rasping radula used in feeding. However, across the seven molluscan clades, there is a great diversity of body plans, and these shared features are modified in each of the groups. For example, aplacophorans are worm-like, and lack a well-developed head, foot, or a solid shell. Chitons, or polyplacophorans, have eight articulating shell plates. Monoplacophorans have a single shell and have serially repeated structures along the body axis, including eight pairs of pedal retractor muscles. Bivalves are laterally compressed, uncephalized, and have two shells but lack a radula. Gastropods, like monoplacophorans, have a creeping foot and a single shell, which may be greatly modified or lost in a number of clades. Gastropods also exhibit a greater degree of nervous system centralization than all the other molluscan clades except the cephalopods. The tusk shells, or scaphopods, have a single, tubular, uncoiled shell.

Given the great morphological disparity in this clade, the phylogeny of the major molluscan groups has been contentious, and many different topologies have been proposed (Giribet et al., 2006; Kocot, 2013; Kocot et al., 2011; Ponder and Lindberg, 2008; Salvini-Plawen and Steiner, 1996; Sigwart and Lindberg, 2015; Smith et al., 2011; Stoger et al., 2013). One of the prominent hypotheses has considered aplacophorans as diverging basally to the rest of the molluscs, with polyplacophorans branching outside of the conchiferans (monoplacophorans, bivalves, scaphopods, and cephalopods) in a clade called the Testaria. Monoplacophorans have

been considered basally diverging conchiferans, with cephalopods as a sister group to the gastropods in the crown (Haszprunar, 2000; Salvini-Plawen and Steiner, 1996). Alternatively, support has also been found for a sister relationship between polyplacophorans and monoplacophorans as the Serialia (Giribet et al., 2006). Still other studies support a clade of scaphopods, aplacophorans, and cephalopods (Stoger et al., 2013). Another topology recovers a clade combining aplacophorans and polyplacophorans as the Aculifera, which is a sister group to the conchiferans (Figure 1.3). The Aculifera hypothesis has received support from recent phylogenomic studies, which also find cephalopods branching basally or in a clade with the monoplacophorans within the conchiferans (Kocot et al., 2011; Smith et al., 2011). A clade of cephalopods and monoplacophorans is also supported by paleontological data (Yochelson et al., 1973).

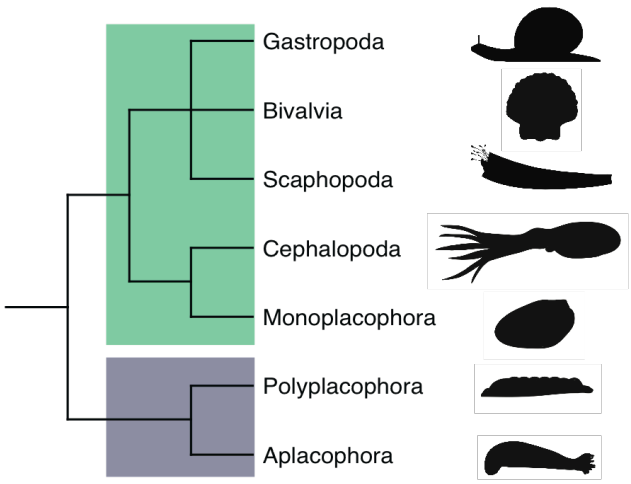


Figure 1.3. Phylogenetic relationships of major molluscan groups.
Current molecular phylogenies recover two major molluscan groups: the aculiferans (Aplacophora+Polyplacophora, in blue), and the conchiferans (in green). Phylogeny is based on Smith et al. (2011). Cartoons for the gastropod and scaphopod were obtained from phylopic.com.

This new arrangement raises many questions, particularly with regard to the ancestral state of a number of molluscan characters (Wanninger and Wollesen, 2015). For example, there

is a great diversity of nervous system organization in adult molluscs. The nervous systems of scaphopods, gastropods, and bivalves are ganglionic: they have discrete clusters of neuronal cell bodies that are connected by cell-body free commissures and connectives. Chitons, on the other hand, have a nervous system organized as pairs of cords, which have neuronal cell bodies throughout. Whether the ancestral state of the molluscan nervous system is ganglionic or cord-based has been a matter of controversy (Bullock, 1965; Wanninger and Wollesen, 2015; Young, 1971).

Spiral cleavage

While adult morphology is vastly different between groups, one characteristic feature of molluscs that is accepted as ancestral is its early developmental program of spiral cleavage. Spiral cleavage is a remarkably stereotyped pattern of cell divisions, which permits the identification of individual cells based on their lineage across taxa (Boyer and Henry, 1998; Freeman and Lundelius, 1992; Henry, 2014). The first two divisions yield the four cells of the A, B, C, and D-quadrants that give rise to structures on the left, ventral, right, and dorsal sides of the larval or adult organism, respectively (Dictus and Damen, 1997). These divisions may be equal, generating four equivalently sized cells, or may be asymmetrical, typically generating a D blastomere that is larger than the others. At the third division (Figure 1.4A), four animal micromeres (1a, 1b, 1c, and 1d) bud off from the vegetal macromeres (1A, 1B, 1C and 1D). Very often, the micromeres are smaller than the macromeres, though this is not always the case. This division takes place along the animal-vegetal axis, but at an angle. The fourth division is also along the animal-vegetal axis, but the angle of the division is in the opposite direction, so divisions alternate between clockwise and counterclockwise, generating a spiral cap of micromeres when viewed from the animal pole.

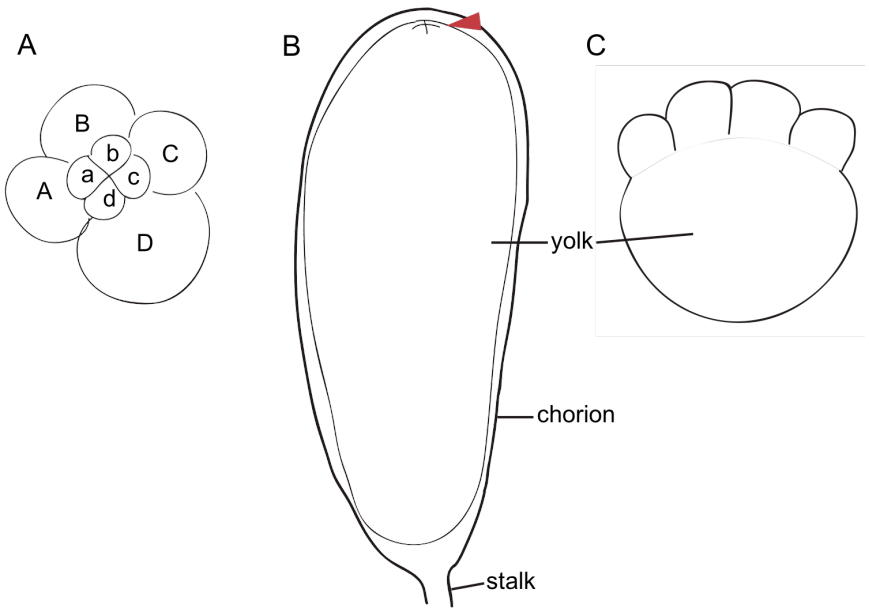


Figure 1.4. Schematics of spiralian, octopus and zebrafish embryos at the 8-cell stage. (A) Animal pole view of a spiralian 8-cell stage embryo with quadrants of micromeres (a, b, c, d) and macromeres (A, B, C, D) indicated. (B-C) Lateral view of 8-cell octopus (B) and zebrafish (C) embryos. Animal pole is oriented to the top. Red arrowhead in (B) indicates the position of the dividing cells in octopus.

The fates of these micromeres and macromeres have been traced in different species. Lineage tracing in a number of molluscs has shown that these divisions occur in a stereotyped manner such that the ultimate fates of specific micromeres and macromeres are largely conserved, suggesting that the spiralian developmental program is homologous across spirally cleaving animals (Boyer et al., 1998; Boyer and Henry, 1998; Henry, 2014). The first quartet of micromeres (generated at the third division) gives rise to anterior ectodermal structures, including the apical organ, the larval eyes, and brain (Lyons et al., 2015; Render, 1997; Wanninger and Wollesen, 2015). The second and third quartets (from divisions 4 and 5) give rise to ectoderm (and neurectoderm) as well as a mesectoderm (a putative spiralian innovation

(Henry and Martindale, 1999; Lyons and Henry, 2014)), and the fourth quartet generates endoderm and mesendoderm.

The D-quadrant has been shown not only to form dorsal structures but also to generate mesendoderm and to act as an organizer. Across spirally cleaving animals the 4d cell gives rise to the majority of mesodermal and endodermal structures, including gut, kidney and muscles, and is known as the mesentoblast. A series of ablation and inhibition studies have shown that either the 3D macromere, the 4d micromere, or both, not only makes significant contributions to larval structures, but also plays a role in patterning the early embryo (Lambert and Nagy, 2001; Sweet, 1998). However, the molecular nature of the patterning signals from the D quadrant organizer is not understood.

Gastrulation takes place after these cleavage stages. There are a variety of gastrulation movements in molluscs, including invagination, epiboly, or a combination of these (Lyons and Henry, 2014). Following gastrulation, molluscs form a trochophore larva, which has a characteristic circumferential band of cilia called a prototroch, which propels the planktonic larva through the water column. They also have an apical sensory organ, which often has a tuft of cilia. This sequence of spiral cleavage followed by a trochophore larval stage is considered plesiomorphic in molluscs (Wanninger and Wollesen, 2015).

The Spiralia and the Mollusca: a common early developmental program generating a diversity of body plans

Spiral cleavage is an ancient and highly conserved mode of development thought to be ancestral to the Mollusca as it is also characteristic of a number of other clades. Indeed, this developmental mode may in fact be ancestral to one of the three main clades of bilaterally symmetrical animals: the Spiralia (=Lophotrochozoa).

The Spiralia clade contains roughly half of extant animal phyla, including Annelida, Brachiopoda, Bryozoa, Cycliophora, Entoprocta, Gastrotricha, Gnathostomulida, Micrognathozoa, Mollusca, Nemertea, Phoronida, Platyhelminthes, and Rotifera (Giribet, 2008) (Figure 1.5). Nemerteans are long, unsegmented worms, while annelids are a large, diverse group of overtly segmented worms that include earthworms, leeches, and polychaetes. Brachiopods, pharonids, and bryozoans have a crown of ciliated tentacles called a lophophore that are used for feeding. The molluscs are the most diverse of these groups, both in terms of number of species as well as morphological disparity among its members (Brusca and Brusca, 2003).

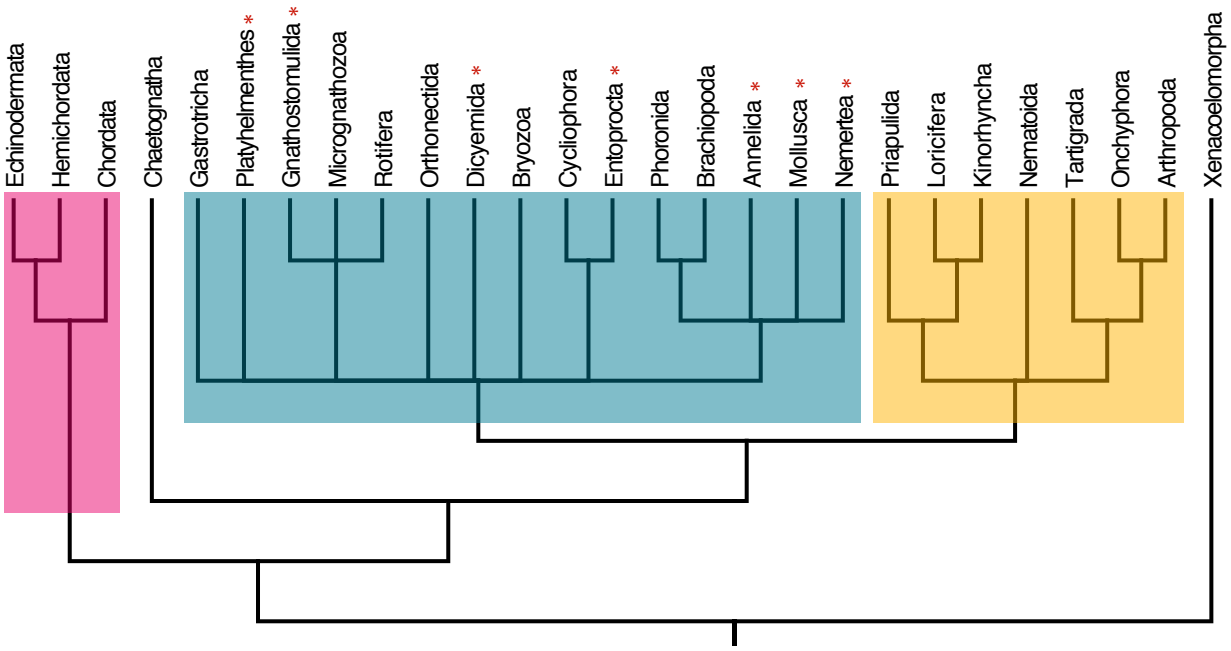


Figure 1.5. Bilaterian phylogeny.
Current understanding of bilaterian relationships, based on Dunn et al. (2014). The deuterostome clade is indicated in red. The spiralians (teal) and ecdysozoans (mustard yellow), along with the chaetognaths, are protostome groups. Groups known to exhibit spiral cleavage are indicated with a red asterisk.

Many of the relationships among the Spiralia have not been fully resolved (Dunn et al., 2014; Giribet, 2008; Laumer et al., 2015). Indeed, even the name of this clade of animals

remains in flux, alternately called the Spiralia or the Lophotrochozoa. The designation Lophotrochozoa represents the conjunction of lophophore-bearing animals with those that possess a trochophore larval stage. Trochophore larvae are characteristic of molluscs and annelids, but has also been described in nemerteans (Maslakova and Hiebert, 2014) and cycliophorans (Funch and Mobjerg Kristensen, 1995). Annelids, molluscs, nemerteans, entoprocts, gnathostomulids, platyhelminth worms, dycemids, and possibly phoronids, undergo spiral cleavage (Henry, 2014). Given the wider taxonomic group encompassed by the presence of spiral cleavage (see Figure 1.5), as well as precedence, many authors prefer the term Spiralia (Dunn et al., 2008; Giribet, 2008; Hejnol, 2010). Here, I will use Spiralia and Lophotrochozoa interchangeably to refer to this clade of animals.

The early developmental program of cephalopods is an evolutionary novelty

While spiral cleavage is a characteristic feature of Spiralia, it is not universal: there is no morphological resemblance between spiral cleavage and early development in cephalopods (Naef, 1928). Early cleavage in cephalopods is meroblastic and bilateral while in non-cephalopod molluscs, cleavage is holoblastic and spiral (compare Figure 1.4A and B). Following fertilization, ooplasm is segregated at the animal pole, forming a blastodisc on top of the yolk (Arnold, 1968; Crawford, 2001). The first cleavage marks the plane of symmetry, dividing the blastodisc into left and right halves. Subsequent cleavages are symmetrical across the plane of the first cleavage furrow. All of the cells are continuous with the yolk until the fourth cleavage, when the four central blastomeres separate completely from the yolk. These central blastomeres are surrounded by twelve "blastocones" (Arnold, 1971), which are sometimes referred to as micro- and macromeres, respectively (Naef, 1928). The early cleavage

pattern is largely similar across all cephalopods, but is clearly distinct from spirally cleaving molluscs (Arnold, 1971; Boyer and Henry, 1998; Naef, 1928; Shigeno et al., 2008).

Blastomeres and blastocones subsequently divide in a loose, somewhat irregular arrangement, and eventually form a blastomere cap at the animal pole. As cell divisions continue, the blastomere cap increases in size, covering the yolk by epiboly. As epiboly is progressing, organ rudiments become apparent as mesendodermal thickenings (Stage 6 of Naef (1928)) with the eye, arm and funnel rudiments arrayed around the centrally located mantle (Boletzky, 2003; Naef, 1928).

Unlike other molluscs, cephalopod embryos develop directly, without ever passing through a larval stage. The primordia of the arms, eyes, mouth, and mantle become apparent before epiboly is complete (Naef, 1928). These primordia appear suddenly and simultaneously, such that the body plan appears all at once, with the mouth at the anterior, the mantle at the center, and the arms at the periphery. As development progresses, the primordia constrict and telescope out, ultimately generating the adult arrangement of the brachial (arm) crown surrounding a beaked mouth, leading to the radula and the esophagus that passes through the middle of the large central brain. The mantle contains the three hearts, gills, kidneys, gonads, and other visceral organs. When swimming, the side of the animal with the mouth and arms is considered to be anterior, while the tip of the mantle is considered posterior, with the top of the animal designated dorsal, and the surface facing the substrate as ventral (Figure 1.6). This "physiological" orientation is orthogonal to the embryonic axes. The embryonic orientation makes sense evolutionarily: in non-cephalopod molluscs the mantle, which secretes the shell, is a dorsal structure, and the foot is ventral. The external shell of *Nautilus* and the internal shells of decapodiform cephalopods develop from the mantle primordium. As this manuscript concerns

octopus development, I use the embryological orientation throughout, with the exception of Chapter 5.3.

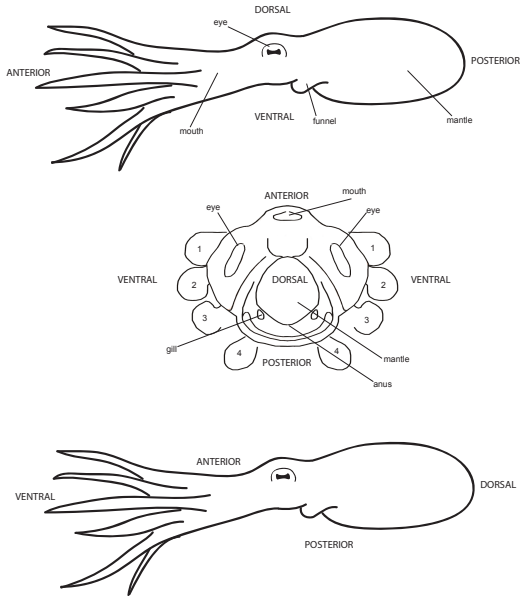


Figure 1.6. Cephalopod axes.
Top panel illustrates the adult octopus morphology while swimming. The beak, surrounded by the arm crown, is considered anterior, while the mantle is posterior, and the side facing the surface of the water is dorsal. The middle panel illustrates a dorsal view of a stage 10 embryo, with arms arrayed at the periphery (marked 1-4) and the mantle at the center, with the mouth at the top and the posterior to the bottom. The bottom panel shows the adult morphology with the "embryonic" axes indicated.

In many respects, the early developmental program of cephalopods more closely resembles that of teleost fish than other spiralians (Figure 1.4). Indeed, both undergo meroblastic cleavage on top of a large yolk, the primary movement of gastrulation is epiboly, and development is direct. This striking convergence in early developmental program between cephalopods and teleosts has been little recognized, but is no less remarkable than the convergence seen between vertebrate and cephalopod eyes.

It is possible that the overt differences between cephalopods and spirally cleaving molluscs may be in part due to physical constraints, such as the meroblastic cleavage of the blastoderm in the context of the large yolk that is found in cephalopods (Buresi et al., 2016; Galis and Sinervo, 2003). However, given the relative conservation of the spiral cleavage program, it may be that loss of spiral cleavage has resulted in profound alteration of developmental constraints, which in turn may be linked to the evolution of novelty in the cephalopod body plan. It is still possible that cephalopods retain elements of the molecular signature, and possibly a cryptic cellular signature, of spiral cleavage. Detailed molecular and embryological studies are required to elucidate early developmental mechanisms in cephalopods, which could then be compared with that of other molluscs and other animals.

A cephalopod model for molecular embryology

While there are many important issues in evolutionary and developmental biology that can be addressed in cephalopods, there are significant barriers to work in this clade. Firstly, cephalopods are exclusively marine and are often quite sensitive to water quality. They present a number of other husbandry difficulties, including their size and lifespan. In particular, many species, although direct developing, undergo a "paralarval" stage, in which the hatchlings are planktonic. These planktonic paralarvae have proven very difficult to maintain in culture (Vidal et al., 2014). Other species have large eggs that produce benthic hatchlings that exhibit adult behaviors on hatching. Some of the latter species can be maintained, some over several generations, in a laboratory (Hanlon and Forsythe, 1985b; Lee et al., 2009). Nevertheless, wild caught embryos are frequently used for developmental studies in a variety of species, including *Sepia officinalis*, *Idiosepius,* and *Octopus bimaculoides* (Buresi et al., 2016; Buresi et al., 2012; Navet et al., 2009; Shigeno et al., 2015; Tarazona et al., 2016; Yoshida et al., 2011). Finally, a

genetically tractable species has not been established, and manipulation of gene expression has not yet been successful in cephalopods.

One barrier to work in cephalopods has been the relative absence of sequence resources. As recently as 2012, there was no sequenced cephalopod genome (Albertin et al., 2012), and only a handful of expressed sequence tags (ESTs) and transcriptomes were available (Bassaglia et al., 2012; Brown et al., 2014). We set out to address this deficiency by sequencing the genome and multiple transcriptomes of the California two-spot octopus, *Octopus bimaculoides*. *O. bimaculoides* (Pickford and McConnaughey, 1949) is a medium-sized octopus native to southern California and is readily available throughout the year. This species has large eggs (~1cm long), which produces benthic hatchlings that can readily be maintained in aquaria and cultured through the lifecycle (Hanlon and Forsythe, 1985a). In addition, this species has a comparable brain organization to *Octopus vulgaris* (Shigeno and Ragsdale, 2015), which has been the subject of detailed study (Wells, 1978; Young, 1971).

The topic of this thesis is the molecular embryology of the octopus. To this end, I present the complement of highly conserved developmental control genes in the genome of this species. I then describe the development and patterning of a key cephalopod innovation: the highly centralized nervous system. Finally, I explore the deployment of the developmental toolkit and the role of taxonomically-restricted genes in octopus embryogenesis.

Of molecules and morphology: the developmental toolkit

A great deal of work has been done in elucidating developmental mechanisms in animals. The bulk of this effort has been concentrated in genetic model organisms, including mice (*Mus musculus),* zebrafish (*Danio rerio*)*,* sea urchins (*Strongylocentrotus purpuratus*)*,* flies *(Drosophila melanogaster*)*,* and nematode worms (*Caenorhabditis elegans*), which are part of

two of the bilaterian superclades, the Deuterostomia and the Ecdysozoa (Figure 1.5). Studies in these systems have found a remarkable similarity in genes that play key roles in developmental processes, especially those involved in patterning the nervous system and body plan (Carroll et al., 2005). Broadly speaking, these genes are often either transcription factors, which regulate the expression of other genes during development, or members of signaling pathways, which mediate cell-cell interactions and spatial pattern formation.

Developmental toolkit: transcription factors

Comparative molecular developmental biology was arguably born with the discovery that the HOX family of transcription factors first identified in *D. melanogaster*, are widely distributed across the Metazoa (McGinnis et al., 1984). This insight prompted a search for other developmental transcription factors shared across bilaterian clades, and returned a number of families not only present in the genomes of flies, nematodes, and mice, but also in cnidarians and even sponges (Putnam et al., 2007; Srivastava et al., 2010). This shared suite of transcription factor families, including the homeodomain, basic helix-loop-helix (bHLH), nuclear hormone receptor (NHR), TBOX, and forkhead box (FOX) proteins, are deployed in similar ways in the development of distantly related bilaterians. For example, Hox genes are expressed in a conserved order along the anteroposterior axis in many animals (Carroll et al., 2005), and the homeobox genes *OTX* and *EMX* are deployed at the anterior end of flies, mice, and hemichordates (Pani et al., 2012; Sen et al., 2013). The NK-class homeobox protein *tinman* is found in the hearts of mice and flies (Harvey, 1996), and *PAX6* expression can drive eye development when ectopically expressed in flies and vertebrates (Chow et al., 1999; Halder et al., 1995). These similarities have led to the suggestion that the conserved expression of these orthologous genes represents a shared positional patterning system in bilaterians (Slack et al.,

1993), and have further led to claims of homology of a complex nervous system through the last common ancestor of bilaterians (Arendt et al., 2001; Denes et al., 2007; Holland et al., 2013; Tomer et al., 2010).

Developmental toolkit: Ligands

Cell-cell interactions mediated by signaling pathways are crucial in embryonic development. Extracellular ligands bind to cellular receptors, which can activate both cellular and nuclear targets, often ultimately activating target genes through the regulation of transcription factors. There are only a handful of signaling pathways that are implicated in most of animal development: WNT, hedgehog (Hh), transforming growth factor β (TGF-β), nuclear hormone receptor (NHR), receptor tyrosine kinase (RTK), Notch and JAK/STAT (Pires-daSilva and Sommer, 2003). The ligands, receptors and downstream mediators of these pathways are conserved at least across the Bilateria, and several groups, including the WNTs, TGFβs, Hhs, and the NHRs have deeper metazoan roots (Koch et al., 2012; Pang et al., 2010; Ryan et al., 2013; Schnitzler et al., 2014; Sebe-Pedros and de Mendoza, 2015).

These signaling pathways are not only conserved in the genome and expressed during development, but some are deployed in strikingly similar ways in the development of distantly related animals. For example, opposing gradients of *dpp/BMP4* and its binding partner *sog/CHORDIN* are found along the dorsal-ventral axis in flies and vertebrates (Arendt and Nubler-Jung, 1994; De Robertis and Sasai, 1996; Holley et al., 1995).

The octopus developmental toolkit

In Chapter 4, I detail our work on the genome and transcriptomes of *O. bimaculoides*, with a particular focus on the core developmental gene repertoire. I found that the inventory of octopus developmental toolkit genes, including transcription factors and ligands, is broadly

similar to that of other invertebrate bilaterians, with two exceptions: the C2H2 zinc finger transcription factors and the protocadherin subfamily of cadherins. These two families exhibited massive expansions in the *O. bimaculoides* genome relative to other animals. In addition, I identified hundreds of cephalopod-specific genes, many of which are expressed in novel cephalopod structures and the nervous system.

The evolution of the cephalopod brain

Cephalopod molluscs have the largest brains among all invertebrates, rivaling in size those of many vertebrates, including elasmobranch fishes and some mammals and birds. The central brain of the adult common octopus, for example, contains more than 40 million neurons and comprises dozens of lobes, many dedicated to so-called higher functions including coordinated movements and learning and memory. The development of the cephalopod brain has, however, been little studied; in particular, it is unknown whether the gene regulatory network thought to underlie patterning in arthropod, annelid, and vertebrate brains are deployed in the development of the cephalopod nervous system. In Chapter 5, we employ gene expression experiments to investigate three key questions in cephalopod molecular neuroembryology.

The arrangement of the cephalopod brain, and its relationship to the nervous systems found in other molluscs, has been the subject of controversy. The greatly elaborated coleoid nervous system does not directly resemble either the ganglia of bivalves, gastropods or scaphopods, or the cords characteristic of polyplacophorans or monoplacophorans. The cephalopod neuroanatomist J. Z. Young described the brain of *Nautilus* as a series of cords surrounding the esophagus (Young, 1971), while the invertebrate anatomist T. H. Bullock (Bullock, 1965) suggested that the secondary spread of neurons from embryonic ganglia into commissures could account for any coleoid cord-like appearance. In the first part of Chapter 5,

we explore whether this highly derived nervous system is initially generated as a series of cords or if it is developed from spatially-restricted ganglia. We studied the expression of neuronal markers in *O. bimaculoides* embryos and found both neural differentiation genes and a neuronal marker distributed in a series of cords rather than as pairs of ganglia, supporting an ancestral cord-based arrangement of the molluscan nervous system.

In the second part of chapter 5, we explore the early molecular patterning of the octopus brain by studying the expression of developmental transcription factors and ligands. A cassette of transcription factor toolkit genes, including Otx, Pax2/5/8, Engrailed, Gbx, and Hox genes, is implicated in anterior-posterior brain patterning in diverse animals (Pani et al., 2012; Reichert, 2005; Urbach, 2007). We found that this cassette is expressed during octopus embryogenesis. The expression of a component of this cassette, Pax2/5/8, Engrailed, and Gbx, coincides with a morphological depression in the ectoderm (the "cleft"). This cleft is situated between the future supraesophageal and subesophageal masses, partitioning the octopus brain into two major divisions. In vertebrates and hemichordates, this molecular cassette defines the isthmus organizer, which expresses *WNT1* and *FGF8*. The cleft in octopus also expresses a suite of signaling ligands, including several not characteristic of the isthmus organizer of vertebrates.

The evolutionary origin of bilaterian brain structures has been an area of great interest and debate. Studies leveraging molecular developmental data have found that elements of the complex brains found in extant vertebrates, flies, and annelids share a common ‘molecular fingerprint’, which has been interpreted as evidence of homology (Denes et al., 2007; Strausfeld and Hirth, 2013; Tomer et al., 2010). In the final section of Chapter 5, we explore the role of these transcription factors and cell type markers in the octopus. We demonstrate that the learning and memory centers in octopus have a different molecular signature from the analogous systems

in vertebrates and annelids, whereas centers with neurosecretory functions exhibit conservation of transcription factor expression. In addition, the basal lobes share the transcription factor profile, but not the cell type, of higher motor centers found in vertebrates. These findings indicate that the higher, integrative centers associated with complex behavior may be divergent and display clade-specific innovations in cephalopod molluscs.

Molecular embryology of the cephalopod body plan

One of the most striking cephalopod novelties is their embryogenesis, which lacks any trace of the spiral cleavage program found in non-cephalopod molluscs and other lophotrochozoans. Instead, cephalopods undergo bilateral, meroblastic cleavage on top of a large yolk. Embryos develop directly, with primordia appearing simultaneously as epiboly progresses. The primordia are arrayed in a circle, with mantle at the center and the arms at the periphery. How this highly derived body plan and its novel developmental program relate to those of other animals has been obscure.

In Chapter 6, I address these questions in two parts. First, I characterize the developmental transcriptome of *O. bimaculoides* using both a candidate gene and a bioinformatic approach. Through my developmental transcriptome analysis, I detected highly conserved developmental toolkit genes primarily during the middle stages that coincide with the emergence of the body plan, but not before. Notably, many of the differentially expressed genes in the early transcriptomes are taxonomically restricted. These results indicate that the highly derived embryogenesis found in octopus, with periods of divergent gene expression at earlier and later stages flanking a mid-developmental period of highly conserved gene expression, is reminiscent of the hourglass model of development (Duboule, 1994; Raff, 1996), and suggests a role for

taxonomically restricted genes in the evolution and development of cephalopod novelty at the early embryonic stages.

In the second half of Chapter 6, I report on the expression of the Hox genes, a family of highly conserved transcription factors implicated in the evolution and development of anteroposterior patterning in a wide range of animals. Our bioinformatic analysis indicated that the expression of octopus Hox genes is coincident with the emergence of the body plan. With *in situ* hybridization, I found that octopus *HOX* genes are expressed in canonical, nested domains along the anteroposterior axis, in contrast to the only other report of *HOX* expression in a cephalopod (Lee et al., 2003). Early expression of the *HOX* genes also reflects the derived arrangement of the cephalopod embryo: the expression domains radiate out from the center of the embryo in nested, bilaterally symmetrical sectors as if in a polar coordinate system, rather than the Cartesian arrangement found in embryos from other species. These data suggest that octopus HOX genes have retained their ancestral role in patterning the anteroposterior axis despite our findings on the octopus genome, which showed atomization of the Hox cluster.

In my final chapter, I consider how the work described in this dissertation advances our understanding of the evolution and development of cephalopods and present future avenues of research, as well as some of the major impediments to work in these fascinating animals.

On contributions from collaborators

The sequencing and annotation of the octopus genome that formed the basis of Chapter 4 was done in collaboration with Dr. Daniel Rokhsar and Dr. Therese Mitros at the University of California at Berkeley, Dr. Oleg Simakov at the Okinawa Institute of Science and Technology, and Yan Wang, Dr. Judit Pungor and Dr. Clifton Ragsdale at the University of Chicago, and was

published in Albertin et al. (2015). The phylogenetic analyses presented in Chapter 4 were done by me, except for the C2H2 transcription factor and the protocadherin annotations, which I did together with Ms. Wang.

The molecular neurodevelopmental studies conducted in Chapter 5 were done in collaboration with Dr. Shuichi Shigeno, Dr. Rahul Parnaik, and Dr. Ragsdale. Sections 5.1 and 5.3 were designed by Dr. Shigeno and Dr. Ragsdale. Chapter 5.2 was conceived by Dr. Parnaik and Dr. Ragsdale. All but one of the figures (Figure 5.2.2D) are my own work.

CHAPTER 2

Materials and methods

Octopus embryo collection

Adult *O. bimaculoides* were obtained from Aquatic Research Consultants (Dr. Charles Winkler, San Pedro, CA). Brooding adult females were shipped with their clutches, and embryos and hatchlings were maintained in artificial seawater (ASW; 32ppt, Tropic Marin®, Aquarientechnik, Wartenberg, Germany) in closed 260 gallon aquaria at 18-22°C. Hatchlings were fed adult brine shrimp (livebrineshrimp.com) and appropriately sized pieces of clam or shrimp obtained from the local supermarket. Juvenile and adult animals were maintained in individual 10-50 gallon aquaria in ASW and were fed fiddler crabs (livemarinebaitfish.com).

Live imaging of embryos stained with SybrGreen I

Embryos were incubated in a solution of 1:20,000 SybrGreen I (Invitrogen) in FASW on a rocker at room temperature. Embryos were imaged on a Leica MZ FLIII fluorescent dissecting microscope or on a Zeiss Lightsheet Z.1. Time-lapse single plane illumination microscopy images were processed in ImageJ.

Scanning electron microscopy

Embryos were dissected out of their chorions and carefully transferred into either Karnovsky's or Hollande's fixative and fixed overnight at room temperature. Fixed embryos were then dehydrated through a series of ethanol washes, before replacing the solution with successive dimethoxypropane (Sigma-Aldrich) washes. Embryos were then transferred to a watch glass under the hood and washed several times in hexamethyldisilazane (Sigma). The hexamethyldisilazane was allowed to evaporate overnight. Dried embryos were mounted on

stubs, sputter coated with platinum (10+5nm), and imaged on a JEOL 840 scanning electron microscope.

Genome size estimate using fluorescence

To estimate the size of the octopus genome experimentally, embryonic cells from *O. bimaculoides* (whole embryos, stage 20) and *Danio rerio* (whole embryos, 24 hours post-fertilization) were harvested, stained with DAPI and separated by flow cytometry on a BD FACSCanto analyzer. Cell isolation, staining and sorting were run in parallel. *D. rerio* was included for calibration.

RNA isolation and transcriptome sequencing

Tissue isolation for transcriptomes for genome annotation

Transcriptomes from twelve tissues were generated to aid with gene prediction and for expression analysis. Tissues were chosen to cover a wide range of transcripts, and to identify genes expressed in structures that feature cephalopod innovations. RNA was isolated from ova, testes, viscera, posterior salivary gland (PSG), suckers, skin, developmental stage 15 (St15; (Naef et al., 2000), retina, optic lobe (OL), supraesophageal brain (Supra), subesophageal brain (Sub), and axial nerve cord (ANC). Adult *O. bimaculoides* were deeply anesthetized in 3% ethanol with 12.5mM $MgCl_2$ and quickly decapitated before tissues were harvested. The ova sample was isolated from a single, mature egg. The testes sample was isolated from a single mature male gonad. The viscera sample combines equal amounts of RNA isolated from the hepatopancreas, kidney, and central heart. Skin was isolated from the eyespots, dorsal and ventral mantle, and the leucophore-rich area between the eyes of one-month-old hatchlings. Suckers were dissected from the distal third of a single adult arm. Embryos were staged

according to Naef (1928), washed in filtered ASW and dissected from the chorion and off the yolk.

Tissue isolation for developmental transcriptomes

Transcriptomes for five stages spanning development were generated. Embryos were staged according to Naef and washed and dissected in filtered ASW. The chorion was carefully pierced and dissected away from the yolk. Embryos were then dissected off the yolk and quickly transferred with as little yolk or filtered ASW possible into a tube containing 20µl TRIzol (Invitrogen) in liquid nitrogen.

RNA isolation

All tissues harvested were quickly dissected and either flash frozen on liquid nitrogen or placed in RNAlater (Ambion) at 4°C overnight (Table 2.1). Samples were stored at -70°C. RNA was isolated using TRIzol following manufacturer's instructions. RNA integrity was analyzed with a Bioanalyzer 2100 (Agilent); only samples with clean rRNA peaks and little to no degradation were used. Total RNA was poly-A selected and directionally sequenced at the University of Chicago Genomics Facility on an Illumina HiSeq2000 per manufacturer's instructions, generating 100bp paired-end reads with an insert size of 300bp.

Table 2.1. *O. bimaculoides* transcriptome sequencing summary.
(*)St15 transcriptome is the same sample used in the developmental transcriptome analysis.

Tissue	Animals	Initial Storage Conditions	Mb	Reads
Ova	Adult #3, female	Liquid Nitrogen	5,626	55,698,688
Testes	Adult #6, male	Liquid Nitrogen	7,422	73,479,938
Hepatopancreas, Kidney, Heart (Viscera)	Adults #2 (male), #3 (female), #5 (female)	RNAlater	7,176	71,057,530
Posterior Salivary Gland (PSG)	Adult #4, male	Liquid Nitrogen	7,195	71,237,812
Skin	7 hatchlings	Liquid Nitrogen	7,925	78,459,268
Suckers	Adult #4, male	Liquid Nitrogen	7,092	70,218,752
Stage 15*	1 clutch, ~15 embryos	Liquid Nitrogen	6,790	67,227,866
Retina	Adult #2, male	Liquid Nitrogen	6,364	63,006,538
Optic Lobe (OL)	Adult #2, male	RNAlater	6,116	60,550,668
Supraesophageal Brain (Supra)	Adult #1, female	Liquid Nitrogen	15,475	154,754,152
Subesophageal Brain (Sub)	Adult #1, female	Liquid Nitrogen	14,103	141,027,874
Axial Nerve Cord (ANC)	Adult #3, female	RNAlater	6,271	62,091,238

Table 2.2. Summary of developmental transcriptome sequencing results.
(*)St15 transcriptome is the same sample used in the genome analysis.

Stage	# embryos	Initial Storage Conditions	Mb	Reads
Stage 1-2	50	Liquid Nitrogen	6,947	68,779,332
Stage 4	10	Liquid Nitrogen	7,532	74,567,818
Stage 7	4	Liquid Nitrogen	6,158	60,961,374
Stage 10	4	Liquid Nitrogen	7,617	75,414,582
Stage 15*	4	Liquid Nitrogen	6,790	67,227,866

De novo transcriptome assembly for genome annotation

Adapters and low-quality reads were removed before assembling transcriptomes using the Trinity *de novo* assembly package (version r2013-02-25; (Grabherr et al., 2011; Haas et al.,

2013). Assembly statistics are summarized in Table 2.3. Following assembly, peptide-coding regions were identified and translated using TransDecoder in the Trinity package. We compared the *de novo* assembled RNA-seq output to the genome to evaluate the completeness of the genome assembly. To minimize the number of spuriously assembled transcripts, only transcripts with ORFs predicted by TransDecoder were mapped onto the genome with BLASTN.

Quantifying gene expression

Following removal of adapters and low quality sequences, RNA-seq reads were mapped to the genome assembly using TopHat 2.0.11 (Trapnell et al., 2009). A range of 76- 90% of reads from different samples could be mapped to the genome. Accepted reads were sorted and indexed with SAMtools (Li and Durbin, 2009). The read counts in each tissue were produced with the bedtools multicov program (Quinlan and Hall, 2010) using the gene model coordinates. The counts were normalized by the total count in each tissue and by the length of the gene. Heatmaps showing expression patterns were generated in R using the heatmap.2 function.

Table 2.3. *O. bimaculoides* transcriptome assembly summary for genome analysis. Statistics are based on all transcript contigs. Total Trinity 'genes' refers to the number of transcript clusters generated by the assembly (isogroups), while Total Trinity Transcripts indicates the number of isoforms (isotigs).

Transcri-ptome	Total Trinity 'genes'	Total Trinity Transcripts	% GC	N10 (bp)	N50 (bp)	Median Contig Length (bp)	Average Contig Length (bp)	Total Assembled Bases (bp)
Ova	57,662	68,363	36	3,455	1,034	377	667.68	45,644,899
Testes	90,425	118,428	37	4,536	1,406	393	780.56	92,440,034
Viscera	95,160	129,052	38	4,701	1,446	415	812.59	104,866,269
PSG	68,598	87,232	37	5,420	1,643	420	865.29	75,481,024
Suckers	102,168	132,422	37	5,947	1,721	388	850.15	112,578,846
Skin	93,008	118,404	37	5,262	1,528	393	808.92	95,779,900
St15	87,932	124,062	37	5,178	1,544	384	799.54	99,192,901
Retina	97,137	124,585	37	5,304	1,508	387	798.77	99,514,962
OL	132,961	167,693	37	5,156	1,356	373	748.79	125,567,386
Supra	177,569	235,295	37	5,465	1,438	369	765.63	180,150,005
Sub	144,721	192,152	37	5,394	1,552	373	793.05	152,385,368
ANC	109,157	142,259	37	5,900	1,612	398	837.28	119,110,298
All	305,458	373,396	36	4,619	1,152	370	702.83	262,434,425

Gene family analysis

Gene families of particular interest, including developmental regulatory genes, neural genes, and gene families that appear to be expanded in *O. bimaculoides*, were manually curated and analyzed. We searched the octopus genome and transcriptome assemblies using BLASTP and TBLASTN with annotated sequences from human, mouse and *D. melanogaster*. Bulk analyses were also performed using Pfam (Finn et al., 2014) and PANTHER (Mi et al., 2013). We used BLASTP and TBLASTX to search for specific gene families in deposited genome and transcriptome databases for *L. gigantea, A. californica, C. gigas, C. teleta, T. castaneum, D. melanogaster, C. elegans, N. vectensis, A. queenslandica, S. kowalevskii, B. floridae, C. intestinalis, D. rerio, M. musculus* and *H. sapiens*. Candidate genes were verified with BLAST

(Altschul et al., 1997) and Pfam (Finn et al., 2014) analysis. Genes identified in the octopus genome were confirmed and extended using the transcriptomes. Multiple gene models that matched the same transcript were combined. The identified sequences from octopus and other bilaterians were aligned with either MUSCLE (Edgar, 2004), Clustal Omega (Sievers et al., 2011), or MacVector 12.6 (MacVector, North Carolina). Phylogenetic trees were constructed with FastTree (Price et al., 2010) using the Jones–Taylor–Thornton model of amino acid evolution, and visualized with FigTree v1.3.1.

Identification of cephalopod-specific genes

Cephalopod novel genes were obtained by BLASTP and TBLASTN searches of the predicted protein set from the octopus genome against the NR database (Pruitt et al., 2005) and a custom database of the mollusc transcriptomes listed in Table 2.3. To ensure that we had as close to full-length sequence as possible, we extended proteins predicted from octopus genomic sequence with our *de novo* assembled transcriptomes, using the longest match to query NR, transcriptome and EST sequences from other animals. Gene sequences with transcriptome support, but without a match to non-cephalopod animals at an e-value cutoff of 1×10^{-3}, were considered for further analysis. Octopus-specific novelties were defined as sequences with transcriptome support but without matches to available sequences from other animals ($<1 \times 10^{-3}$), including nautiloid and decapodiform cephalopods.

Table 2.4. Non-cephalopod mollusc transcriptomes used to identify candidate cephalopod-specific genes.

Species Name	Reference
Antalis entalis	Smith et al. (2011)
Aplysia californica	Moroz et al. (2006)
Cadulus tolmeiei	Smith et al. (2011)
Chaetopleura apiculata	Smith et al. (2011)
Crepidula fornicata	Sadamoto et al. (2012)
Laevipilina hyalina	Smith et al. (2011)
Lingula anatina	Smith et al. (2011)
Littorina littorea	Smith et al. (2011)
Lymnea stagnalis	Henry et al. (2010)
Neomenia megatrapezata	Smith et al. (2011)
"Neomeniomorph"	Smith et al. (2011)
Nucula expansa	Smith et al. (2011)
Perotrochus lucaya	Smith et al. (2011)
Siphonaria pectinata	Smith et al. (2011)
Yoldia limatula	Smith et al. (2011)

Developmental transcriptome differential expression analysis

Reads from the five developmental stages were trimmed with Trimmomatic (Bolger et al., 2014) were assembled as described above. Reads were mapped to the assembled developmental transcriptomes and transcript abundance was calculated with RSEM (Li and Dewey, 2011) bundled with the Trinity toolkit. Differentially expressed transcripts were identified with edgeR (Robinson et al., 2010) with a p-value cutoff of 1xe-5 and a minimum of a 16-fold change in expression values. Transcripts were partitioned into 100 clusters of genes with similar expression patterns. Cluster plots were used to identify clusters of transcripts that warranted detailed study based on their high stage-specific patterns of expression. These expression patterns were visualized in R with heatmaps generated from the TMM-normalized expression matrix produced with RSEM. The transcripts in these clusters were then blasted against several custom databases, as well as the Uniprot database with an e-value cutoff of 1xe-5.

Candidate gene developmental transcriptome analysis

The developmental transcriptome assembly was queried with the highly conserved developmental transcription factors, ligands and other genes of interest identified in Chapter 3. Heatmaps showing expression patterns across developmental stages were generated in R using the heatmap.2 function using data from the TMM-normalized expression matrix.

PCR isolation of *O. bimaculoides* genes

We isolated octopus sequences of interest using degenerate PCR and 5'RACE, as well as standard PCR. Degenerate PCR primers were designed by R. Parnaik as described in Shigeno et al. (2015). Non-degenerate PCR primer pairs were designed using sequences identified in the genome or transcriptome with MacVector software. For degenerate PCR, primers were used at a concentration of 5mM, whereas for 5'RACE and standard PCR, primers were used at a concentration 1.5mM. cDNA was generated with the SMART RACE kit (Clontech) using *O. bimaculoides* embryos between stages 10 and 12. For degenerate PCR, reactions were incubated at 94°C for 2 minutes, followed by 40 cycles of 94°C for 35 seconds, 42°C for 35 seconds, and one minute at 72°C. For 5'RACE, a gene specific reverse primer was used with the universal primer provided with the SMART RACE kit. The 5'RACE reactions were incubated at 94°C for 2 minutes, followed by 35 cycles of 94°C for 1 minute, 50°C for one minute, and 72°C for 2 minutes. Standard PCR reactions were incubated at 94°C for 4 minutes, followed by 35-40 cycles of 94°C for 45 seconds, 52°C for 45 seconds, and 72°C for one minute. After the last cycle, the 72°C incubation was extended by 5-10 minutes.

Table 2.5. PCR primers used for octopus cDNA isolation.
Primers are listed in 5' to 3' orientation.

Octopus cDNA	Primers
ELAV (Degenerate pair 1)	Forward: GTGAACTACCTGCCCCARAMNATGAC Reverse: GCGGCCTCGTCGTARTKNGYCAT
ELAV (Degenerate pair 2)	Forward: GGGCTACGGCTTCGTNAAYTAYRT Reverse: CACGGCGCCGAANGGNCCRAA
NEUROD	Forward: TTGGTGACAGCAATAGCG Reverse: GCGGATGATGACAGAATCTC
PAX1/9 (Degenerate Pair 1)	Forward: CARYTNGGNGGNGTNTTYATHAAYGG Reverse: ACCANACTYGDATNCKNGCYTTCNGT
PAX1/9 (Degenerate Pair 2)	Forward: CARYTNGGNGGNGTNTTYATHAAYGG Reverse: YTCNCKNGTRTADATRTCNGGRTA
PAX6	Forward: AAYCARYTNGGNGGNGTNTTYGTNAAYGG Reverse: GKRTCNCKDATYTCCCANGMRAANAT
NKX2.1 (Degenerate)	Forward: CARGCNCARGTNTAYGARYTNGA Reverse: CCADATYTTNACYTGNGTNGG
NKX2.1 (5'RACE)	Reverse: AAGATGTTCCCGTTCAGGCG
NKX2.2 (Degenerate)	Forward: AARGCNCARACNTAYGARYTNGA Reverse: CCADATYTTNACYTGNGTNGG
NKX2.2 (5'RACE 1)	Reverse: GCGAATGATGCTAGCGAGAT
NKX2.2 (5'RACE 2)	Reverse: AGGTAGCGTTGCCGGCGGAA
SOXE (Degenerate pair 1)	Forward: ATGAAYGCNTTYATGGTNTGG Reverse: GGNGGNGTNGGNCCNTG
SOXE (5'RACE)	Reverse: TGGGGGTATTGGTCAGCAAG
DPP	Forward: TTYGAYGTNCAYCCNGCNGT Reverse: GGRAANGGRCAYTTNCCRTCRCA
MSX	Forward: AARCAYAARMCNAAYMGNAARCC Reverse: TTYTGRAACCADATYTTNAC
MSX (5'RACE)	Reverse: ACTGTGAGGTTGTGAAGGG
DACH	Forward: ATCAACATACAGTTCCCCAC Reverse: TGGTGTCTGGCATTCTCTG
SCR	Forward: ACAACAGTCATAGTCCGAAC Reverse: AGGGGGTCAAGATAGCAAG
LOX5	Forward: GCTTACTGTTTCAATGGACCG Reverse: AAGGATTTGTGGGGAAGGGGTG
LOX4	Forward: CCACACTTACAACAGATGTCTC Reverse: GTTTTCCTCCGAGTCACTG

Table 2.5. ***continued.***

LOX2	Forward: TGGGGTGGAAGAACTTGAGG Reverse: GTCGTGTCAGATAACGGTTG
POST2	Forward: CCATTTTATCCCGCCATCAC Reverse: TTACGCTTGGTCCCTGTTCTCC
POST1	Forward: TTCGGGAAATAACCACACG Reverse: CTCTCACAACTTCTACATCACC
WNT2	Forward: ACGAGCCTGTAGTAAAGGG Reverse: GCATTTCACTTGACAACACC
WNT4	Forward: ACCTTCTAACTTTGCTCGC Reverse: ATCGCTCATTGGAAATCCTC
WNT4/5	Forward: TTCTTACTGCCACGACAGCGAG Reverse: GTCCTCCGTTACAACCTTTG
WNT16	Forward: ACCTGGATTAGTGTGGAGGC Reverse: TGGGAATAGGCAAGTGACG
FGF8	Forward: TTCGTGCTGGTTCAGTCAG Reverse: TGGGTGCTGTATTTGTGC

Riboprobe preparation

PCR products of the expected length were ligated into either pCR2.1-TOPO (Invitrogen) or GEM-T EASY (Promega). Sanger sequencing to identify the cDNAs and their orientation was carried out by the University of Chicago Comprehensive Cancer Center DNA Sequencing Facility. Plasmids were linearized using the appropriate restriction enzyme to generate the antisense template. Antisense digoxigenin (Roche) -labeled riboprobes were transcribed with RNA polymerase (SP6 or T7, New England Biolabs). Following the transcription reaction, the remaining template was removed with RNase-free DNase I (Roche) and riboprobes were precipitated twice and stored in 100% formamide at -20°C until use.

Whole mount *in situ* hybridization

Embryos were staged according to Naef (2000) and were immersion fixed in 4% PFA (paraformaldehyde [Sigma-Aldrich] in phosphate-buffered saline [PBS], pH7.4) overnight at

4°C. For late stage embryos and early hatchlings, animals were deeply anesthetized in 5% ethanol in ASW before fixation. After fixation, embryos were dissected out of their chorions and away from the yolk in PBT (DEPC-treated PBS with 1% Tween 20), and postfixed overnight in 4% PFA.

All washes were performed on a rocker at room temperature, unless otherwise indicated. Embryos were washed twice in PBT for five minutes and incubated in 6% hydrogen peroxide (Sigma Aldrich) in PBT for one hour. Embryos were then washed twice for five minutes in PBT and three times for 30 minutes in Detergent Mix (50mM Tris-HCL [pH 7.5], 1mM EDTA [pH 8.0], 150 mM NaCl, 1% SDS in DEPC-treated water) before being post-fixed in 4% PFA for 20 minutes. Embryos were washed twice in PBT before being stored in embryo hybridization solution (50% formamide, 5x SSC, 1% SDS, 200 µg/ml heparin, 12.5 mg/ml yeast RNA) at -20°C.

Embryos were incubated at 72°C for at least one hour before adding 0.2-0.75 mg antisense riboprobe to 5mL embryo hybridization solution. Probes were allowed to hybridize overnight at 72°C (12-16h). Embryos were washed in 72°C Solution X (50% formamide, 5x SSC, 1% SDS) for 30 minutes four times, and once for one hour. Embryos were then washed three times for five minutes in TBST (Tris buffered saline with 1% Tween 20) and blocked for two hours in 20% DIG buffer (Roche) in TBST. Riboprobes were detected with anti-DIG Fab fragments coupled to alkaline phosphatase (Roche Diagnostics). The anti-DIG antibody was preadsorbed with octopus embryo powder in 5% DIG buffer in TBST for at least one hour. The preadsorbed antibody was diluted to a final concentration of 1:5000 in 5% DIG buffer in TBST and incubated with the embryos on a rocker overnight at 4°C.

Embryos were washed three times for five minutes in TBST, five times for one hour, and overnight. The next day, embryos were washed three times for ten minutes in NTMT (100 mM Tris-HCl [pH 9.5], 100 mM NaCl, 50 mM MgCl2, 1% Tween 20) and incubated in nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP, Denville Scientific) for one to five days at room temperature. Color reactions were stopped with several washes in StopTE (10 mM EDTA, 10 mM Tris pH 7.5). Following multiple TBST washes, the stained wholemounts were stored in 10% formalin in PBS at 4°C. Selected embryos were embedded in 30% gelatin with 10% sucrose in PBS and were incubated overnight at 4°C in 30% sucrose/PFA. 24 micron sagittal sections were cut a sledge microtome and mounted on Superfrost Plus slides (Fisherbrand). Slides were coverslipped with 80% glycerol or were dehydrated through ethanol washes into Histoclear (National Diagnostics) and were coverslipped with Eukitt (Sigma-Aldrich).

Section *in situ* hybridization

Two-week-old hatchling *O. bimaculoides* were deeply anesthetized in 5% ethanol in filtered artificial ASW. Arms, mantle, eyes and skin were dissected away and the central brain was immersion-fixed in cold PFA overnight at 4°C. Tissue was cryogenically protected in PFA with 20% sucrose overnight at 4°C and embedded in 10% gelatin with 20% sucrose in PBS at 60°C. After cooling on ice, embedded brains were incubated in 20% sucrose/PFA overnight and sagittal sections were cut at 24 microns on a sledge microtome. Sections were mounted on Superfrost Plus slides. Tissue was allowed to dry on slides and was stored at -80°C.

Slides were allowed to warm to room temperature and postfixed for 15 minutes in PFA. Slides were washed three times for five minutes in PBS and incubated at 37°C for 30 minutes in proteinase K solution (Roche Diagnostics; 1 μg proteinase K per milliliter of 100 mM Tris-HCl

[pH 8.0], 50 mM EDTA [pH 8.0]). Sections were postfixed for 15 minutes in PFA, washed three times for five minutes in PBS, and stored in hatchling hybridization solution (50% formamide, 5x SSC, 1% SDS, 200 µg/ml heparin, 125 mg/ml yeast RNA, 0.5mg/ml acetylated BSA) at -20°C.

Tissue was allowed to prehybridize for at least one hour at 72°C in hatchling hybridization solution. Slides were then transferred to mailers with 1-2 mg antisense riboprobe in 15mL hatchling hybridization solution, and were incubated overnight at 72°C. Slides were rinsed once quickly in preheated Solution X and then washed three times for 45 minutes at 72°C in Solution X. The sections were washed three times for 15 minutes in room temperature TBST and blocked for two hours in 20% DIG buffer in TBST. Anti-DIG Fab fragments were preadsorbed as described above and added to the tissue at a 1:5000 dilution in 5% DIG buffer in TBST. The incubation was carried out overnight at 4°C on a rocker. Slides were washed three times for five minutes, three times for thirty minutes in TBST and three times for ten minutes in freshly prepared NTMT. Slides were then incubated in NBT/BCIP in NTMT and the color reaction was allowed to proceed for 1 to 5 days at room temperature. Reactions were stopped in StopTE and washed in TBST overnight. Slides were dehydrated through a series of ethanol washes into Histoclear and were coverslipped with Eukitt.

Microscope and image analysis

Images were captured on an Axioskop 40 microscope (Zeiss) or a MZ FLIII fluorescence dissecting microscope (Leica) with a Zeiss AxioCam HRc camera and Zeiss Axiovision 4.8 software. Images were processed with Photoshop CS5 and CS6 (Adobe Systems) and ImageJ (National Institutes of Health).

CHAPTER 3

Developmental staging of the California two-spot octopus, *Octopus bimaculoides*

Background information

The California two-spot octopus (*Octopus bimaculoides*, Pickford and McConnaughey (1949); Cephalopoda, Coleoidea, Octopoda, Incirrata), is a medium sized, hardy octopus species named for its two bright blue eyespots. *O. bimaculoides,* along with its sister species *Octopus bimaculatus*, are members of the *O. vulgaris* species complex found along the southern California coast (Söller et al., 2000). *O. bimaculoides*, with an adult mantle length of up to 18cm, is considerably smaller than *O. vulgaris*, and all life stages (embryos, hatchlings, juveniles and adults) can be readily maintained in closed, artificial seawater systems, which serves to make this species attractive for laboratory study (Hanlon and Forsythe, 1985b). *O. bimaculoides* embryos are large with a clear chorion, and do not pass through the paralarval stage characteristic of many other species, including *O. vulgaris*. Instead, *O. bimaculoides* hatchlings are benthic and readily engage in hunting and other adult behaviors shortly after hatching, making them straightforward to culture.

Timing of *O. vulgaris* and *O. bimaculoides* development

The detailed monograph by Adolf Naef (1928) describing embryogenesis in a range of squids, cuttlefish, and octopods remains the definitive work on cephalopod development. Here, we adapt his staging system for *O. vulgaris* for use in *O. bimaculoides*. Relative to *O. vulgaris,* the embryos of *O. bimaculoides* are very large, and take longer to develop: while *O. vulgaris* embryos hatch in approximately one month at 20°C, *O. bimaculoides* development lasts ~70-90 days, depending on water temperature and oxygenation. It was reported by Naef that *O. vulgaris* embryos developed poorly when removed from the care of the brooding female, as she provides

water flow over the embryos and cleans them. However, we have found that this is not necessary if the embryos are maintained in well-oxygenated water with continuous flow, and the descriptions below are based on observations of embryos under these conditions.

The developmental staging system in *O. vulgaris* described by Naef progresses approximately one stage per day until stage 15, where it increases to roughly three days per stage. This timing differs greatly in *O. bimaculoides.* In particular, the early cleavage stages in *O. bimaculoides* take place over a week, while in *O. vulgaris* they last a single day. At 22°C with strong oxygenation, the epibolic and organogenic stages in *O. bimaculoides* proceed with relatively similar timing to *O. vulgaris*, with each stage lasting roughly one day. Maintaining embryos at 18°C can lengthen these stages several fold, depending on oxygenation. By stage 15, the timing of stages in *O. bimaculoides* lengthens relative to that of *O. vulgaris* and become progressively longer as the stages progress. Stage 16 lasts for several days in *O. bimaculoides* at 22°C, and stage 20 can last a month or more. Given these differences, and particularly the sensitivity of *O. bimaculoides* embryos to temperature and oxygenation, morphological markers are more consistent markers of developmental stages than age. As these stages span long periods of time and often encompass a number of developmental events, we use a (-) and (+) to indicate early and late appearing features for the stage in question.

Cleavage Stages

At roughly one centimeter long and 2 millimeters wide, *O. bimaculoides* eggs are large and yolky. The egg is contained in a translucent, leathery chorion and long stalks extend from the vegetal side of the chorion. The stalks of dozens of eggs are wound together and are fixed onto the walls of the den by the brooding female. The germinal disc is restricted to the rounded animal end of the egg and sits at the opposite end of the egg from the stalk.

Cleavage is meroblastic and bilaterally symmetrical. The first two cleavages divide the germinal disc into roughly equal quadrants (Figure 3.1), and take approximately 24 hours at 22°C. The third cleavage is parallel to the plane of the first cleavage, generating two pairs of lateral cells and two pairs of medial cells (Figure 3.2D). This represents a striking departure from the highly conserved pattern of spiral cleavage found in non-cephalopod molluscs. The arrangement of the eight blastomeres in *Octopus* is also quite different from what is observed in decapodiform cephalopods: in squids and cuttlefish, the (presumably) posterior medial cells are quite narrow, but in *Octopus* they are more similar in shape to the anterior medial cells (Naef, 1928).

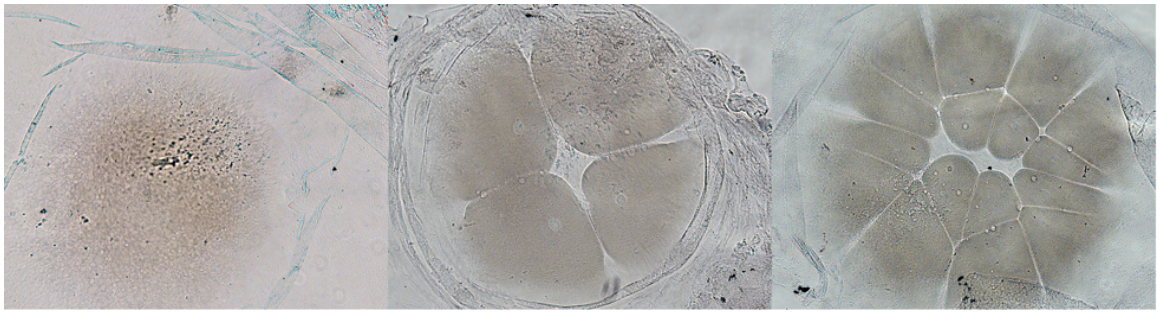


Figure 3.1. Lightfield photomicrographs of *O. bimaculoides* embryos at early cleavage stages (one, four, and sixteen cells).

Divisions are synchronous through at least the 64-cell stage (Figure 3.2G). Cells continue to proliferate, forming a cap of blastomeres at one end of the large yolk by the end of **stage 1** (Figure 3.2H). At the beginning of **stage 2**, we see this blastomere cap begin to compact, with cells at the margin forming a ring (Figure 3.2I). While Naef suggests that these cells are involuting from the margin, time-lapse light sheet microscopy suggests that cells may delaminate and ingress to form a second layer of cells. Interestingly, DAPI-stained nuclei radiate out from the **stage 2** blastodics (Figure 3.2I-J). This observation was not reported by Naef for *O. vulgaris*.

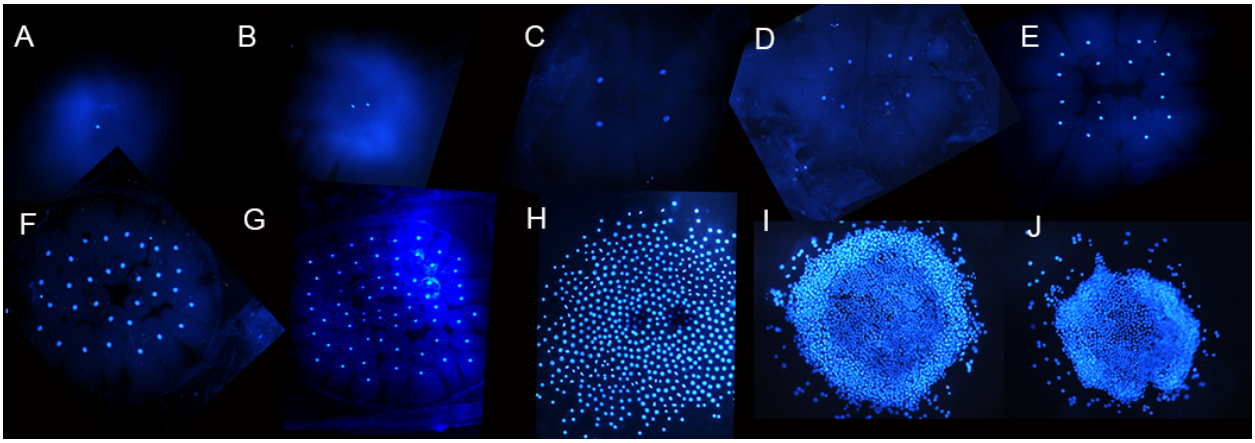


Figure 3.2.DAPI stained nuclei highlight the arrangement of cells in early *O. bimaculoides* embryos.
A) One cell. (B) Two cells. (C) Four cells. (D) Eight cells. (E) 16 cells. (F) 32 cells. (G) 64 cells. (H) ~400 cells (I) Stage 2. (J) Stage 2+.

Gastrulation and epiboly

The embryo begins to expand again at **stage 3** with the onset of epiboly. By **stage 4**, the embryonic cap extends approximately 5% of the egg length (Figure 3.3). While at this stage the embryo looks quite uniform using light microscopy (Figure 3.3), we can clearly distinguish the embryo proper from the extraembryonic ectoderm with a nuclear stain such as DAPI or SYBR Green I (Figure 3.4). The portion of the embryonic disc that will contribute to the embryo has more densely packed cells, whereas the extraembryonic ectoderm features larger cell nuclei that are spaced further apart (Figure 3.4H). Time-lapse SPIM suggests that the spreading of the cellular cap is due to cell divisions, as cells do not appear to be migrating relative to each other.

The cellular cap continues to envelop the yolk by epiboly. At epibolic stages, few morphological landmarks are discernable. Starting at **stage 3**, the center of the cellular cap is marked by a spot (Figure 3.4H-I, Figure 3.5A-E, red arrowheads), and there is a slight arc visible

in the embryonic tissue starting at **stage 4** (Figure 3.4H). Both the embryonic and the extraembryonic tissues continue to undergo cell division. At the end of stage 4, the cellular cap, which had been fairly flat, begins to curve down over the yolk. As *O. bimaculoides* embryos are much larger than the *O. vulgaris* embryos described by Naef, the amount of yolk covered by the extraembryonic ectoderm does not coincide with Naef's staging: by the middle of **stage 5**, the cap covers roughly 10% of the yolk in *O. bimaculoides*, where it covers approximately 25% of the yolk in *O. vulgaris* (Naef, 1928).

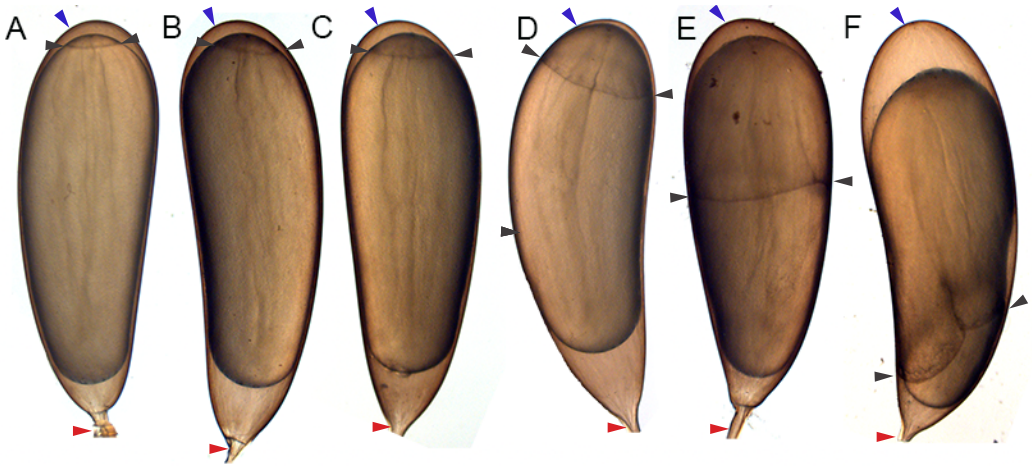


Figure 3.3. *O. bimaculoides* embryos photomicrographed in lateral views to demonstrate the progression of epiboly over stages 3-9.
(A) Stage 3. (B) Stage 4. (C) Stage 5. (D) Stage 6+ (E) Stage 8. (F) Stage 9. Blue arrowheads are at the rounded end of the chorion, where the initial cleavage and epibolic stages take place. Red arrowhead points to the stalk. Black arrowheads indicate the extent of epiboly.

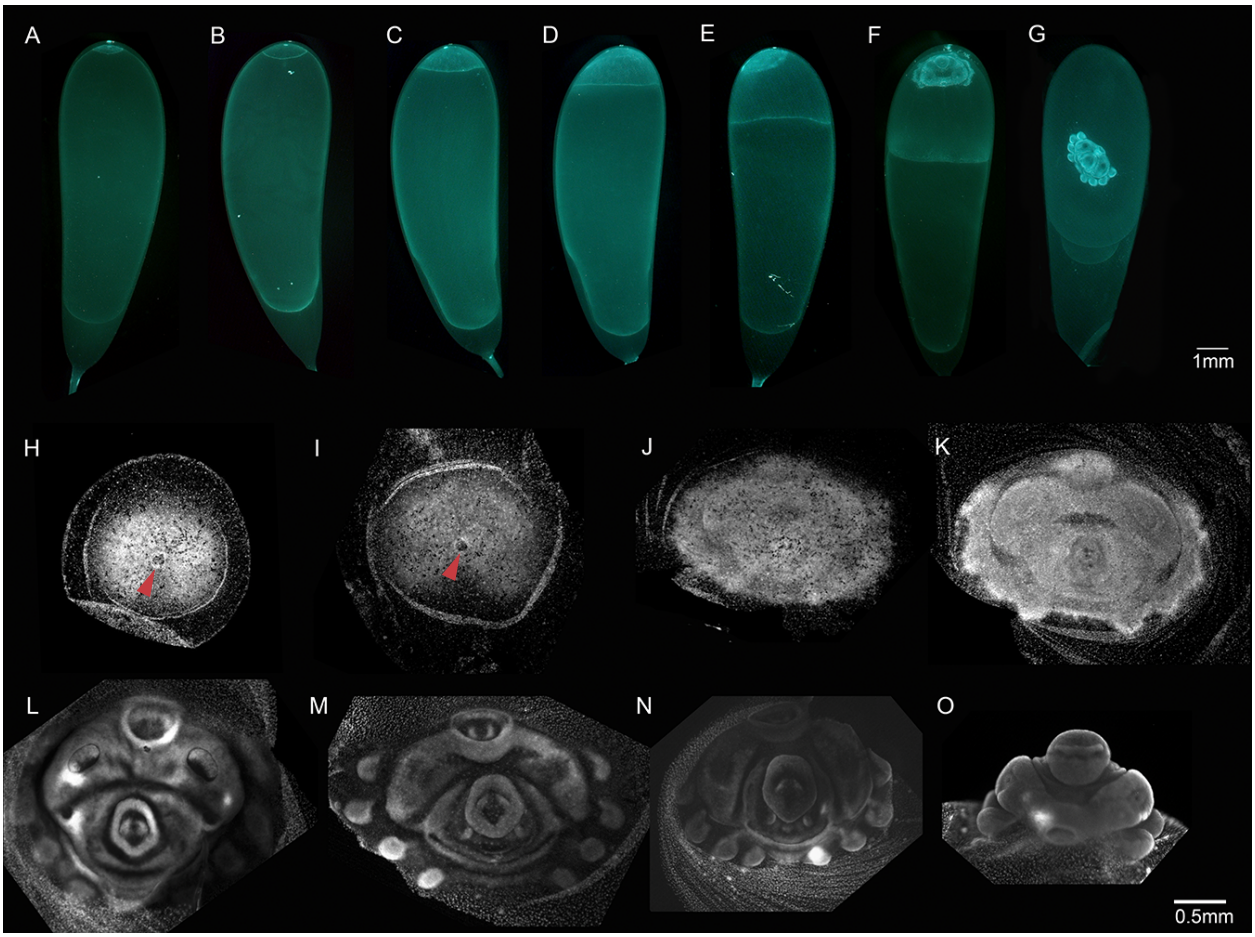


Figure 3.4. Nuclear staining at epibolic and organogenic stages of *O. bimaculoides* development.
(A-G) Lateral view of live *O. bimaculoides* embryos stained with SYBR Green I. (A) Stage 3. (B) Stage 4. (C) Stage 5. (D) Stage 6. (E) Stage 7. (F) Stage 8. (G) Stage 9. Note that blastokinetic movements are underway at stages 8 and 9. (H-N) Dorsal views of DAPI-stained fixed *O. bimaculoides* embryos. (H) Stage 4. (I) Stage 5+. (J) Stage 6+. (K) Stage 7. (L) Stage 7+. (M) Stage 8. (N) Stage 9. (O) Oral view of a fixed stage 10 embryo stained with DAPI. Red arrowheads in H and I indicate the central spot. The morphology of later stages of *O. bimaculoides* development (stages 13 and 18-20) are photodocumented in Figure 2 of Shigeno et al. (2015).

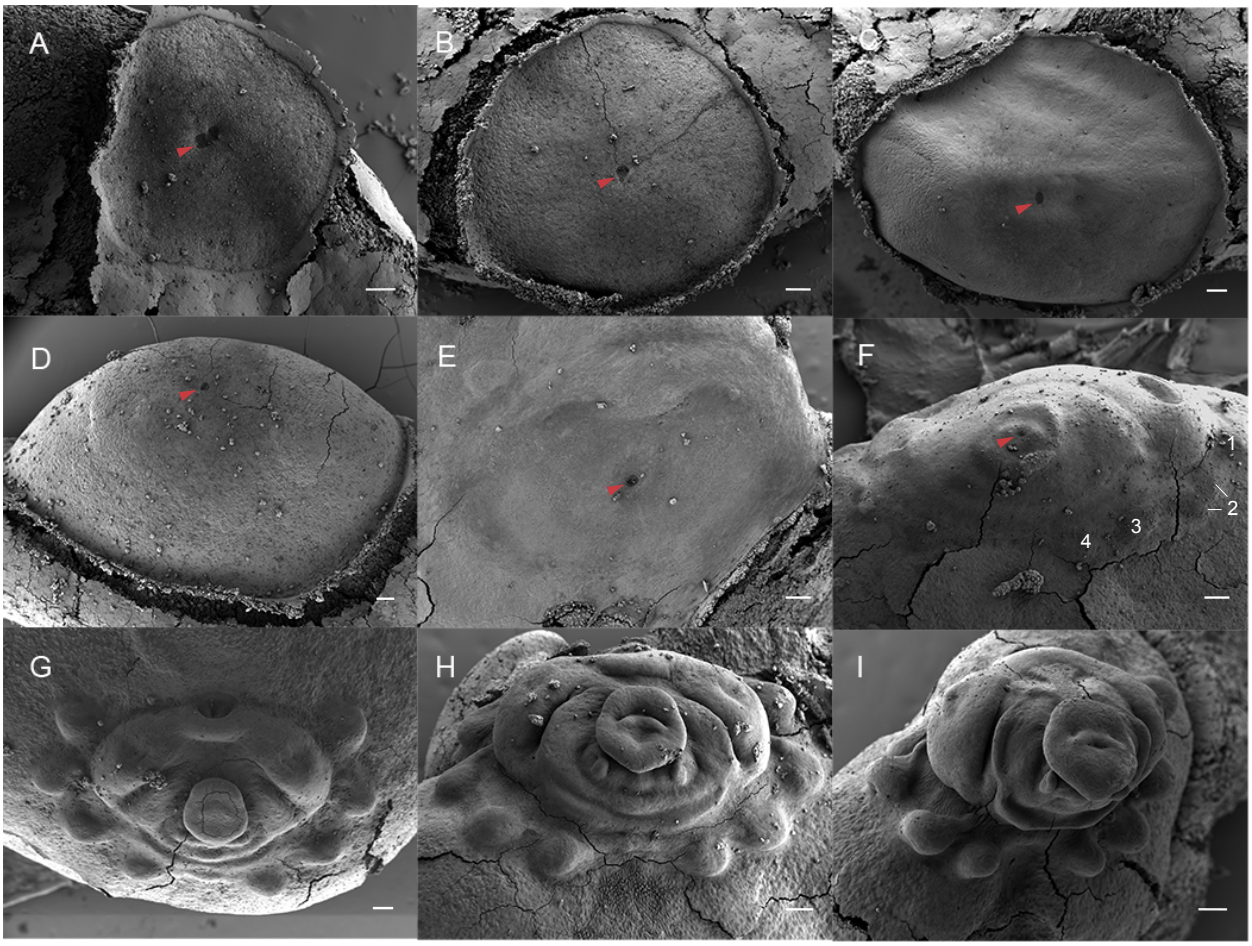


Figure 3.5. SEM images of stages 3-10 *O. bimaculoides* embryos.
Embryos in panels E-I are oriented with the prospective mouth to the top, except for panel D, where the prospective mouth is facing to the right. (A) Stage 3. (B) Stage 4. (C) Stage 5. (D) Stage 6. (E) Stage 6+. (F) Stage 7. (G) Stage 8. (H) Stage 9. (I) Stage 10. Red arrowheads indicate the central spot. Arms primordia are indicated 1-4 in panel F. Scale bars: 100uM.

Emergence and elaboration of organ primordia

At **stage 6** (20% epiboly), organ primordia are not yet distinct, and scanning electron microscopy (SEM) indicates the surface of the embryo is smooth (Figure 3.5D). While the primordia are not yet evident using light microscopy, they can be discerned using nuclear dyes or SEM at stage 6+. Nuclear dyes reveal the prospective arms as patches of nuclei at the periphery, and the eyes, mantle and palliovisceral cord are discernable (Figure 3.4J). The prospective

primordia appear in SEM to be made up of cells with numerous protrusions, but surface of the embryo is flat (Figure 3.6).

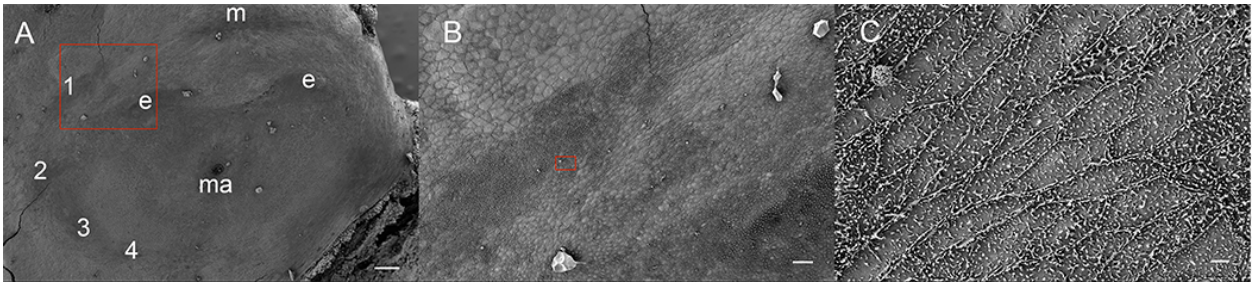


Figure 3.6 SEM of organ primordia at late stage 6.
(A) Stage 6+. e, eye; m, mouth; ma, mantle; 1, 2, 3, 4 mark the arm primordia. Scale bar 100μM. (B) Higher magnification view of region between the left eye and arm 1 marked by the box in (A). Scale bar 20μM. (C) Higher magnification view of region of arm 1 in box in (B). Scale bar 2μM.

As **stage 7** begins (~30% epiboly), primordia begin to emerge as protrusions from the surface of the embryo. The mantle and the eyes protrude first, followed by the arms. Notably, the primordia of the second arm appears to have two parts at stage 7 (Figure 3.5F), making it appear as if there are five arms on each side, as found in decapodiforms. The fold of the eye primordia is apparent. As stage 7 progresses, the primordia become more clearly demarcated (Figure 3.4L).

By **stage 8** (~50% epiboly), the body plan is present and primordia are well-established, with the mouth and arms at the periphery, the mantle at the center, and the eyes, neural cords, and gills located in-between (Figure 3.4M, Figure 3.5G). The eye fold continues to constrict. A depression is clearly evident between the prospective supraesophageal and subesophageal brain (the "cleft" described in Chapter 5.2). The clefts first appear at late stage 7 (Figure 3.4L, Figure 3.5F). The bilateral clefts, which first appear at stage 7, if extended to the midline, would meet at an obtuse angle. By stage 8, the angle is decreased and the clefts form an incomplete chevron. By

stage 8, the clefts have become prominent, and the angle of the chevron decreases. The clefts progressively narrow through stages 9 and 10.

By **stage 9** (75% epiboly), the primordia become more prominent, particularly the mouth, eyes, supraesophageal and subesophageal brain, mantle and gills. The embryo, which began development at the rounded end of the chorion, opposite the stalk, begins to migrate over the yolk, proceeding to the stalked end of the chorion (Figure 3.4G). The embryo will completely circumnavigate the yolk by the end of stage 10.

Epiboly is complete by **stage 10**. The most frequent developmental anomaly observed is that epiboly will end prematurely, cutting into the yolk and pinching some of it off. Development can proceed normally under these circumstances, simply producing a commensurately small embryo. However, incomplete epiboly is often accompanied by elevated mortality.

The organ primordia begin to contract at stage 10, with the arm crown constricting towards the center of the embryo resulting in the central primordia being positioned dorsal to the arms (Figure 3.5I, Figure 3.7A-C). The retina primordia have an orange hue. The embryo also begins a second rotation around the yolk, this time remaining at the pointed, stalked end until after stage 16. By **stage 11**, the organ primordia fully telescope out from the embryo and the axes of the embryo begin to resemble the adult orientation. The arm buds project out from the embryo forming the brachial crown, though the mouth is still located between the first pair of arms. The mantle also projects prominently away from the yolk. The tissue surrounding the eyes (white body) becomes thick and the retina is orange in color. Waves of contraction (peristalsis) along the extraembryonic tissue (yolk sac) are evident.

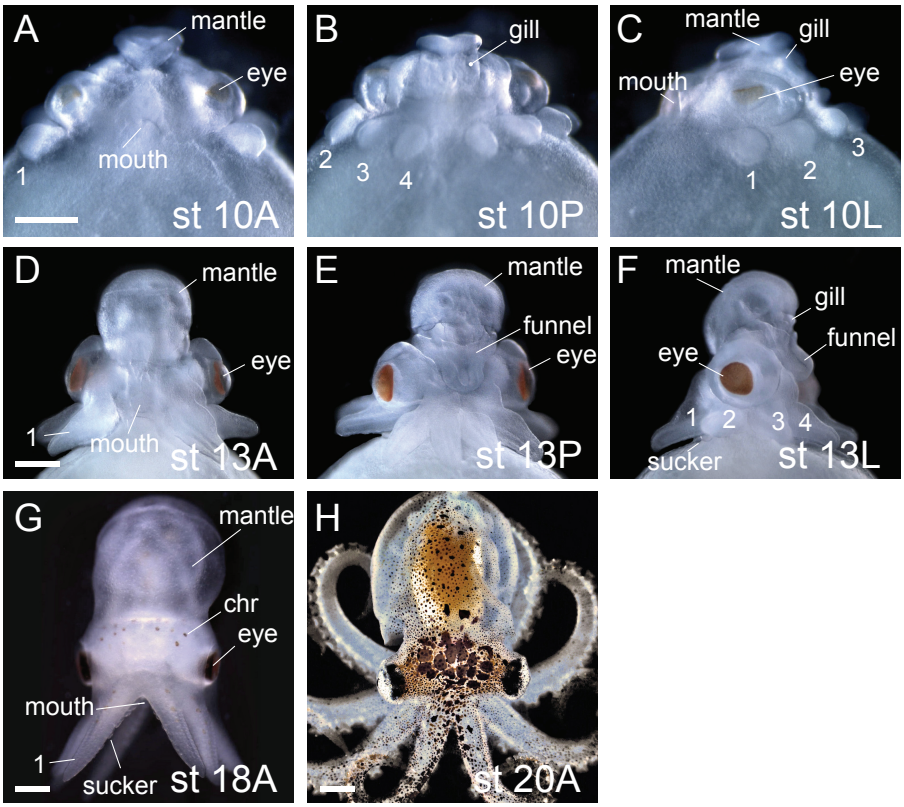


Figure 3.7. Later stages of octopus development.
(A) Stage 10, anterior view. (B) Stage 10, posterior view. (C) Stage 10, lateral view. (D) Stage 13, anterior view. (E) Stage 13, posterior view. (F) Stage 13, lateral view. (G) Stage 18, anterior view, with chromatophores indicated (chr). (H) Stage 20, anterior view. Arms are indicated as 1-4, starting at the anterior. Images were supplied by Dr. Shuichi Shigeno, and are modified from Shigeno et al. (2015).

At **stage 12**, the pillar-like arms elongate and have recognizable suckers on the ventral side. The mantle begins to extend over the gill primordia. The posterior ends of the funnel primordia meet medially, forming a tube. The eye-optic lobe-white body complex begins to form a stalk, bulging out from the central brain. The embryo is also increasing in size, and is distinctly larger than at stage 11. At **stage 13**, the elongating arms are visibly pointed at the tips (Figure 3.7D-E). Contrary to Naef's finding in *O. vulgaris*, which develop only three suckers prior to hatching as a paralarva, sucker development in *O. bimaculoides* is extensive. At **stage 14,** the

mantle also continues to extend over the visceral primordia, and the distal portion of the funnel has closed, forming a tube.

By **stage 15**, the arms each have roughly a dozen suckers. The arms have begun to twitch and flex. The retinas are a deep orange-red color, and the statocysts are prominent bulbs posterior to the subesophageal brain. Within the mantle cavity, the hearts can be distinguished. The mantle extends further from the yolk, and the embryo continues to increase in size. The **stage 16** embryo is visibly larger than the stage 15 embryo, and the retinas have darkened to a reddish-brown. The gills also are elaborate and elongated, and the mantle has begun to twitch. By **stage 17**, the hearts have begun to beat and the mantle to pump. The mouth is found ventrally and the gap between the first arm pair is less prominent. At **stage 18** brown chromatophores are evident on the dorsal aspect of the head complex and the ventral edge of the mantle (Figure 3.7G). The embryo continues to increase in size, and has absorbed more than half the yolk. At this point the embryo has reversed its orientation and now sits opposite from the stalk.

Stage 19 embryos continue to grow, absorbing the yolk. Brown, yellow, and red chromatophores cover the skin and begin expanding and contracting. The eyes are covered with iridophores, yielding a green appearance. The embryo remains largely transparent. The embryo will react to visual stimuli at these stages, particularly changes in light levels or movement. The arms continue to elongate and are lined with more than two-dozen suckers. Premature hatching can occur at stage 19 if the embryo is disturbed, for example by agitating the strands of embryos. If the external yolk sack has not been completely absorbed, it will remain attached for several hours. By **stage 20**, the yolk is nearly depleted and embryo becomes very active in the chorion. Upon hatching, the octopuses rapidly engage in a number of adult behaviors, including inking,

swimming, changing color, walking and hiding. The eponymous eyespots are not evident until several weeks post-hatching.

CHAPTER 4

Cephalopod developmental control genes in the octopus genome

INTRODUCTION

Cephalopods diverged from the rest of the conchiferan molluscs more than 500 million years ago. They have since branched into a number of extant clades comprising the Nautiloidea (*Nautilus* and *Allonautilus*) and Coleoidea (squid, cuttlefish and octopus). Coleoid cephalopods such as the octopus show remarkable morphological departures from the basic molluscan body plan (Kroger et al., 2011). The molluscan foot evolved into flexible arms lined with hundreds of suckers, which function as specialized tactile and chemosensory organs, and the muscular mantle enables the jet-propelled way of life of coleoid cephalopods. The cephalopod nervous system is vastly modified in size and organization from the ancestral molluscan arrangement of nerve cords and ganglia (Wells, 1978; Young, 1971). In the common octopus, for example, the nervous system comprises a circumesophageal brain, paired optic lobes, and axial nerve cords in each arm. Together these structures contain nearly half a billion neurons, more than six times the number in a mouse brain (Herculano-Houzel et al., 2006; Young, 1971). Extant coleoid cephalopods demonstrate sophisticated behaviors including complex problem solving, task-dependent conditional discrimination, observational learning and spectacular displays of camouflage (Fiorito and Scotto, 1992; Grasso and Basil, 2009; Hanlon and Messenger, 1996; Hvorecny et al., 2007).

Perhaps as dramatic is the extent of convergent evolution with vertebrates, and in particular, fishes (Packard, 1972). While the camera eye is the most well known example of convergence, cephalopods and fish also feature elaborate nervous systems, giant axons, lateral line systems, a blood-brain barrier, and a closed circulatory system. In addition, the early

development is strikingly similar between cephalopods and fish, with meroblastic cleavage on top of a large yolk followed by epiboly, a process strikingly different from the spiral cleavage program of other molluscs. Given their remarkable features, cephalopods have been used as models to study a number of biological processes but, until recently, little sequence information was available from this clade.

Sequencing data will greatly aid the study of cephalopod biology. Sequencing the genome produces a comprehensive list of genes and reveals the adjacent regulatory sequences, as well as overall genome architecture. To date, genome-scale analyses have been overwhelmingly focused on two of the three bilaterian clades: the deuterostomes and the ecdysozoans; in 2015, more than four out of five published genomes were from vertebrates and arthropods (Dunn and Ryan). This work has revealed that animals share a core set of genes, including developmental transcription factors and ligands, as well as genes involved in nervous system function, in addition to the housekeeping genes typical of eukaryotes. The sequencing of lophotrochozoan taxa has expanded our knowledge of this toolkit, as many gene families appear to have been lost in a number of ecdysozoan clades (Putnam 2007).

These bilaterian toolkit genes are often found in multiple copies in vertebrates, lending support to Ohno's hypothesis (Ohno, 1970) of two rounds of whole genome duplication along the vertebrate lineage (Holland et al., 1994). Some researchers have attributed the morphological and behavioral complexity of vertebrates to this expanded gene repertoire (Nithianantharajah et al., 2013). It has been proposed that whole genome duplications were also important in cephalopod evolution (Bonnaud et al., 2004; Hallinan and Lindberg, 2011; Yoshida et al., 2011), paralleling the independent whole genome duplication events that occurred early in vertebrate evolution (Holland et al., 1994).

To uncover the genetic features of these remarkable and highly derived animals, we sequenced the genome and transcriptomes of the California two-spot octopus, *Octopus bimaculoides*. Several mechanisms can drive genomic novelty, including expansion of existing gene families, evolution of novel genes, modification of gene regulatory networks, and reorganization of the genome. We found evidence for all of these mechanisms in the *O. bimaculoides* genome, including expansions in several gene families, a suite of octopus- and cephalopod-specific genes, and extensive genome shuffling. While paleopolyploidy is an attractive framework for both gene family expansion and increased regulatory complexity, we saw no indication of a whole genome duplication event in *O. bimaculoides* (Table 4.1).

The focus of this chapter is a complete inventory of the developmental control genes in octopus. A portion of these results have been reported in Albertin et al. (2015).

RESULTS

Fluorescence-based genome size estimate

To estimate the size of the octopus genome experimentally, I compared the fluorescence of DAPI-stained nuclei from octopus embryos and embryonic zebrafish (*Danio rerio*). The ratio of the *O. bimaculoides* peak to the *D. rerio* peak is 1.83-1.85 across multiple runs (Figure 4.1). Accepting a haploid genome size estimate of 1.454 Gb for *D. rerio* (Freeman et al., 2007; Howe et al., 2013), I estimate the genome size of *O. bimaculoides* to be 2.66-2.68 Gb. This estimate of genome size accords well with the 2.87 Gb *O. bimaculoides* genome estimate based on counting k-mers, given an assumption that genome sequencing reads sample the genome uniformly (Albertin et al., 2015). This estimate is several times larger than other sequenced molluscan and lophotrochozoan genomes (Simakov et al., 2013).

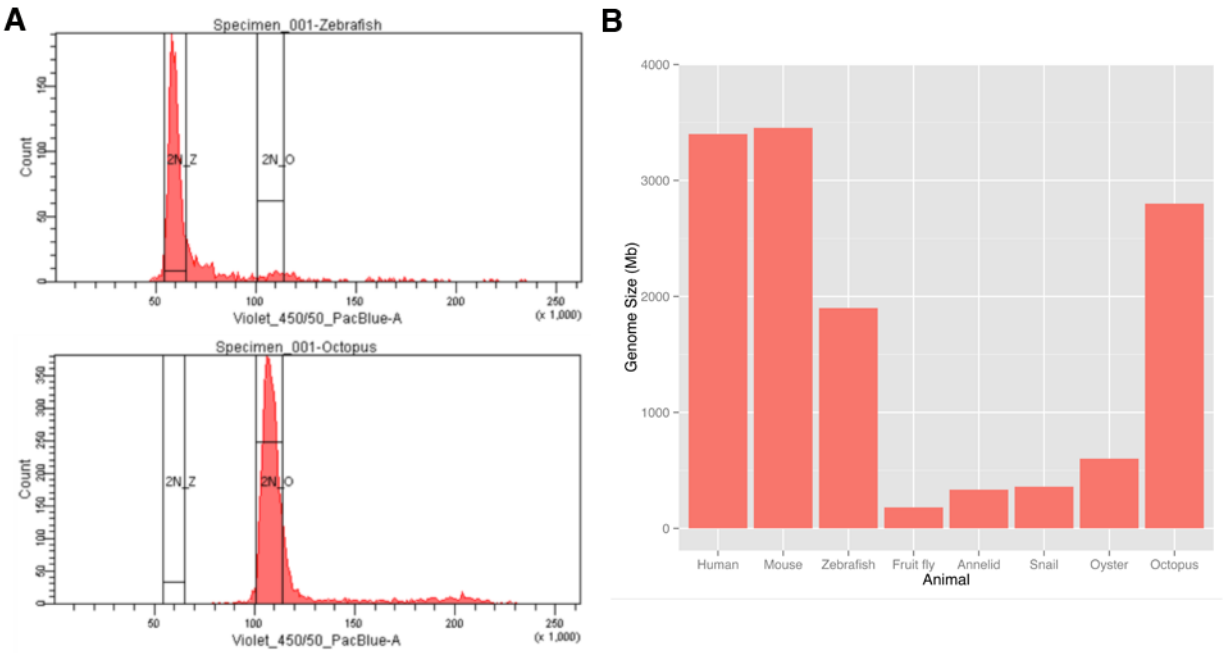


Figure 4.1. Octopus genome size relative to other bilaterians.
(A) Fluorescence peaks from genomic DNA in *D. rerio* (top) and *O. bimaculoides* (bottom) embryonic cells. Relative DAPI content of the cells is indicated in red, with the sharp peaks corresponding to 2C (diploid content) (Vinogradov, 1998). (B) Genome size of selected bilaterians.

Genome and transcriptome sequencing.

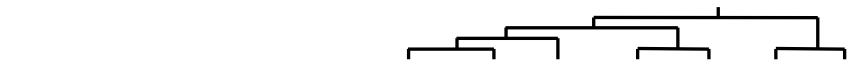
A draft genome sequence for *O. bimaculoides* was generated from 60-fold redundant shotgun sequencing of a single diploid individual and we annotated it using extensive transcriptome sequence from 12 tissues (totaling 100 gigabases). The shotgun assembly captures 83% of the estimated genome size and has a contig N50-length of 5.4 kb and a scaffold N50-length of 470 kb. The assembly captures local linkages: half of all genes are on scaffolds with 8 or more genes, with the longest scaffold containing 99 genes. We compared the *de novo* assembled RNA-seq output to the genome to evaluate the completeness of the genome assembly and found that it accounts for more than 97% of transcripts with open reading frames: only 1,130

out of 48,259 transcripts (2.34%) did not have a match in the genome with a minimum identity of 95%.

Gene complement

We predicted 33,638 protein-coding genes and found alternate splicing at 2,819 loci. In gene family content, domain architecture, exon-intron structure, and functional content, the octopus genome encodes a "typical" invertebrate bilaterian gene set similar to that of the cephalochordate *Branchiostoma floridae* (Putnam et al., 2008), the polychaete annelid *Capitella teleta*, and the limpet *Lottia gigantea* (Simakov et al., 2013). As developmental control genes are often highly conserved across phylogenetically disparate groups of animals, and may serve as indicators of genomic events, such as duplications, or dramatic losses, we catalogued in detail members of developmental signaling ligand and transcription factor families. Families surveyed include the Fibroblast Growth Factor, Wnt, Transforming Growth Factor β, and Hedgehog developmental ligands, as well as the transcription factors of the C2H2-ZNF, Homeodomain, High Mobility Group, basic Helix-Loop-Helix, Nuclear Hormone Receptor, Fox, and Tbox families. The results are summarized in Table 4.1 and discussed in detail below. The broad conclusion from this work is that, with the key exception of the C2H2-ZNF and protocadherin subfamily of cadherins, the complement of octopus developmental control genes is very similar to that seen in other invertebrates, especially lophotrochozoan, and greatly reduced from that present in vertebrates.

Table 4.1. Selected metazoan developmental control genes.
Number of members of selected developmental ligand and transcription factor families from *O. bimaculoides* (Obi), *L. gigantea* (Lgi), *C. teleta* (Cte), *D. melanogaster* (Dme), *C. elegans* (Cel), *B. floridae* (Bfl), and *H. sapiens* (Hsa). Dendrogram above species names reflects their evolutionary relationships.



		Obi	Lgi	Cte	Dme	Cel	Bfl	Hsa
Ligands	Hedgehog	1	1	1	1	0	1	3
	Fibroblast Growth Factor	3	2	1	3	3	8	22
	Delta/Jagged	4	1	1	2	4	2	7
	Wnt	12	10	12	7	5	17	19
	TGFβ/BMP	12	9	14	6	5	22	33
Transcription Factors	Tbox	9	9	7	8	21	9	18
	Fox	16	28	26	17	18	42	43
	High Mobility Group	23	15	14	13	16	51	125
	Nuclear Hormone Receptor	40	44	45	16	274	33	48
	Helix Loop Helix	50	63	64	59	42	78	118
	Homeodomain	114	121	111	104	99	133	333
	C2H2 Zinc Finger	1790	413	222	326	211	1338	764
	Cadherins	19	55	24	17	12	28	42
	Protocadherins	168	16	25	0	0	2	71

DEVELOPMENTAL SIGNALING LIGANDS

Hedgehog

Hedgehog (Hh) signaling is an ancient part of the bilaterian developmental toolkit: *Hh*s have been found in the sea anemone *N. vectensis*, as well as most bilaterian animals. *Hhs* participate in a range of developmental processes, including segment polarity regulation and patterning of the gut, wing disc and eyes in *D. melanogaster* (Lee et al., 1992; Tabata and Kornberg, 1994), patterning of the gut, neural tube and limbs in vertebrates (Hooper and Scott, 2005; McGlinn and Tabin, 2006), and gut formation in *H. robusta* (Kang, 2003). While mammals have three *Hh*s (*Shh, Ihh,* and *Dhh*), a single *Hh* is present in the invertebrate bilaterian genomes examined, with the exception of *C. elegans*, which lacks a clear *Hh* gene (Burglin and Kuwabara, 2006) (Table 1). We found a single *Hh* in the *O. bimaculoides* genome (Figure 4.2), which codes for both the

N-terminal Hh signaling domain and the C-terminal Hint domain characteristic of authentic *Hh* proteins.

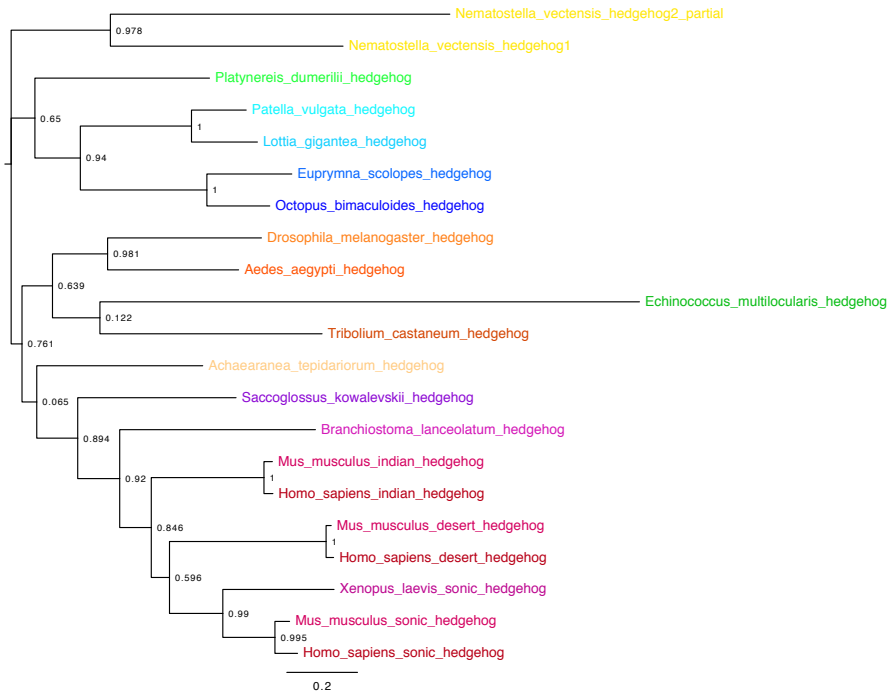


Figure 4.2. Phylogenetic tree of eumetazoan hedgehogs.
Human (red), *Mus musculus* (pink), *Xenopus laevis* (plum), *Drosophila melanogaster* (light orange), *Tribolium castaneum* (dark orange), *Aedes aegypti* (orange), *Nematostella vectensis* (mustard yellow), *Platynereis dumerilii* (green), *Lottia gigantea* (teal), *Patella vulgata* (sky blue), *Octopus bimaculoides* (dark blue), *Euprymna scolopes* (blue), *Branchiostoma lanceolatum* (magenta), and *Saccoglossus kowalevskii* (purple). These colors will be maintained throughout.

FGFs

Fibroblast growth factors (FGFs) are a family of small, extracellular signaling peptides that play an important role a number of biological processes. FGFs act through binding to transmembrane tyrosine kinase receptors (FGFRs), which activate cytoplasmic cascades. FGFs appear to be absent from sponges, *Trichoplax adhaerens*, and ctenophores, but are present in eumetazoans, including cnidarians (Bertrand et al. 2013, Matus et al. 2007). A few FGFs have been identified in model ecdysozoan species: there are three in *D. melanogaster* (*branchless,*

pyramus, and *thisbe*), four in *T. castaneum* (*FGF1A, FGF1B, FGF8, BNL*), and two in *C. elegans* (*Egl-17* and *Let-756*). A recent search for FGFs in lophotrochozoan taxa also retrieved two in molluscs and one in annelids (Oulion et al. 2012). At least in part through whole genome duplication, vertebrates have greatly expanded the FGF family: humans and mice have 22 FGFs, which are subdivided into several subfamilies. How these FGFs from distantly related taxa are related, however, has been difficult to resolve (Oulion et al. 2012).

Our search of the *O. bimaculoides* genome recovered three *FGF* family members. To identify these FGF genes, we built a number of phylogenetic trees, but found that trees generated with relatively few taxa were unable to resolve FGF relationships. However, the addition of many invertebrate taxa recovered several well-supported branches that illuminate the evolution of this signaling ligand family.

We robustly recovered an FGF8/17/18 family, with a single member in invertebrates (including the deuterostome *S. kowalevskii*), and three in vertebrates (Figure 4.3). *D. melanogaster* has two members of the FGF8/17/18 family, however, other arthropods appear to have just one. While *C. elegans* does not have an FGF8 ortholog, we did find a single FGF8/17/18 gene in other nematodes, suggesting the absence in *Caenorhabditis* is from gene loss. A single FGF8/17/18 gene was confirmed in lophotrochozoans, including molluscs (cephalopods, bivalves, and gastropods), brachiopods (*Terebratalia transversa*) and annelids (*Capitella teleta*).

We recovered an FGF1/2 family in bilaterians, with a single member in nematode (LET-756), arthropod, and mollusc species, and two in vertebrates. There is no FGF1/2 in *Drosophila*, but it is present in other insects.

We recovered a large clade of FGFs that are deuterostome-specific. This clade contains the FGF9/16/20, FGF4/6/11/12/13/14, FGF3/5/7/10/22, and FGF19/21/23 subfamilies. However, a single protostome sequence, an FGF from octopus, routinely resolves within this clade as a member of the 9/16/20 subfamily. Finally, we obtained phylogenetic evidence that Branchless is an insect-clade specific FGF.

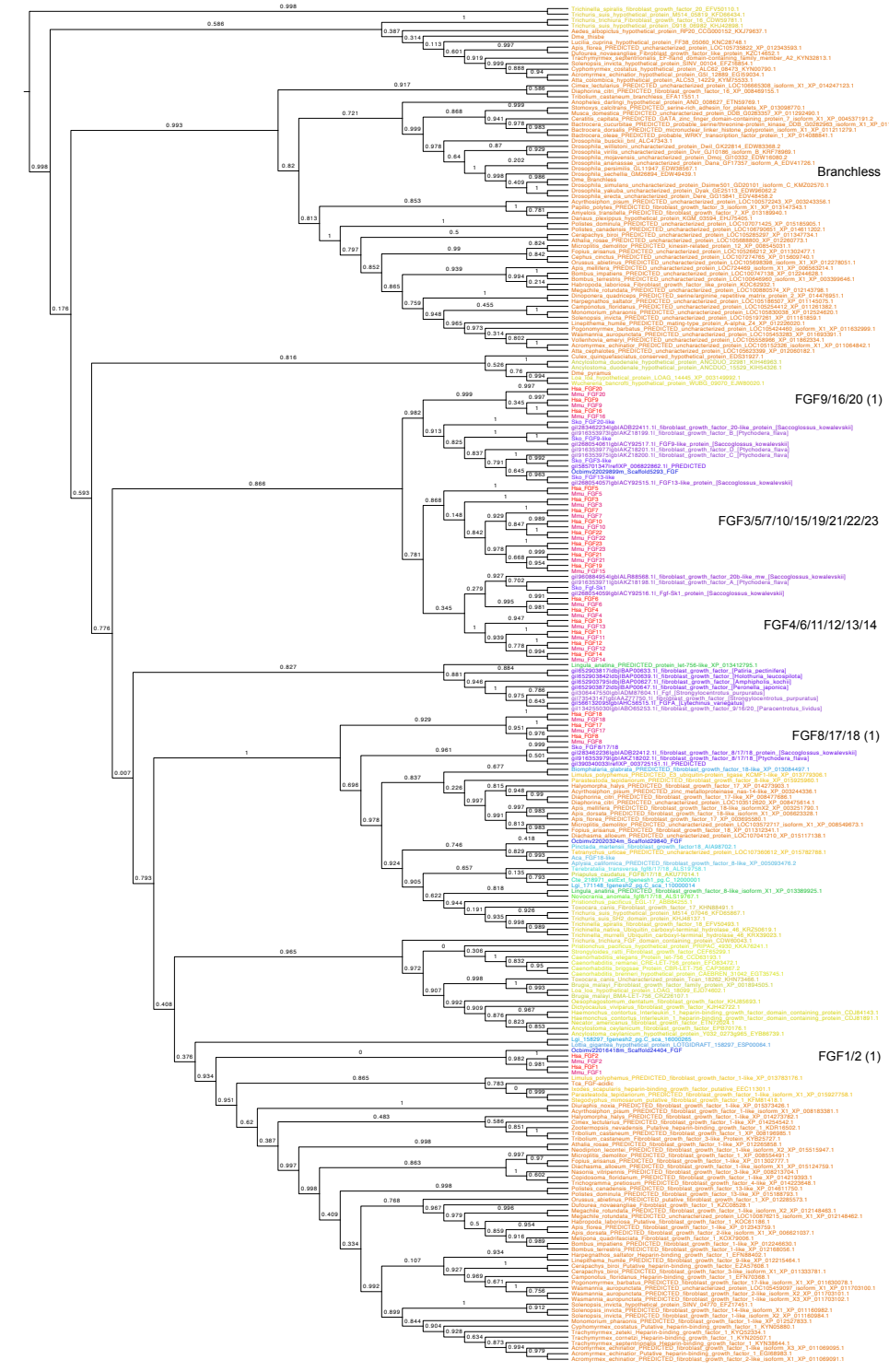


Figure 4.3. Metazoan FGF phylogenetic tree.
Human genes are indicated in red, ambulacrarian in purple, arthropod in orange, chelicerate in mustard, nematode in chartreuse, non-mollusc lophotrochozoan in green, mollusc in blue, and octopus in dark blue. *Homo sapiens* abbreviated to Hsa; *Mus musculus* abbreviated Mmu.

WNTs

WNTs are secreted intercellular signaling glycoproteins that play important roles in developmental processes in metazoans. Thirteen *Wnt* subfamilies have been identified in eumetazoans. Apart from chordates, each subfamily, when present, is represented by one gene. Twelve of the 13 subfamilies are present in the cnidarian *N. vectensis* (Kusserow et al., 2005), and 11 and 12 of the subfamilies are found in the lophotrochozoans *L. gigantea* and *C. teleta*, respectively (Cho et al., 2010). This bilaterian *Wnt* gene set is expanded in vertebrates (humans have 19), and reduced in ecdysozoans (*D. melanogaster* and *C. elegans* have 7 and 5 *Wnts*, respectively) (Eisenmann, 2005; Rubin et al., 2000). *O. bimaculoides*, like *C. teleta*, has 12 *WNTs* (Figure 4.4). All protostomes examined thus far, including octopus, lack *WNT3* family members (Bolognesi et al., 2008; Cho et al., 2010). While *WNT8* is absent in *L. gigantea*, it is present in *O. bimaculoides* and in *C. gigas*, indicating that it is part of the ancestral molluscan *Wnt* complement.

Unlike the Hox cluster and other transcription factor families (see below), three of the *Wnts* have maintained genomic linkage in the *O. bimaculoides* genome. *WNT1*, *WNT6*, *WNT9* and *WNT10* are found on the same chromosome in a number of animals, including *B. floridae, D. melanogaster, C. teleta*, *L. gigantea,* and *N. vectensis* (only *WNT10* and *WNT6*) (Cho et al., 2010). In humans, *WNT1* and *WNT10B* are on chromosome 12q13 and *WNT6* and *WNT10A* are on chromosome 2q35. In *O. bimaculoides*, *WNT1*, *WNT6*, and *WNT10* are all located on the same scaffold. While *WNT5* and *WNT7* are also linked in *N. vectensis, B. floridae* and *L. gigantea*, we do not find evidence that this linkage is retained in octopus.

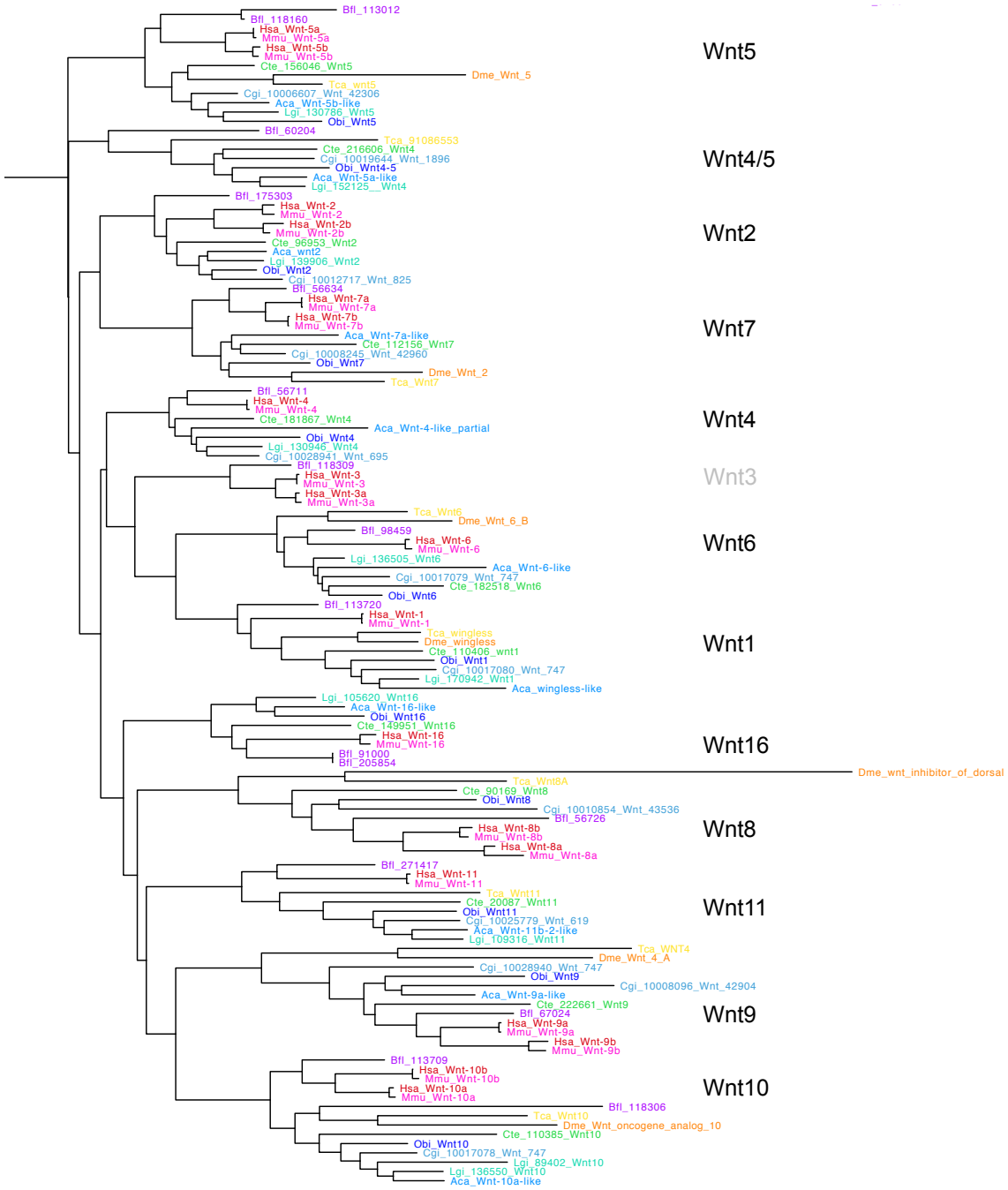


Figure 4.4. Phylogenetic tree showing distribution of the 13 eumetazoan Wnts. Branch labels are colored according to species: Hsa (red), Mmu (pink), Tca (*Tribolium castaneum,* yellow), Dme (*Drosophila melanogaster,* orange), Cgi (*Crassostrea gigas,* cornflower blue), Aca (*Aplysia californica,* sky blue), Lgi (*L. gigantea,* teal), Obi (*O. bimaculoides,* dark blue), Cte (*Capitella teleta,* green). Gene families are listed to the right. The WNT3 label is faded to reflect its absence in protostomes, including *O. bimaculoides.*

Transforming Growth Factor-β

The Transforming Growth Factor-β (TGF-β) family of ligands includes TGF-β *sensu stricto*, Activins, Leftys, Growth Differentiation Factors (GDFs), and Bone Morphogenetic Proteins (BMPs). The TGF-βs play important roles during animal development. Mammalian genomes encode 33 TGF-β family members (Wharton and Derynck, 2009) while only handfuls are reported in ecdysozoans (Gesualdi and Haerry, 2007; Gumienny and Savage-Dunn, 2013). We found 12 *TGF-β* genes in *O. bimaculoides*, including homologs of *BMP2/4*, *BMP3*, *BMP5-8*, *GDF2*, *GDF8/11*, *Nodal*, and *ADMP* (Figure 4.5). While other lophotrochozoans appear to have *BMP10/GDF2* homologs, we did not retrieve an octopus homolog in this analysis. Overall, lophotrochozoan genomes have roughly twice as many TGF-βs than do model ecdysozoans, and have homologs of genes absent in ecdysozoan clades, including *Nodal, BMP3,* and *ADMP*.

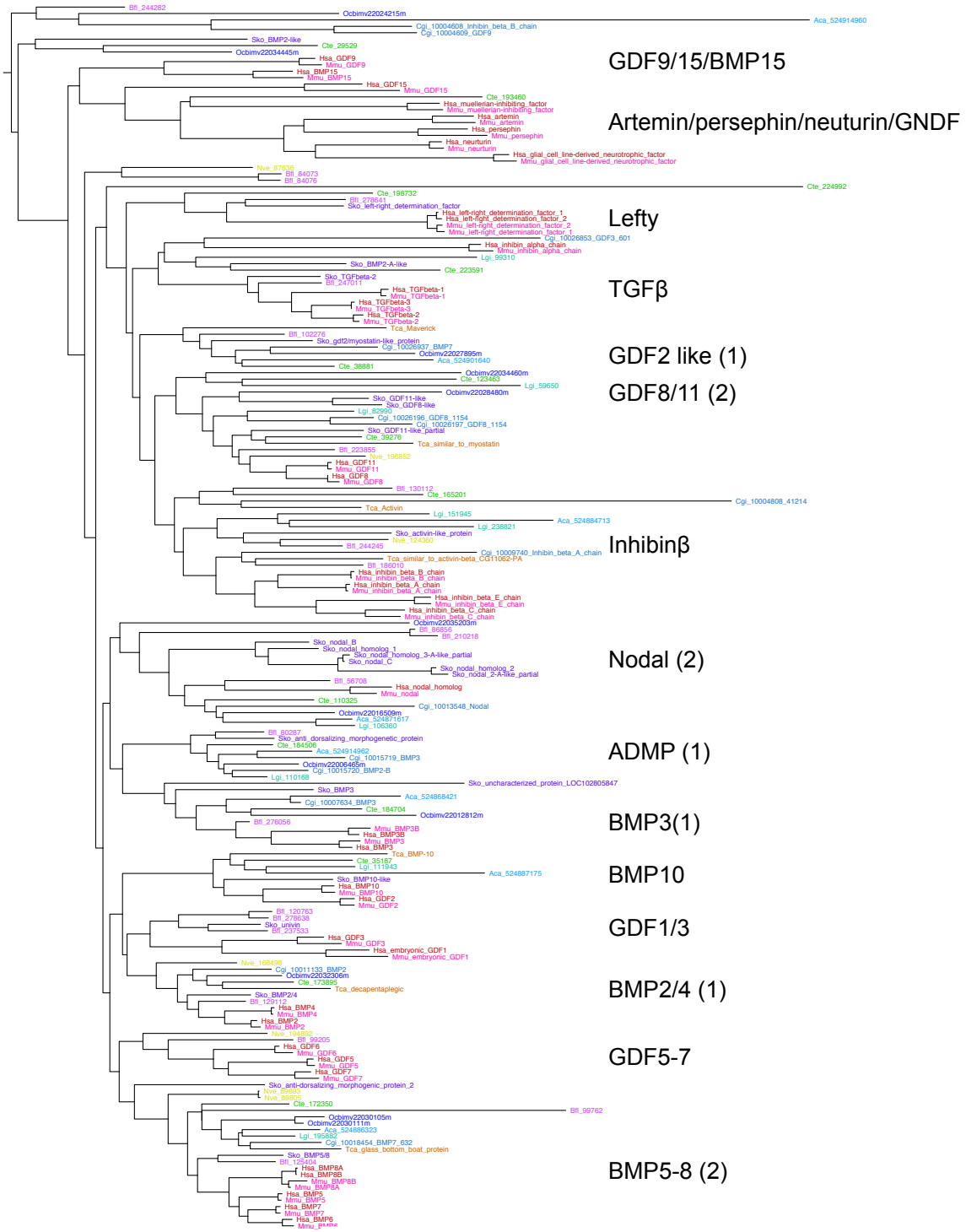


Figure 4.5. Phylogenetic tree showing distribution of eumetazoan TGFβ superfamily proteins.

Branch labels are colored according to species: Hsa (red), Mmu (pink), Nve (yellow), Tca (orange), Cgi (cornflower blue), Aca (sky blue), Lgi (teal), Obi (dark blue), Cte (green). Gene families are listed to the right; parentheses indicate number of homologs found in the *O. bimaculoides* genome.

DEVELOPMENTAL TRANSCRIPTION FACTORS

Tbox transcription factors

T-box genes are a family of transcription factors found throughout the animal tree that play crucial roles in development. The T-box domain is encoded by a 180-200 amino acid sequence and binds DNA. Brachyury (also known as T) plays an important role in the development of the notochord and posterior mesoderm in vertebrates. Mutation of a number of Tbox family members result in a range of congenital human syndromes, including defects of the heart and limbs, as well as craniofacial abnormalities. The octopus genome contains single orthologs of *BRACHYURY/TBX19, EOMESODERMIN/T-BRAIN/TBX21, TBX1/10,* and *TBX4/5,* two orthologs of *TBX20*, and three from the *TBX2/3* family (Figure 4.6). We did not recover a *TBOX6* or *TBX15/18/22* orthologs in octopus. Neither *TBX6* nor *TBX15/18/22* appear to have protostome orthologs, though *TBX15/18/22* can be found in *C. intestinalis* and *N. vectensis*, suggesting that this family was lost in protostomes.

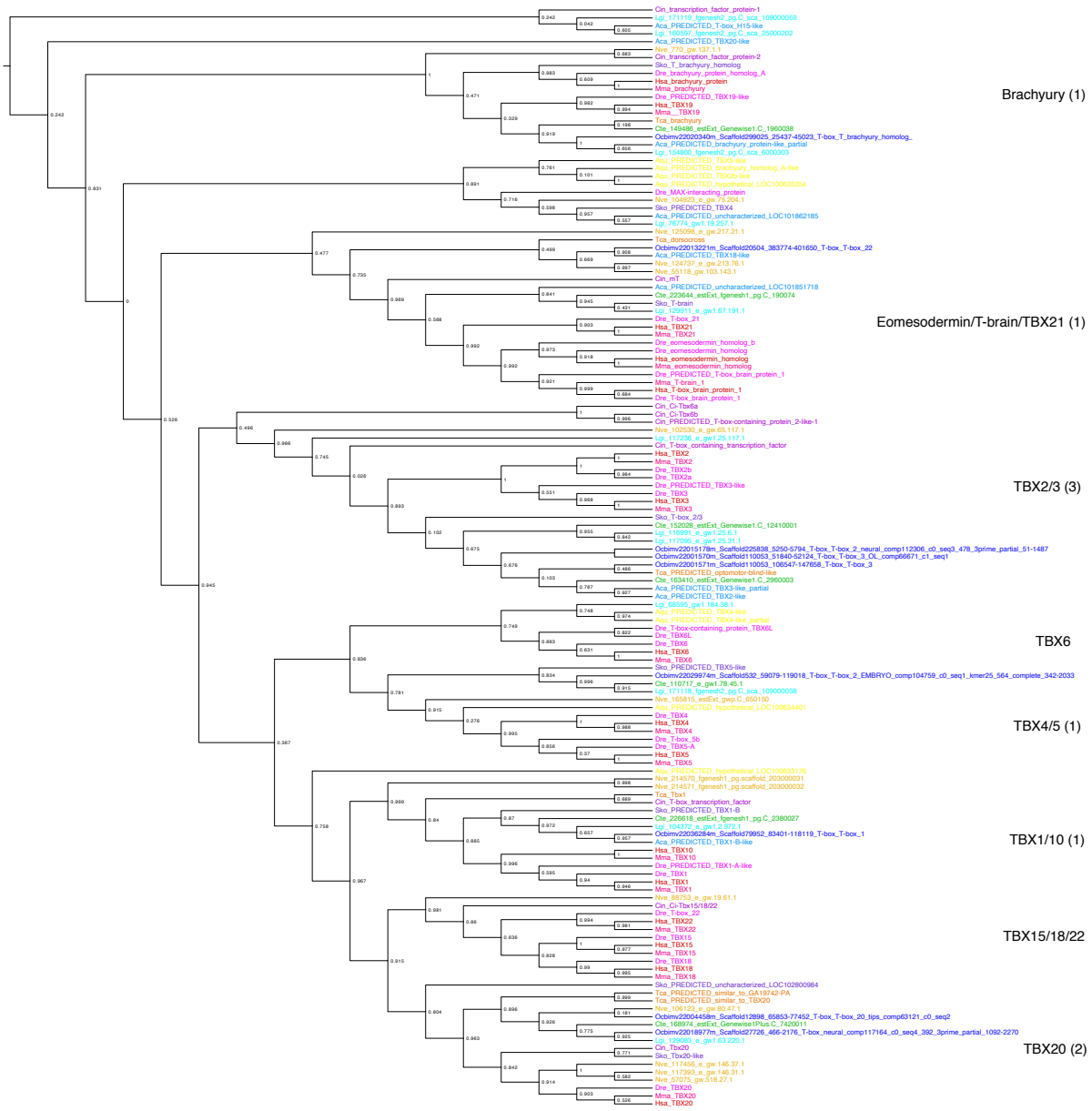


Figure 4.6. Phylogenetic tree of the TBOX gene family in eumetazoans.
The number of octopus orthologs is indicated in parentheses.

Fox transcription factors

Fox genes are found throughout eukaryotes, and are characterized by the presence of a roughly 100 amino acid winged helix DNA-binding domain. There are 43 Fox genes in humans, and they are divided into 19 subfamilies (FOXA to FOXS). We find a single member of the FOXA, B, C, D, F, H, I, J, K, L, N, O, and P subfamilies in the octopus genome (Figure 4.7). While there are lophotrochozoan members of the FoxE and G subfamilies, there was no match in the *O. bimaculoides* genome assembly. There were no lophotrochozoan orthologs for FoxH, and only the only lophotrochozoan FoxQ we retrieved was from *C. teleta*. There are apparently two eumetazoan FoxL subfamilies, each with an octopus ortholog. Subsets of FoxL, FoxC, FoxF, and FoxQ genes are clustered in flies, mammals, and amphioxus, but such clusters are not present in the *O. bimaculoides* genome.

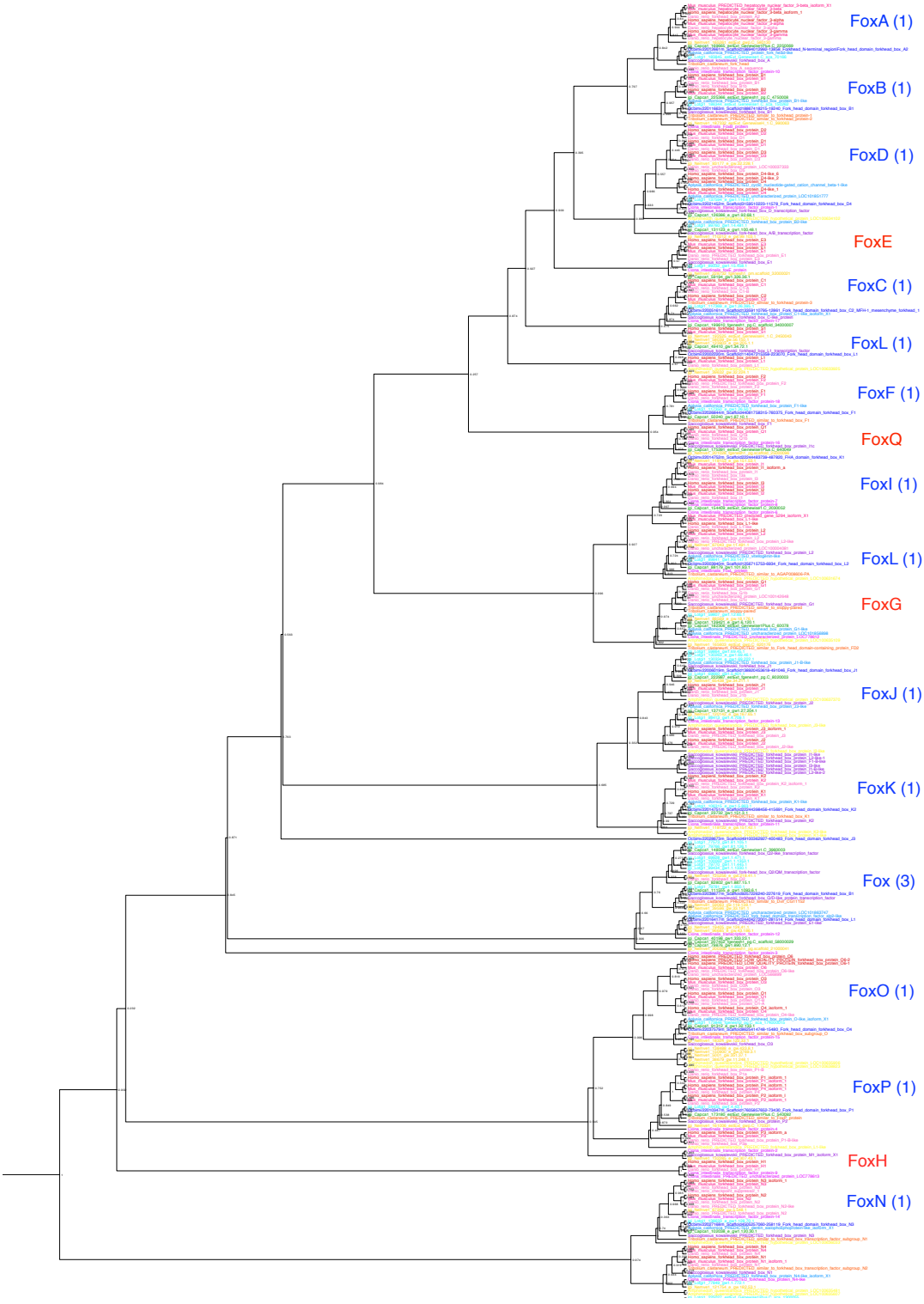


Figure 4.7. Forkhead box transcription factors in metazoans.
Subfamilies without octopus orthologs are indicated in red. The number of octopus orthologs is indicated in parentheses.

High Mobility Group transcription factors

The High Mobility Group (HMG) is a collection of proteins involved in the regulation of transcription, replication, recombination, and DNA repair. The HMG transcription factor family includes the canonical and non-canonical HMG members, as well as a group of proteins with a HMG-box, such as the SRY-box (SOX) and TCF transcription factors. The HMG box is approximately 75 amino acids, and codes for three alpha helices that can bind to, and bend, DNA (Malarkey and Churchill 2012). We recovered single octopus orthologs for many of the HMG-box transcription factor families, including *HMG1, BBX, HMG20, HMGB, MAELSTROM, CAPICUA, TCF, TOX, SOX1/2/3, SOX5/6/13, SOX7/17/18,* and *SOX30* (Figure 4.8). *SOX8/9/10, SOX4/11/12, TOX2/4, UBF,* and *SMARCE* have lophotrochozoan orthologs, but appear absent in octopus. We recovered multiple octopus orthologs for *PMS* and for *TFAM*, as well as two octopus genes containing a HMG box that grouped with other invertebrate sequences but do not have an ortholog in vertebrates.

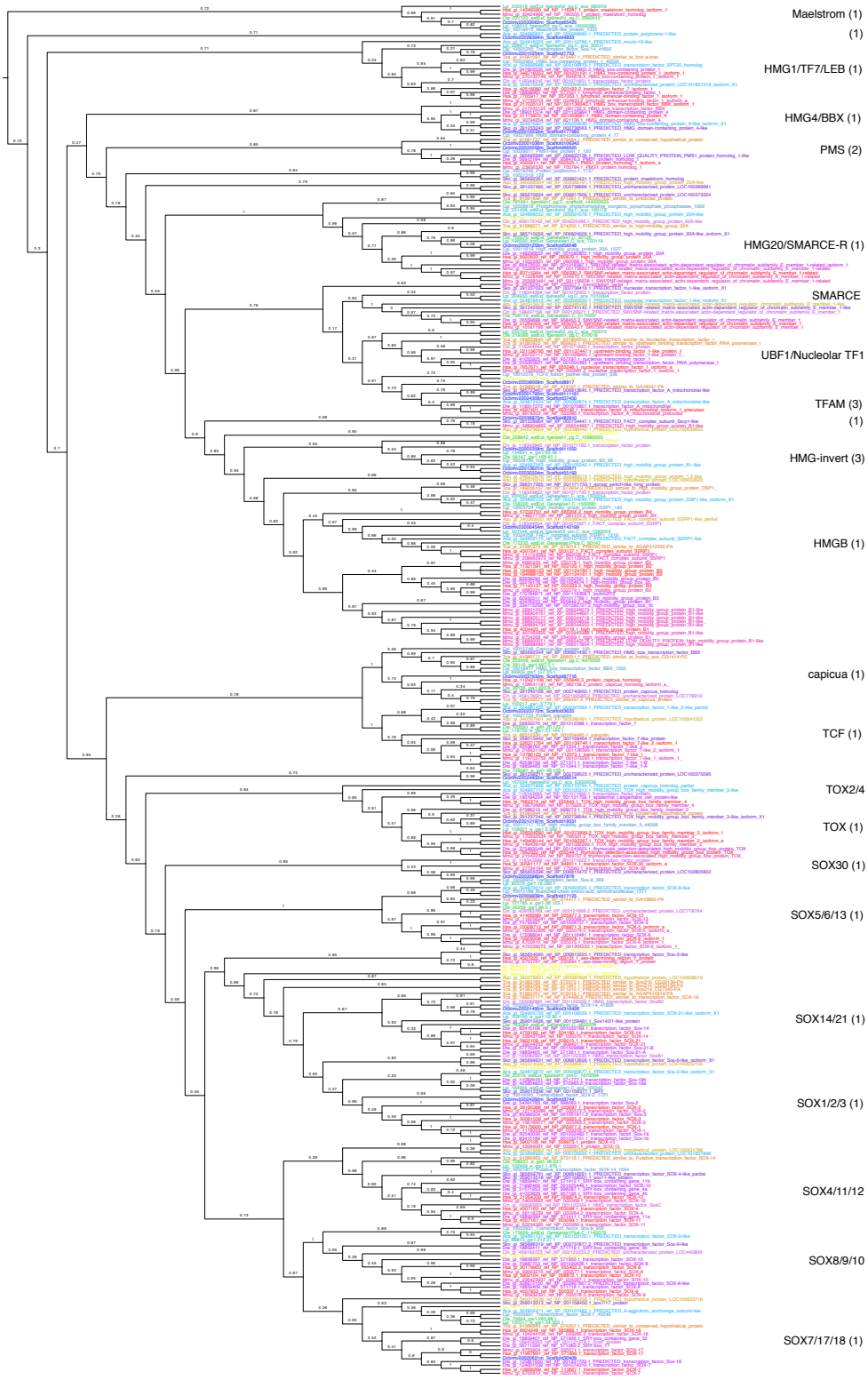


Figure 4.8. Phylogeny of metazoan HMG-box proteins.
The number of octopus orthologs found in each family is indicated in parentheses.

Nuclear Hormone Receptors

Nuclear Hormone Receptors (NHRs) are a family of related transcription factors, many of which are activated by small lipophilic molecules such as steroid hormones, retinoic acid and thyroid hormone (Sever and Glass, 2013). NHRs share a common structure, with a DNA-binding domain, a linker region and a ligand-binding domain, as well as a variable N-terminal region (Escriva et al., 2004). While many NHRs act through identified ligands, there are also a large number of orphan receptors, which do not have known ligands (Sever and Glass, 2013). The NHR superfamily is not found outside of the Metazoa. NHRs can be broken into seven subfamilies (NR0-6): 49 have been identified in humans, and here we describe 40 octopus, 44 snail, and other invertebrate orthologs for many of the NHR subfamilies.

The type I subfamily of NRs (NR1) includes thyroid hormone receptor (THR), retinoic acid receptor (RARs), vitamin D receptor (VDR), ecdysone receptor (EcR), peroxisome-proliferator activated receptor (PPAR), and a number of orphan receptors. The NR1B subfamily of retinoic acid receptors has three human, one snail and two octopus orthologs (Figure 4.9). There are three *PPAR* genes in humans, but only one in *O. bimaculoides* and *L. gigantea*, and none in *D. melanogaster.* We detected a subfamily closely related to the NR1C genes with three octopus members, six in *L. gigantea* and one in *S. kowalevskii* that is not present in vertebrates. NR1D, the human subfamily that includes *NR1D1/ear-1* and *NR1D2*, has one ortholog present in the genomes of *O. bimaculoides, L. gigantea* and *D. melanogaster*. We find one ROR (NR1F1) in and one NR1H/ecdysone receptor *O. bimaculoides* and *L. gigantea*. There are four NR1I (vitamin D receptor subfamily) genes in octopus and seven in *L. gigantea*. We do not recover a thyroid hormone receptor (NR1A) in the *O. bimaculoides* genome, but it is present in the *L. gigantea* and *C. gigas* genomes.

Hepatocyte nuclear factor (HNF4/NR2A), retinoid X receptor (RXR/NR2B), TLX (NR2E), COUP-TF (NR2F), and a number of other orphan receptors are in the Type II subfamily (NR2). We find single representatives of the HNF4/NR2A, RXR/NR2B and NR2C subfamilies in both the *O. bimaculoides* and the *L. gigantea* genomes (Figure 4.9). There are seven octopus and five snail members of the NR2E subfamily, and two octopus and one snail NR2Fs.

The steroid receptors, such as the estrogen (ESR/NR3A), estrogen related (ESRR/NR3B) progesterone (PGR/NR3C), and androgen receptors (AR/NR3C) are in the NR3 subfamily. We find a single member of the NR3A and two members of the NR3B subfamilies in the *O. bimaculoides* and *L. gigantea* genomes (Figure 4.9). We were unable to identify invertebrate orthologs of NR3C members AR or PGR. In addition, there is an octopus gene that groups with a snail gene that branches just outside the rest of the NR3 family. Finally, we find a single representative of NR4, two NR5/FTZF1s, and one NR6, and one NR0B in both *O. bimaculoides* and *L. gigantea.* The NR0 subfamily does not have a DNA binding domain, only the ligand-binding domain that also serves as a dimerization interface.

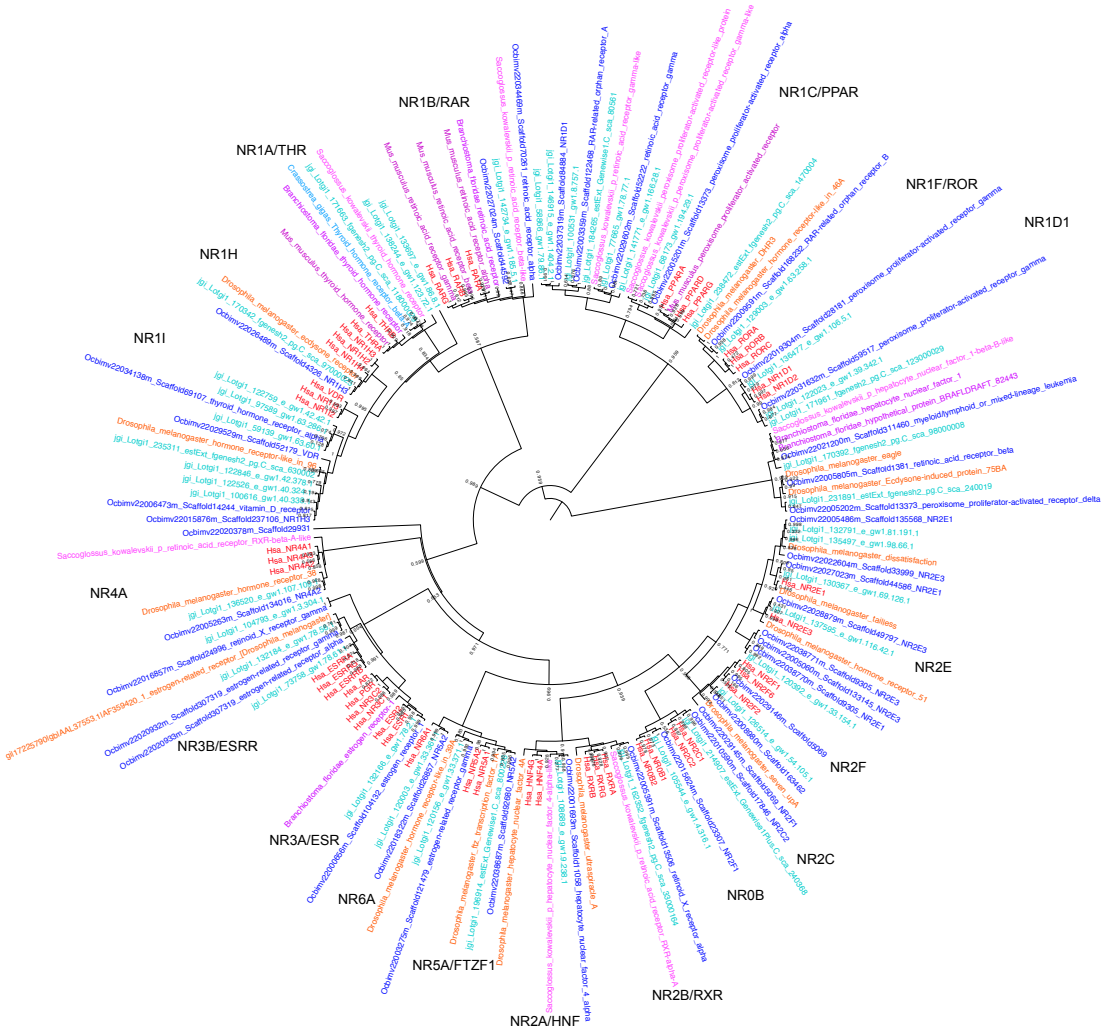


Figure 4.9. Nuclear Hormone Receptors in selected bilaterians.

Helix Loop Helix transcription factors

Basic helix-loop-helix (bHLH) transcription factors share a dimerization motif characterized by two alpha helices connected by a loop. Most members have, in addition, an adjoining sequence of basic residues that binds to DNA. There are 110 members in the human genome, including a number of well-known developmental regulators such as *ACHAETE-SCUTE*, *ATONAL*, *MYOGENIN*, *TWIST*, *SCLERAXIS*, *NEUROGENIN*, *OLIG* and *NEUROD*. For most gene families, including the ones listed above, there is a single ortholog in the octopus genome (Figure 4.10), for a total of 50 octopus bHLH genes. This is in contrast to observations in *D. melanogaster*, which lacks a number of bilaterian bHLH families, including *NEUROD* and *OLIG* (Simionato et al., 2007). Exceptions include the spermatogenesis and oogenesis specific bHLH transcription factor family, which has three octopus members and two in vertebrates, and the hairy/enhancer of split family, which has seven human orthologs, and five to eight in other lophotrochozoans, but appears to only have a single octopus representative. There is a protostome-specific bHLH group with nine members in octopus, and a lophotrochozoan-specific clade related to the achaete-scute family with one octopus ortholog. While there are lophotrochozoan members of the *MUSCULIN*/*TCF21* and *TFAP4* families, no octopus orthologs were identified.

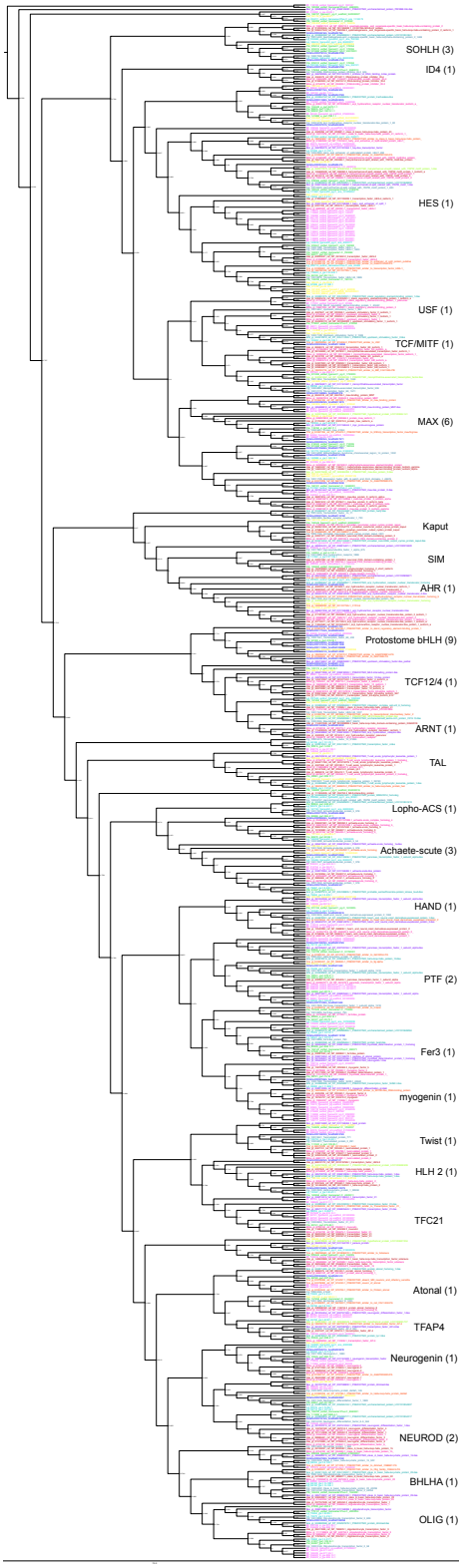


Figure 4.10. Basic Helix-Loop-Helix transcription factors in metazoans.
The number of octopus orthologs for each family is noted in parentheses.

Homeodomain transcription factors

The homeodomain is a highly conserved 60-amino acid helix-turn-helix DNA binding motif encoded by a 180-nucleotide homeobox sequence, and is found in most eukaryotes. Homeodomain transcription factors all have at least one homeodomain, and make up a large superclass with many members. In animals, the homeodomain superclass is divided into the ANTP (including the HOX and NKX homeodomain proteins), CERS, CUT, HNF, LIM, POU, PRD, PROS, SINE, TALE, and ZF classes (Holland). Invertebrate bilaterians have roughly 100 homeodomain proteins, while vertebrates have more than 200 (Table 4.1). We identified 114 homeobox genes in *O. bimaculoides*.

The antennapedia class (ANTP) of homeodomain proteins includes the HOX genes (discussed above) as well as the NKL genes. In octopuses, we find 48 ANTP class homeodomain proteins, including 8 HOX and 2 PARAHOX genes (Figure 4.11). As with the HOX genes, *O. bimaculoides* has a single ortholog of many of the NKL genes, while vertebrates have 2 or more. Exceptions include *IRX2/5* and *NKX3*, where we did not find an octopus ortholog, and Rough, which is present in *O. bimaculoides* and in *T. castaneum,* but appears to be absent in vertebrates.

The PRD class includes transcription factors with a paired domain and a homeodomain (PAX genes excluding Pax1/9), as well as proteins with a closely related homeodomain but without a paired domain (sometimes referred to as PRD-like). We find representatives of *PAX2/5/8, PAX3/7*, and *PAX6* in octopus. We recovered 15 other members of the PRD-like genes, including *OTX*, *PITX*, *GSC*, and *PHOX2* (Figure 4.11).

There are six members of the SINE class of homeodomain transcription factors in humans, corresponding to three invertebrate homologs: *SIX1/2*, *SIX3/6*, and *SIX4/5*. While a *SIX1/2* ortholog has been found in other invertebrates including the longfin inshore squid

Doryteuthis pealeii (Koenig, 2015), we did not recover one in octopus. However, we did find octopus orthologs of *SIX3/6* and *SIX4/5* (Figure 4.11).

Our homeodomain supertree did not fully resolve a single TALE class node: the *IRX* genes branched with the ANTP class and *TGIF* grouped with the ZF homeodomains, while the remaining TALE genes grouped together (Figure 4.11). We find members of the *IRX*, *MEIS*, *MNX*, *PBX*, *PKNOX*, and *TGIF* families. The CUT class also did not resolve together, with *ONECUT* nested in the TALE class, *CUX* on a branch with members of other classes that did not group with their own class, and *SATB* with the ZF class. We find two *ONECUT* and one *CUX*, but no *SATB* orthologs in octopus. The POU class resolved reasonably well, with only part of the *POU6* family sitting apart from the other POU members. We recovered octopus orthologs of *POU2*, *POU3/4*, *POU4*, and *POU6*. A *POU1* ortholog was not detected in any mollusc, though a cnidarian and an annelid sequence branched with the vertebrate *POU1* genes.

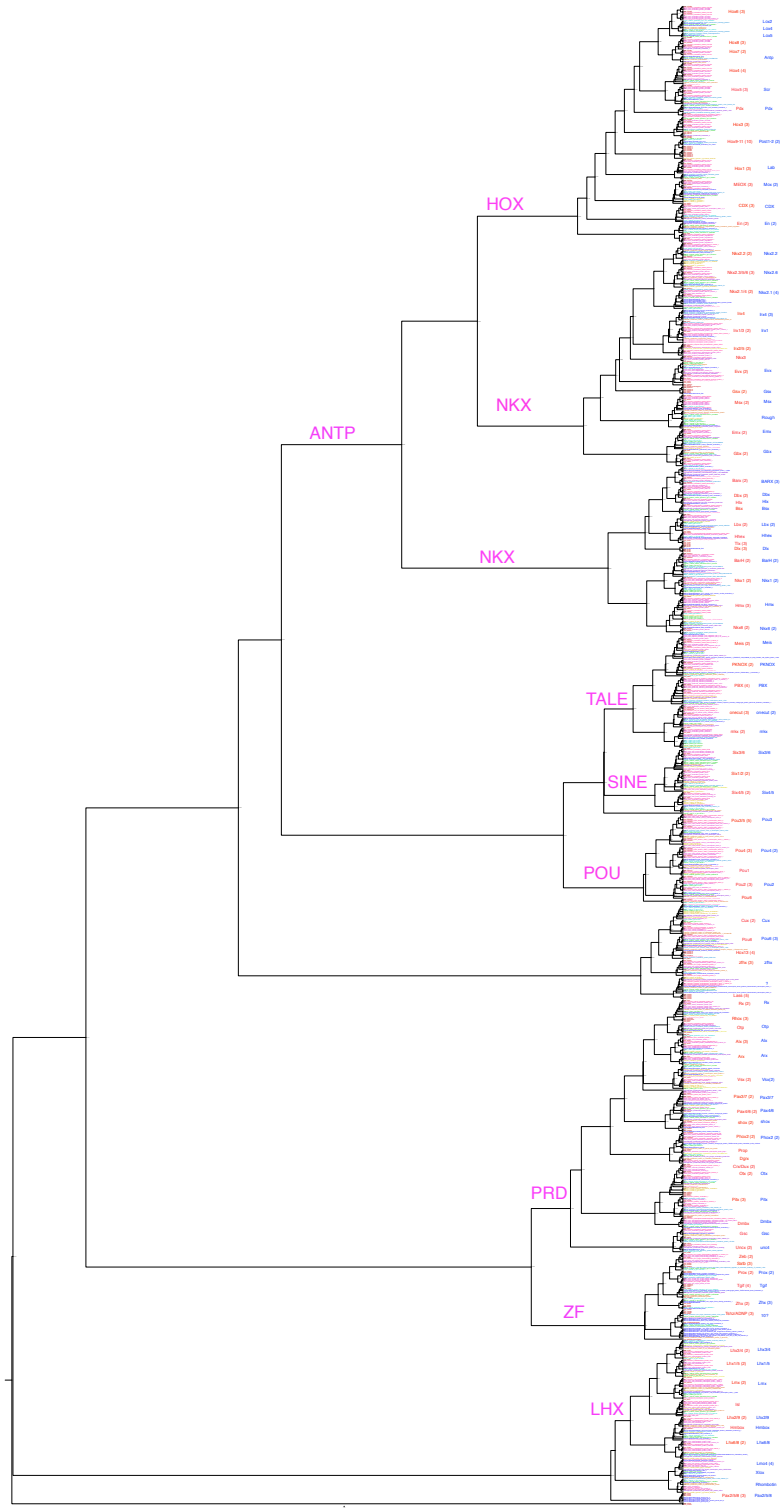


Figure 4.11. Eumetazoan homeodomain transcription factor phylogeny.
Numbers of human orthologs are indicated to the right of the tree in red. The numbers of octopus orthologs are listed in blue. Major subfamilies are indicated on the branches in pink. Octopus does have a PAX1/9 gene, but it is not included here as it does not contain a homeodomain.

Hox genes. Hox genes encode a family of homeodomain transcription factors that play critical roles in animal development, such as specifying anterior-posterior identity. In clades with whole genome duplications in their history, Hox genes are typically retained in multiple copies (Holland et al., 1994; Putnam et al., 2008). We found a single Hox complement in *O. bimaculoides*, consistent with the absence of whole genome duplication in cephalopods and the detection of a single Hox complement in the bobtail squid *Euprymna scolopes* by PCR(Callaerts et al., 2002). We identified eight Hox genes in *O. bimaculoides*: *LAB*, *SCR*, *LOX5*, *ANTP*, *LOX2*, *LOX4*, *POST2* and *POST1*, as well as the ParaHox genes *GSX* and *CDX* (Figures 4.11-4.12). We did not detect genes resembling Hox paralog groups (PG) 2, 3, or 4. While no PG2 (*PB*) Hox gene was identified in either *E. scolopes* (Callaerts et al., 2002) or *S. officinalis* (Pernice et al., 2006), one has been described for *Nautilus pompilius* (Iijima, 2006), which diverged from octopus ~400 mya (Kroger et al., 2011). Genes for *ZEN* (PG3) and *DFD* (PG4) were identified in *E. scolopes,* while *LOX2*, a lophotrochozoan-specific central class gene, was not found (Callaerts et al., 2002). These findings may indicate that there are different patterns of Hox gene retention or loss across cephalopods.

In many bilaterians, the Hox genes cluster in the genome (Figure 4.12), and the order of genes in the cluster reflects the timing and pattern of their expression along the anterior-posterior axis in developing embryos; this feature is called collinearity. Clustering has been described in a wide range of animals, including at least one mollusc, *L. gigantea* (Simakov et al., 2013). Remarkably, octopus Hox genes are not organized into a cluster as in other bilaterian genomes, but are atomized, with no two Hox genes found on the same scaffold (Figure 4.12). Even if the two smallest scaffolds (those containing *ANTP* and *LOX2*) were linked, the shortest distance between these Hox genes would be at least 50kb if they were transcribed on different strands,

and 100kb if they were transcribed on the same strand, as documented for Hox clusters. With the exception of the short scaffold containing *ANTP*, all of the scaffolds containing Hox genes have a number of short open reading frames outside of the Hox locus.

All eight *O. bimaculoides* Hox genes are detected in the transcriptome of stage 15 embryos, as would be expected for developmental regulatory genes. Hox genes were also expressed in adult tissues, particularly the subesophageal brain, axial nerve cord, skin and suckers (Figure 4.13).

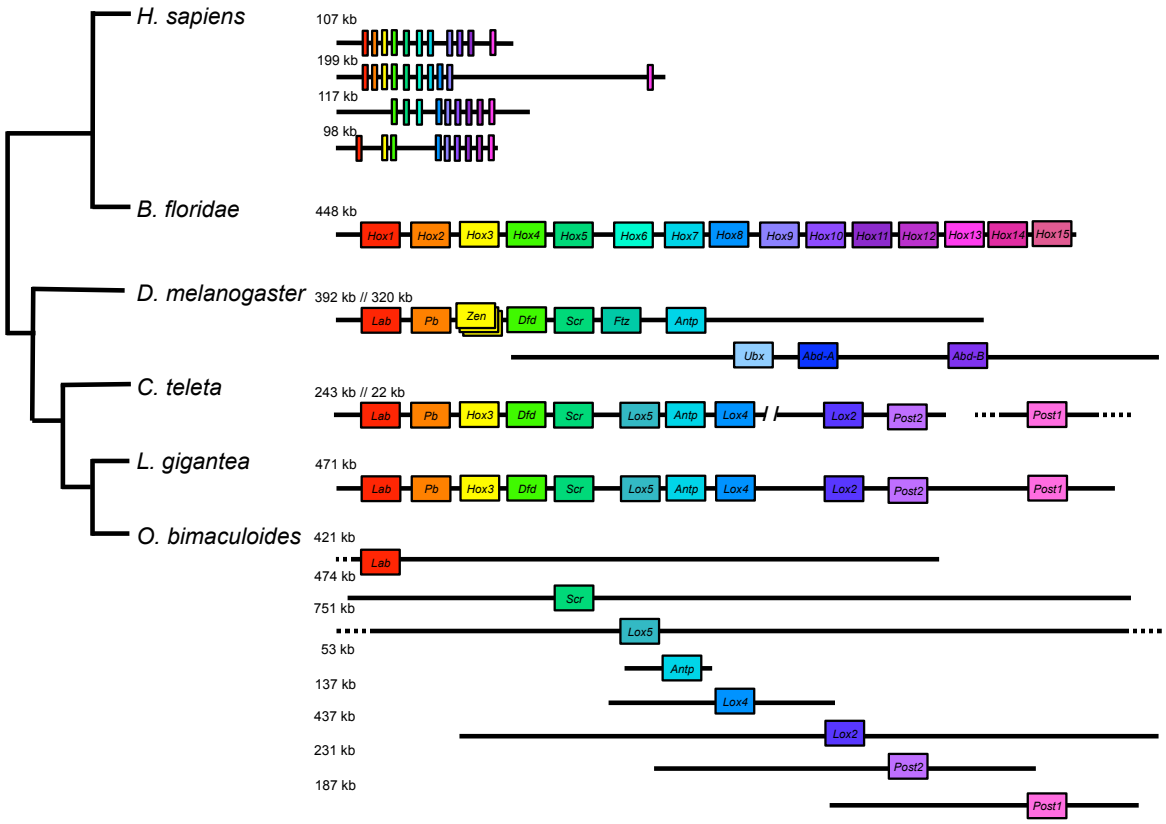


Figure 4.12. Bilaterian Hox clusters.
At the top, the four compact Hox clusters of *H. sapiens* and the single *B. floridae* cluster are depicted. The *D. melanogaster* Hox complex is split into two clusters. We included genes in the *D. melanogaster* locus that are homologues of Hox genes but have lost their homeotic function, such as *fushi tarazu (ftz)*, *bicoid*, *zen* and *zen2* (the latter three are represented as overlapping boxes). Hox genes in *C. teleta* are found on three scaffolds (Simakov et al., 2013). *L. gigantea* has a single cluster with the full known lophotrochozoan gene complement. In *O. bimaculoides* many of the scaffolds are several hundred kb in length, and no two Hox genes are on the same scaffold. *Continued on next page.*

Figure 4.12 *continued.* The positions of *O. bimaculoides* genes approximate their locations on scaffolds. Dashed lines indicate that the scaffold continues beyond what is shown. Scaffold length is depicted to scale with size noted on the left. Genes are positioned to illustrate orthology, which is also highlighted by color.

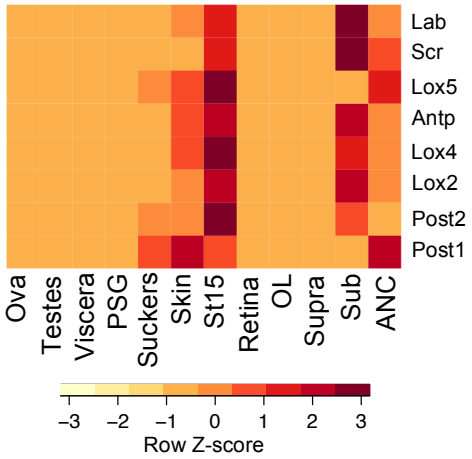


Figure 4.13. Expression of *O. bimaculoides Hox* genes detected in development, and in adult skin, suckers and neural tissues.
Absence of supraesophageal *Hox* gene expression was also reported for *E. scolopes* (Lee et al., 2003). Cells are colored according to number of standard deviations from the mean expression level. Abbreviations: posterior salivary gland, PSG; embryonic stage 15, St15; optic lobe, OL; supraesophageal brain, Supra; subesophageal brain, Sub; axial nerve cord, ANC.

C2H2-ZNFs

The zinc finger (ZNF) domains are small peptide domains that bind zinc ions at Cys and His residues to form "finger-like" protrusions (Iuchi, 2001). The best characterized and most prevalent of the ZNFs, the Cys2His2 type (C2H2), coordinates a single zinc ion to stabilize its secondary structure. Two or more fingers, connected by short linker sequences, are needed to bind to unique DNA or RNA sequences. Binding specificity, or the protein's "fingerprint," depends on both the identity of the ZNFs and that of the linker sequences (Liu et al., 2014). C2H2-ZNFs have been implicated in many cellular processes, including cell fate determination and early development (Liu et al., 2014). However, much remains unclear about the function and evolution of these proteins.

ZNFs are prevalent among eukaryotes, forming several large superfamilies throughout the animal kingdom. Humans have over 700 C2H2-ZNF proteins (Figure 4.14), making up 3% of the protein-coding genes in the genome (Klug, 2010). Clade-specific ZNF gene expansions have been described in a number of vertebrates, raising the interesting possibility that they function in species-specific innovations (Liu et al., 2014; Shannon et al., 2003). For example, the KRAB ZNF gene family descended from a single ancestral sequence and has expanded dramatically in many tetrapod clades to become the largest family of transcription factors in humans, with roughly 400 members. A subset of these KRAB ZNFs is the result of independent expansion in primates (Nowick et al., 2010).

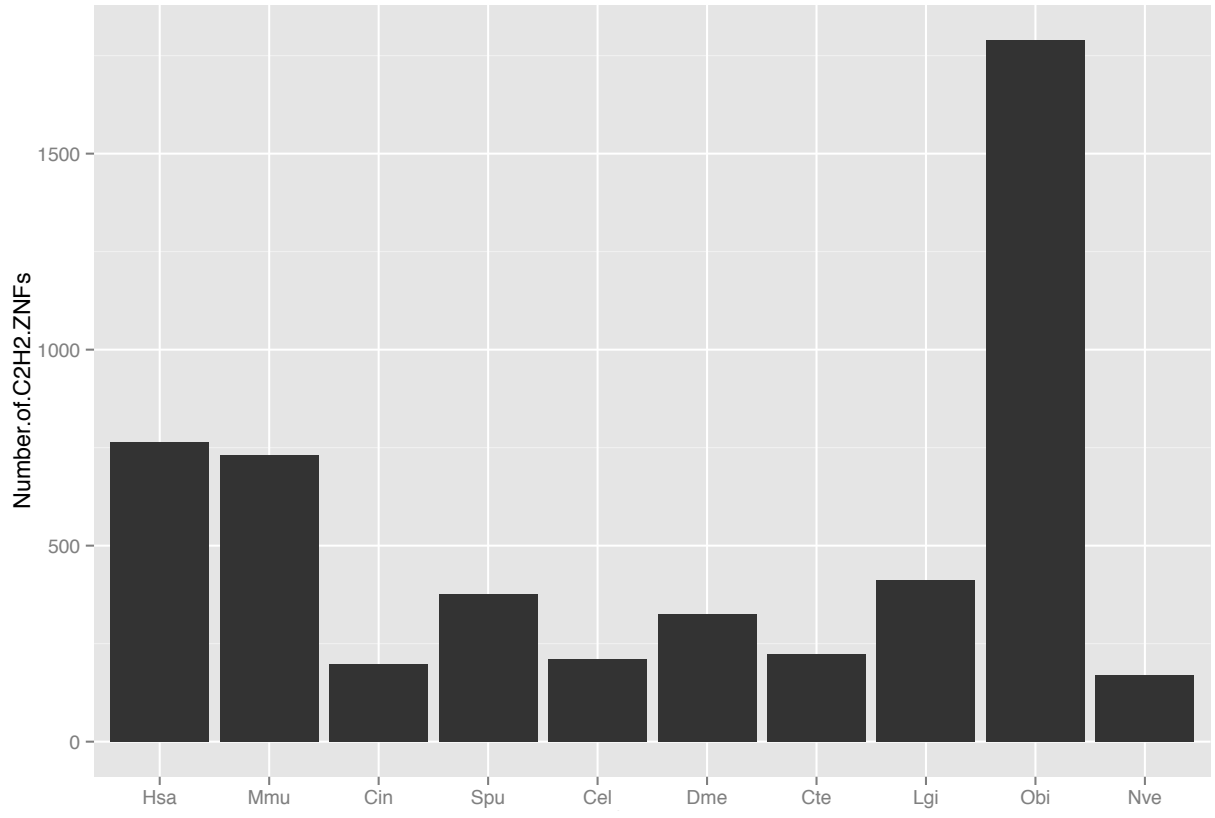


Figure 4.14. C2H2 genes in select eumetazoan species.
While most invertebrate species investigated have far fewer C2H2-ZNFs than do mammals, *O. bimaculoides* presents a dramatic expansion with almost 1,800 individual C2H2 genes.

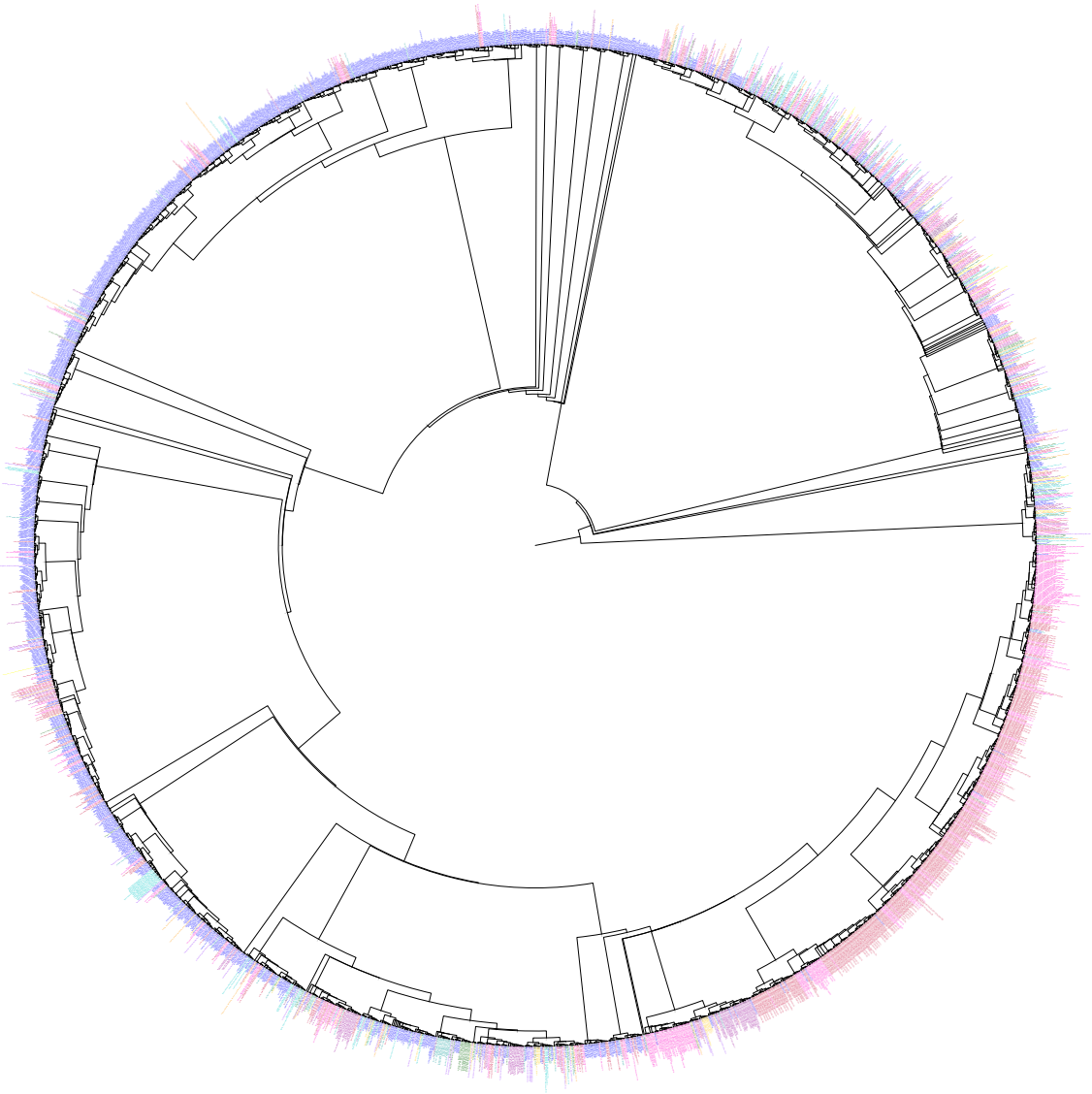


Figure 4.15. Eumetazoan C2H2 phylogenetic tree.
A eumetazoan C2H2 phylogenetic tree was built using 1,452 sequences from octopus, 539 from human, 388 from mouse, 159 from zebrafish, 7 from *A. queenslandica*, 88 from *A. californica*, 24 from *C. intestinalis*, 65 from *S. kowalevskii*, 44 from *T. castaneum*, 39 from *N. vectensis*, 87 from *L. gigantea*, and 72 from *C. teleta*. The alignments for this tree were made with CLUSTALO and the tree was built with FastTree.

We found 1,790 C2H2-ZNF proteins in the genome of *O. bimaculoides* (Table 4.1; Figure 4.15) Octopus ZNFs are clustered along scaffolds but transcribed in different directions (Figure 4.16), which has also been found in vertebrate ZNF clusters (Shannon et al., 2003). Whereas the human ZNF family is dominated by an expansion of the KRAB-type ZNFs, the octopus ZNF complement appears to be characterized by the elaboration of multiple C2H2 types. All invertebrates we investigated, from well-characterized genetic models like *D. melanogaster* and *C. elegans* to other lophotrochozoans, only have a complement of, at most, a few hundred C2H2-ZNF genes, indicating that the expansion in octopuses is unusually dramatic.

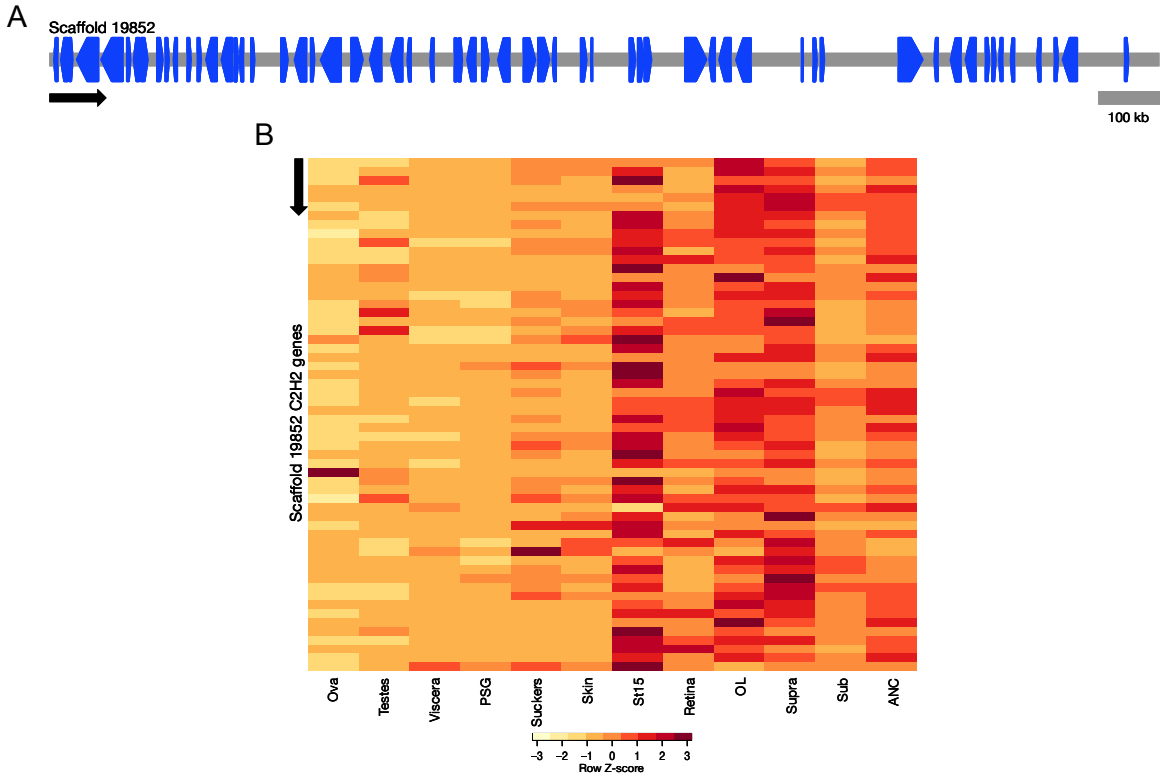


Figure 4.16. Genomic organization of the largest octopus C2H2 cluster.
(A) Scaffold 19852 contains 58 C2H2 genes that are transcribed in different directions. (B) Expression profile of C2H2 genes along Scaffold 19852 in 12 octopus transcriptomes. Neural and developmental transcriptomes show high levels of expression for a majority of these C2H2 genes. In **a** and **b**, arrow denotes scaffold orientation.

ZNF proteins require at least two or three finger domains to bind DNA selectively; ZNFs with multiple C2H2 arrays may have multiple binding sites mediated by different finger sets (Iuchi, 2001). We identified 1,790 octopus C2H2s with three or more C2H2 domains (Figure 4.15-17). Some of these C2H2-ZNF genes also contained other ZNF motifs such as Elf1 and FYVE. 52% (931/1,790) of these genes contain over 20 C2H2 domains (Figure 4.17). We found the helix capping linker sequence, TGEKP, to be present in octopus C2H2-ZNF sequences. An analysis of the C2H2 gene sequences did not reveal any octopus-specific linker sequences or effector domains shared among multiple genes.

Among this large and dynamic family of transcription factors are a number of highly conserved genes that play critical roles in bilaterian development. The *GLI* genes, which include four orthologs in humans (*GLI1/2/3/4)* and *cubitus interruptus (Ci)* in *D. melanogaster*, are involved in the hedgehog signaling pathway. We identify two *GLI* octopus orthologs. The related *GLIS* and *ZIC* genes are represented by three and five orthologs in the human genome. Our analysis identifies a *GLIS1/2* and a *GLIS3* ortholog, and two octopus *ZIC1/2/3/4/5* genes, though other protostomes only have one *ZIC* gene (Figure 4.15). *SNAIL* genes have been implicated in mesoderm formation and epithelial-to-mesenchyme transitions in a number of taxa (Barrallo-Gimeno and Nieto, 2005). The human genome contains three *SNAIL* orthologs, *SNAIL1, SNAIL2* and *SNAIL3.* We find two *SNAIL1/2/3* genes in *O. bimaculoides* and *L. gigantea*, and one in *C. teleta*. The *KROX/EGR* (early growth response) genes did not resolve well on this tree, although single orthologs of *KROX24/EGR1* and *KROX20/EGR/2/3* were identified. Overall, these highly conserved C2H2 family members appear to not be expanded in the octopus genome.

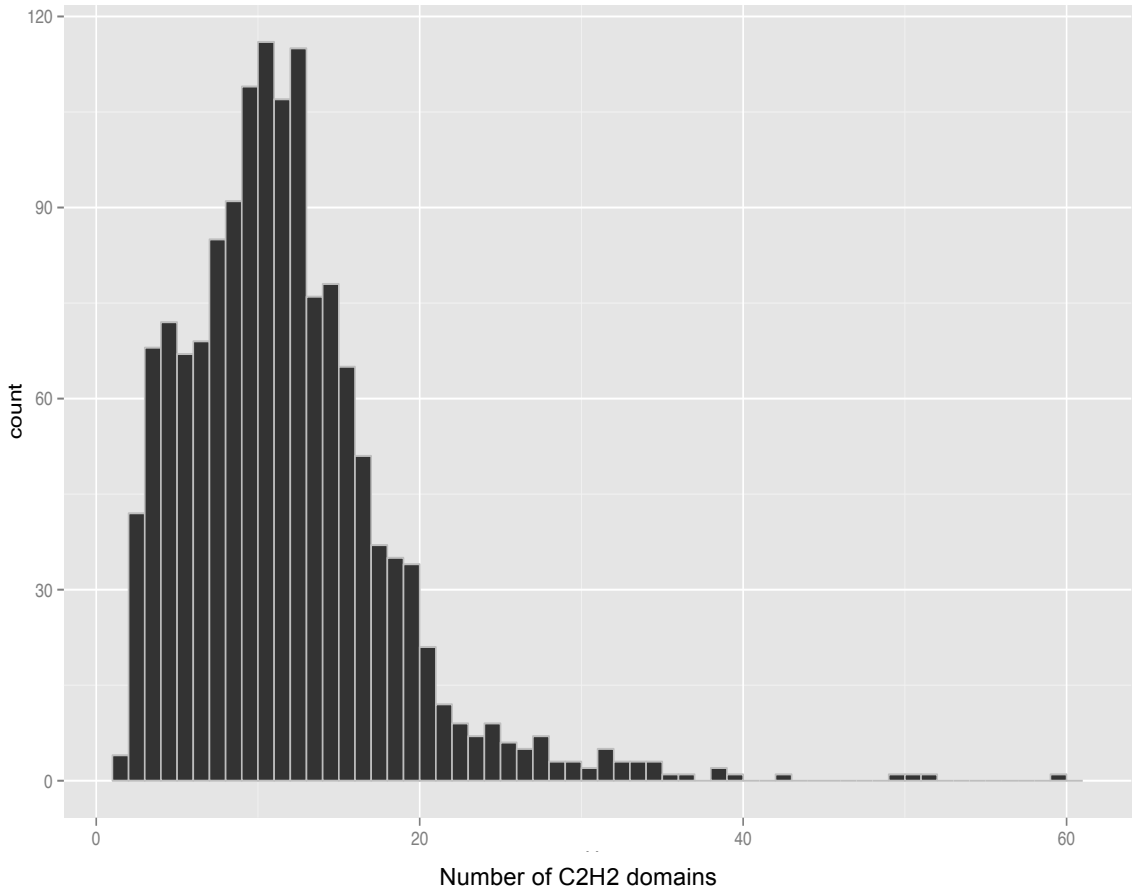


Figure 4.17. Distribution of C2H2 Domains in octopus ZNFs.
Octopus ZNF genes have between 2 and 60 C2H2 domains.

Of the 1,790 octopus C2H2 genes, 1,429 were confirmed as expressed in our stage 15 and adult tissue RNA-Seq dataset. Transcriptome analyses revealed strong enrichment of C2H2 gene expression in neural tissues, especially the optic lobes, the supraesophageal brain, and axial nerve cord (Figure 4.15). Our data support the long-standing idea that ZNFs function in neural complexity and development (Layden et al., 2010; Liu et al., 2014). The optic lobes and supraesophageal brain are major centers of sensory processing, learning and memory. The axial nerve cords of the arms together contain nearly 350 million cells and represent a unique octopus neural innovation (Young, 1971). This expression profile is consistent with the observation that a greater number of C2H2-ZNFs is associated with greater neural complexity in vertebrate species

(Vinogradov, 2013). We also found C2H2 expression in St15 embryos. High activity of the unusually large C2H2 repertoire during embryogenesis may indicate a role in the development of cephalopod-specific innovations. The octopus C2H2-ZNF complement is one of the largest gene family expansions in the animal kingdom, and our evidence lends support to the importance of these genes in conserved transcriptional pathways.

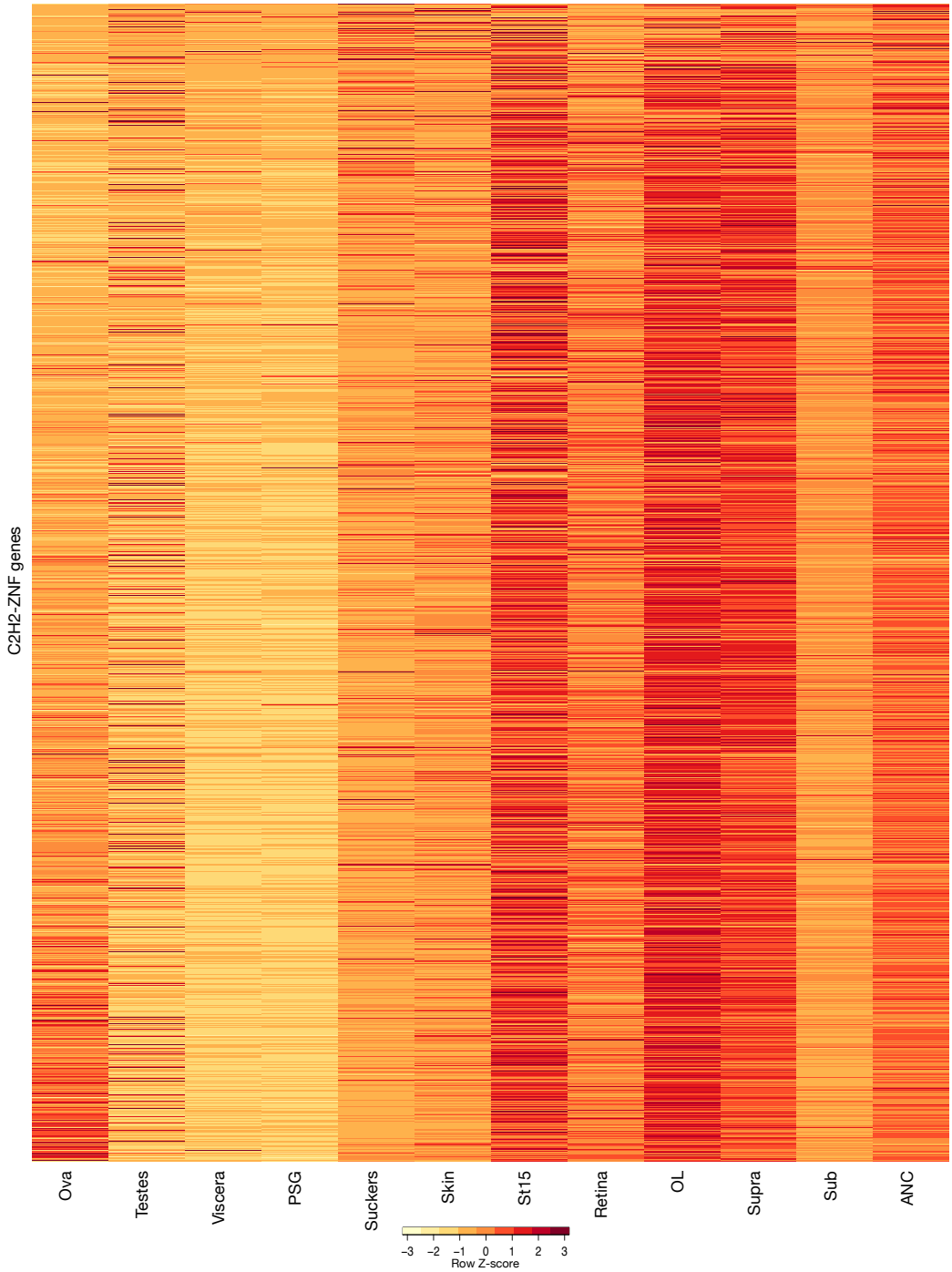


Figure 4.18. Expression of 1,429 C2H2-containing genes in 12 octopus transcriptomes. Cells are colored according to number of standard deviations from the mean expression level.

Protocadherins

Protocadherins are cell adhesion molecules of the cadherin superfamily. In vertebrates, they are the largest subgroup, with more than 60 isoforms in humans and mice (Frank and Kemler, 2002). Protocadherins share a structure of 6 or 7 extracellular cadherin repeats (EC), a single transmembrane region (TM), and a cytoplasmic domain (CD) (Hulpiau and van Roy, 2009; Morishita and Yagi, 2007) (Figure 4.19). In cell culture experiments, protocadherins have been shown to interact promiscuously to form tetramers within the same cell (in *cis*) while the extracellular (EC) domain tetramers have been shown to mediate strictly homophilic intercellular binding (in *trans*) (Schreiner and Weiner, 2010). Given these combinatorics, the combined EC domains of the tetramer potentially allow for hundreds of thousands of specific interactions. As protocadherins generate combinatorial complexity at the cell surface and are predominantly expressed in the nervous system, this gene family may provide a molecular basis for the circuit diversity and specificity of the vertebrate neuronal circuitry, analogous to the role proposed for DSCAM in flies (Chen and Maniatis, 2013). Genetic studies in mice have suggested that protocadherins may be important in dendritic patterning in cortical neurons and neurite self-avoidance in Purkinje and starburst amacrine cells (Lefebvre et al., 2012).

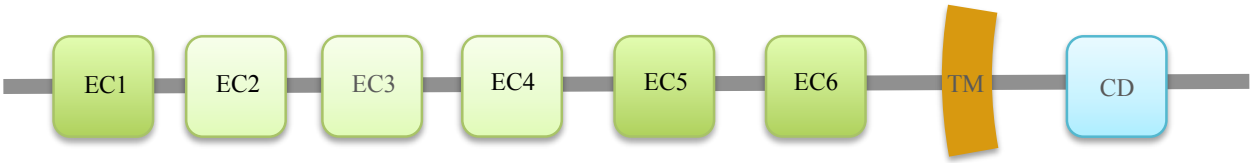


Figure 4.19. Structure of a typical protocadherin.
Protocadherin proteins have 6-7 extracellular cadherin domains (EC) and a transmembrane domain (TM). Some vertebrate protocadherins also have a highly conserved cytoplasmic domain (CD).

Of the roughly 60 protocadherins present in mammalian genomes, more than 50 are found on a single chromosome in three clusters: α, β, and γ. α and γ protocadherins are generated by alternative splicing. Each gene is made up of one variable exon (containing the extracellular ECs, TM, and part of the intracellular domain) and three constant exons that are shared by all genes in the cluster (Chen and Maniatis, 2013). This intron-exon structure, with the ECs and TM being encoded by one or two large exons, is found in many of the non-clustered protocadherins as well (Vanhalst et al., 2005).

Until recently, protocadherins were thought to be a vertebrate innovation. While they are absent from the genomes of the ecdysozoan model organisms *D. melanogaster* and *C. elegans*, solitary protocadherins have previously been identified in *A. californica* and *N. vectensis*, indicating that this family predates the Bilateria (Hulpiau and van Roy, 2011). While this gene was apparently lost in the lineage leading to the Ecdysozoa, here we find a substantial number in lophotrochozoan genomes, including *O. bimaculoides*, *L. gigantea*, *C. gigas,* and *C. teleta* (Figure 4.20).

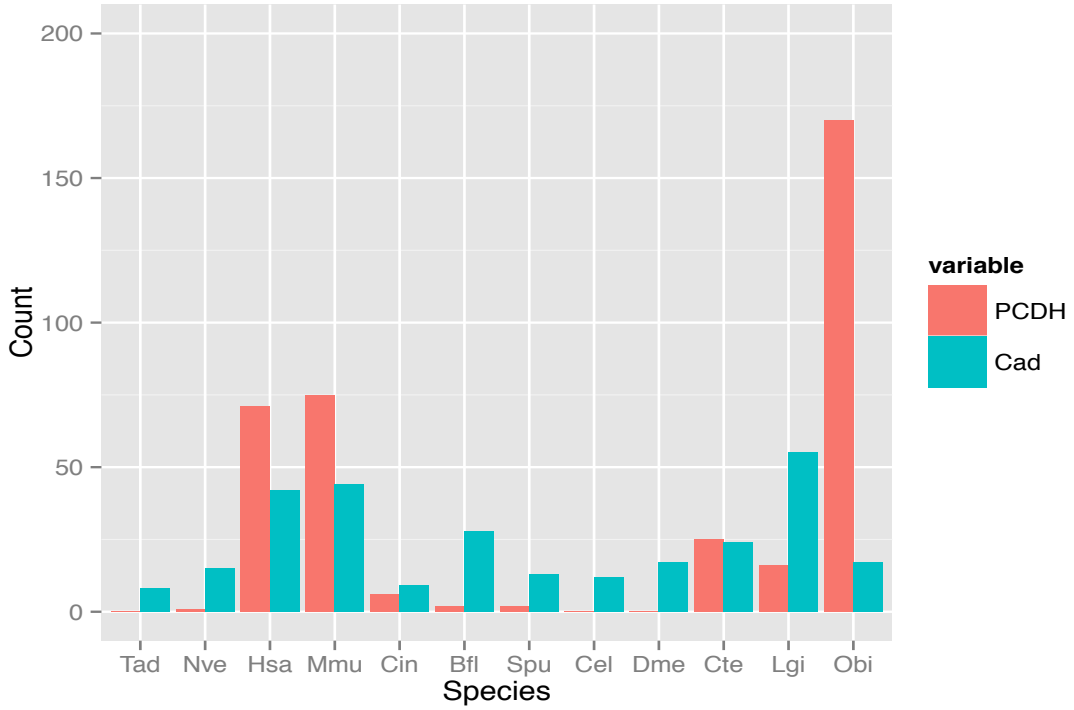


Figure 4.20. Distribution of cadherins and protocadherins across metazoans. Numbers of protocadherins and other cadherins identified in the genomes of *H. sapiens, M. musculus, C. intestinalis, B. floridae, S. purpuratus, C. elegans, D. melanogaster, C. teleta, L. gigantea, O. bimaculoides, N. vectensis,* and *T. adhaerens.*

These lophotrochozoan protocadherins group with vertebrate protocadherins on a cadherin superfamily tree with separate, clade-specific expansions in vertebrates, annelids, molluscs, squid, and octopuses (Figure 4.21-22). In annelids, protocadherins from different species interdigitate on the tree, suggesting the protocadherins expanded prior to the divergence of the polychaetes. Snail and oyster protocadherins also interdigitate, which indicates that the gene family expanded before the divergence of the bivalves and the gastropods, but independently from cephalopods and annelids. The topology of this tree also supports the affinity of the bivalves and the gastropods, as proposed by Smith et al. (2011) and Kocot et al. (2011).

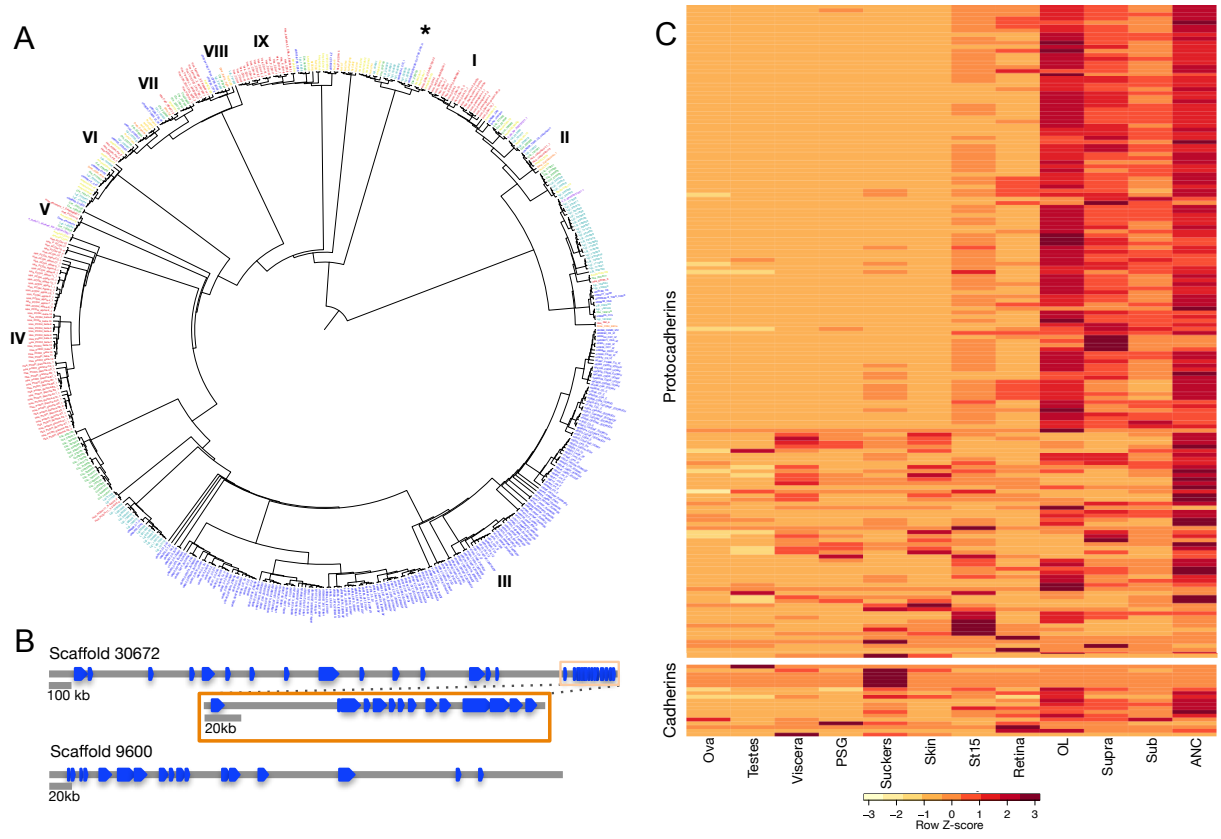


Figure 4.21. Protocadherin expansion in octopus.
(A) Phylogenetic tree of cadherin genes in Hsa (red), Dme (orange), *Nematostella vectensis* (mustard yellow), *Amphimedon queenslandica* (yellow), Cte (green), Lgi (teal), Obi (blue), and *Saccoglossus kowalevskii* (purple). I, Type I classical cadherins; II, calsyntenins; III, octopus protocadherin expansion (168 genes); IV, human protocadherin expansion (58 genes); V, dachsous; VI, fat-like; VII, fat; VIII, CELSR; IX, Type II classical cadherins. (B) Scaffold 30672 and Scaffold 9600 contain the two largest clusters of protocadherins, with 31 and 17, respectively. Clustered protocadherins vary greatly in genomic span and are oriented in head-to-tail fashion along each scaffold. (C) Expression profiles of 161 protocadherins and 19 cadherins in 12 octopus tissues; not shown are the 7 protocadherins that were not detected in the tissues sampled. Cells are colored according to number of standard deviations from the mean expression level. Protocadherins have high expression in neural tissues. Cadherins generally show a similar expression pattern, with the exception of a group of sucker-specific cadherins. A novel cadherin with over 80 extracellular cadherin domains found in octopus and *C. teleta* is denoted with an asterisk in (A).

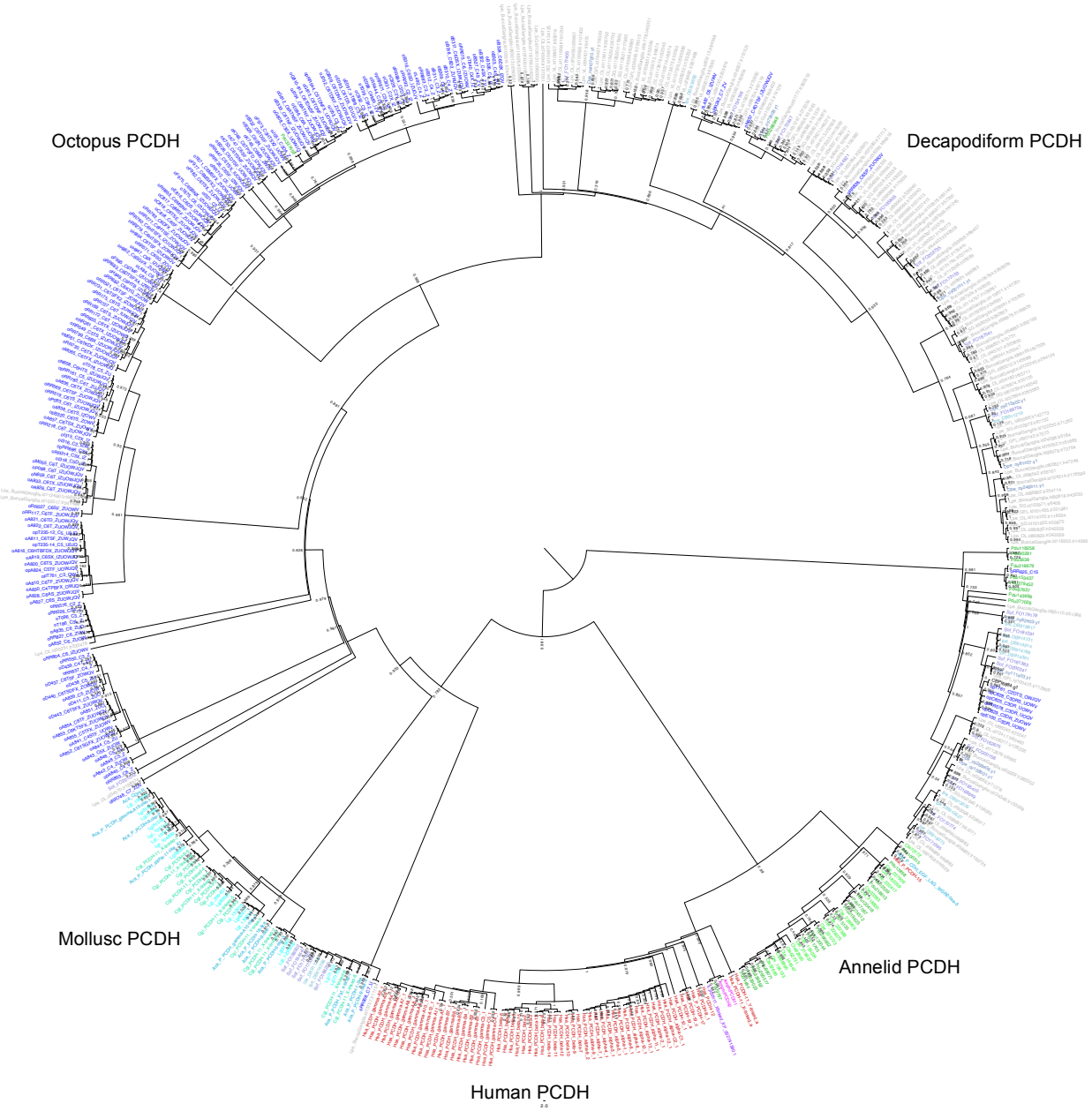


Figure 4.22. Metazoan protocadherin expansions.
Phylogenetic tree of protocadherin genes in *H. sapiens* (red), *C. teleta* (dark green), *P. dumerilii* (light green), *L. gigantea* (teal), *A. californica* (sky blue), *C. gigas* (turquoise) *O. bimaculoides* (blue), *D. pealeii* (grey), *S. officinalis* (slate), and *I. paradoxus* (light blue-grey). We assigned the 15 Scaffolds with 3+ PCDHs an alphabetic name from A-P (skipping O for clarity). Protocadherins found in the transcriptome but not identified in the genome are designated T. Protocadherins on scaffolds with 2 or fewer PCDHs are designated RR.

The expansion of the protocadherins in cephalopods is particularly striking: we identified 168 protocadherins in the *O. bimaculoides* genome, more than twice the number found in mammals. The vast majority of these genes are contained in a single clade-specific expansion, with many of them arranged in clusters in the genome (Figure 4.21). The three largest scaffolds have 31, 17, and 10 protocadherins clustered together, respectively, and at least 25 other scaffolds have two or more protocadherins. Given the fragmented nature of the genome assembly, the number of clustered cadherins is likely much larger than what we describe here. Protocadherins that are clustered on the same scaffold are transcribed in the same direction (Figure 4.21), and many have a similar intron-exon structure to that found in vertebrates: the first exon contains the ECs, TM, and part of the intracellular domain, and precedes three smaller exons. We were unable to find evidence of a splicing mechanism similar to the one found in vertebrates in our *de novo* assembled transcriptomes.

The human γ protocadherin cluster contains 22 genes. If all of these genes can assemble into tetramers in *cis*, they could produce more than 230,000 specific molecular interfaces (Schreiner and Weiner, 2010), which is far more than the roughly 19,008 possible extracellular domains for *Drosophila* DSCAM (Zipursky and Sanes, 2010). If the octopus protocadherins assemble into tetramers, then the largest cluster (Scaffold 30672) alone could generate nearly one million different tetramers. Together, all octopus protocadherins could generate hundreds of millions of different tetramers, an order of magnitude greater than the 25 million tetramers that could be produced by 71 human protocadherins. As the *O. bimaculoides* protocadherins are predominantly expressed in nervous tissues (Figure 4.21), this gene family could play a general role in providing a diverse molecular substrate for the development of large, complex nervous systems.

We also identified 155 protocadherins in the transcriptomes of the longfin inshore squid, *Doryteuthis pealeii* (Brown et al., 2014), as well as a number of other decapodiform protocadherins from cephalopod ESTs deposited in GenBank. Maximum likelihood trees show segregated octopus and decapodiform protocadherin expansions (Figure 4.22). This finding may reflect a massive expansion in parallel after the two major clades of coleoid cephalopods diverged. Alternatively, it may be a signature of concerted evolution.

One mechanism of concerted evolution is gene conversion, which has been described for vertebrate protocadherins, largely at the 3' end of the sequence: in EC5, EC6, and in the CD (Noonan et al., 2004). In contrast to this vertebrate arrangement, trees of **each** of the EC domains show that the octopus and decapodiform expansions remain largely distinct (Figures 4.24-26). Interestingly, a number of protocadherins found on the same scaffold cluster together in these trees (asterisks, Figures 4.24-26), which is also seen in trees based on full-length sequence (Figures 4.21-22). Alignments demonstrated regions of these sequences with few nucleotide and amino acid differences. For example, only ~4% of the nucleotides differ over a 560nt stretch across EC4 and EC5 of 14 genes found on Scaffold 30672 (Figure 4.23). Other regions of sequence, particularly in EC2 and EC3, show much greater sequence variation. The finding that protocadherins with highly similar sequences are located in clusters is consistent with a model of gene conversion. However, as each of these genes have different variations at the nucleotide level, it could also indicate that these are the result of recent duplications.

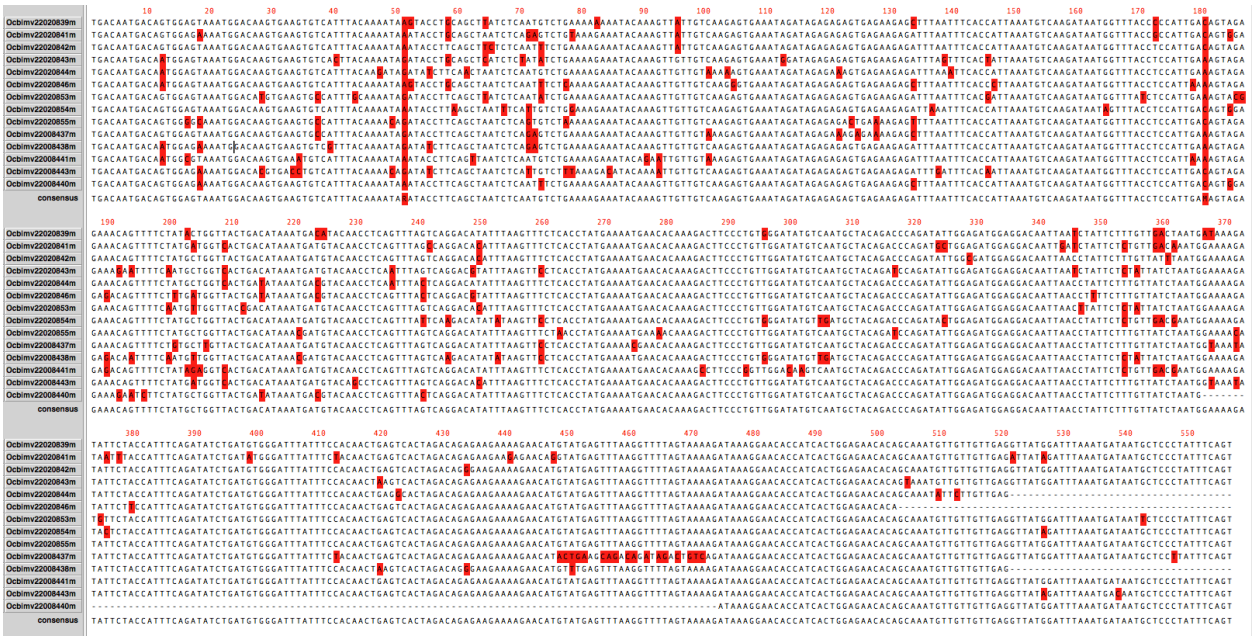


Figure 4.23. Multiple sequence alignment of EC4 and EC5 of 14 protocadherins located on Scaffold 30672.

Nucleotides marked in red do not match the consensus sequence.

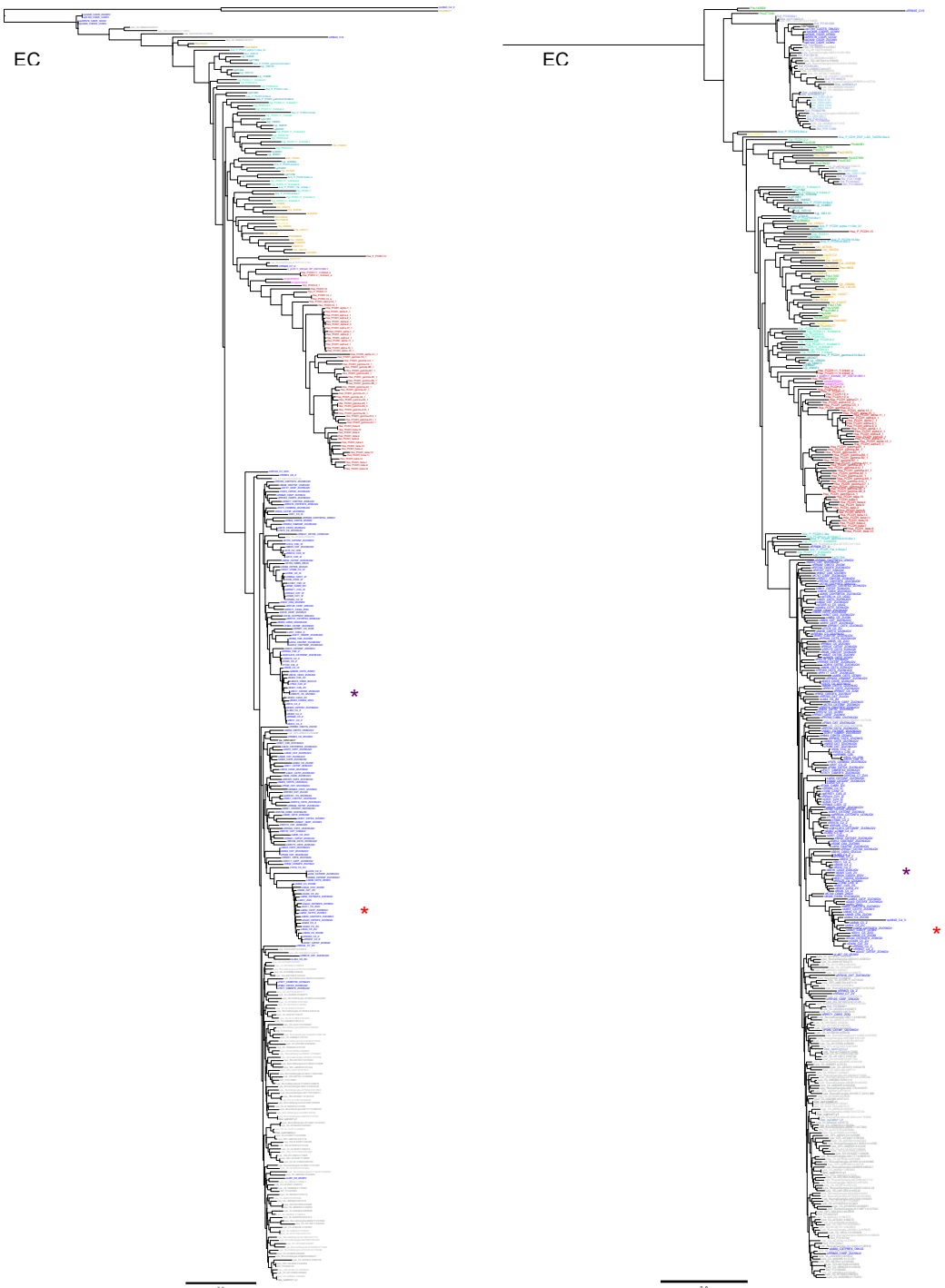


Figure 4.24. Maximum likelihood (ML) trees of EC1 and EC2 domains of metazoan protocadherins.
Phylogenetic trees of protocadherin EC1 and EC2 in *H. sapiens* (red), *C. teleta* (orange), *P. dumerilii* (yellow), *L. gigantea* (teal), *A. californica* (sky blue), *C. gigas* (turquoise) *O. bimaculoides* (blue), *D. pealeii* (grey), *S. officinalis* (slate), and *I. paradoxus* (light blue-grey). Red asterisk: clustering of sequences on Scaffolds 30672 and 159035. Purple asterisk: clustering of sequences on Scaffold 9600.

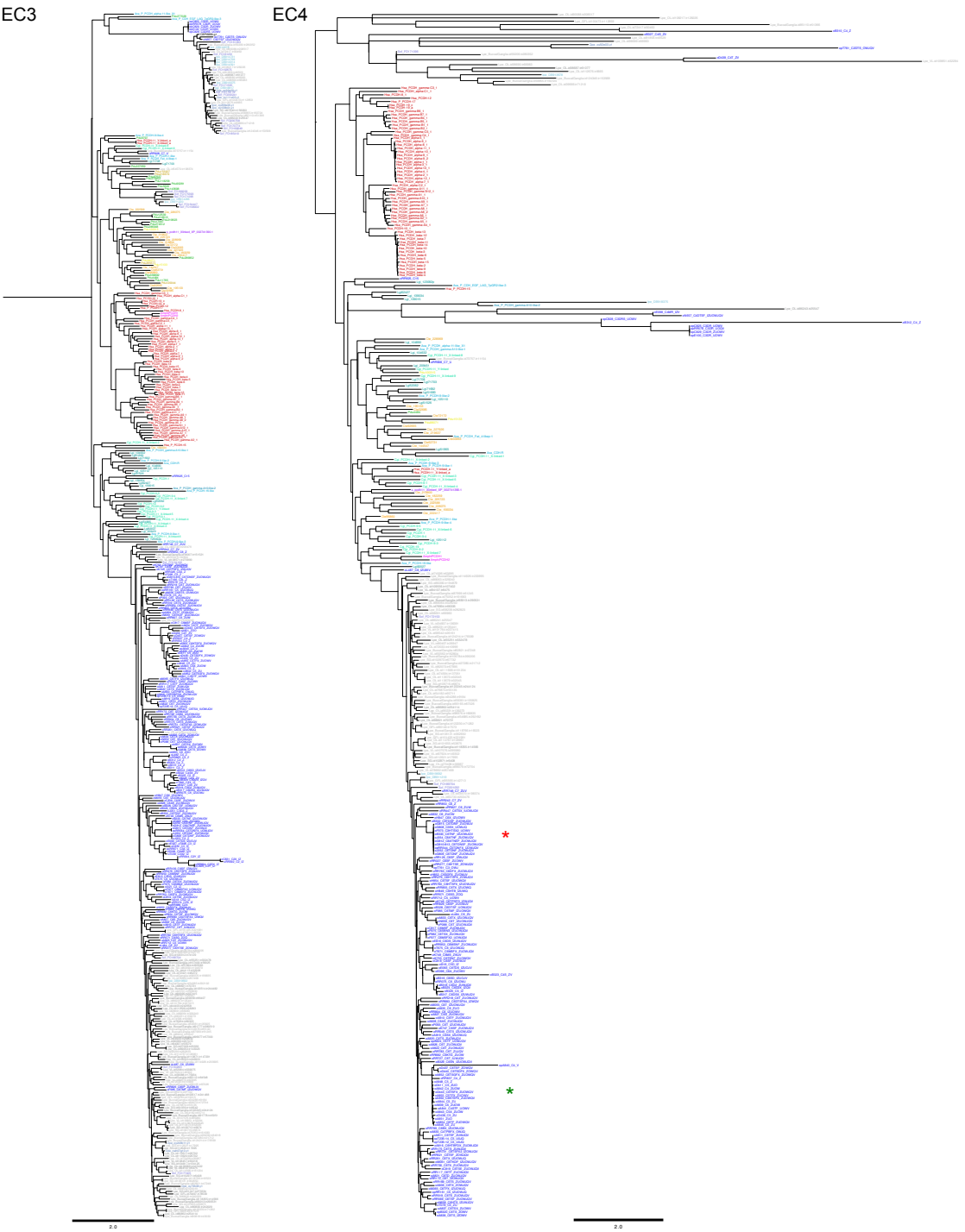


Figure 4.25. ML trees of EC3 and EC4 domains of metazoan protocadherins.
Phylogenetic trees of protocadherin EC3 and EC4 in *H. sapiens* (red), *C. teleta* (orange), *P. dumerilii* (yellow), *L. gigantea* (teal), *A. californica* (sky blue), *C. gigas* (turquoise) *O. bimaculoides* (blue), *D. pealeii* (grey), *S. officinalis* (slate), and *I. paradoxus* (light blue-grey). Red asterisk: clustering of sequences on Scaffolds 30672 and 159035. Green asterisk: clustering of sequences on Scaffolds 93179 and 309453.

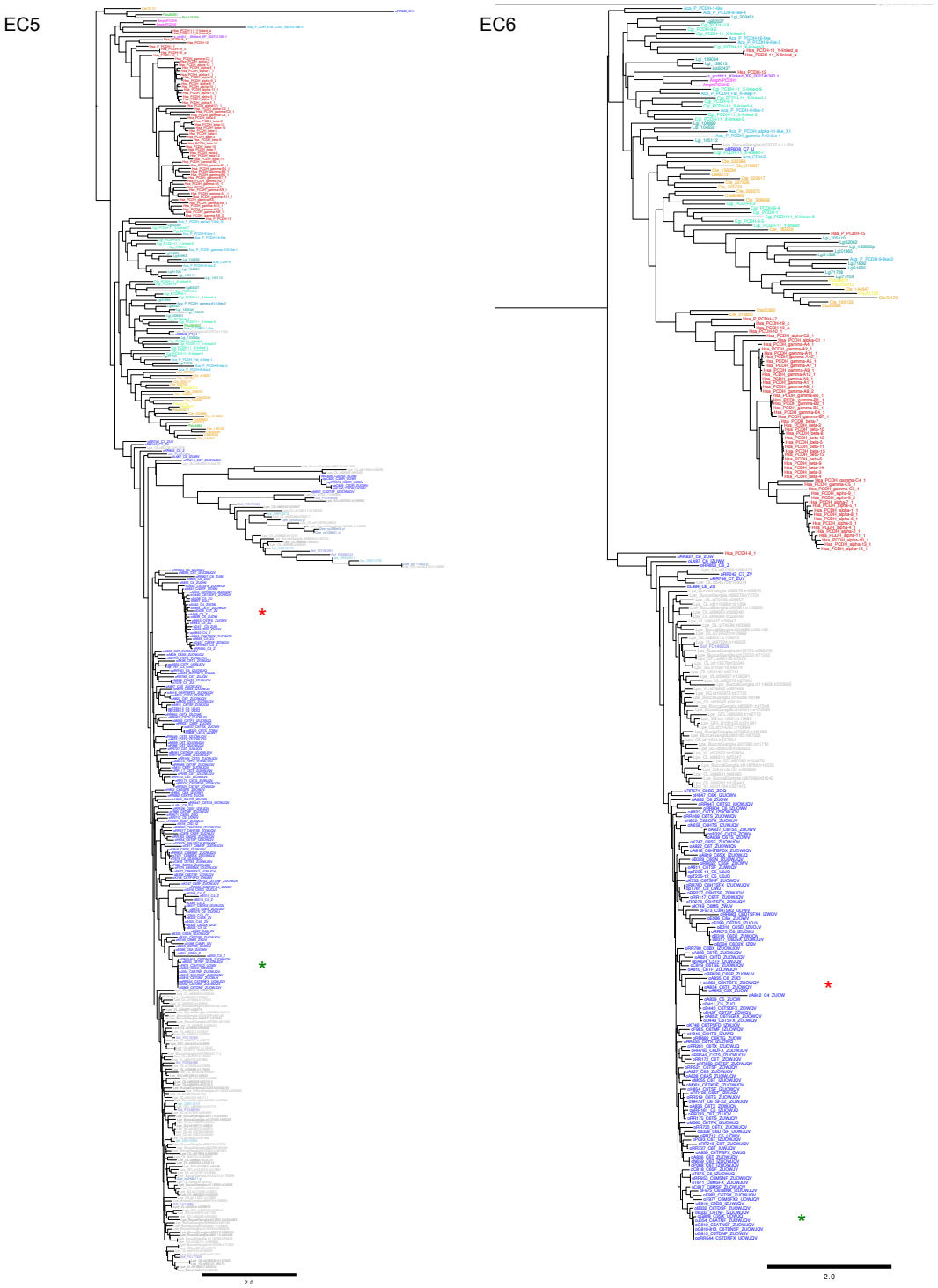


Figure 4.26. ML trees of EC5 and EC6 domains of metazoan protocadherins. Phylogenetic trees of protocadherin EC5 and EC6 in *H. sapiens* (red), *C. teleta* (orange), *P. dumerilii* (yellow), *L. gigantea* (teal), *A. californica* (sky blue), *C. gigas* (turquoise) *O. bimaculoides* (blue), *D. pealeii* (grey), *S. officinalis* (slate), and *I. paradoxus* (light blue-grey). Red asterisk: clustering of sequences on Scaffolds 30672 and 159035. Green asterisk: clustering of sequences on Scaffolds 93179 and 309453.

As noted above and in Figure 4.20, we identified multiple protocadherins in non-cephalopod lophotrochozoans, specifically the polychaete annelids *C. teleta* and *P. dumerilii*, and the molluscs *L. gigantea* and *C. gigas*. *L. gigantea* and *C. gigas* each have 17-25 protocadherins; in *L. gigantea* 14 are clustered on one scaffold, while *C. gigas* has two clusters of four and three doublets. Unlike the clusters in *O. bimaculoides*, these clustered protocadherins in other molluscs are transcribed in alternating directions. The non-cephalopod molluscan protocadherins group together in phylogenetic analyses of full-length sequence, but show a very different pattern from the cephalopod genes. *L. gigantea* and *C. gigas* protocadherins interdigitate, which could indicate that the protocadherins started to expand before bivalves and snails diverged. We see a similar pattern in the annelids, in which the polychaete protocadherins group together on the tree, with the *C. teleta* and *P. dumerilii* genes interdigitating. Unlike the molluscan protocadherins, the largest grouping of the 25 *C. teleta* protocadherins that we could identify is two genes on a single scaffold. These mollusc and annelid protocadherins remain segregated on the tree (Figure 4.22). Similar clade-specific patterns of protocadherin expansions have also been described in vertebrates (Hulpiau and van Roy, 2009; Morishita and Yagi, 2007; Yu et al., 2008).

Cephalopod novel genes

Taxonomically restricted genes can be important in the evolution of clade-specific traits (Johnson and Tsutsui, 2011; Tautz and Domazet-Loso, 2011). Many of the predicted proteins identified in the octopus genome without a match to other animal proteins also did not have a match to EST or transcriptome sequences from any other cephalopod with an e-value cutoff of 1e-5. The available cephalopod sequences that we searched are predominantly from the decapodiforms *E. scolopes* and *S. officinalis*, so many of these sequences could represent

octopod-specific genes. We found 3,557 putative octopus-specific protein-coding genes expressed in the transcriptomes, 1,020 of which have a match to an *Octopus vulgaris* transcriptome (Smith et al., 2011). Of these candidate octopus-specific transcripts, 1,520 are expressed in a tissue-specific manner, which we defined as having more than 75% of the total expression, when normalized to transcriptome size, in a single tissue (Figure 4.27). Because many of these genes are expressed in the central nervous system (ANC, OL, supraesophageal and subesophageal brain), expression values for these four transcriptomes were combined for this analysis. Many octopus-specific genes are expressed primarily in the testes, which have been described as a site for orphan gene expression in other animals (Begun et al., 2007; Levine et al., 2006; Palmieri et al., 2014).

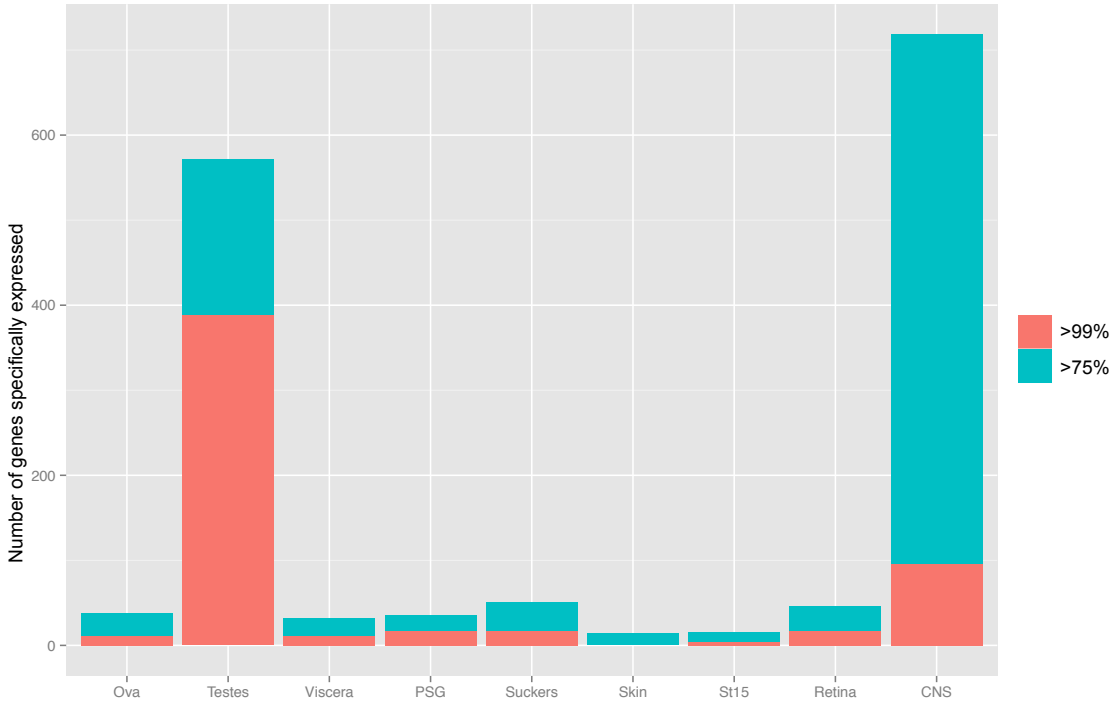


Figure 4.27. Tissue-specific expression of octopus-specific genes.
Genes with >75% of total expression in a single tissue are represented. The subset with >99% expression in a tissue is shown in red. A large number of octopus-specific genes are found in the central nervous system and the testes.

DISCUSSION

Our analysis of the octopus genome sheds light on the genetic underpinnings of cephalopod novelties and raises many new questions. Whole genome duplication has been suggested to result in a relaxation of constraint on genes through redundancy, which allows for subfunctionalization or neofunctionalization, candidate substrates for evolutionary innovations (Ohno, 1970). Several researchers have proposed that whole genome duplications were important in the evolution of the cephalopod body plan (Bonnaud et al., 2004; Hallinan and Lindberg, 2011; Yoshida et al., 2011), paralleling the role ascribed to the independent whole genome duplication events that occurred early in vertebrate evolution (Holland et al., 1994). The proposal of one or more rounds of whole genome duplication in coleoids derives from the chromosome number in octopus species (N = 28 or 30) ((Adachi et al., 2014), and references therein), which is approximately double the putative chromosome number of the molluscan ancestor (N=15-20) (Hallinan and Lindberg, 2011; Simakov et al., 2013). We did not find, however, any evidence in octopus for the genome-wide retention of gene duplicates expected after such an event, even in gene classes that are usually retained after paleopolyploidy in other clades, such as developmental control genes (Table 4.1; (Holland et al., 1994; Putnam et al., 2008)). The gene family expansions we do observe are organized in clusters along the genome, rather than distributed in doubly conserved synteny (Dietrich et al., 2004; Kellis et al., 2004). While we cannot rule out whole genome duplication followed by considerable gene loss, the extent of loss needed to support this claim would far exceed that which has been observed in vertebrates, making it extremely improbable that the octopus lineage experienced such an event (Albertin et al., 2015). It is more plausible that other mechanisms, such as chromosome fragmentation, acted to increase chromosome number in coleoids. Although whole genome

duplication is an attractive framework for both gene family expansion and increased regulatory complexity across multiple genes, we found no evidence for it, ruling out a grand convergence of genomic mechanisms for octopus and vertebrate morphological complexity.

The evidence for coleoid polyploidy is derived predominantly from work in decapodiform cephalopods. While Hallinan and Lindberg (2011) reconstructed a whole genome duplication event at the base of the coleoid cephalopods, they also found some evidence for an independent duplication in decapodiform cephalopods. This suggestion will be addressed as further cephalopod genomes become available.

In gene family content, domain architecture, exon-intron structure, and functional content, the octopus genome encodes a "typical" invertebrate bilaterian gene set similar to that of the cephalochordate *Branchiostoma floridae* (Putnam et al., 2008), the polychaete annelid *Capitella teleta*, and the limpet *Lottia gigantea* (Simakov et al., 2013). We detected a fairly standard set of developmentally important transcription factors and signaling pathway genes relative to other invertebrate bilaterians, emphasizing that the evolution of the cephalopod body plan did not require notable elaborations of these “toolkit” genes. We recovered single octopus orthologs of most of the transcription factors and developmental ligands that have been widely characterized in other animals. The *WNT, HH, TGF-β* and *FGF* ligand families of *O. bimaculoides* all appear to have a basic lophotrochozoan gene inventory. The octopus genome encodes a single member of most of the homeodomain, basic helix-loop-helix and nuclear hormone receptor families. Notably, the gene complements present in the *D. melanogaster* and *C. elegans* model systems often showed dramatic departures from those seen in lophotrochozoans and vertebrates, making them unreliable models of any general invertebrate

genome condition (Table 4.1), and highlighting the importance of greater taxon sampling when drawing conclusions across wide phylogenetic distances.

In clades with whole genome duplications in their history, Hox genes are typically retained in multiple copies (Holland et al., 1994; Putnam et al., 2008). We found only a single Hox complement in *O. bimaculoides*, consistent with the absence of whole genome duplication in cephalopods. Nine of eleven known lophotrochozoan Hox genes have been isolated by PCR from the Hawaiian bobtail squid, *Euprymna scolopes* (Callaerts et al., 2002). The expression of these Hox genes in *E. scolopes* was examined using *in situ* hybridization, but whether *E. scolopes* Hox genes are clustered was not investigated. The combinatorial expression patterns found did not conform to spatial collinearity under the assumption of a conventionally organized cluster (Lee et al., 2003). The detection of a single Hox complement from the bobtail squid *E. scolopes* by PCR (Callaerts et al., 2002) is consistent with our findings. Notably, octopus Hox genes are not organized into a cluster as in other bilaterian genomes: no two Hox genes were found on the same scaffold (Figure 4.2). This finding suggests that the forces maintaining a compact Hox cluster in other genomes are not operating in the octopus and that the *cis*-regulatory regions surrounding these highly conserved transcription factors has been greatly altered.

While most gene families are found in single copy in the octopus genome, we recovered massive expansion through clustered duplications of two gene families implicated broadly in morphological and neural development (C2H2-ZNFs and protocadherins), in principle providing a basis for the development of a very atypical invertebrate nervous system. The C2H2-ZNF transcription factors are the most abundant transcription factor families in animal genomes, and show taxonomically-restricted expansions that may contribute to evolutionary innovation across

these species (Liu et al., 2014; Shannon et al., 2003). While other lophotrochozoans examined have 200-400 C2H2-ZNFs, we identified nearly 1,800 multi-exonic C2H2-containing genes in *O. bimaculoides* (Table 4.1). C2H2-ZNF transcription factors contain multiple C2H2 domains that, in combination, result in highly specific DNA binding. The octopus C2H2-ZNFs typically contain 10-20 C2H2 domains but some have as many as 60 (Figure 4.14). The majority of the transcripts are expressed in embryonic and nervous tissues (Figure 4.15). This pattern of expression is consistent with roles for C2H2-ZNFs in cell fate determination, early development, and transposon silencing, which have been demonstrated in genetic model systems (Nowick et al., 2010).

The octopus genome encodes 168 multi-exonic protocadherin genes, nearly three-quarters of which are found in clusters on the genome (Figure 4.22), a dramatic expansion relative to the 17 genes found in each of the *Lottia, Crassostrea,* and *Capitella* genomes. Single protocadherin genes are found in the invertebrate deuterostomes *Saccoglossus kowalevskii* and *Strongylocentrotus purpuratus,* so their absence in ecdysozoans like *Drosophila* and *C. elegans* is due to gene loss. Vertebrates also show a remarkable expansion of the protocadherin repertoire, but this is generated by complex splicing from a clustered locus (Chen and Maniatis, 2013; Takeichi, 2007; Yagi, 2008). Thus both cephalopods and vertebrates have independently evolved a diverse array of protocadherin genes.

A search of available transcriptome data from the longfin inshore squid *Doryteuthis* (formerly, *Loligo*) *pealeii* (Brown et al., 2014) also demonstrated an expanded number of protocadherin genes (Figure 4.23). Surprisingly, our phylogenetic analyses suggest that the squid and octopus expansions may have arisen independently, as octopus and squid protocadherins

also segregate in phylogenetic trees (Figures 4.23-27). Further sequencing of cephalopod species should yield interesting insights into the history of this gene family.

The expression of protocadherins in octopus neural tissues (Figure 4.22) is consistent with a central role for these genes in establishing and maintaining cephalopod nervous system organization. Protocadherin diversity provides a mechanism for regulating the short-range interactions needed for the assembly of local neural circuits (Zipursky and Sanes, 2010), which is where the greatest complexity in the cephalopod nervous system appears (Young, 1971). The importance of local neuropil interactions, rather than long-range connections, is likely due to the limits placed on axon density and connectivity by the absence of myelin, since thick axons are required for rapid high fidelity signal conduction over long distances. The sequence divergence between octopus and squid protocadherin expansions may in turn reflect the dramatic differences between decapodiforms and octopuses in brain organization, which has been most clearly demonstrated for the vertical lobe, a key structure in cephalopod learning and memory circuits (Shomrat et al., 2011; Young, 1971). Finally, the independent expansion and nervous system enrichment of protocadherins in octopus and vertebrates offers a striking example at the molecular level of convergent evolution between these clades.

Novel genes are also likely to contribute to the unique biology of the octopus. Of the hundreds of gene families without recognizable similarity outside cephalopods, the only family studied thus far (the reflectins) is involved in a specialized cephalopod behavior. More than 10% of the predicted protein coding genes in the octopus genome are expressed but have no detectable match to available sequences from other animals, including squid, cuttlefish, and *Nautilus*. Over 1,000 of these genes could be confirmed in publicly available transcriptome data from *Octopus vulgaris*. Of the 3,557 octopus-specific genes, more than 40% are expressed in a

tissue-specific manner, most notably in the testes and central nervous system (Figure 4.27). While the lack of similarity of these genes to those found in other animals makes it difficult to speculate on their function, those expressed in a tissue- or stage- specific manner are particularly compelling candidates for future study. In addition, those genes that are found to be conserved with other cephalopods, like the reflectins, may lend insight into the molecular bases of morphological novelty.

CHAPTER 5

Development and molecular patterning of the octopus brain

SECTION I

Evidence for a cordal, not ganglionic, pattern of cephalopod brain neurogenesis

INTRODUCTION

Cephalopods have the largest brains of any invertebrate; by a measure of brain to body weight, the coleoid cephalopod central brain rivals that of vertebrates (Packard, 1972), and that measure only accounts for approximately 40% of the neurons in the central nervous system (Young, 1971). How the organization of the highly centralized cephalopod nervous system corresponds to that of other molluscs remains obscure.

Molluscan relationships and nervous systems

Given their vast diversity of form, molluscan relationships have been difficult to resolve using morphology-based methods. Historically, the cephalopods and gastropods have been considered sister groups within the crown group Conchifera, along with bivalves, scaphopods, and monoplacophorans (Salvini-Plawen and Steiner, 1996). The wormlike Chaetodermomorpha and Neomeniomorpha, which lack shells, have been considered as "primitive" or basally branching clades, with polyplacophorans alternately placed as a sister group to the aplacophorans (Chaetodermomorpha + Neomeniomorpha) or the conchiferans. However, the recent application of molecular phylogenomics supports the union of the polyplacophorans and the aplacophorans as the Aculifera, and retrieves a monophyletic Conchifera, but with a different internal arrangement (Kocot et al., 2011; Smith et al., 2011). Here, the cephalopods are the sister group to

either the monoplacophorans (Smith et al., 2011) or to the remaining conchiferans (Scaphopoda + Bivalvia + Gastropoda) (Kocot et al., 2011).

Broadly considered, molluscan nervous systems are organized as either medullary cords or ganglia. The nervous systems of polyplacophorans, aplacophorans and monoplacophorans have medullary cords (Figure 5.1.1; (Bullock and Horridge, 1965)), which have cell bodies encasing the neuropil along the length of the cord (Richter et al., 2010). The nervous systems of scaphopods, gastropods, and bivalves are arranged quite differently: these groups possess clusters of neuronal cell bodies and neuropil called ganglia that are connected via commissures or connectives that are cell body-free (Bullock and Horridge, 1965). It is difficult to relate the elaborate, highly centralized brain of adult cephalopods to either the ganglia of gastropods and bivalves or the cords of chitons. The traditional phylogenetic arrangement of molluscan taxa, which places cephalopods as sister to gastropods or scaphopods, would suggest a ganglionic plesiomorphic condition. However, as new analyses recover a basally-branching cephalopod clade within the Conchifera, the ancestral state of their nervous systems is not polarized: they could either reflect the cordal arrangement found in the aculiferans, or the ganglionated condition of bivalves, gastropods, and scaphopods.

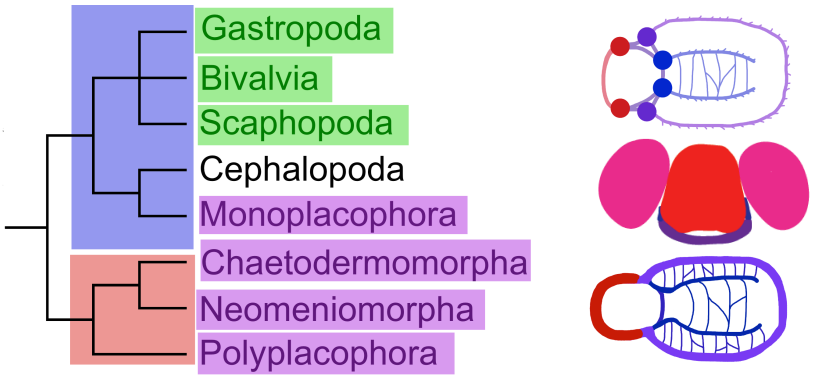


Figure 5.1.1. Relationships of extant molluscan clades and schematic of their adult nervous systems.
Phylogenetic tree after Smith et al. (2011). Red box highlights aculiferan taxa, while blue box indicates conchiferans. Taxa highlighted in purple have a cord-based nervous system, while green indicates clades with a ganglionic nervous system. To the right are schematics of adult nervous systems found in a basally-branching gastropod (*Lottia,* top), a cephalopod (*Octopus*, middle), and a polyplacophoran (bottom). Cerebral, palliovisceral, and pedal cords and ganglia (circles) are indicated in red, purple, and blue, respectively. The polyplacophoran and gastropod nervous systems are represented in a dorsal view, oriented with anterior to the left, after Bullock (1965). The schematic of the cephalopod brain is a dorsal view, with the mouth to the top, and the left and right optic lobes (pink) flanking the supraesophageal brain (red) with subesophageal brain in blue. Connectives lacking cell bodies are indicated in the gastropod schematic as lighter lines connecting the ganglia.

Cordal versus ganglionic models of cephalopod brain development

While the adult cephalopod brain is highly centralized, obscuring its underlying arrangement, the developmental origins of this neural tissue could be informative. If the developing nervous system is arranged as isolated clusters of neural primordia that do not cross the midline, this would support a ganglionic arrangement. Indeed, previous studies of cephalopod nervous system development based on histological sections described four pairs of ganglia (Marquis, 1989; Marthy, 1987; Shigeno et al., 2001a; Shigeno et al., 2001b; Yamamoto et al., 2003). The cerebral ganglia, which give rise to the supraesophageal brain, appear as clusters flanking the mouth (Figure 5.1.2 *Right panel*). Lateral to the cerebral ganglia are the optic ganglia, which resemble medullary cords. The spherical palliovisceral and pedal ganglia

appear at the posterior. As these ganglia increase in size and condense as the embryo grows and undergoes compaction, these initially separate territories fuse across the midline.

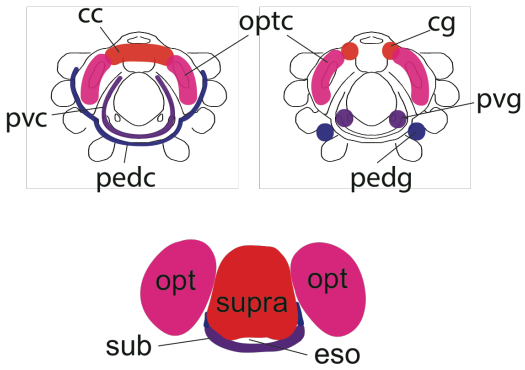


Figure 5.1.2. Competing hypotheses of cephalopod brain development.
Upper panels: cartoons of a stage 10 octopus embryo with cerebral (red), optic (magenta), palliovisceral (purple), and pedal (blue) primordia indicated. Mouth is to the top, arms are lateral, mantle is central. *Left upper panel* illustrates a cordal development of the octopus nervous system, where developing neurons are found in cord-like territories spanning the midline. *Upper right panel* depicts the traditional model of cephalopod neural development, where developing cerebral, palliovisceral and pedal neurons are found in spatially restricted ganglia (circles). *Lower panel* illustrates the arrangement of the adult cephalopod brain in a top-down view, with the central supraesophageal (supra; red) and subesophageal (sub; purple) masses flanked to the left and the right by optic lobes (opt; pink). Brain is oriented with the anterior (mouth) to the top. Other abbreviations: cc, cerebral cord; cg, cerebral ganglion; eso, esophagus; opt, optic lobe; optc, optic cord; pedc, pedal cord; pedg, pedal ganglion; pvc, palliovisceral cord; pvg, palliovisceral ganglion.

If, however, the cephalopod nervous system develops from cords, we would expect to see elongated, rather than nodular, neural territories (Figure 5.1.2 *Left panel*). The rope-like primordia of the cerebral, palliovisceral and pedal cords would cross the midline from early stages. As the support for a ganglionic nervous system in cephalopod embryos stems largely from the reconstruction of histological data, definitive identification of neural territories in cephalopod embryos has not been possible. Here, we use wholemount *in situ* hybridization of pan-bilaterian neural markers in octopus embryos to evaluate these opposing hypotheses.

Neural markers in bilaterians

elav (embryonic lethal, abnormal visual system) is a RNA binding protein gene expressed by young neurons, but not neuronal progenitors (Kim et al., 1996; King et al., 1994; Robinow and White, 1991; Wakamatsu and Weston, 1997). *ELAV* exhibits highly conserved neuron-specific expression during development across animals as phylogenetically distant as flies, worms, hemichordates, and vertebrates (Denes et al., 2007; Meyer and Seaver, 2009; Nomaksteinsky et al., 2009). *NEUROD* and *NEUROGENIN* (*NEUROG*) are highly conserved members of the atonal basic helix-loop-helix transcription factor family found throughout the Metazoa. These genes have been shown to be involved in the neuronal differentiation and cell-type specification of the developing nervous systems of vertebrates and worms (Paridaen and Huttner, 2014).

We describe here the pattern of neurogenesis in octopus embryos and assess the processes by which the octopus brain is assembled during embryogenesis. Our findings, based on *ELAV, NEUROD,* and *NEUROG* gene expression, provide support for the conclusions that (1) the cephalopod brain develops from concentric cords, and (2) a system of neuronal cords is a feature of molluscan neuroembryogenesis shared by basally branching conchiferans (octopuses and monoplacophorans), and aculiferans.

RESULTS

ELAV mRNA labels neuronal cell bodies but not processes in octopus brain

To study the early patterns of neuronal development, we isolated an *ELAV* cDNA fragment from *O. bimaculoides* (Figure 5.1.3). Using the human *ELAV3* as bait, we isolated three

octopus ELAV-like sequences through BLAST of proteins identified in the octopus genome. Phylogenetic analysis indicated that *O. bimaculoides*, along with other invertebrates sampled, has a single *ELAV* that clusters with the vertebrate *ELAV* sequences. Lophotrochozoans, along with the cnidarian *Nematostella vectensis*, also have an ELAV-like sequence, which appears to have been lost in ecdysozoans and deuterostomes. In *O. bimaculoides*, this ELAV-like sequence has been duplicated. However, a search of ESTs deposited for the European cuttlefish *Sepia officinalis* revealed only one *ELAV-like* sequence, which may indicate that this duplication is octopus-specific.

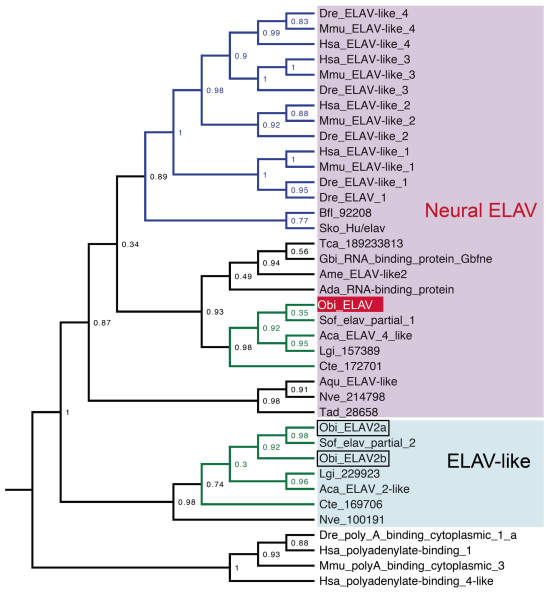


Figure 5.1.3. Phylogenetic analysis of *ELAV* and *ELAV-like* genes.
Invertebrates, including sponges, cnidarians and placozoans, have a single neural *ELAV* gene, while jawed vertebrates have four or more. In addition, *ELAV-like* genes are present in cnidarians and lophotrochozoans, but not deuterostomes and (apparently) ecdysozoans. In *O. bimaculoides*, we recovered two *ELAV-like* genes and one neural *ELAV* gene. Relationships of metazoan *ELAV* and *ELAV-like* genes are illustrated by an approximately maximum likelihood tree with genes from *O. bimaculoides* (Obi), *Sepia officinalis* (Sof), *Lottia gigantea* (Lgi), *Aplysia californica* (Aca), *Capitella teleta* (Cte), *Saccoglossus kowalevskii* (Sko), *Branchiostoma floridae* (Bfl), *Danio rerio* (Dre), *Mus musculus* (Mmu), *Homo sapiens* (Hsa), *Tribolium castaneum* (Tca), *Gryllus bimaculatus* (Gbi), *Apis mellifera* (Ame), *Anopheles darlingi* (Ada), *Amphimedon*

queenslandica (Aqu), *Trichoplax adhaerens* (Tad) and *Nematostella vectensis* (Nve). Likelihood support indicated at nodes.

ELAV expression in an octopus hatchling brain was restricted to the outer rind of cell bodies, and did not stain the neuropil (Figure 5.1.4). This expression pattern highlights the architecture of the octopus brain, with the supraesophageal and subesophageal brain further subdivided into functionally distinct lobes. *ELAV* expression was detected in all brain lobes examined, though staining in the nascent frontal-vertical system was lighter than other lobes. This region of the octopus brain greatly expands during the early juvenile period, which might be reflected in the lower expression of a marker of postmitotic neurons such as *ELAV*

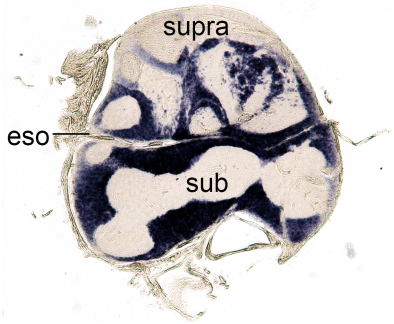


Figure 5.1.4. *ELAV* expression in *O. bimaculoides* hatchling brain.
In situ hybridization for *ELAV* on a midline section through a 2 week old hatchling brain. Mouth is to the left and the supraesophageal brain is to the top. *ELAV* expression marks the cell bodies at the periphery of the lobes. The central neuropils of the lobes are not stained. eso, esophagus; sub, subesophageal; supra, supraesophageal.

ELAV gene expression in octopus brain development

Using *ELAV* as a neuronal marker, we studied the distribution of *ELAV*-positive cells in octopus embryos using wholemount *in situ* hybridization. *ELAV* expression was first detected at stage 7, when it is present in broad swaths of the embryo, encircling the eye primordia on each side, and extending across the midline. It is also expressed in condensations flanking the mantle, extending around the mantle to the posterior and spanning the midline to form a U shape. From this earliest time point of detection, anterior and posterior bands of *ELAV* expression were found spanning the midline rather than spatially restricted bilateral pairs of ganglia as would be

expected if the octopus nervous system was based on a series of ganglia (compare Figure 5.1.5A-B with upper panels of Figure 5.1.2).

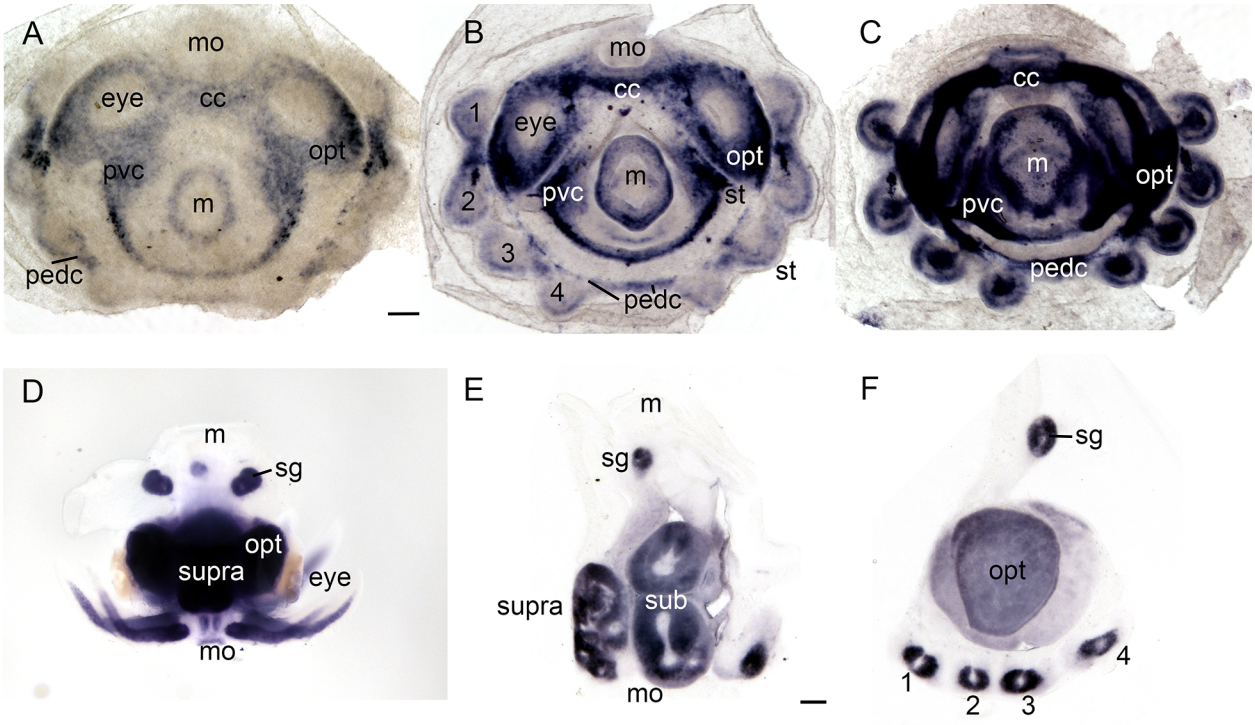


Figure 5.1.5. *ELAV* gene expression in octopus embryos demonstrates a cordal development to the nervous system.
Wholemount *in situ* hybridization of *ELAV* in *O. bimaculoides* embryos at stage 7 (A), stage 8 (B), stage 10 (C), and stage 16 (D-E). (A-C) Embryos are oriented with the mouth (mo) at the top and mantle (m) at the center. (D) Stage 16 embryo oriented with the mouth to the bottom and the mantle at the top, with anterior to the front. (E-F) Parasagittal sections of stage 16 wholemount with the mouth oriented to the bottom, mantle at the top, and supraesophageal brain to the left. cc, cerebral cord; opt, optic cord; pedc, pedal cord; pvc, palliovisceral cord; sg, stellate ganglion; st, statocyst; 1, 2, 3, 4 indicate the developing axial nerve cords of the arms.

By stage 8 (Figure 5.1.5B) *ELAV*-positive cells were clearly distributed in a system of cords: an anterior cerebral cord and posterior palliovisceral and pedal cords, with the palliovisceral cord nested between the mantle and the pedal cord (compare Figure 5.1.5B with Figure 5.1.2 *Left Upper Panel*). These nascent cords show a gradient of *ELAV* expression, with the center of the broad swath expression more strongly stained than the edges (Figure 5.1.5B),

suggesting that there are neurogenic gradients in these early cords. At this stage we can also detect expression in the developing pedal cord. *ELAV* expression at stage 10 not only shows prominent cerebral, palliovisceral and pedal cords, but is also detected in the developing axial nerve cords of the arm buds (Figure 5.1.5C). By stage 16, the neural tissues have condensed into the adult arrangement, with the central supraesophageal and subesophageal brain flanked by optic lobes (Figure 5.1.5D)

The cerebral and optic cords

The cerebral cord gives rise the supraesophageal brain, which contains the learning and memory centers of the frontal–vertical lobes and the higher motor centers found in the basal lobes. The cord-like arrangement of the cerebral neurogenic territory was apparent from the first stage at which we were able to detect *ELAV* expression, which flanks the mouth and crosses the midline. The cerebral cord expression territory is largely continuous with that of the optic lobes, which surround the eye primordia. By stage 16, this anterior neural territory has been organized into many of the lobes of the supraesophageal mass (Figure 5.1.5E) that are visible in hatchling (Figure 5.1.4) and adult brains (Shigeno and Ragsdale, 2015; Young, 1971). At this stage the optic cords resemble the optic lobe arrangement found in adults (Young, 1971).

The palliovisceral cord

The palliovisceral cord gives rise to the posterior subesophageal mass, which innervates the mantle, visceral mass, and collar (Young, 1971). *ELAV* gene expression at stages 7 and 8 demonstrated concentrations of prospective palliovisceral neurons in an arc flanking the mantle primordium (Figure 5.1.5A-B). A U-shaped palliovisceral cord is clearly in evidence by stage 10, encircling the mantle (Figure 5.1.5C). At later stages this cord condenses around the esophagus to form the posterior subesophageal lobe (Figure 5.1.5E).

The pedal cord

The pedal cord gives rise to the anterior and middle subesophageal masses in the mature brain, and innervates the arms and the anterior funnel. Although not readily detected at stage 7, by stage 8 *ELAV* expression was present in the prospective pedal cord and in the axial nerve cords of the arm primordia. This was quite prominent at stage 10 (Figure 5.1.5C). Pedal cord *ELAV* expression can be seen extending across the posterior midline even at stages 8 and 10, supporting a cordal rather than ganglionic arrangement. At later stages, the *ELAV*-rich pedal cord centralizes to form the compact structure comprising the anterior and middle subesophageal masses (Figure 5.1.5E).

Markers of neurogenetic territories

In addition to *ELAV* staining, we studied the expression of neuronal determination genes *NEUROD* and *NEUROG* with wholemount *in situ* hybridization. Both of these markers were strongly expressed at stage 8 in territories congruent with *ELAV* expression patterns. *NEUROG* was strongly expressed in the tissue overlying the *ELAV*-rich territories encircling the mantle and the arm buds. the prospective cerebral, palliovisceral and pedal cords, as well as in neural territories in the mantle and arm buds (Figure 5.1.6A). We detected a similar cordal expression for *NEUROD* (Figure 5.1.6B), which is a transcriptional target of Neurogenin in *Xenopus* and mouse (Seo et al., 2007). These observations establish a cordal pattern of neurogenesis in *O. bimaculoides*.

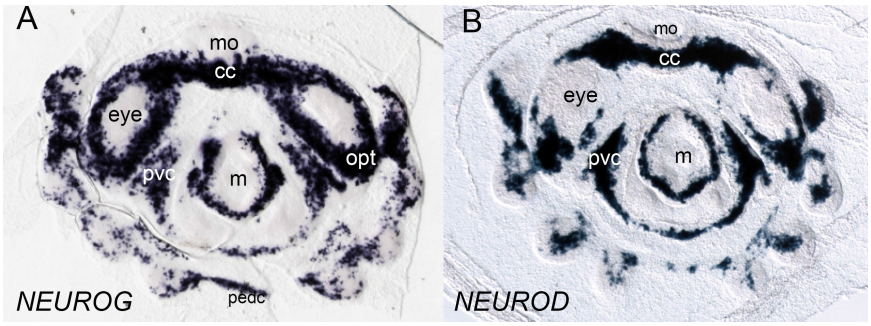


Figure 5.1.6. Expression of neuronal differentiation genes at stage 8.
Dorsal views of wholemount *in situ* hybridization of *NEUROG* (A) and *NEUROD* (B).

Cross sections following wholemount *in situ* hybridization demonstrated an apical-basal layering of these neuronal markers (Figure 5.1.7). *NEUROG* expression is found superficially in the ectoderm, whereas *ELAV* and *NEUROD* are expressed in deeper cells. These results are also consistent with a pro-neural role for *NEUROG*, with *ELAV*- and *NEUROD*- positive neurons migrating to a basal position. A similar pattern was found in the ventral nerve cord of *Platynereis dumerilii,* where *Neurogenin* was found in apical cells, while *elav* expression was detected in basal cells. Thus an apical-to-basal mode of neurogenesis is shared between annelids and molluscs (Hartenstein and Stollewerk, 2015).

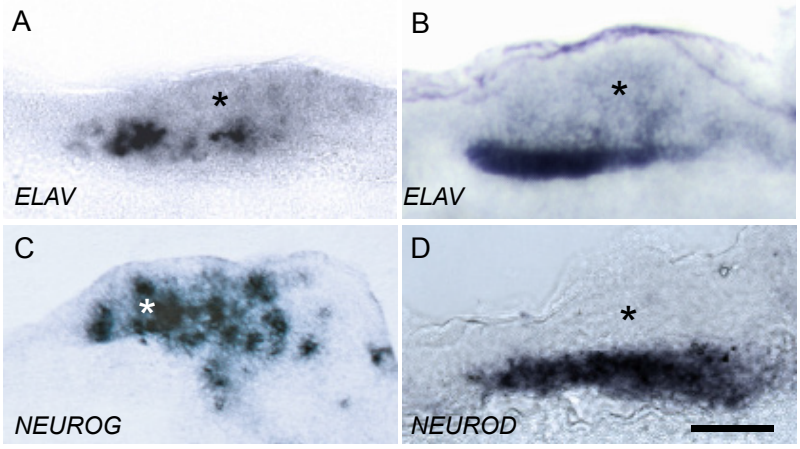


Figure 5.1.7. Apical-basal organization of octopus neurectoderm.
Sagittal sections of the palliovisceral cord of embryos stained for *ELAV* at stages 7 (A) and 8 (B), and *NEUROG* (C) and *NEUROD* (D) at stage 8. Asterisks indicate the putative neurogenic ectoderm, apical to the basal site of neuronal differentiation. Scale bar: 20µm.

DISCUSSION

The expression patterns of neurogenetic and neural marker genes in octopus embryo wholemounts support a cord-based arrangement in early embryogenesis of the exceptional cephalopod nervous system. This underlying cordal pattern is strikingly different from the multiple spherical ganglia that have been described in other cephalopod species from reconstructions of tissue sections (Marquis, 1989; Shigeno et al., 2001a; Shigeno et al., 2001b; Yamamoto et al., 2003). However, these studies were unable to specifically label neuronal cells, making the conclusion of a ganglionic arrangement less secure. Our results also differ somewhat from findings in the cuttlefish *S. officinalis*, which also leveraged *elav* expression to examine neurogenesis (Buresi et al., 2013). Buresi and colleagues describe *sof-elav* expression in discrete ganglia in *S. officinalis* embryos, and they do not report the presence of cords spanning the midline. Close examination of the panels suggest that this may be due to a detection problem in the *in situ* hybridization molecular biology, although this could represent a species difference. In *S. officinalis*, *sof-elav* was also expressed asynchronously in the neurogenic territories: expression in the palliovisceral territory was detected earlier than in the cerebral or optic cords. In contrast to the findings in cuttlefish, the cerebral, optic, and palliovisceral cords are first detected coincidently in *O. bimaculoides*, suggesting that the timing of neurogenesis may be different in these species.

These data indicate the development of the cephalopod nervous system is quite different from the ganglionic architecture of the adult nervous systems of gastropods, which had long been hypothesized to be the cephalopod sister group. Neurodevelopmental studies in gastropods suggest that the nervous system first appears as spatially-restricted ganglia, but as of yet, no

study of gastropod neurogenesis has leveraged a pan-neuronal marker such as *ELAV* (Croll and Voronezhskaya, 1996; Jacob, 1984; Kristof et al., 2016; Page, 2006; Wanninger and Wollesen, 2015). The multiple-cord arrangement of the embryonic cephalopod nervous system, along with the recent molecular phylogenies that place cephalopods as a sister group to monoplacophorans within the Conchifera (Kocot et al., 2011; Smith et al., 2011), therefore support a multiple cord morphology as the plesiomorphic state for both molluscs and conchiferans. These results indicate that the large, elaborate, and highly compacted brains of cephalopods retain the ancestral molluscan neural architecture of multiple cords.

SECTION II

Conservation of transcription factors, but not ligands, in an octopus brain signaling center

INTRODUCTION

Eubilaterians share a common molecular toolkit of genes expressed along the anteroposterior axis, many of which have roles in patterning the body plan and brains of such distantly related animals as vertebrates (Lumsden and Krumlauf, 1996; Wada et al., 1998), hemichordates (Pani et al., 2012), insects (Arendt and Nubler-Jung, 1999; Reichert, 2005; Urbach, 2007), and annelids (Denes et al., 2007; Frobius et al., 2008). A prominent example of the deployment of this conserved molecular toolkit can be found in the patterning of the vertebrate midbrain-hindbrain boundary. This territory is defined by anterior expression of *OTX*, with *EN* and *PAX2/5/8* expression at the boundary between the developing mid- and hindbrain, and *GBX* and *HOX* genes expressed posterior to this region. In vertebrates, the MHB boundary is

also a signaling center known as the isthmus organizer (IsO), where abutting domains of *FGF8* and *WNT1* signal to nearby neural structures. The IsO, along with the associated transcription factors, appears to be present in hemichordates (Pani et al., 2012). In *Drosophila melanogaster*, *otd* (an *OTX* ortholog)*, unpg* (a *GBX* ortholog)*,* and *labial* (a *HOX1* ortholog) are expressed in the deutocerebral and tritocerebral neuromeres (Reichert, 2005; Urbach, 2007). This striking similarity of transcription factor expression during brain development in both deuterostomes and protostomes has led to the suggestion that this arrangement of conserved developmental control genes represents a shared tripartite brain ground plan in bilaterians (Reichert, 2005).

Coleoid cephalopods (octopuses, squid, and cuttlefish) are highly derived molluscs with the largest nervous systems found in invertebrates. They present a suite of innovations absent in other clades, including prehensile, sucker-laden arms and a sophisticated adaptive coloration system. Notably, they have the largest brain of any invertebrate, with dozens of functionally distinct lobes. How this unusual body plan, with this extraordinary nervous system, is patterned at a molecular level is only starting to be understood.

To investigate the molecular basis of brain and body patterning in the cephalopods, we studied the expression of a number of highly conserved transcription factors and ligands during the development of the California two-spot octopus, *Octopus bimaculoides*. We found a striking correspondence in the expression of the MHB transcription factors along the highly derived octopus body plan to what has been described in other, distantly related animals. In fact, these transcription factors molecularly identify a morphological boundary in the developing cephalopod embryo, the cleft, which we describe here. Strikingly, the cleft, like the isthmus organizer, is a signaling center, but it expresses a different suite of ligands than what has been described at the equivalent MHB/IsO in other animals.

RESULTS

The clefts are morphological landmarks on the anterior-posterior axis of the octopus embryo

Octopuses and other cephalopods have large, yolky eggs, and undergo bilaterally symmetric, meroblastic cleavage that produces a cap of blastomeres at one end of the yolk. The embryonic cells continue to divide, covering the yolk through epiboly. As epiboly progresses, a central disc that will give rise to the embryo becomes distinct from the extraembryonic ectoderm. The organ primordia appear from this central disc with the body plan fully laid out: the mouth and arm primordia are located at the periphery and the mantle is at the center.

The developing central nervous system is found in the tissue between the central mantle and the peripheral arm primordia. The cerebral and optic cords are located anteriorly (Figure 5.2.1, purple), while the palliovisceral and pedal cords are located posteriorly (Figure 5.2.1, red and blue). These two regions of the developing embryo are separated by a furrow, extending from the developing statocyst at the edge of the embryo towards the central midline on either side, arranged like the wings of a chevron, which we refer to as the clefts.

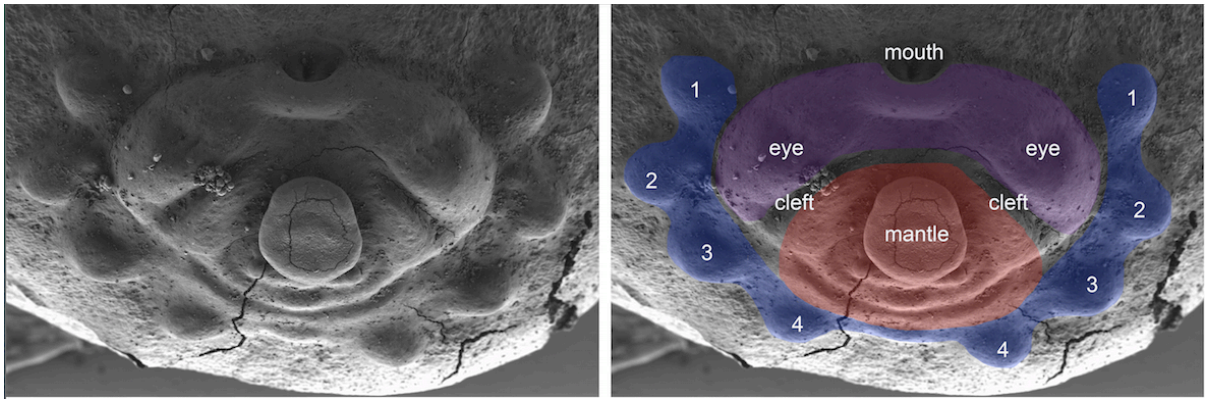


Figure 5.2.1. The octopus body plan in stage 8 embryos.
A scanning electron micrograph of a stage 8 octopus embryo, highlighting the arrangement of organ primordia on top of the large yolk, with the mouth at the anterior oriented to the top, and the mantle at the center. Arms are arrayed at the periphery, highlighted in blue and marked as 1-4. The anterior structures are highlighted in purple, including the eyes and optic and cerebral cords, and the posterior structures of the central body are highlighted in red. The clefts are the pronounced depression in the ectoderm separating these two territories.

The clefts are not neuronal

The clefts pass between the anterior cephalic and optic cords and the posterior palliovisceral cord. Examination of the expression of the panneural marker *oELAV*, as well as the expression of the neural differentiation markers *oNEUROD* and *oNEUROG,* indicate, however, that the cleft itself is free of neurons.

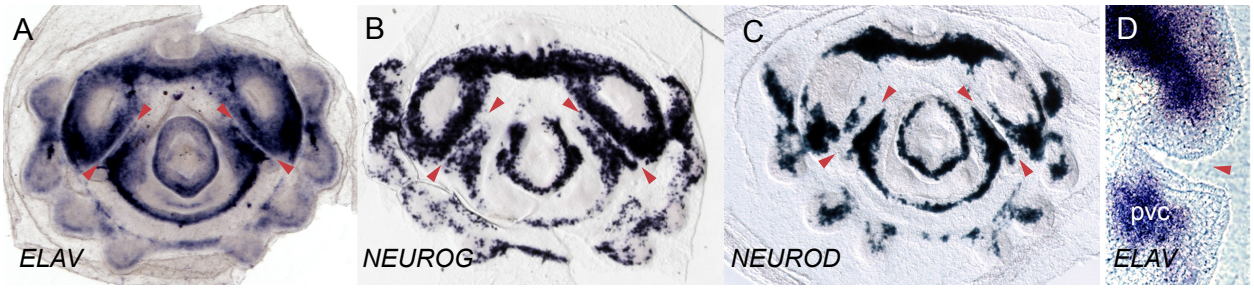


Figure 5.2.2. Expression of neural determination genes.
ELAV (A), *NEUROG* (B), and *NEUROD* (C) expression in stage 8 octopus embryos mark neural territories. (D) Section of stage 9 *ELAV* wholemount expression, with anterior to the top, and the palliovisceral cord (pvc) at the bottom. The image in (D) was supplied by Dr. Rahul Parnaik. Red arrowheads mark the clefts, which lack expression of these genes.

The clefts mark a molecular boundary for transcription factor expression

To study the molecular underpinnings of body plan patterning in octopus, we analyzed the expression of a number of transcription factors that show restricted expression along the anteroposterior axis in other bilaterians, including *OTX, PAX2/5/8, PAX6, EN, GBX,* and *HOX.*

Anterior markers

By stage 8, *OTX* is detected throughout the anterior region of the embryo, and is expressed strongly in the eye and optic lobe primordia, as well as the cephalic cord (Figure 5.2.3A). The expression stops immediately anterior to the cleft (Figure 5.2.3D). *PAX6* is also

strongly expressed anteriorly in the embryo, particularly in the cerebral cord and throughout the eye fields, and forms a sharp boundary with the clefts (Figure 5.2.3B).

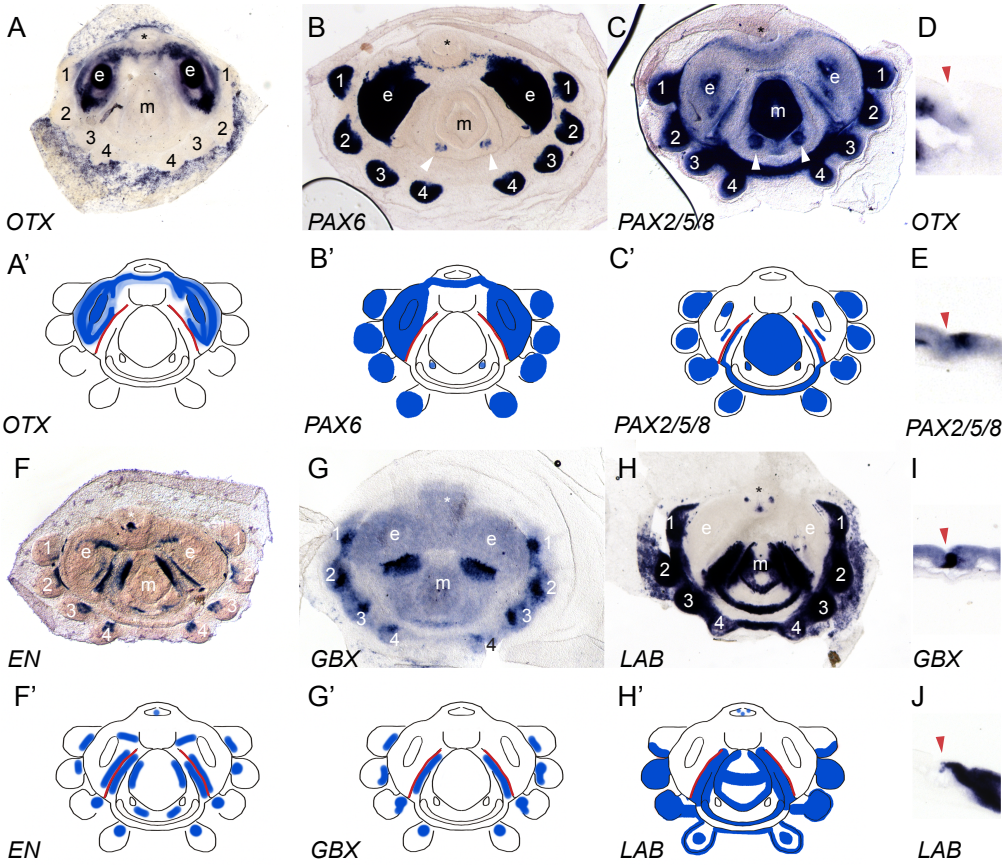


Figure 5.2.3. Transcription factor expression in the clefts at organogenic stages. Expression of *OTX* (A, A' D) is present anteriorly in the embryo, especially in the eyes, and the cerebral and optic cords. In addition, *OTX* is expressed throughout the extraembryonic ectoderm. *PAX6* (B, B') is strongly expressed throughout the embryo anterior to the cleft, in the arms and in the gills. *PAX2/5/8* (C, C', E) labeling is intense in the arms, mantle, gills, pedal cord, and posterior to the clefts. Expression of *EN* (F, F') is found posterior to the cerebral cord, in the clefts, and posterior to the clefts, as well as in the arm buds. *GBX* (G, G', I) is expressed in and behind the clefts and in the arm primordia. *LABIAL* (*LAB,* H, H', J) is intense in the palliovisceral cords and throughout the territory posterior to the clefts, and in the arms, and in three spots in the mouth. Parasagittal cross sections through the cleft (D, E, I, J) illustrate the expression of these genes at this boundary. Red arrowheads in the parasagittal sections mark the site of the cleft, with the tissue to the right of the arrowhead being posterior to the cleft. Red lines in the schematics indicate the location of the clefts. Arms are marked 1-4; e, eyes; m, mantle; *, mouth. White arrowheads in B and C indicate expression in the gills.

Cleft markers

EN is expressed initially in a broad swath just posterior to the future clefts. It later refines into a series of stripes radiating out from the center of the embryo: one posterior to the cephalic cord, one just anterior to the cleft, one in the cleft, and in the developing palliovisceral cord (Figure 5.2.3F). *GBX* is in the cleft immediately to the posterior in the palliovisceral cord (Figure 5.2.3G).

At organogenic stages, *PAX2/5/8* is expressed in the developing eyes, in a small stripe between the eyes and the clefts, and at the posterior of the cleft, as well as in the arms, pedal cord, gills and mantle (Figure 5.2.3C). *PAX2/5/8* appears in the cleft later than *EN* or *GBX*.

LABIAL is expressed posterior to the cleft from the earliest stages of detection. At organogenic stages, it is expressed in a broad swath posterior to the clefts, including in the mantle, palliovisceral and pedal cords, and the arm primordia (Figure 5.2.3H).

The clefts are a signaling center in the octopus embryo

This arrangement of genes expressed around the cleft bears a striking resemblance to the transcription factor profile of the MHB in vertebrates. In vertebrates, *OTX* and *GBX* are expressed in abutting domains, meeting at the MHB, where *PAX2*, *PAX5* and *PAX8* are coexpressed with *EN1* and *EN2*. The MHB is the location of the isthmus organizer in vertebrates, which expresses *WNT1* and *FGF8*. However, when we examined the expression of these signaling ligands, we found that *WNT1* was not expressed in the cleft region, and was instead found in the appendage primordia and the developing anal region (Figure 5.2.4).

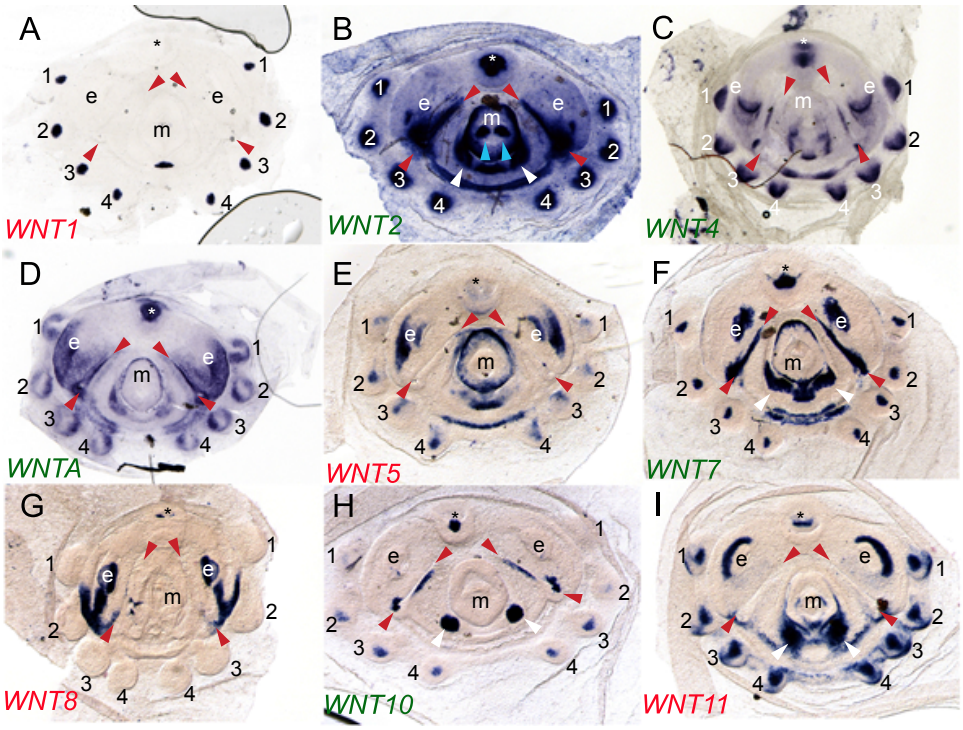


Figure 5.2.4. Expression of octopus *WNT* genes at stage 8.
Anterior is to the top of all panels. *WNT1* (A) is expressed in the arm and anus primordia, but is absent from the clefts (marked by red arrowheads throughout). *WNT2* (B) is expressed in the mouth, arms, within and anterior to the clefts, in the palliovisceral cord, statocysts (located at the lateral edge of the cleft), mantle and gills. *WNT2* is also expressed in the transient fin buds, indicated by blue arrowheads (Naef, 1928). Expression of *WNT4* (C) is found in the mouth, posterior eye fields, anus, mantle, clefts, statocysts, and distal half of the prospective arms. *WNTA* (D) is detected laterally within the optic territories, mouth, clefts, and in anterior and posterior stripes in the arm buds. *WNT5* (E) is expressed in the optic cords, retina, anus, palliovisceral cord, mantle, and in the center of the arm buds. Detection of *WNT7* (F) is observed in the mouth, eyes, arm buds, cleft, anus, and central portion of the palliovisceral and pedal cords. *WNT8* (G) expression is restricted to the optic territories. *WNT10* (H) is detected in the clefts, statocysts, gills and arm buds, and *WNT11* (I) is expressed in the developing eyes, arm buds and gills. Gene names in green indicate presence of expression in the clefts, while red indicates absence of expression in the clefts. Arms are marked 1-4; e, eyes; m, mantle; *, mouth. Red arrowheads mark the clefts and white arrowheads point to expression in the gills.

FGF8/17/18, on the other hand, is broadly expressed in the ectoderm throughout the embryo (Figure 5.2.5A,C), including the clefts. Examination of the only FGF receptor in the octopus genome revealed a similar low-level expression throughout the mesoderm underlying the ectoderm, but also exhibited elevated expression posterior to the clefts (Figure 5.2.5B,D).

FGFRs are receptor tyrosine kinases that activate intracellular cascades, including the mitogen activated protein kinase pathway. A key member of this pathway, ERK, is phosphorylated in response to signaling. We found elevated phosphorylated ERK staining posterior to the clefts (Figure 5.2.5F), as well as expression of the FGF responsive gene *SPROUTY* in the clefts and in other tissues where *FGFR* is expressed (Figure 5.2.5E). These results suggest that a FGF might play a role in signaling from the cleft. It will be important to study the expression patterns of the two other octopus FGFs, *FGF1/2* and *FGF9/16/20*, in octopus embryos.

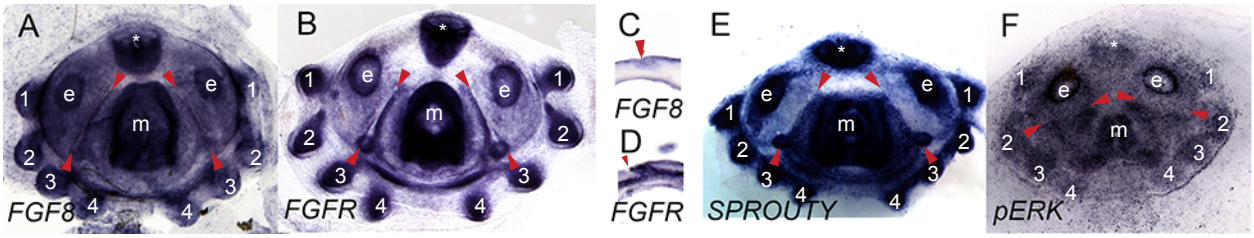


Figure 5.2.5. FGF signaling during organogenic stages of *O. bimaculoides* development. (A) *FGF8* expression is detected broadly throughout stage 8 embryos. *FGF8* expression is particularly strong in the mouth, eyes, and mantle and arm buds. (B) The single FGF receptor in the octopus genome (*FGFR*) is also expressed throughout the octopus embryo, most prominently in the eyes, mouth, mantle, arms, clefts and statocysts. Sagittal sections of *FGF8* (C) and *FGFR* (D) through the clefts (marked with a red arrowhead) illustrate the expression patterns of these genes. *FGF8* is expressed superficially, with slightly deeper expression at the cleft. *FGFR* is expressed more deeply in the embryo, except at the cleft, where it is expressed at the surface and in deeper tissue. (E) *SPROUTY* expression in octopus is detected in the mouth, eyes, arms, statocysts and in the tissue posterior to, and including, the clefts. (F) phospho-ERK staining in stage 7 embryo is detected in the mouth, eyes, arm buds, mantle and clefts. Immunostaining was carried out as described in Gabay and Shiloh (1997). Arms are marked 1-4; e, eyes; m, mantle; *, mouth. The location of the cleft is indicated by red arrowheads.

In addition to the canonical isthmus organizer ligands, we examined the expression eight of the ten remaining *WNT* genes found in the octopus genome (Albertin et al., 2015). While *WNT1* is not expressed in the clefts, we found strong expression of *WNT2, WNT4, WNTA, WNT7,* and *WNT10* in these structures. *WNT2* and *WNTA* are expressed in a wedge adjacent to

the anterior edge of the clefts, while *WNT4* and *WNT7* are expressed in the clefts. *WNT10* is found in a stripe along the medial posterior edge of the clefts.

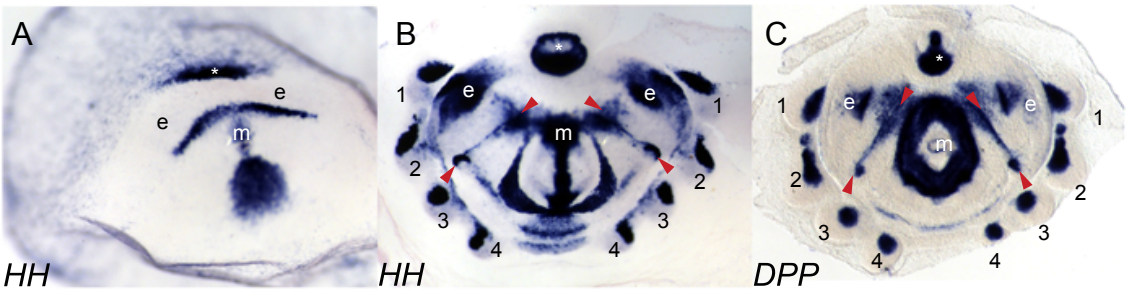


Figure 5.2.6 Expression of *HH* and *DPP* during octopus development.
(A) Expression of *HH* in stage 6 octopus embryos shows expression in the mouth and prospective clefts and mantle. (B) At stage 8, *HH* is expressed in these territories, as well as in the anterior side of the arm buds, eyes, and palliovisceral cord. (C) *DPP* expression at stage 8 is detected in the mouth, eyes, mantle, statocysts and arm buds. Arms are marked 1-4; e, eyes; m, mantle; *, mouth. Clefts are indicated with red arrowheads.

Outside of the WNTs, we surveyed the expression of two other, highly conserved developmental signaling ligands, *HH* and *DPP*. *HH* is expressed from before the appearance of organ primordia in the prospective mouth and gut, as well as in an intermediate territory that presages the site of cleft formation (Figure 5.2.6A). This expression pattern becomes stronger and more elaborate as organogenesis proceeds, and is later in the eyes, arms and throughout the palliovisceral cord (Figure 5.2.6B), as was also demonstrated in the cuttlefish *Sepia officinalis* (Navet et al., 2009). At these stages, we also detect *DPP* expression in the clefts as well as in the eyes, mantle, mouth, statocysts and arm primordia (Figure 5.2.6C). The *DPP* expression is found anterior to the clefts medially, as well as in the posterior of the clefts.

DISCUSSION

Our results show that the molecular patterning system underlying the remarkable cephalopod brain is largely congruent with observations from other bilaterians. The anterior neurogenic domain is characterized by the expression of *OTX* and *PAX6* while the trunk is patterned by expression of *HOX* genes, with an intermediate domain of *PAX2/5/8* and *GBX* expression separating the supraesophageal brain primordia from the subesophageal brain and trunk. This overall pattern of anterior expression of *OTX/PAX6*, intermediate *EN, PAX2/5/8* and *GBX*, and posterior *HOX* genes, has been observed in deuterostomes and ecdysozoans (Reichert, 2005), and has been interpreted as indicative of the homology of a tripartite (or three part) brain. Here, we report the similar expression of these genes in a mollusc, dividing the anterior from the posterior brain, while the intermediate territory appears non-neurogenic. This supports a homology of the "tripartite" developmental gene regulatory network.

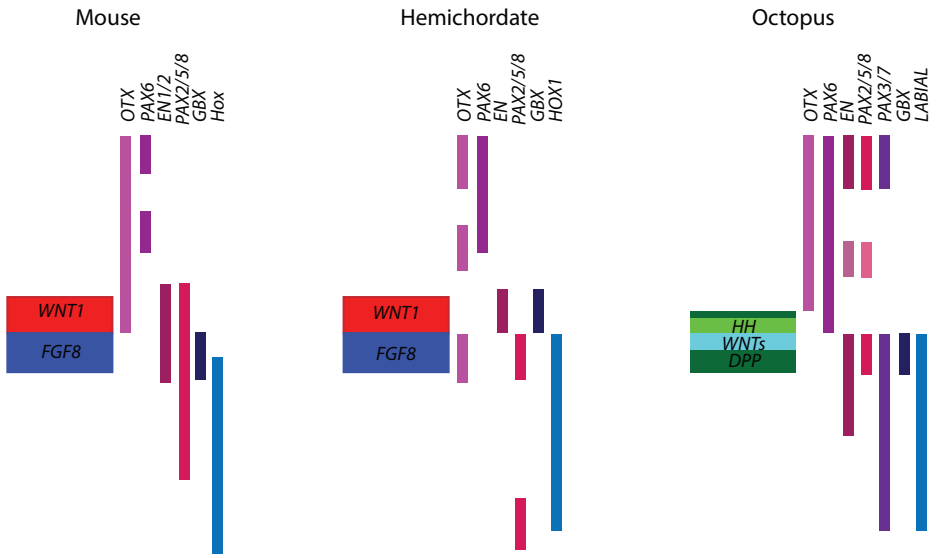


Figure 5.2.7. Schematic of expression of the anteroposterior patterning cassette in mice, hemichordates, and octopus.
Anterior is to the top. Signaling ligands are indicated to the left, and transcription factors to the right. Mouse and hemichordate data after Pani et al. (2012).

The intermediate territory separating the anterior and posterior neural structures is a morphological depression in cephalopods, which we refer to as the clefts. The clefts are thought to give rise to the nuchal crest, a non-neural structure that is the junction between the head complex and the visceral mantle. More importantly, the cleft is a morphological boundary between the anterior cerebral and optic cords which become the supraesophageal mass and the optic lobes (Shigeno et al., 2015) and the palliovisceral cord, which contributes to the posterior subesophageal mass. This is the major division in the cephalopod brain.

The clefts are first morphologically evident at stage 7, as the body plan begins to emerge, coincident with the expression of *ELAV* in the neural placodes. These placodes thicken as nascent neurons ingress, but the cleft, as a non-neuronal structure, does not, giving it the appearance of a depression in the emerging octopus body plan. However, the clefts are presaged by the expression of *HH* at stages 5 and 6 as wings extending laterally from the midline (Figure 5.2.6A) before they become apparent morphologically. We can detect transcription factor expression in the clefts at stage 7, as the body plan is emerging. Early transcription factor expression along this boundary includes *PAX6*, *EN,* and *GBX.* Other ligands and transcription factors expressed in the clefts are detected soon thereafter.

Notably, the territory identified in hemichordates by this suite of transcription factors also delineates a morphological boundary between the collar and the trunk (Pani et al., 2012). In brachiopods, a spiralian clade, this signal is found in the apical/mantle (or head/trunk) boundary as well (Vellutini and Hejnol, 2016). In both of these animals, this region does not identify a transition in the nervous system, but a boundary in the body plan. This may suggest a plesiomorphic role in patterning the anteroposterior axis in bilaterians, rather than nervous system patterning. In cephalopods, the neurogenic ectoderm is widespread (Figure 5.2.2), so

delineating between a role in axial patterning as opposed to neural patterning will be difficult to resolve in this group, though the cleft itself appears to be non-neuronal. Studies of the expression of *EN* and *PAX2/5/8* in other molluscs also support a role for these transcription factors in body patterning, as they are not detected in the nervous system (Jacobs et al., 2000; Wollesen et al., 2015), but broader taxonomic sampling will be crucial in resolving the ancestral role of these patterning genes.

This territory is not only a morphological boundary, but also a molecular one, separating anterior and posterior gene expression territories. We found that the clefts are a signaling center, expressing at least eight ligands. Some of these signals, like *FGF8*, appear to be conserved with other animals, including a brachiopod (Vellutini and Hejnol, 2016), vertebrates, and hemichordates (Pani et al., 2012). Others, such as *HH* and *DPP*, have not been described at this molecular boundary in other animals. These results provide support a signaling center role for this boundary in the last common ancestor of bilaterians animals, and suggest a role for lability of signaling ligands in the evolution of novel body plans.

PART III

The molecular architecture and evolution of the octopus brain

INTRODUCTION

Of the many striking features of cephalopods, their brains may be the most impressive: they have the largest and most complex central nervous systems among invertebrates. The central brain of *Octopus vulgaris*, comprising the supraesophageal, subesophageal, and optic lobes, contains more than 120 million neurons, and the axial nerve cords have another 350

million. This remarkable nervous system is divided into dozens of functionally distinct lobes: the vertical lobe, for example, has been implicated in visual learning and memory (Young, 1971), while the anterior basal lobe plays a role in the generation of complex movements.

How this most complex of invertebrate brains evolved, and how it relates to brains found in other clades, has not been explored using modern molecular techniques. In the last decade, a number of studies have found correspondence at the molecular level between the brains of vertebrates, annelids, and arthropods, prompting the assertion of homology (Denes et al., 2007; Strausfeld and Hirth, 2013; Tessmar-Raible et al., 2007; Tomer et al., 2010). These studies conclude that the sensory-associative territories of the vertebrate pallium and the mushroom bodies of arthropods and annelids (Tomer et al., 2010), as well as structures associated with higher motor control such as the vertebrate basal ganglia and the arthropod central complex (Strausfeld and Hirth, 2013), are derived by common descent from the brain of the last common ancestor of bilaterally symmetrical animals. These conclusions are based on work in three of the four major metazoans clades with complex brains (chordates, arthropods, and annelids; Figure 5.3.1); the fourth is the molluscs.

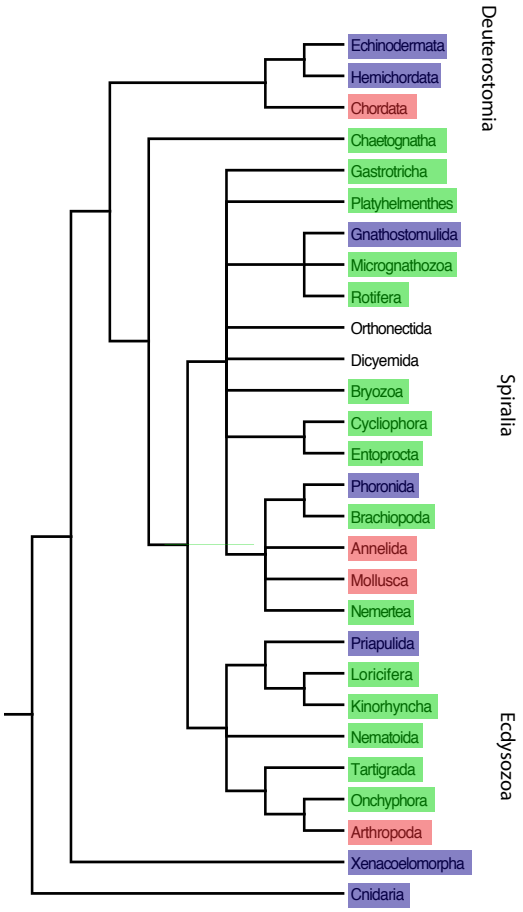


Figure 5.3.1. Distribution of nervous system types across the eumetazoa.
Type of nervous system characteristic of the phylum (brain, ganglia, or diffuse) highlighted in red, green and blue, respectively, based on Northcutt (2012). Phylogeny based on Dunn (2015). Xenacoelomorphs demonstrate both diffuse and ganglionic nervous system, though the ancestral state is thought to be diffuse (Perea-Atienza et al., 2015; Semmler et al., 2010). Molluscs also have a range of nervous system centralization, including cords, ganglia, and large, complex brains. Orthonectids and dicyemids are secondarily simplified and lack nervous systems (Aruga et al., 2007; Hanelt et al., 1996).

The plesiomorphic state of the molluscan nervous system is considered to be an anterior cord with four longitudinal cords (tetraneury) (Brusca and Brusca, 2003; Bullock and Horridge, 1965; Young, 1971). New molluscan phylogenies suggest that cephalopods and monoplacophorans form a sister group within the conchiferans, outside of a clade that contains bivalves, gastropods and scaphopods (Kocot et al., 2011; Smith et al., 2011). This separates the traditionally allied

cephalopods and gastropods, which both express some degree of cephalization, suggesting that this feature is convergent between these two groups, and also convergent with the centralized brains found in other clades. The lobes of the cephalopod brain are not reflected in the nervous systems of other molluscs, suggesting that this large, highly differentiated nervous system is a cephalopod novelty.

Here, we study the expression of molecular markers of three areas of the octopus brain: the neurosecretory centers, the basal lobes, and the frontal-vertical system. We found evidence that the neurosecretory and higher motor centers of the octopus brain share a molecular signature with those of other bilaterians, while the sensory-associative learning and memory systems appear to be quite different.

RESULTS

Neurosecretory centers

Neurosecretory centers secrete neuropeptides and other signals into the surrounding tissue or the vasculature, controlling a range of processes, including growth, metabolism, sexual maturation and reproduction. Neurosecretory structures are found in a broad range of animals, including vertebrates, insects and annelids. *Orthopedia (OTP)* has been shown to play an important role in the development of the hypothalamus, an endocrine structure in vertebrates (Acampora et al., 2000; Del Giacco et al., 2008). A shared transcription factor profile (including *OTP*, *NKX2.1,* and *RX*) in the hypothalamus in vertebrates and neurosecretory cells in annelids, as well as the expression of cell type markers such as vasopressin-neurophysin, has led to the

suggestion of a common bilaterian origin of the neurosecretory cell type (Tessmar-Raible et al., 2007).

The octopus has several neurosecretory structures, including the buccal and subpedunculate lobes in the supraesophageal brain, as well as a number of other tissues and glands located in other regions of the nervous system and the body. While we were unable to identify an *RX* homolog in octopus, we did examine the expression of *OTP* and *NKX2.1* in embryos and hatchling brains. We detected *OTP* expression in stage 16 embryos in the prospective superior buccal and the subpedunculate lobes (Figure 5.3.2A-B). This expression continued in the hatchling brain in the superior buccal lobe (Figure 5.3.3A-B), and is also found in the anterior and posterior subesophageal brain. *NKX2.1* is broadly expressed in the supraesophageal brain, particularly in the inferior frontal lobe, the buccal lobe, the basal lobes, and the vertical lobe. In addition, we were able to detect *ISLET* expression, which is expressed in these regions in at both stage 16 and in hatchling brains. This combinatorial expression of *OTP* and *NKX2.1* in centers known to have endocrine functions in the octopus brain suggests a conserved molecular signature of neurosecretory cells shared with other animals.

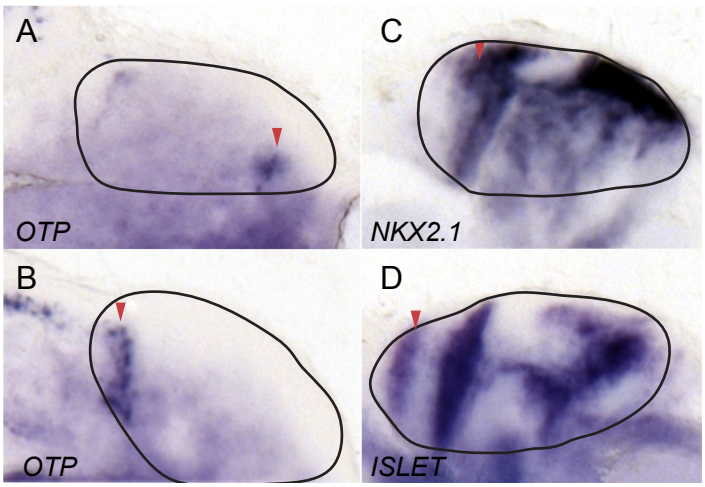


Figure 5.3.2. Expression of neurosecretory markers in octopus embryos.
Sagittal sections of whole mount *in situ* hybridization of stage 16 embryos for *OTP* (A, B), *NKX2.1* (C), and *ISLET*. Red arrowhead in (A) indicates position of subpedunculate lobe identified with *OTP*. Red arrowheads in (B-D) indicate the position of buccal lobe. (A) is a lateral section through the same embryo as (B). Embryos are oriented with the mouth to the left and the supraesophageal brain (circled) facing to the top.

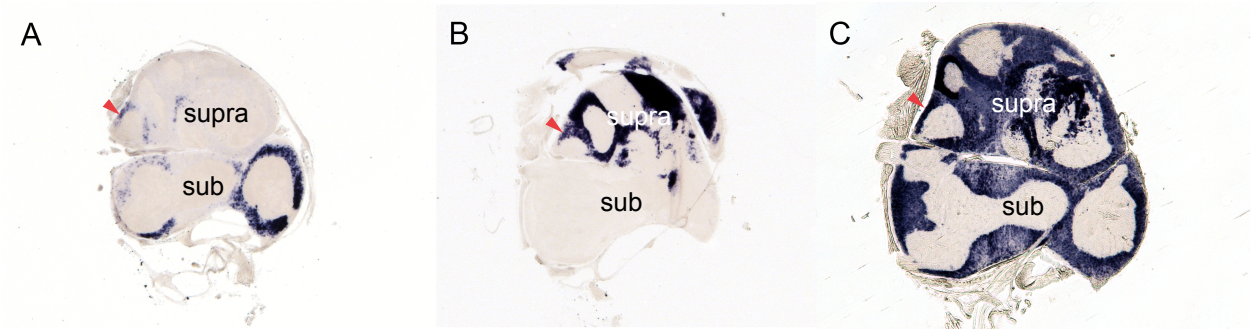


Figure 5.3.3. Expression of neurosecretory markers in octopus hatchling brains.
Sagittal section *in situ* hybridization of 2 week old hatchling brains for *OTP* (A), *NKX2.1* (B), and *ISLET* (C). Red arrowheads indicate the position of the buccal lobe. Brains are oriented with the mouth to the left and supraesophageal brain to the top.

Higher motor centers

From the jet-propelled escape response and their fierce attacks to their remarkable camouflage abilities, cephalopods can elicit spectacular movements. The control of these movements is carried out at several levels: so-called 'lower' motor centers contain the neurons that directly innervate the muscles, such as the axial nerve cords (Wells, 1978). 'Intermediate'

motor centers can produce movement from groups of muscles: stimulation of the brachial lobe in the subesophageal brain results in contraction of all of the arms at once. The 'higher' motor centers, found in the basal lobes of the supraesophageal brain, are concerned with controlling complex patterns of movement. Young (1971) recognized distinct lobes within this complex, including the anterior basal, interbasal, median basal and lateral basal. The anterior basal lobe controls posture and movement of the head and eyes during locomotion, and also plays a role in the movements of the arms during prey capture and feeding. The median basal lobes control respiration and swimming, and skin patterning is controlled by the lateral basal lobes (Young, 1971).

Higher motor centers are present in the brains of other bilaterians, including the central complex of arthropods and the basal ganglia of vertebrates. In vertebrates, the structures of the basal ganglia arise from the ventral forebrain, and are characterized by the expression of a number of highly conserved transcription factors, including *DLX* and *NKX2.1*. We assayed these markers in the developing octopus brain, and found *DLX* and *NKX2.1* are expressed at stage 16 in the anterior basal lobe (Figure 5.3.4), and *NKX2.1* expression persists in the hatchling brain (Figure 5.3.3B). These data indicate a molecular similarity with the higher motor centers of other animals.

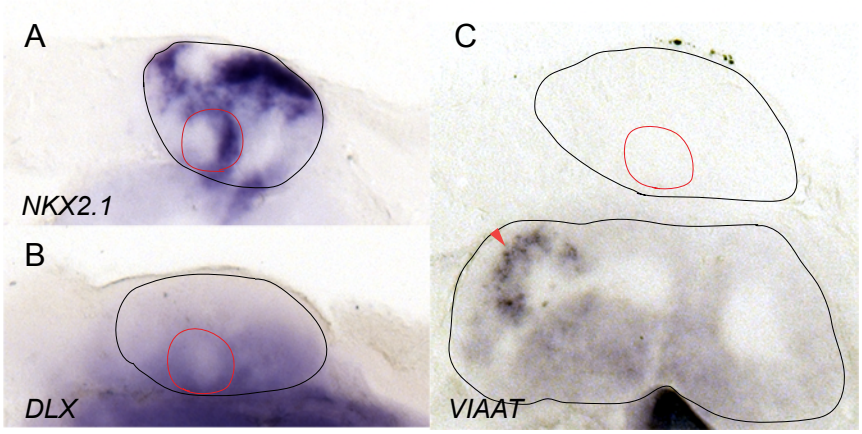


Figure 5.3.4. Expression of markers of higher motor centers in the central brain of stage 16 octopus embryos.
Expression of *NKX2.1* (A) and *DLX* (B) in the supraesophageal brain (outer circle. The anterior basal lobe is indicated in the red circle. *VIAAT* (C) is not expressed in the supraesophageal brain, but it is expressed in the anterior subesophageal lobe (indicated with a red line).

Inhibitory GABAergic neurons are a prominent feature of the vertebrate basal ganglia (Reiner et al., 1998). To investigate if the cephalopod basal lobes share not only a molecular signature, but also the cell type, characteristic of higher motor centers in vertebrates, we examined the expression of two markers of GABAergic cells: glutamic acid decarboxylase (Wissler et al., 2013), which is an enzyme involved in GABA synthesis, and the GABA vesicular transporter *VIAAT*. *VIAAT* is absent from the supraesophageal brain of stage 16 octopus embryos, though it is expressed in the anterior subesophageal lobe (Figure 5.3.4C). While we were able to detect expression of both *GAD* and *VIAAT* in overlapping areas of hatchling brains, both are detected in scattered cells of the dorsal basal lobes but are absent from all the other motor lobes of the supraesophageal brain, including the anterior basal lobe (Figure 5.3.5). These results are consistent with studies of GABA-immunoreactivity in the northern octopus *Eledone cirrhosa*, which found an absence of GABA-positive cell bodies in the anterior, median, and lateral basal lobes (Cornwell et al., 1993). This suggests that while the octopus basal lobes share the

transcription factor profile of the higher motor centers found in other animals, they differ from the insect central complex and the vertebrate basal ganglia, which are greatly enriched in GAD+ fibers.

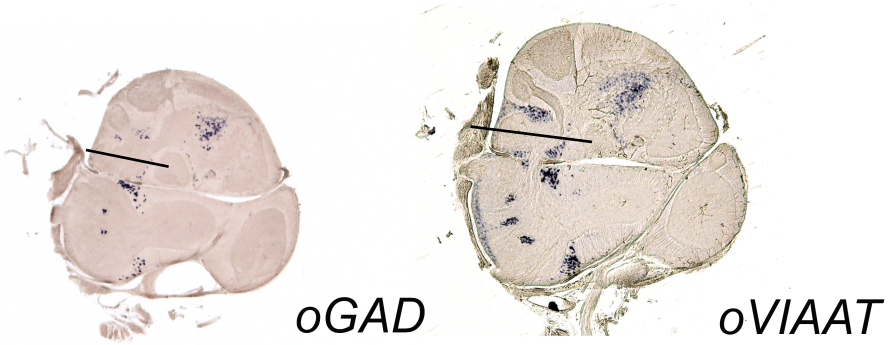


Figure 5.3.5. Markers of higher motor centers in octopus hatchling brain.
GAD and *VIAAT* are detected in hatchling brains in the supraesophageal and subesophageal brain, but are absent from the anterior basal lobe (indicated by a black line). Brains are oriented with the mouth to the left and anterior to the top.

The learning and memory centers

Cephalopods demonstrate formidable learning and memory capabilities (Hanlon and Messenger, 1996; Wells, 1978; Young, 1971). Lesion experiments in the 20th century demonstrated that the frontal-vertical system plays a critical role in these abilities, analogous to the cerebral cortex and hippocampus of mammals. Both the vertebrate pallium and the sensory-associative centers of annelids (mushroom bodies) express *PAX6* and *EMX*, both homeodomain transcription factors, which has led to the suggestion that this shared molecular 'fingerprint' supports the homology of these structures (Tomer et al., 2010). While *D. melanogaster* mushroom bodies also express *PAX6*, they lack *EMX*. The fly and annelid mushroom bodies and the vertebrate pallium also all express *DACH*.

To test whether the cephalopod learning and memory centers share this signature, we examined the expression of these genes in embryonic and hatchling brains. While *PAX6* and *EMX* are expressed in the supraesophageal brain, they do not overlap, and they are not expressed in the lobes of the frontal-vertical system. *EMX* is expressed in the buccal lobes, while *PAX6* and *DACH* are expressed in the anterior basal lobe (Figure 5.3.6) We did not detect any of the pallial and mushroom body markers in the nascent frontal or vertical lobes in octopus. Examination of the expression of these genes in hatchling brains also found that they are absent in the frontal-vertical system (Figure 5.3.7). However, expression of a suite of other genes, including *NKX2.1* and *ISLET* are present in these territories in both embryos and hatchlings (Figure 5.3.2-3).

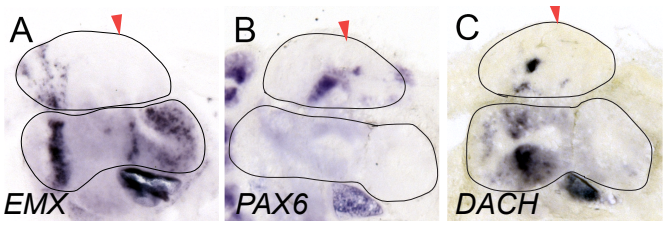


Figure 5.3.6. Expression of higher sensory associative markers in stage 16 octopus embryos.
(A) *EMX*, (B) *PAX6*, and (C) *DACH* expression in the brain of octopus embryos. Supraesophageal brain is outlined on top, subesophageal brain is outlined on the bottom, and the red arrowhead indicates location of the vertical lobe.

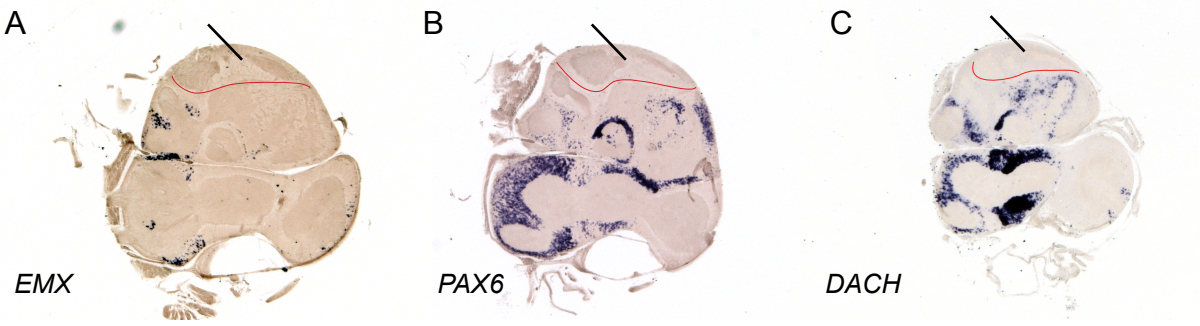


Figure 5.3.7. The octopus hatchling frontal-vertical system does not express *EMX* (A), *PAX6* (B) or *DACH* (C).
Black line indicates location of the vertical lobe, which is outlined with a red line.

DISCUSSION

The large, complex cephalopod brain has long been considered an example of convergent evolution with vertebrates (Packard, 1972). We investigated the molecular underpinnings of this remarkable structure, studying the expression of highly conserved developmental transcription factors as well as markers of cell function that have been found to be expressed in bilaterian brain structures. Our examination of areas of the octopus brain involved in endocrine functions, motor control, and sensory associative learning and memory illustrated both conservation and convergence.

The identified neurosecretory centers located in the octopus central brain demonstrated expression of transcription factors, including *OTP* and *NKX2.1*, that are also known to be expressed in neurosecretory cells in annelids and vertebrates (Tessmar-Raible et al., 2007). Conservation of *OTP* expression in neurosecretory centers between *O. bimaculoides* and other bilaterians is consistent with the interpretation of homology at the level of developmental control genes (Tessmar-Raible et al., 2007). While the only study of *OTP* in a non-cephalopod mollusc (the snail *Patella* (Nederbragt et al., 2002*)*) did not describe a neurosecretory role for this gene, further studies in diverse molluscs, including aculiferans, will be informative.

The higher motor centers found in the octopus basal lobes also share a transcription factor signature, notably of *DLX* and *NKX2.1*, which are also found in the vertebrate basal ganglia and the insect central complex (Strausfeld and Hirth, 2013). However, the motor centers of the octopus appear devoid of the inhibitory GABAergic neurons characteristic of the basal ganglia (Cornwell et al., 1993; Reiner et al., 1998). These results may be interpreted in several ways. The presence of these transcription factors and cell types may represent homology (Shubin et al., 2009) of the higher motor structures across the Bilateria, with loss of the cell type in the octopus

basal lobes. Alternatively, these results may suggest the presence of a deep homology of higher motor centers: that is, that the molecular and developmental underpinnings (in this case, the transcription factor signature) may be homologous, but they pattern phylogenetically distinct structures. Support for this explanation may be found in the differences in the architecture and the role of GABAergic cells between the insect central complex and the vertebrate basal ganglia (Farries, 2013). While these interpretations are not mutually exclusive, better phylogenetic sampling may help resolve this issue.

Of key interest will be work in non-cephalopod molluscs and other lophotrochozoans, the study of which may help elucidate the role of these genes in patterning the brains and bodies of this diverse group of animals. In particular, the molecular studies finding genealogical correspondence between the brains of vertebrates, annelids, and arthropods are largely based on evidence from these groups, all of which have highly differentiated brains. However, these represent only three of the roughly 35 major clades of bilaterians. Many of the other clades have less centralized nervous systems (Figure 5.3.1), which could represent either a plesiomorphic (Northcutt, 2012) or a secondarily reduced state (Strausfeld and Hirth, 2013). In the few taxa with less centralized nervous systems where developmental expression of toolkit genes has been explored, genes involved in neural patterning in vertebrates and flies tend to be expressed along the body axis (Pani et al., 2012; Santagata et al., 2012). This has been interpreted as supporting a common bilaterian coordinate system (Farries, 2013) or as secondary modification (Strausfeld and Hirth, 2013). Further sampling of these clades will be critical to evaluating homology and homoplasy in bilaterian nervous system evolution.

Regardless of the state of the urbilaterian brain, the brains of cephalopods clearly represent a departure from the ancestral molluscan nervous system of an anterior cord with four

longitudinal cords (tetraneury) (Moroz, 2009). Broad molecular characterization of molluscan taxa, particularly the aculifera, will aid in polarizing these characters in this vast, diverse group of animals. For example, the shared expression of *DLX* and *NKX2.1* in the cephalopod basal lobes, vertebrate basal ganglia and the arthropod central complex may be the result of homology or homoplasy, as discussed above. By understanding the molecular patterning of a wider range of molluscs, we may be able to infer if this 'molecular fingerprint' identifies higher motor centers in molluscs and other animals, which would support homology by descent of these structures, or if perhaps this shared expression is indicative of a shared gene regulatory network that provides positional information that patterns evolutionarily distinct structures.

Finally, we found that the cephalopod frontal-vertical system does not share diagnostic features of the developmental regulatory network found in the learning and memory centers in other animals. Homology of the annelid and arthropod mushroom bodies and the pallium of vertebrates has been suggested based on the shared expression of *EMX, PAX6* and *DACH.* While the pallium of vertebrates and the frontal-vertical system of cephalopods are both populated by millions of small amacrine cells, we found that the molecular signature of these two learning and memory centers are quite different: *EMX, PAX6* and *DACH* are absent (Figure 5.3.6-7). Instead, we detect expression of developmental markers in the frontal-vertical system that are absent in the vertebrate pallium or the annelid mushroom bodies, including *NKX2.1* and *ISLET* (Figures 5.3.2-3). The 'molecular fingerprint' of the developing and hatchling frontal vertical system of octopus suggests that this structure represents a clade-specific novelty.

As an evolutionary innovation that apparently does not share common suite of developmental and molecular markers with other animals, the frontal-vertical system of cephalopods represents an opportunity to explore functional, developmental, and evolutionary

constraints on learning and memory centers. Our findings in earlier in Chapter 5 have highlighted the conservative expression of transcription factors across the anteroposterior axis in *O. bimaculoides*, indicating that the early patterning of the nervous system is not drastically different from what has been described in other bilaterians. However, this initial conservation underlies structures, like the frontal-vertical system, that have no counterpart in other animals.

CHAPTER 6

Gene expression in embryonic *Octopus bimaculoides* reveals conserved and novel mechanisms for establishing the body plan

SECTION I

Octopus bimaculoides embryonic transcriptomes support the developmental hourglass model

INTRODUCTION

Cephalopods are extraordinary molluscs. They possess complex, camera eyes, the largest brains among the invertebrates, flexible, sucker-lined arms, and skin that can rapidly alter both color and texture to permit camouflage and communication. The developmental program underlying these novel features is, in itself, a striking departure from the rest of the molluscs. Molluscs, as well as other members of the Spiralia, share a common pattern of initial cell divisions known as spiral cleavage as well as a cell lineage fate map that is conserved between distantly related groups (Boyer and Henry, 1998; Hejnol, 2010; Henry, 2014). Following gastrulation, the embryo passes through one or more larval stages before taking on the adult morphology.

This highly conserved developmental program appears to be absent in cephalopods. Cephalopods have large, yolky eggs with meroblastic cleavage that have lost all trace of the spiral cleavage pattern (Naef, 1928; Wadeson and Crawford, 2003). The initial cell divisions create a blastomere cap that spreads over the large yolk by epiboly. As epiboly is progressing, the primordia of the adult body plan emerge simultaneously, without passing through a larval stage. These primordia are arrayed in a circle, with the dorsal mantle at the center and the ventral

arms at the periphery. How this highly derived body plan and its novel developmental program relate to those of other animals has been obscure.

Studies in distantly related animals have found that many developmental regulatory genes are conserved, and have shared roles in development. The best known among these are the Hox genes, a family of homeodomain transcription factors that are expressed in a conserved fashion along the anteroposterior axis in many bilaterians. Many of the "toolkit" genes that regulate developmental processes are transcription factors, including other homeodomain, Forkhead box (FOX), Tbox, and basic helix-loop-helix (bHLH) proteins, as well as highly conserved signaling ligands, such as the WNTs, Transforming Growth Factor βs (TGFβs), and Fibroblast Growth Factors (FGFs). These genes have been found to be present and play a role in the development of such morphologically and phylogenetically diverse groups as hemichordates and flies, vertebrates and annelids. However, relatively few developmental toolkit genes have been studied in molluscs, a group with the greatest range of body plans among all extant phyla. Thus far, these studies have largely highlighted divergent roles for these highly conserved developmental regulatory genes in molluscs (Baratte et al., 2007; Hinman et al., 2003; Lee et al., 2003; Navet et al., 2009; Navet et al., 2008; Perry et al., 2015; Samadi and Steiner, 2009; 2010). Only a handful of molecular studies of developmental control genes have been done in cephalopods (Baratte et al., 2007; Buresi et al., 2012; Lee et al., 2003; Navet et al., 2009; Navet et al., 2008), leaving their role in the development of the novel cephalopod body plan largely unexplored.

Fundamental questions regarding the molecular basis of the novel developmental program in this highly derived group remain to be answered. What role do developmental regulatory genes that are highly conserved among other bilaterian groups play in the establishment of the cephalopod body plan? How is the body plan set up at a molecular level,

and how can it be related to what we see in other animals? Did the evolution of cephalopod embryogenesis, and its departure from the ancestral molluscan spiral cleavage program, require the evolution of new genes?

To begin to characterize the genes involved in the development and elaboration of the cephalopod body plan, we have sequenced transcriptomes spanning *Octopus bimaculoides* embryogenesis. Transcriptomes not only provide the sequence of the genes expressed in the sampled tissue, they can also quantitatively measure gene expression, which can identify genes that are differentially deployed, and can provide an unbiased assessment of transcripts present. We selected five stages that represent distinct embryonic events: gastrulation (stage 2), epiboly (stage 4), emergence (stage 7) and elaboration of organ primordia (stage 10), and later development when a number of cephalopod-specific structures are present (stage 15). Using our transcriptome data, we took a two-pronged approach to identify genes that may play a role in establishing the octopus body plan. First, we explored the expression of candidate genes across stages. Leveraging the highly conserved developmental transcription factors and ligands identified in the genome (Chapter 4), we found that they are primarily expressed during the organogenic stages, but not before. Then, we bioinformatically identified differentially expressed transcripts across stages to explore suites of genes that show dynamic and stage specific expression profiles. Many of the stage-specific transcripts in the early transcriptomes are clade-specific, which may support a role for novel genes in this highly derived embryonic program.

RESULTS

The developmental toolkit is deployed during O. bimaculoides *organogenesis*

We performed a directed search for transcription factors and ligands genes that regulate development in other animals, focusing on the gene families identified in our exploration of the octopus genome in Chapter 4. The vast majority of these genes are present in the developmental stages sampled from *O. bimaculoides*. For example, we found expression of all eight octopus Hox genes in our transcriptomes. These genes are primarily expressed during and after organogenesis, but not before (Figure 6.1.1). These results correspond well with *in situ* hybridization studies, which did not detect expression of Hox genes at stage 4 or before (Chapter 6.2).

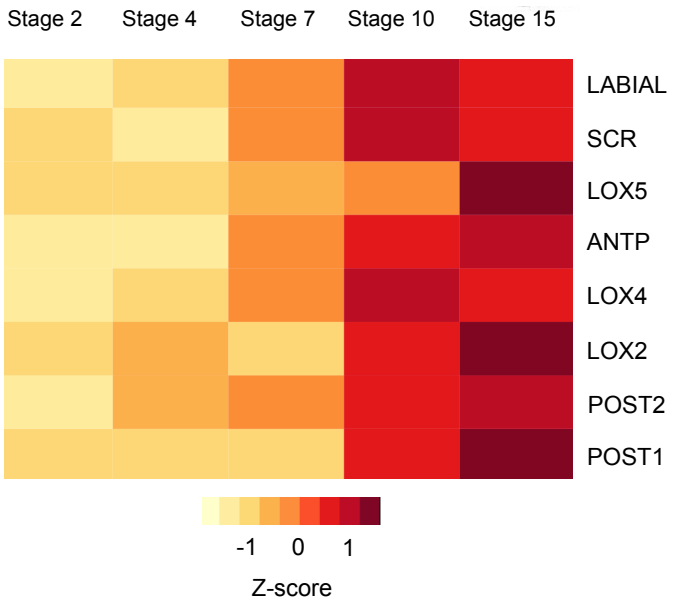


Figure 6.1.1 Heat map of Hox gene expression across five developmental transcriptomes. Expression profile of the 8 octopus Hox genes across stages 2, 4, 7, 10 and 15. Drawings from Naef (1928) of the selected stages are provided at the top. Cells are colored according to number of standard deviations from the mean expression level.

We profiled the expression of the other developmental transcription factor families, including the TBOX, homeodomain, FOX and the bHLH families (Figure 6.1.2). The majority of these transcription factors appear to be primarily expressed during or after stage 7, suggesting that most do not play pivotal roles in developmental events prior to the onset of organogenesis. This analysis does identify a handful of candidates from these families that may be involved in early developmental events. For example, the highly expressed TBOX gene at stage 2 is the octopus ortholog of *EOMESODERMIN*, a gene that plays an early role in mesoderm patterning in vertebrates (Ryan et al., 1996; Showell et al., 2004). As stage 2 is a gastrulation stage in octopus, this may suggest a conserved role for *EOMESODERMIN* in mesoderm development in this species. Other highly conserved transcription factors expressed at stage 2 include the homeobox genes *PAX2/5/8* and *PROSPERO*, as well as *FOXK,* and *FOXL*. *SOX14* and *SOHLH* (spermatogenesis and oogenesis specific basic helix-loop-helix) are also detected at stage 2. In mice, *Sohlh2* is a expressed in ovaries and testes (Ballow et al., 2006), and is critical factor for oogenesis (Choi et al., 2008), so its presence in early octopus embryos may support a similar role in cephalopods. Overall, however, many of the highly conserved metazoan developmental transcription factors appear to be expressed during or after the appearance of organ primordia, but not before.

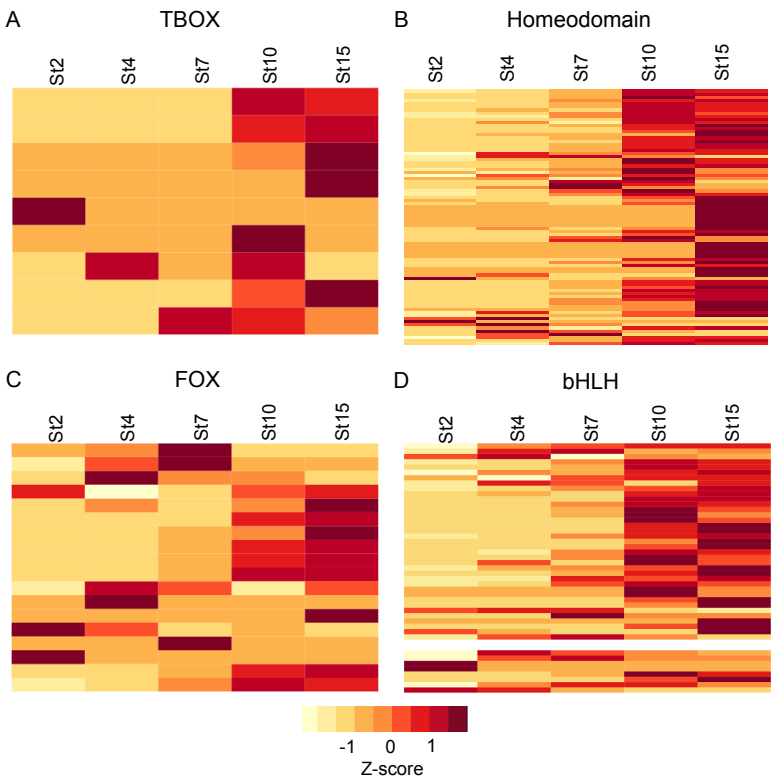


Figure 6.1.2 Heatmaps of expression of members of developmental transcription factor families.
Expression profile of members of the TBOX (A), homeodomain (B), FOX (C), and bHLH (D) transcription factor families. Cells are colored according to number of standard deviations from the mean expression level.

An exception to this pattern may be the nuclear hormone receptors (NHRs). We found 38 of the 40 NHRs in the octopus genome in our developmental transcriptome assembly. Of these, 10 NHRs demonstrated strong expression at stage 2 (Figure 6.1.3). Their expression in early transcriptomes may support a role for these genes in early patterning, and present attractive candidates for further study.

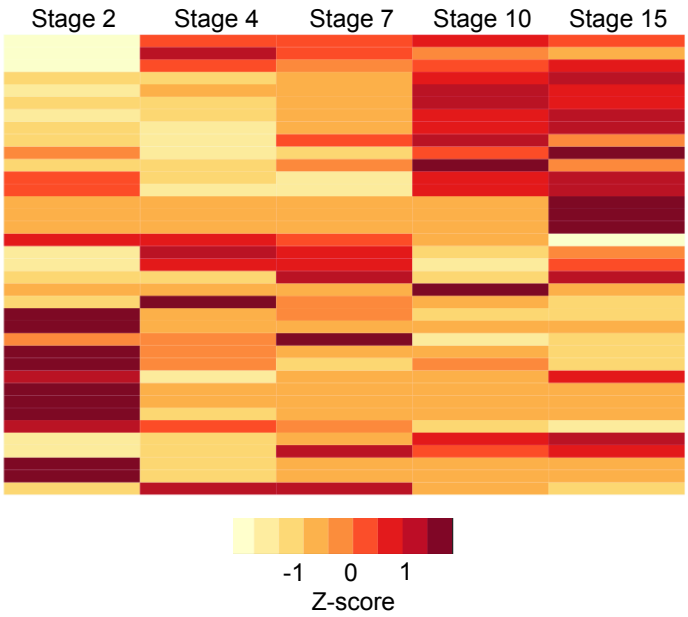


Figure 6.1.3 Expression of the nuclear hormone receptors during octopus development
Cells are colored according to number of standard deviations from the mean expression level.

The C2H2 transcription factors are the largest transcription factor family in animal genomes (Klug, 2010; Schmitz et al., 2016), and have been suggested to be important in the evolution of clade-specific traits (Liu et al., 2014; Nowick et al., 2010; Shannon et al., 2003). While there are a number of highly conserved C2H2 containing genes, the family as a whole appears to be quite dynamic in a number of animal genomes. The octopus has an exceptionally large C2H2 family: we recovered 1,790 C2H2 domain-containing genes (Albertin et al., 2015). Nearly 800 of these C2H2-ZNFs are expressed in the five different developmental stages sampled (Figure 6.1.4). As with the other developmental transcription factors investigated, we found that many of the C2H2-containing genes expressed during octopus development are expressed at stage 7 or later, but the family as a whole appears to be more dynamic than the more highly conserved families.

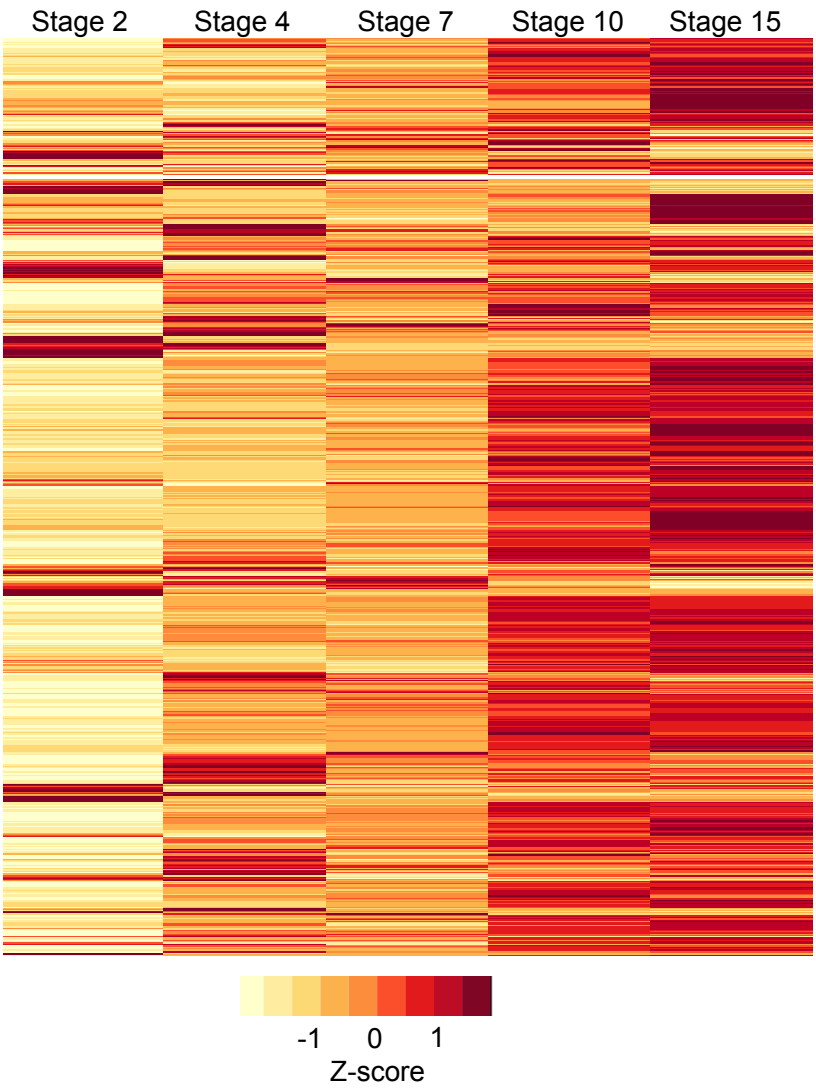


Figure 6.1.4 Expression of the C2H2 transcription factors during octopus development. Cells are colored according to number of standard deviations from the mean expression level.

We also evaluated the expression of members of several signaling ligand families across octopus development. The WNT family, for example, is found in all bilaterians examined. The family has 13 members in the Bilateria, 12 of which are found in lophotrochozoans, including octopus (Table 4.1). We found that all twelve are expressed during development, using both our transcriptome analysis and through *in situ* hybridization histology. Expression of all of the *WNTs*

examined with *in situ* hybridization was detected in our transcriptomes at stage 7, but only *WNT2* is robustly detected in stages prior to organogenesis. *WNT1* may also be faintly expressed at stage 5 with *in situ* hybridization, and is also weakly detected in the bioinformatics analysis at stage 4. These data parallel the onset of expression as detected using *in situ* hybridization: the earliest WNT gene we have been able to detect is *WNT2*, which is expressed from stage 4 onwards (Figure 6.1.5). The remaining *WNTs* are not detected until stage 7, as the body plan becomes apparent.

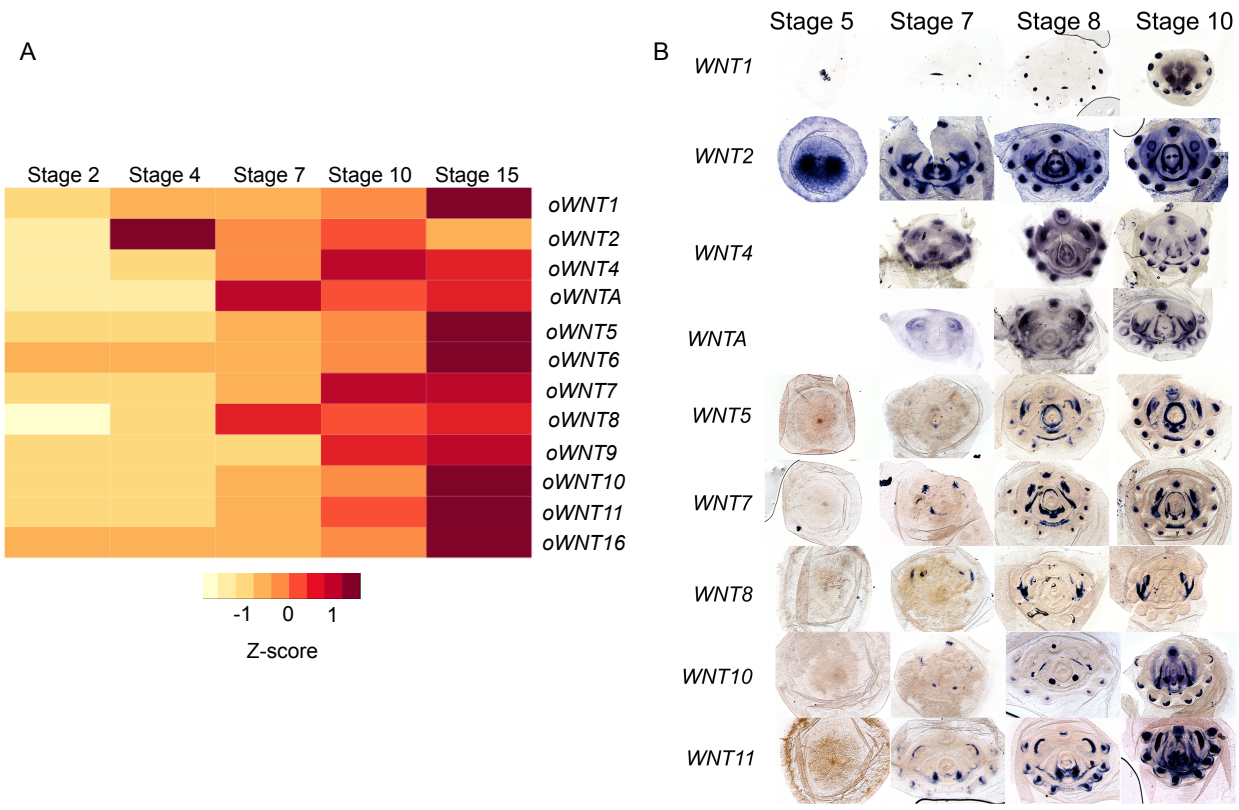


Figure 6.1.5. Developmental expression of WNT genes in octopus.
(A) Expression profile of members of the WNT ligands as detected by our transcriptome analysis. Cells are colored according to number of standard deviations from the mean expression level. (B) *In situ* hybridization of 9 octopus *WNT* genes at stages 5, 7, 8 and 10. *WNT1* exhibits weak staining at the center of the embryo at late stage 5. *WNT1* is expressed in the arm primordia and anus at stages 7 through 10. *WNT2* is detected at late stage 4, and earlier (not shown). At stages 7 and 8, *WNT2* is expressed in the mouth and anterior to the clefts, in the palliovisceral cord, around the mantle and in the fin buds, as well as in the arms. *WNT4* is expressed throughout the embryonic territory at stages 7 and 8. At stage 10, *WNT4* is expressed in the posterior part of the eyes, in the clefts and statocysts, around the mantle and in the arm primordia. *WNTA* is detected laterally in the eye fields at stage 7. At stage 8, *WNTA* is expressed in the mouth and arms as well as the lateral eye fields, in and behind the clefts, and adjacent to the mantle. Expression in these primordia continues at stage 10, but the boundaries are sharper. *WNT5* is staining is faint lateral to the eyes and in the mantle at stage 7, but not detected at earlier stages. At stages 8 and 10, *WNT5* can be robustly detected lateral to the eyes and in the retina, as well as around the mantle, in the center of the arm primordia, around the mantle, in the anus and along the central posterior midline. *WNT7* is faintly detected in the eyes and posterior to the mantle at stage 7. At stages 8 and 10, *WNT7* is robustly detected in the mouth, eyes, clefts, surrounding the posterior of the mantle, in the central portion of the pedal cord and in the arms. *WNT8* is first detected at stage 7 in the eyes and lateral to the head complex. Later, it is found medially in the eye primordia and in the posterior lateral regions of the optic primordia. At stage 7, *WNT10* is expressed in the mouth, lateral eye primordia, and in the gills. At stages 8 and 10, it is expressed in the statocysts, clefts, gills, arms and mouth. *WNT11* is first detected in the arms, eyes, and gills at stage 7. Expression continues in these territories at stages 8 and 10.

Many of the TGF-β ligands appear to be primarily expressed later in development (Figure 6.1.5). Early-expressed TGFβs include an TGFβ-like gene that nests with other lophotrochozoan sequences on a long branch at the base of the TGFβs, which may suggest a divergent, clade-specific history for this TGFβ subfamily. Members of the FGF family appear to be absent at stage 2, but may also be a good candidate for early roles in octopus development. For example, *FGF8/17/18* (Figure 6.1.6B) is expressed at stage 4. At organogenic stages, octopus *FGF8/17/18* is expressed in ectodermal tissue, while its receptor is expressed in mesodermal tissue (Figure 5.2), supporting a role for this family in mesoderm development, as has been described in other invertebrates (Green et al., 2013; Stathopoulos et al., 2004).

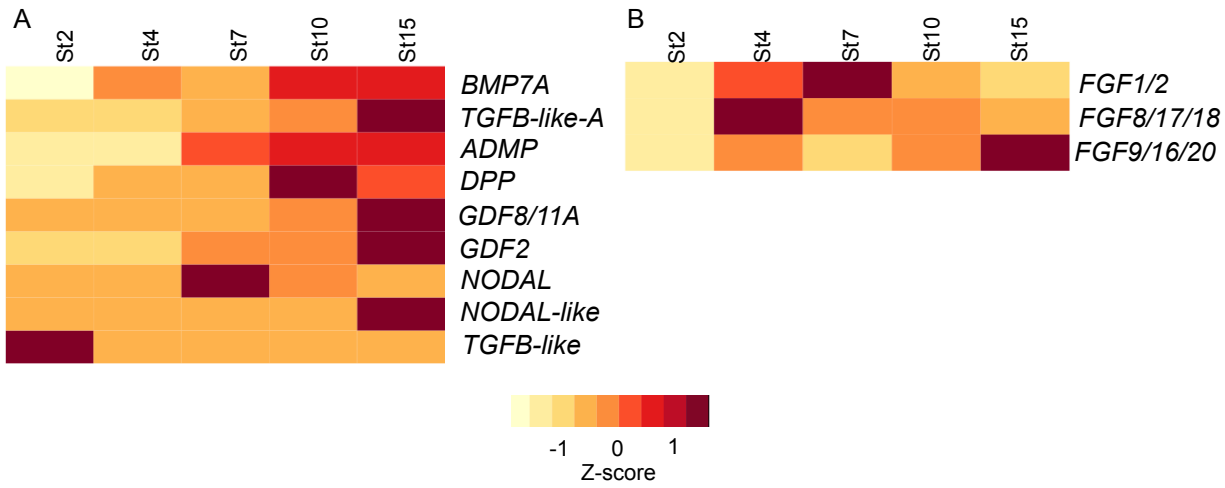


Figure 6.1.6. Expression of TGFβ and FGF genes across five stages of octopus development. (A) TGFβ ligands and (B) FGF ligands are expressed across octopus developmental stages. Cells are colored according to number of standard deviations from the mean expression level.

Stage-specific genes are taxonomically-restricted

In addition to our candidate gene approach, we also employed an unbiased search of the transcriptome, focusing on genes that are differentially expressed across the five stages sampled. In particular, we sought genes with stage specific expression profiles in order to identify

candidates that may play a role in distinct phases of octopus development. Using edgeR, we identified 6,525 transcripts that are differentially expressed across the five stages sampled ($p>0.0001$). We partitioned this pool based on transcript expression profiles into 100 clusters, and grouped these clusters based their expression, focusing on those clusters of genes that are strongly expressed at only a single stage. We identified 979 genes that are expressed at stage 2, 98 at stage 4, 405 at stage 7, 38 at stage 10, and 195 at stage 15 (Figure 6.1.7). To rule out spurious assemblies, we retained only transcripts that had matches in the genome (98.2%).

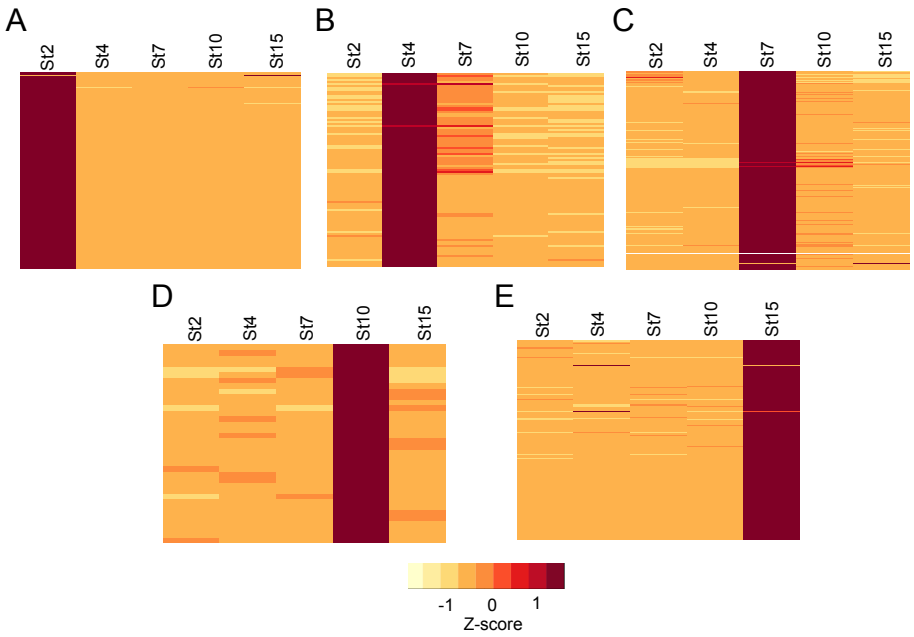


Figure 6.1.7. Heatmaps of expression of stage specific genes.
Clusters of genes identified as having a stage-specific profile of expression for stages 2 (A), 4 (B), 7 (C), 10 (D) and 15 (E). Cells are colored according to number of standard deviations from the mean expression level.

We performed Blast searches of Uniprot (2015) and our own databases, which include representative lophotrochozoan, ecdysozoan, deuterostome, and cnidarian sequences, to identify transcripts expressed in a stage-specific manner. A number of candidates were identified by this search. Among the genes expressed at stage 2 are *NANOS* and *FRIZZLED*, as well as the C2H2s, bHLH, and homeodomain transcription factors identified through the candidate gene family

directed search. At stage 15, many of the stage-specific genes are protocadherins and homeodomain transcription factors, as well as extracellular matrix proteins, including fibrillin and collagen.

Strikingly, many of the stage-specific transcripts did not have matches to genes found in other animals; that is, they are taxonomically restricted (Figure 6.1.8). This is particularly evident at early stages, where 59% of the 979 stage 2-specific genes did not retrieve matches in Uniprot, even with a relaxed e-value of 1x10-4. A search of spiralian genomes and transcriptomes (including *Capitella teleta*, *Lottia gigantea, Crassostrea gigas*, and *Aplysia californica*) found that nearly half of these do have matches in other lophotrochozoans, while the remaining genes appear to be cephalopod-specific. At stage 4, 61% of stage-specific transcripts have a match in Uniprot, while only an additional 5% have a match in our lophotrochozoan database. Intriguingly, our results indicate that very few of the transcripts specifically expressed during epiboly (stage 4) are of "lophotrochozoan" origin, while they have a similar proportion of putatively cephalopod-specific transcripts. At stages 7 and 10, nearly 90% of the transcripts are found in other animals, with three quarters of them present outside of the Lophotrochozoa. Less than 65% of stage 15-specific transcripts recover matches in Uniprot, with the remaining 36% split nearly evenly between cephalopod-specific and lophotrochozoan-specific transcripts.

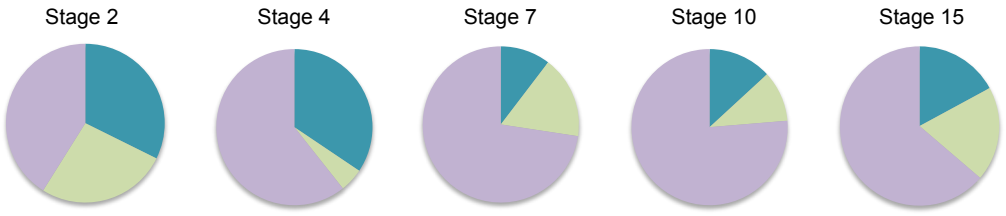


Figure 6.1.8. Proportion of stage-specific genes with and without matches in animal and lophotrochozoan protein databases.
Proportion of stage-specific genes with matches in Uniprot (purple), our lophotrochozoan-enriched database (green), and without matches in non-cephalopod species (teal), across 979 genes stage 2, 98 at stage 4, 405 at stage 7, 38 at stage 10, and 195 at stage 15.

DISCUSSION

Transcriptome analysis spanning five stages of octopus development has provided new insights and resources for studying this species. Our assembled transcriptome provides differential expression data that identified important new candidates for further study. Our approach was two-pronged, leveraging differential expression data to explore both the role of candidate genes as well as searching for genes based on their expression profile.

Our candidate gene approach found that the highly conserved developmental toolkit genes that have been characterized in other animals are also expressed over the course of octopus development. Surprisingly, we found that developmental transcription factors, including Hox and other homeodomain proteins, TBOX, FOX, and bHLH genes, appear to be expressed during or after the onset of organogenesis, which begins approximately two weeks post fertilization at 21°C. Exceptions to this trend include the NHRs, *EOMESODERMIN*, *PAX2/5/8, SOX14,* and *SOHLH,* which are highly expressed at stage 2.

Similarly, we found that many of the signaling ligands examined are dynamically expressed over development, with many of the WNTs and TGFβs expressed at later stages. These results, in both the transcription factors and the signaling ligands examined, correspond well with data from our *in situ* hybridization studies, which have found very few of these toolkit genes expressed before organogenic stages. Given the good correspondence between our bioinformatics and molecular analyses, these toolkit genes detected at stages 2 and 4 in our bioinformatics analysis will be interesting candidates to pursue in the future.

We examined in detail transcripts that are differentially expressed across the developmental stages sampled. Many of the transcripts that exhibit stage-specific expression are taxonomically restricted, particularly in stages prior to organogenesis. The majority of stage 2-

specific transcripts did not have a match in the Uniprot database. Uniprot, the Universal Protein Database, contains protein sequences from many animal genomes, but is heavily biased in favor of ecdysozoans and deuterostomes. We did find that roughly half of the genes that do not have a match in Uniprot are found in our custom database, which includes a number of lophotrochozoan taxa. This, in essence, is a measure of the "age" of the sequence, as has been used in phylostratigraphy (Domazet-Loso et al., 2007). Sequences that evolved before the bilaterians diverged, for example, will be found in deuterostomes and protostomes, where as sequences that evolved after this split will be restricted to a particular clade. By contrast, genes present across animal groups make up a greater proportion of the stage-specific transcripts during organogenic stages (stages 7-10). These transcripts code for transcription factors, including C2H2, homeodomain, and coil-coil domain containing genes, as well as genes involved in other biological processes, including cilia-related genes and kinases.

Both the candidate-gene and the stage-specific transcript approach recovered a similar pattern: genes shared across bilaterian groups are found at organogenic stages. The candidate gene approach found that highly conserved members of the developmental toolkit are expressed at or after stage 7, which is when organ primordia begin to appear. Similarly, a greater proportion of the transcriptome expressed in a stage-specific manner at organogenic stages is shared across bilaterian clades, while earlier and later stages are more divergent. This is reminiscent of the hourglass model of development proposed by several authors (Duboule, 1994; Raff, 1996). This mid-embryonic period of conservation, often called the phylotypic period, is when the basic body plan is defined. In octopus, this conserved period has many properties ascribed to the phylotypic stage of other animals, including Hox gene (Figure 6.1.1-2, Chapter 6.2; (Duboule, 1994)) and signaling ligand (Figure 6.1.3-4; (Raff, 1996)) expression.

Most studies of this model have focused on within-phylum comparisons (Irie and Kuratani, 2011; Kalinka et al., 2010); here, however, we retrieve a developmental hourglass across phyla. Such inter-phylum comparisons between ecdysozoans and vertebrates have also found a central period of conservation during organogenic stages (Levin et al., 2012), suggesting that the signal of conservation can be detected at least within bilaterians, irrespective phylogenetic distance. The recent reanalysis (Dunn and Hejnol, 2016) of a study of developmental transcriptomes spanning animals (Levin et al., 2016) suggests that deeper comparisons, particularly with ctenophores, may be less straightforward.

This central period of conservation is all the more striking given the divergent nature of cephalopod embryogenesis. Not only have they lost the early pattern of cell divisions for which the Spiralia get their name, but also the characteristic trochophore larval stage, which also contributed to the name of the Lophotrochozoa. Whether such a central stage is even present in other molluscs, and, if so, what stage it corresponds to in cephalopods, remains to be determined.

The similarity of the organogenic stages underscores the divergence of early octopus development. Our analysis has found only a few known developmental regulatory genes that may play a role in early embryogenesis. What it has revealed instead is a possible role for taxonomically-restricted genes in the highly divergent early embryogenesis in cephalopods. As the cephalopod-specific genes show no similarity to sequences in other animals, it is difficult to speculate on their function. Further study of these genes, particularly across cephalopod groups, will be essential to understanding the molecular basis of this evolutionarily novel developmental program.

SECTION II

Conserved anteroposterior expression of *Octopus bimaculoides HOX* genes is arranged in a polar coordinate geometry

INTRODUCTION

Hox genes are a family of highly conserved transcription factors have been primarily characterized for their role patterning the anteroposterior (AP) axis in bilaterally symmetrical animals. The Hox genes were first characterized in *Drosophila melanogaster*, where Hox mutants generate spectacular phenotypes in which body segments take on a new identity, as in the *Antennapedia* mutant (Gehring, 1998). Further work has highlighted a role for this family in patterning a diverse range of animals, including vertebrates, hemichordates, crustaceans, insects, annelids, and molluscs (Aronowicz and Lowe, 2006; Averof and Patel, 1997; Hinman et al., 2003; Lee et al., 2003; Samadi and Steiner, 2009; 2010).

The Hox genes in many animals also exhibit a conserved arrangement in the genome: they are often present in a cluster, and their arrangement in the cluster reflects the order of expression along the AP axis, a feature known as collinearity. Thus, the 3'-most gene, *labial* in flies or *HOX1* in vertebrates, is expressed anteriorly, where the 5'-most gene is expressed posteriorly. While present in distantly related groups, including chordates, echinoderms, arthropods, annelids, and molluscs, the Hox cluster shows different degrees of organization (Cameron et al., 2006; Duboule, 1994; 2007; Frobius et al., 2008; Simakov et al., 2013). In vertebrates, for example, there are multiple Hox clusters, the result of multiple rounds of whole genome duplication. These clusters are very tightly organized, extending not more than 200 kb (Figure 4.2; (Duboule, 2007)). There is a single Hox cluster in amphioxus, though at roughly 450kb, it is more than twice the size of the clusters found in vertebrates (Figure 4.2; (Ferrier et

al., 2000)). The tunicate *Oikopleura*, on the other hand, has an atomized Hox cluster (Seo et al., 2004). The insect *Drosophila melanogaster* and the annelid *Capitella teleta* have broken clusters (Frobius et al., 2008; Simakov et al., 2013), while the gastropod *Lottia gigantea* has an intact cluster of a comparable size to that of amphioxus. In octopus, we described a single complement of Hox genes (Chapter 4) that appear to be distributed throughout the genome: no two were found on the same scaffold (Figure 4.2). This accords well with the report in *Euprymna scolopes*, the Hawaiian bobtail squid, of a single complement of Hox genes that do not exhibit the canonical anteroposterior "collinear" domains of expression, though their arrangement in the bobtail squid genome is unknown.

While the Hox genes clearly have a common origin, the numbers and direct relationships of the different paralog groups (PGs) differ between clades. These PGs are considered to belong to anterior, PG3, central, and posterior classes. In vertebrates, for example, PG1 and PG2 belonging to the anterior class, PG4-8 are central class members, and PG9+ are posterior class. While there are multiple Hox clusters in vertebrates, none of the individual clusters has all of the ancestral vertebrate PGs. In ecdysozoans, the anterior class consists of *labial* (PG1) and *proboscopedia* (PG2). In *D. melanogaster,* the PG3 group has lost its homeotic function, and has been duplicated into three genes, *bicoid*, *zen*, and *zen2*, though it is represented in other ecdysozoan genomes as a single gene, *zen*. The ecdysozoan central class Hox genes are *deformed, sex-combs reduced (scr), fushi tarazu (ftz* – also no longer a homeotic gene in *D. melanogaster), antennapedia (antp), ultrabithorax (ubx),* and *abdominal-A (abd-A)*, and the posterior class gene is *abdominal-B (abd-B).* Lophotrochozoans ancestrally possess between 9 and 11 PGs (Barucca et al., 2016): PG1-5 displaying clear homology to the vertebrate and ecdysozoan paralog groups, *Lox5, Antp, Lox4,* and *Lox2,* and the lophotrochozoan-specific posterior genes *Post1* and *Post2*.

The posterior class genes appear to exhibit clade-specific duplications, particularly in the deuterostomes and the lophotrochozoans, but the relationships of several of the central class genes of deuterostomes, ecdysozoans and lophotrochozoans remain murky.

As Hox genes play a conserved role in patterning the AP axis, they are of particular interest in the study of clades with a diversity of body plans. The Mollusca is the second largest metazoan clade after the arthropods, and exhibits a remarkable array of different body plans, including the worm-like aplacophora, the uncephalized bivalves, the eight shelled chitons, and the eight armed octopuses. The homeodomains of Hox genes have been identified in seven of the eight molluscan clades from each of the lophotrochozoan paralog groups (Barucca et al., 2016; Barucca et al., 2003; Biscotti et al., 2007; Callaerts et al., 2002; Canapa et al., 2005; Fritsch et al., 2015; Fritsch et al., 2016; Iijima, 2006; Perez-Paralle et al., 2005; Pernice et al., 2006; Samadi and Steiner, 2009; 2010). All eleven lophotrochozoan Hox genes are present and arranged in the canonical order in a single genomic cluster in the gastropod *L. gigantea* (Simakov et al., 2013), supporting this arrangement as the ancestral state in molluscs.

While homeodomain sequences have been identified in all major molluscan clades except monoplacophorans, expression studies of the Hox genes have been limited to the gastropods and cephalopods, and recently, aplacophorans. The studies in gastropods and cephalopods did not find a strong relationship with a role in AP patterning for the Hox genes: although some degree of collinearity was found in the gastropod *Haliotis asinina* (Hinman et al., 2003) and the squid *E. scolopes* (Lee et al., 2003), many of the Hox genes appeared to violate collinearity. Instead, many of them were described as being coopted into clade-specific novelties, such as the mantle margin in the gastropods *H. asinina* (Hinman et al., 2003) and *Gibbula varia* (Samadi and Steiner, 2010) or the brachial crown of *E. scolopes* (Lee et al., 2003). However, recent work in

an aculiferan found evidence of the canonical, collinear anteroposterior pattern of expression (Fritsch et al., 2015).

Thus far, no cephalopod appears to retain the complete molluscan complement of Hox genes. However all eleven molluscan paralog groups have been identified in cephalopods, suggesting the last common ancestor of extant cephalopods had the ancestral molluscan Hox gene complement (Barucca et al., 2016). Nine Hox genes have been identified in *E. scolopes*: all except *proboscopedia* (PG2) and *lox2*. Six of the eight Hox genes examined in *E. scolopes* are expressed in the central nervous system, including the palliovisceral and pedal neurogenic domains. However, the expression pattern of the Hox genes in *E. scolopes* do not reflect the nested domains found in other animals, as they appear restricted to specific ganglia. In addition to expression in the developing central nervous system, *E. scolopes* Hox genes were also detected in a number of novel morphological structures, including the arm crown, funnel, and the light organ. Within the nervous system of *E. scolopes*, Hox expression does not violate collinearity, however, Hox expression in cephalopod novelties is described as a combinatorial code that is not collinear.

Our findings on the Hox gene inventory in the *O. bimaculoides* genome accord well with observations in other cephalopods. We detected a single Hox complement with relatively minor differences from *E. scolopes*: both *E. scolopes* and *O. bimaculoides* lack PG2, *E. scolopes* lacks a *lox2* ortholog (Callaerts et al., 2002), while no PG3 sequence was identified in *O. bimaculoides.*

We found that the genomic organization of the Hox genes, which has been thought to be important in regulating the expression of these developmental regulators in other animals, has been disrupted in *O. bimaculoides* (Albertin et al., 2015). The arrangement of the Hox genes in *E.*

scolopes has not yet been characterized, but alterations in genomic environment could be responsible for the novel expression domains described for this species.

Here, we characterize the expression of the eight Hox genes detected in the *O. bimaculoides* genome before and during the emergence of the body plan. Contrary to what has been described in *G. varia* and *E. scolopes*, we found an early pattern that accords well with the canonical, nested expression described in other bilaterians. We found this feature of "collinearity" presumed to have existed in the last common ancestor of bilaterians is present in the spatial and temporal pattern of Hox expression in *O. bimaculoides* despite the fact that the Hox locus is fragmented. Furthermore, the early expression of *O. bimaculoides* Hox genes highlights the novel geometry of the cephalopod embryo.

RESULTS

HOX1/LABIAL

The PG1 gene *LABIAL* is detected in *O. bimaculoides* at early stage 6, before the body plan is morphologically evident, as bilateral wedges radiating out from the center of the embryo (Figure 6.2.1). As the body plan emerges, the anterior edge of expression is clearly at the posterior edge of the cleft, and strong expression is found in the palliovisceral and pedal cords. In addition to the developing neural tissues, *LABIAL* can be detected in all four of the arm primordia, as well as the mantle.

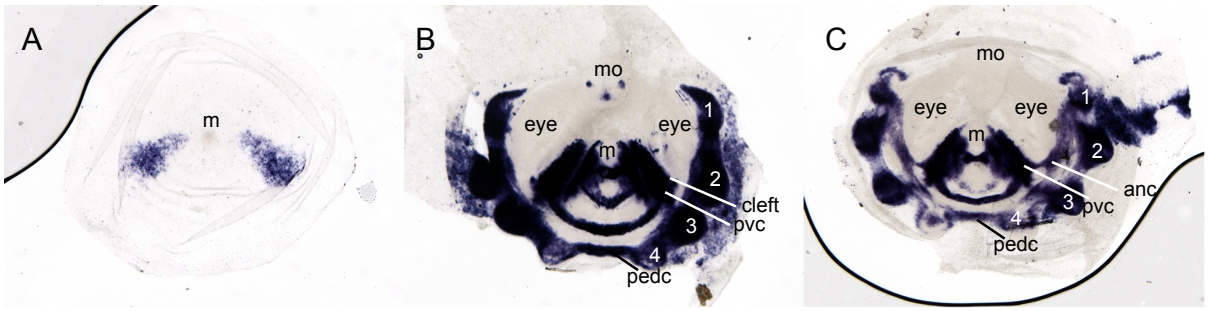


Figure 6.2.1. *LABIAL* expression pattern in *O. bimaculoides.*
Dorsal view, anterior faces up. (A) *LABIAL* is detected at stage 6 as bilateral wedges radiating out from the central mantle (m). (B) Expression of *LABIAL* at late stage 7 and stage 8 (C) is detected posterior to the clefts in the palliovisceral cord (pvc), pedal cord (pedc), mantle, and in the arms (1-4).

HOX5/Sex combs reduced

Similar to *LABIAL, SCR* is expressed in bilaterally symmetrical, radiating wedges at stages 5 and 6 (Figure 6.2.2). The anterior edge of expression appears to be slightly posterior to that of *LABIAL*. As the body plan is emerging, *SCR* is detected in a stripe across the embryo, through the anterior palliovisceral cord, and in arms 2 and 3 (Figure 6.2.2D). As anterior organ primordia expand and the condense posteriorly, this initial stripe forms a "W" in the central body, with bilateral expression in the palliovisceral and pedal cords extending from the clefts back to the anterior limit of the fourth arm, and then extending medially and anteriorly towards the center of the mantle. Expression in arms 2 and 3 also continue.

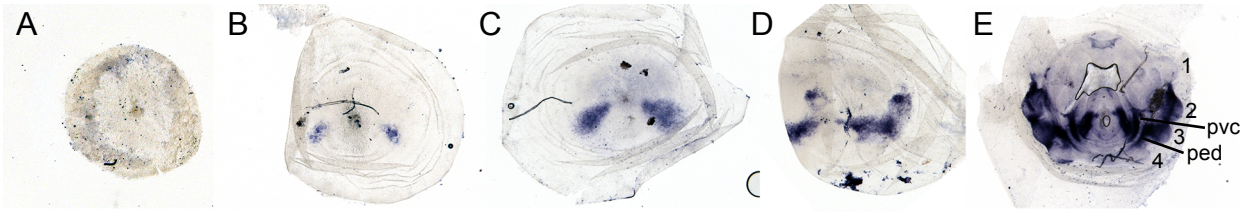


Figure 6.2.2. *SCR* expression during octopus development.
Dorsal view, anterior faces up. (A) *SCR* is not expressed at stage 4. Expression of *SCR* transcripts are detected as bilateral spots posterior to the mantle at stage 5 (B) and at stage 6 (C) take on a teardrop shape. (D) At stage 7, *SCR* is expressed in a stripe extending from the mantle to the third arm, as well as in arms 2 and 3. (E) Expression of *SCR* at stage 9 in arms 2 and 3, as well as segments of the pedal and palliovisceral cords at the level of arms 2 and 3. A stripe of expression extends up to the mantle.

LOX5

LOX5 is expressed at stage 6 in bilateral patches found laterally within the embryonic territory (Figure 6.2.3). At stage 7 and later, *LOX5* is detected in the gills, arm primordia, and the pedal cords. At stage 9, expression is also detected in the buccal territory.

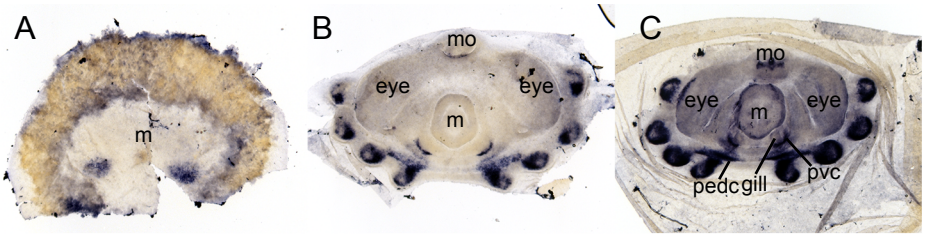


Figure 6.2.3. Expression of *LOX5* in *O. bimaculoides* embryos.
Dorsal view, anterior faces up. (A) Transcripts of *LOX5* are detected in bilateral patches posterior to the mantle at stage 6. These embryos were counterstained with SYBR green, which produces a yellowish tinge, particularly evident in the yolk still attached to the embryo following staining. The staining in the yolky areas is likely an artifact. (B) *LOX5* is expressed in all four arm primordia, as well as in bilateral patches along the gills and palliovisceral cord at stage 7. (C) *SCR* expression at stage 9 is similar to stage 7. Expression is also detected in the mouth.

ANTENNAPEDIA

ANTP is first detected at early stage 5 as a triangular-shaped swath extending from the center of the embryo posteriorly (Figure 6.2.4). At stage 6, the expression pattern has shifted laterally to become two symmetrical patches with additional labeling towards the center of the embryo. By stage 7, *ANTP* is expressed in arms 2 and 3, as well as the anterior half of arm 4, and in the gill primordia. This pattern is maintained at stage 10, though expression in the arms has resolved to anterior and posterior stripes, as well as in the axial nerve cord territory and the adjacent pedal cord. *ANTP* transcripts are also detected in the mouth at this stage.

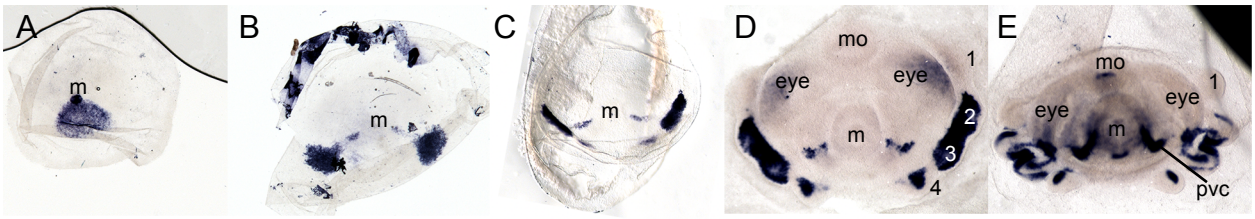


Figure 6.2.4. *ANTP* is detected before and during organogenic stages in *O. bimaculoides.*
Dorsal view, anterior faces up. (A) Stage 5 expression of *ANTP* is found in a quadrant posterior to the mantle. (B) At early stage 6, *ANTP* transcripts are detected in lateral, bilaterally symmetrical patches, with a stripe extending from these patches towards the central mantle. (C) At late stage 6, expression in the lateral patches has transitioned to the primordia of arms 2, 3, and 4. (D) *ANTP* is detected in arms 2 and 3, as well as the anterior portion of arm 4 at stage 7. Expression is also detected in a bilateral pair of patches adjacent to the mantle. (E) At stage 10, *ANTP* is detected in the nascent axial nerve cords, as well as the anterior and posterior edges, of arms 2 and 3, as well as the anterior of arm 4. Expression is also detected in a portion of the palliovisceral and pedal cords, as well as a patch lateral to the mantle.

LOX4

At stage 5, *LOX4* is expressed in bilaterally symmetrical wedges at the posterior (Figure 6.2.5). As the body plan becomes apparent, this expression is detected in a stripe from the fin folds on the mantle, through the gills and palliovisceral cords, to the pedal cord. It is also expressed in the posterior of arm 2, and throughout arms 3 and 4.

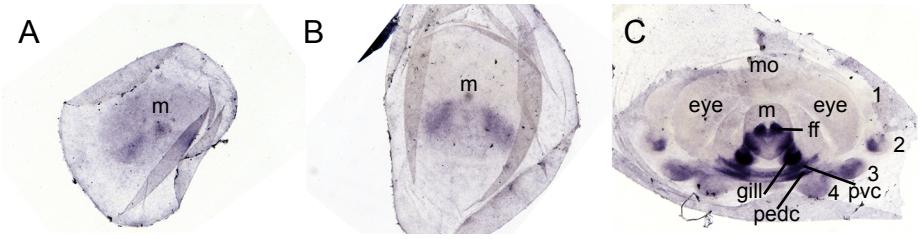


Figure 6.2.5. *LOX4* expression in octopus embryos.
Dorsal view, anterior faces up. (A) Expression is detected at stage 5 as weakly stained wedges posterior to the mantle. Light staining anterior to the mantle may be background. (B) The wedges of expression posterior to the mantle are stronger at stage 6. (C) At stage 8, expression is detected in the fin folds (ff), gill, and segments of the palliovisceral and pedal cords, as well as arms 2, 3, and 4.

LOX2

LOX2 appears to come on later than the more anterior Hox genes, as expression is first detected at stage 7 as bilateral stripes, with light expression in the fin folds that extends through the gills, palliovisceral, and pedal cords, a pattern much like that of *LOX4* (Figure 6.2.6). At stage 8, expression can also be detected in the arms crown, largely restricted to the tissue between arms 3 and 4.

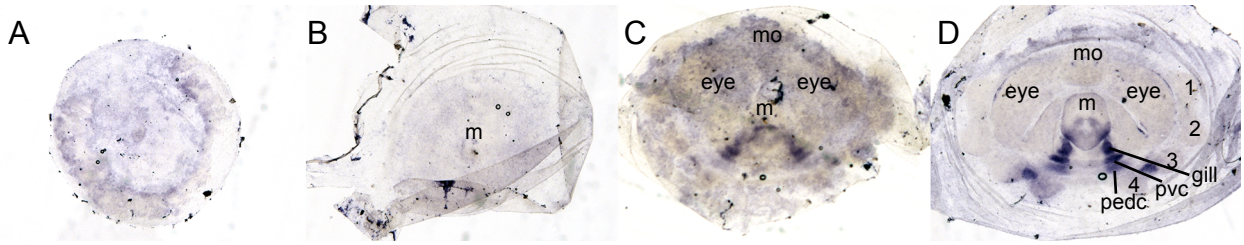


Figure 6.2.6. *LOX2* expression in octopus embryos
Dorsal view, anterior faces up. *LOX2* is not detected at stage 4 (A) or stage 6 (B). *LOX2* is expressed at stage 7 in bilateral, posterior wedges that encompass segments of mantle, gill, palliovisceral and pedal primordia. (D) At stage 8, *LOX2* is clearly expressed in the gill primordia, as well as segments of the pedal and palliovisceral cords. It can also be detected between the buds of arms 3 and 4, as well as in the fin folds.

POST2

POST2 transcripts are detected at stage 6 in a posterior sector of the embryo, one that extends from the center of the embryo (Figure 6.2.7). As organ primordia become apparent, this expression is strongest at the anterior edges of the wedge and extends through the posterior mantle, gills, palliovisceral and pedal cords, and arm 4, as well as all of the tissue between these discrete structures.

Figure 6.2.7. Expression of *POST2* in octopus embryos
Dorsal view, anterior faces up. (A) Expression of *POST2* at stage 6 shows a wedge of expression posterior to the mantle. Strong staining in the extraembryonic ectoderm is an artifact. (B) *POST2* expression at stage 7 is similar to stage 6 (A), but shows stronger expression at the edges of the wedge. Expression is detected in the posterior mantle, and central swaths of palliovisceral and the pedal cords, as well as the primordia in the intermediate regions. (C) Expression of *POST2* at stage 8 includes the expression pattern of stage 7 (B), but now extends to arm 4.

POST1

POST1 is expressed very faintly at late stage 6 as bilaterally symmetrical spots posterior to the future mantle (Figure 6.2.8). By late stage 7, *POST1* is strongly detected at the midline of the posterior of the mantle, as well as in the palliovisceral and pedal cords. *POST1* is also found in all four arm primordia from stage 7 onwards.

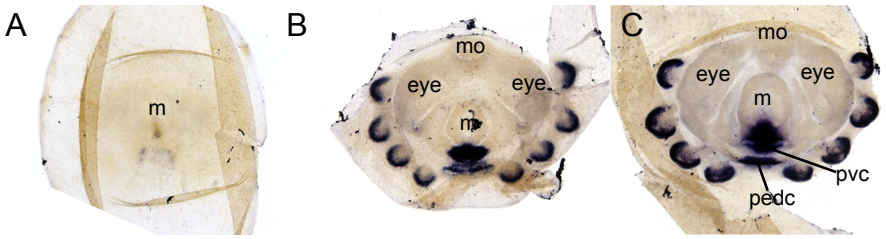


Figure 6.2.8. Expression of *POST1* in octopus development.
Dorsal view, anterior faces up. (A) Expression of *POST1* at stage 6 is weakly detected posterior to the mantle. At stage 7 (B) and stage 8 (C), *POST1* is expressed in all four arm primordia, as well as in a narrow region at the posterior center of the embryo, including the mantle, and central portions of the palliovisceral and pedal cords.

DISCUSSION

Unexpected "collinearity" of Hox gene expression in octopus embryos

Octopus Hox genes have lost the clustered arrangement found in other animals (Duboule, 2007) but nonetheless have retained the ancestral nested, "collinear" expression along the anteroposterior axis. Interestingly, octopus Hox genes are detected in stages immediately prior to the appearance of organ primordia, consistent with a role for these genes in patterning the highly derived octopus body plan. Molluscs are considered to be unsegmented, indicating that the Hox genes, which can be considered "segment identity genes", exhibit a similar pattern of expression in animals without overt segmentation.

While studies of Hox gene expression in gastropods and another cephalopod have primarily detected transcripts in the central nervous system and novel morphological structures, we have found an initial role in patterning the body plan, before the nervous system emerges (Chapter 5.1). Even after neuronal markers can be detected in the developing neural cords, expression of the Hox genes is present in both neural and non-neural structures, and not restricted to the domains of the primordia, which is the pattern described for *G. varia* (Samadi and Steiner, 2009; 2010) and *E. scolopes* (Lee et al., 2003). In many respects, the Hox expression patterns in octopus resemble those of the polyplacophoran *Acanthochitona crinita* (Fritsch et al., 2015; Fritsch et al., 2016) and the basally branching gastropod *H. asinina* (Hinman et al., 2003), both of which exhibit nested, anteroposteriorly-restricted Hox expression.

Hox gene expression in cephalopod novelties

While the early pattern of Hox gene expression in octopus is consistent with the ancestral nested domains along the anteroposterior axis, the later pattern is more complex. The most striking deviation from a collinear pattern is found in the arms. It was this pattern that led Lee

and colleagues (Lee et al., 2003) to describe the expression of Hox genes in *E. scolopes* arm primordia as a "combinatorial code". Our results in *O. bimaculoides* arm buds are consistent with the patterns described in *E. scolopes*. Both *O. bimaculoides* and *E. scolopes* show expression of *LABIAL, LOX5*, and *POST1* throughout the arm crown. But, although the expression of *POST1* and *LOX5* throughout these structures is a violation of collinearity, the remaining Hox genes in octopus exhibit the canonical restriction along the anteroposterior axis, even in this novel cephalopod structure (Figure 6.2.9). This interpretation fundamentally differs from that of a combinatorial code as described in *E. scolopes*.

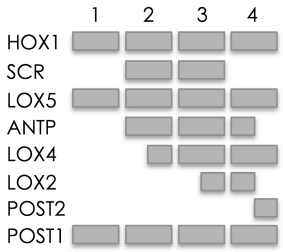


Figure 6.2.9. Schematic of Hox expression in *O. bimaculoides* arm buds.
All eight Hox genes in *O. bimaculoides* are expressed in the arm crown. Presence of expression is indicated by the gray boxes. Expression of *ANTP* in arm 4, *LOX4* in arm 2, *LOX2* in arms 3 and 4, and *POST2* in arm 4 is restricted to half of the arm bud, as indicated by a smaller box.

Hox gene expression reflects the layout of the octopus body plan

The large embryos that undergo direct development characteristic of cephalopods are a striking departure from the molluscan developmental sequence of spiral cleavage followed by trochophore and veliger larval stages. Cephalopod eggs are filled with yolk, and the embryos that develop on top of the yolk are arranged very differently from other animals. The ventral structures in cephalopods, the arms, are splayed out at the periphery of the embryos, while dorsal structures are central. The embryos are therefore circular, with anterior and posterior occupying opposite sections of the circle. This radial arrangement is highlighted by the early expression of

the Hox genes: anterior genes are expressed in wedges radiating out from the center, like sectors of a polar coordinate system, with the posterior genes expressed in a sectors that include the midline (Figure 6.2.10). This pattern becomes more complex as the octopus embryo continues to develop and the organ primordia compact, with arm primordia constricting to form the arm crown. Understanding how these derived embryos with their novel body plan are arranged will facilitate future molecular developmental comparisons with other animals.

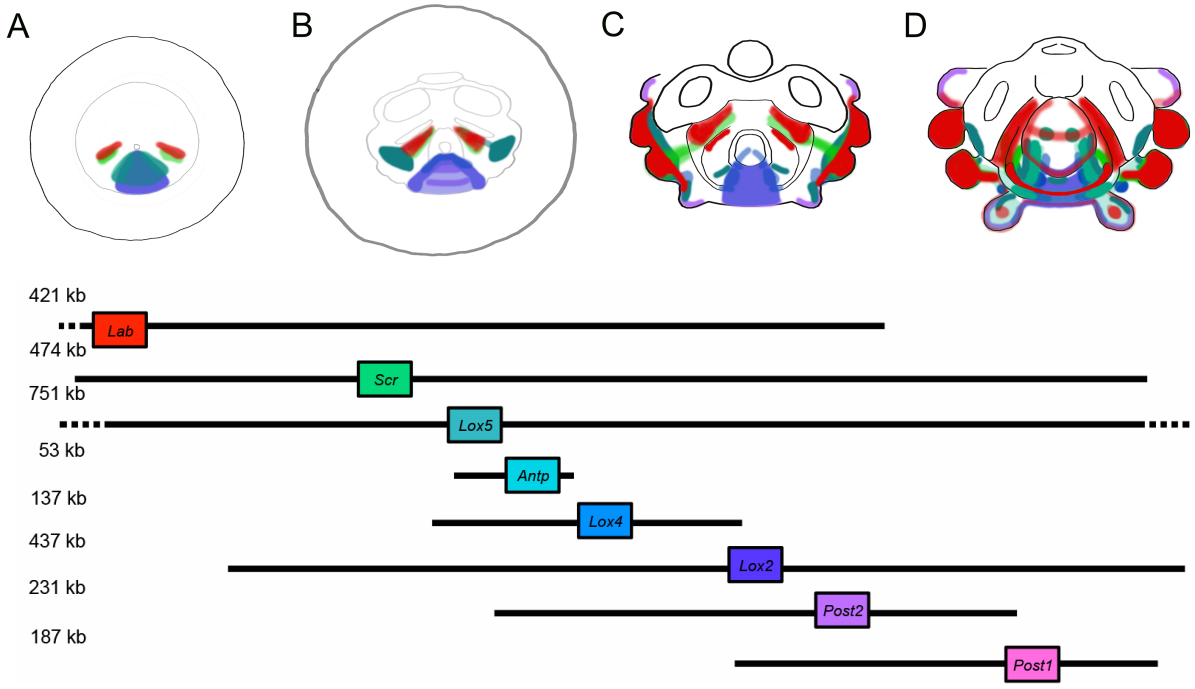


Figure 6.2.10. Schematic of Hox gene expression and arrangement in the genome in *O. bimaculoides*.
(A-D) Dorsal view of Hox gene expression in stage 5 (A), stage 6 (B), stage 7 (C), and stage 9 (D) octopus embryos. Embryos are oriented with anterior to the top. Hox genes are color coded: *LABIAL* is red, *SCR* is green, *LOX5* is teal, *ANTP* is sky blue, *LOX4* is cerulean, *LOX2* is royal blue, *POST2* is purple, and *POST1* is magenta. At the bottom, the genomic arrangement of Hox genes in octopus is depicted, as described in Chapter 4.

Implications for the understanding of the novelties in the octopus body plan

Lee and colleagues (2003) studied Hox gene expression in embryos of the Hawaiian bobtail squid. Their broad conclusions were that "collinearity" is absent in cephalopods, and that the Hox genes, either singly or in combination, underlie the development of novel cephalopod structures, including the arm crown. My data support a very different scenario, where octopus Hox genes show a conserved anteroposterior pattern of expression along a very different body plan. We find octopus Hox genes are expressed in wedges, radiating out from the center of the embryo, reminiscent of a polar coordinate system, and very different from the Cartesian arrangement found in other species. While these results and our interpretation differs greatly from the work done in *E. scolopes*, the differences in expression pattern are likely due to the sensitivity of the *in situ* hybridization methods employed by Lee and colleagues rather than a difference between these two species. In addition, we were able to examine expression at earlier stages than were reported for *E. scolopes.* The implication of these findings is that the molecular underpinnings of the highly derived cephalopod body plan are established before organogenic stages, which has been the focus of our bioinformatics analyses. These data strongly suggest that cephalopod-specific novel genes are candidates for construction of this body plan, since developmental regulatory genes seen throughout the animal kingdom appear to be deployed only after the body plan has been established.

CHAPTER 7

Conclusions and future directions

The fundamental evolutionary puzzle of cephalopods is that they show not parallel evolution but true convergence, not innovation but true novelty. The modern field known as evolutionary developmental biology, at its core, proposes to study the basis of evolutionary change. The profound differences in early cephalopod development relative to other molluscs represents a true novelty. Whether these changes underlie other cephalopod evolutionary novelties can only start to be addressed by this thesis. Using a molecular approach, we explored the complement of highly conserved developmental genes in the octopus genome and their deployment in the development of the nervous system and the body plan, which demonstrates both conserved and novel patterns. Our bioinformatics analysis also identified a large number of clade-specific genes deployed in early development. In this final chapter, I discuss our findings and areas of future study raised by these results.

Molecular conservation and innovation in octopus development

Broadly, we find that the developmental toolkit repertoire in *O. bimaculoides* is largely similar to that of other invertebrate bilaterians. We did not find evidence for expansion across multiple families of developmental transcription factors and ligands in the octopus genome, suggesting that the evolution of the cephalopod brain and body plan did not require significant elaborations of these genes (Chapter 4). There are two notable exceptions: the massive expansions of the C2H2s and the protocadherins. The C2H2-containing genes in octopus comprise one of the largest metazoan gene families. Clade-specific expansions of these genes have been suggested to play a role in evolutionary innovations in other clades (Liu et al., 2014;

Nowick et al., 2010; Shannon et al., 2003). As transcription factors, they may be a source of innovation in the regulatory landscape in the octopus, and our study of developmental transcriptomes indicates that they are dynamically expressed across developmental stages. What role these genes may play in octopus development, and the development of other cephalopods, will be an important area for future study.

The expansion of the protocadherins in the octopus genome was among our most surprising results. Protocadherins are absent from ecdysozoan genomes, and only a handful of these genes had previously been detected outside of the chordates (Hulpiau and van Roy, 2011). That there are a number of protocadherins found in spiralians, and a massive expansion in octopus was therefore quite unexpected. It is tempting to speculate that the parallel expansion of protocadherins in vertebrates and cephalopods reflects a requirement for protocadherins, or protocadherin-like molecules, in the construction of large nervous systems. Elucidating the roles of octopus protocadherins will provide important insights into the general function of these genes. Current models of protocadherin function point to a role similar to that of *Dscam* in the *D. melanogaster* nervous system. Like the vertebrate protocadherins, the *Dscam* locus generates an extraordinary number of isoforms that exhibit homophilic binding, and *Dscam* is deployed in the nervous system, where it is required for neuronal wiring and self-avoidance (Chen et al., 2006; Wojtowicz et al., 2007). Our finding that coleoid cephalopod evolution has specifically driven the expansion of protocadherins suggests that this molecular mechanism may have important additional and specific roles in controlling the wiring of large nervous systems. Unfortunately, functional studies of octopus protocadherins are hindered by the absence of gene manipulation and cell culture techniques in these animals (see below).

The distribution of protocadherins in animal genomes also highlights notable deviations in gene content of the model ecdysozoans *C. elegans* and *D. melanogaster* relative to vertebrates and spiralians, making them unreliable representatives of the general invertebrate genome condition. Well-known examples of the dramatic departure in gene family content in the *C. elegans* genome include expansion of nuclear hormone receptors and absence of a hedgehog ligand. We also find larger complements of *WNTs* and *TGFβ* family members in spiralians relative to ecdysozoans. While developmental regulatory genes generally appear to be expanded in vertebrates relative to invertebrates, this is not true across all gene families. An important example lies in the discs, large (DLG) protein family, which is a key postsynaptic scaffold component. Vertebrates have five DLG members, while *D. melanogaster* harbors one *DLG1-4* gene and a *DLG5*. This is consistent with two genome duplications in the vertebrates and has suggested a unique complexity of the synaptic machinery in vertebrates (Nithianantharajah et al., 2013). Octopus, however, has three *DLG1-4*s and 2 *DLG5s*, and the snail *L. gigantea* also has 3 *DLG1-4s* and a single *DLG5* (Albertin et al., 2015). These results underscore the importance of comparisons across a broad phylogenetic range for correctly interpreting patterns across distantly related animals, and particularly emphasize the importance of the Spiralia when considering the bilaterian generalizations.

While in overall developmental gene content the octopus genome is similar to other invertebrate bilaterians, it harbors extensive rearrangements as documented by a loss of ancestral linkages that are maintained in other bilaterians (Albertin et al., 2015). However, the Hox genes, a prominent bilaterian cluster that has been atomized in *O. bimaculoides*, exhibit an anteroposterior expression pattern in the embryos that is broadly similar to the pattern seen in other animal embryos (Chapter 6.2). These findings suggest that, at least at organogenic stages,

the genomic rearrangements did not alter the spatial "read-out" of these developmental transcription factors.

Other homeobox genes also exhibit expression patterns consistent with those observed in other bilaterians. Our examination in *O. bimaculoides* of members of a gene regulatory cassette involved in patterning the midbrain-hindbrain boundary in vertebrates revealed an overall topology that is strikingly similar. Furthermore, this cassette identifies a morphological boundary, the cleft, between the supraesophageal and subesophageal brain. These data support a conserved cassette of transcription factors involved in the development of the octopus body plan along the anteroposterior axis.

This interpretation is at odds with previous studies of homeobox genes in cephalopods, which described divergent deployment and novel roles for these genes (Baratte et al., 2007; Lee et al., 2003; Navet et al., 2009; Navet et al., 2008). These differences in interpretation likely arise from two factors. The first is a difference in detection strength, which may, at least in part, be due to the presence of yolk in the tissue preparations employed in the studies of other cephalopods.

The second is that the arrangement of the cephalopod body plan is strikingly different from other animals, with the dorsal structures located at the center and lateral structures found at the periphery. Once the "Cartesian" arrangement of gene cassettes is translated to the "polar coordinate" geometry of the cephalopod body plan, the overall patterns appear quite conservative. An example of this can be found in the interpretation of the role of *pax6* in *Sepia officinalis*. An initial study described the expression of *pax6* as exhibiting both conserved expression in the optic territories and novel expression in the arm crown, however, a more recent study finds that the expression in the arms is in fact part of a bilaterian mediolateral patterning

cassette (Buresi et al., 2016; Navet et al., 2009). The developing arm buds, which express *nkx2.2, pax6, pax3/7*, and *pax2/5/8*, have the molecular coordinates of the ventral midline of other bilaterians (Lauri et al., 2014), molecularly identifying the laterally positioned cephalopod arm buds as ventral structures.

While we found a great deal of conservation in the expression of these transcription factors at organogenic stages, we found notable differences in the deployment of highly conserved developmental regulators as well. The transcription factor cassette shared between the midbrain-hindbrain boundary of vertebrates and hemichordates identifies the isthmus organizer, a signaling center that expresses *FGF8* and *WNT1*. This transcription factor cassette appears to be conserved in octopus, and it identifies a signaling center with an expanded complement of signaling ligands relative to vertebrates, including *HH* and *DPP,* as well as other *WNTs*, though it does not express *WNT1*. These results suggest that while homeodomain transcription factors are expressed in a highly conserved manner along the octopus body plan, other elements of the developmental toolkit may be more flexible.

Recent comparative developmental studies have found striking similarities in the expression of developmental regulatory genes in distantly related animals. A prominent example is the shared expression of a suite of genes between vertebrate pallium and annelid and fly mushroom bodies (Tomer et al., 2010). When we examined the expression of these genes in the analogous octopus brain structure, the frontal-vertical lobes, we found no correspondence: the diagnostic genes were not expressed in this structure. These data indicate that the sensory associative centers in octopus represent genuine evolutionary novelties, and present an important opportunity to study the underlying developmental and structural aspects of these convergent brain structures.

We have also detected a large number of clade-specific genes in the octopus developmental transcriptomes. Taxonomically restricted genes regularly make up 10-30% of the protein coding genes in metazoan genomes. Of course, the number of clade-specific genes detected will depend on how distant the comparisons are: closely related species are expected to have a greater proportion of genes in common than more distantly related groups. As there are few sequence resources available for molluscs and other spiralians, it is difficult to evaluate how taxonomically restricted a given octopus gene is. As more sequences become available, this resolution will greatly improve. Nonetheless, these taxonomically restricted genes, particularly those expressed in novel cephalopod structures (see below) and in early development will be exciting candidates for future studies.

Comparing cephalopod development to the molluscan developmental program

The expression of highly conserved developmental toolkit transcription factors at organogenic stages is all the more striking given the highly derived early embryonic stages in cephalopods. While spiral cleavage followed by a trochophore larval stage is likely the ancestral mode of development in molluscs, cephalopod embryogenesis is bilateral, meroblastic and direct, and completely departs from typical molluscan sequence. That the expression of developmental regulators along the novel octopus body plan at the organogenic stage should so closely resemble patterns observed in distantly related bilaterians is therefore rather a surprise. Further studies in non-cephalopod molluscs will be important to understand how these genes are deployed in indirect developing species. Comparisons of developmental toolbox gene expression between cephalopods and non-cephalopod molluscs may also yield insights into the developmental consequences of direct versus indirect developmental programs.

The Hox genes, for example, have been examined in gastropods and polyplacophorans, in addition to cephalopods (Fritsch et al., 2015; Fritsch et al., 2016; Hinman et al., 2003; Lee et al., 2003; Samadi and Steiner, 2009; 2010). In both gastropods and polyplacophorans, Hox genes are first expressed in the trochophore larval stages. This stage, which is common to not only non-cephalopod molluscs, but also annelids, nemerteans, and others (Henry, 2014), may therefore correspond molecularly to the organogenic stages in cephalopods. Indeed, a recent study in the Pacific oyster *Crassostrea gigas*, found that novel homeobox genes were deployed at early and late stages of development, but not during the trochophore stage (Paps et al., 2015). Exploration of more developmental transcription factors in molluscan embryogenesis will be key to understanding how it relates to the exceptional developmental program in cephalopods.

The early stages of cephalopod development have been characterized morphologically, but little is known about the molecular events that take place. Our bioinformatics data find only a few candidate developmental toolkit genes that are expressed prior to organogenesis. Instead, we recover numerous taxonomically restricted genes at these early stages. What role these cephalopod- and lophotrochozoan- specific genes may play in early developmental events will be an important area of future study. Overall, our exploration of octopus developmental transcriptomes found expression of novel genes at early stages and conserved genes at organogenic stages, which is consistent with the hourglass model of development.

The large, yolky eggs of cephalopods and fish exhibit a number of superficial similarities, among them, meroblastic cleavage and epibolic gastrulation. These similarities are likely due to the physical constraints placed on large, yolky eggs (Galis and Sinervo, 2003). The molecular and cellular events in cephalopod morphogenesis have not been characterized, and the underlying mechanisms may have convergent or novel elements. For example, radial

intercalation is a widely used morphogenetic process involved in fish, fly, and mammalian gastrulation and neural folding in amphibians and fish. Recent studies have found that a complement component normally associated with immune function provides a critical chemotactic signal in radial intercalation in gastrulation and neural crest migration in amphibians and fish (Carmona-Fontaine et al., 2011; Szabo et al., 2016). Whether radial intercalation takes place in cephalopods, and if so, what molecular mechanisms are deployed, is unknown. Comparisons between these convergent early developmental programs at the cellular and molecular levels may yield insights into the processes governing early development.

Tool development for work in cephalopods

While genome sequencing is important for advancing the study of cephalopods, there are a number of other impediments to research in this clade. As discussed in the introduction, cephalopod husbandry is challenging. A few species have been able to be maintained over several generations in the laboratory, and these will be important to pursue for further tool development.

A major barrier to cephalopod research is the absence of gene manipulation in this group. The embryos, while accessible to researchers from early stages, are difficult to manipulate. Octopus embryos have a thick, leathery chorion that is difficult to penetrate. Similarly, squid and cuttlefish have chorions and additional jelly layers that also make embryo access difficult. Holes in the chorion often result in yolk leakage or infection that results in mortality. Culturing embryos outside of the chorion is also a challenge, and embryos are particularly sensitive and vulnerable prior to the end of epiboly. These features have presented significant impediments to developing gene manipulation techniques in cephalopods. However, these barriers will have to be overcome for future studies.

In the absence of gene manipulation, pharmacological manipulation might be an attractive means to test the role of signaling pathways. For example, the earliest gene we have been able to detect as expressed in the clefts is *HH*. Cyclopamine is a widely used small molecule inhibitor of cellular responses to *HH* signaling that has been shown to work in the annelid *Platynereis* (Dray et al., 2010). Assaying the role of *HH* signaling in the cleft could yield important insights into the molecular events important in setting up the octopus body plan.

Lineage tracing studies have been important in the study of diverse spiralians (Boyer and Henry, 1998; Henry, 2014). These studies have found that particular blastomeres give rise to similar tissues at later stages. For example, the progeny of the D quadrant generates dorsal structures as well as mesendoderm in a wide range of spiralians. To date, there are no lineage tracing studies in cephalopods, due to many of the physical factors that have impeded the development of gene manipulation in these animals. Therefore, such questions as whether there is an equivalent of the spiralian quadrants in cephalopods has not been assessed.

Lineage tracing in cephalopods could also help clarify the origin of the different germ layers. For example, molluscs and other spiralians have two sources of mesoderm: endomesoderm and ectomesoderm (Henry and Martindale, 1999; Lyons and Henry, 2014; Perry et al., 2015). In many bilaterians, mesoderm is derived from an endomesodermal precursor. However, spiralians also have a second source of mesoderm that is derived from cells that will give rise to ectoderm (ectomesoderm). How different germ layers appear and whether ectomesoderm is present in cephalopods are open questions.

There are, however, methods of studying early development that do not require egg entry. Light sheet microscopy, also referred to as single plane illumination microscopy (SPIM), is an attractive technology for time-lapse imaging, as tissues are exposed to only a thin plane of light,

minimizing phototoxicity and photobleaching. Nuclei in live octopus embryos can be labeled with a fluorescent dye such as SYBR Green I (Molecular Probes), which does not require physical manipulation and does not result in appreciable abnormalities during development, and can be followed for days using light sheet microscopy, yielding exciting insights into developmental events. For example, my observations using time-lapse light sheet microscopy of stage 2 octopus embryos suggests ingression and compaction may be characteristic of gastrulation in this species, contrary to what has been classically described in other cephalopods (Naef, 1928). Time-lapse light sheet microscopy during epiboly suggests that the embryonic cap spreads mainly through cell divisions, and that there are no major cell rearrangements at these stages. These preliminary results are very promising, however, the slow pace of *O. bimaculoides* development means that imaging over several days will only capture a single developmental epoch. Faster developing species may therefore be more attractive candidates for such studies. While non-invasive labeling by a nuclear dye has been quite effective, the development of gene manipulation coupled with time-lapse light sheet microscopy would be a powerful tool for the study of the novel cephalopod developmental program.

Ablation techniques have been important for the study of developmental potential. While traditional cell ablation techniques required physical access to the targeted embryonic cells, laser cell ablation can precisely target cells for destruction through a clear chorion. My preliminary experiments in *O. bimaculoides* indicate that this technique is feasible in this species and will be an exciting avenue to pursue in the future.

Finally, the absence of cell culture in this group is a major problem. Indeed, after decades of effort, there are no established marine cell culture lines (Cai and Zhang, 2014; Rinkevich, 2011), though culture from a freshwater snail has been achieved (Hansen, 1976). Cultured cell

lines could be an important resource for exploring gene functions, as they have been in mammals and insects. For example, protocadherins play a role in cell adhesion in mammals, and *in vitro* assays have been developed to study their role in cell-cell interactions (Schreiner and Weiner, 2010). One pressing set of experiments is to test members of the octopus protocadherin expansion for homophilic adhesion and repulsion; this is only one of the many questions that would be best studied with cephalopod cell culture.

Cephalopod genomics

Having a sequenced and annotated genome is a key tool in 21st century biology. It provides a complete list of genes and their adjacent regulatory sequences, as well as their arrangement in the genome. This inventory of genes provides researchers with a much larger set of candidate conserved genes, and can also provide insights into innovations in the genome. Cephalopods and other lophotrochozoans have been largely left behind in the sequencing frenzy for other clades: as of 2015, more than 200 animal genomes have been published, with more than fourth-fifths of them belonging to vertebrates and arthropods (Dunn et al., 2014).

The sequencing of the octopus genome has now moved cephalopod research into a "post-genomic" era. While this lone genome provided important insights into the genetic features of these animals, additional sequenced cephalopod genomes will be necessary to address many of the questions raised. For example, the protocadherins were greatly expanded in the octopus genome. Ms. Wang and I mined the available transcriptomes for the longfin inshore squid, *Doryteuthis pealeii* and recovered nearly as many protocadherins from this dataset (Albertin et al., 2015; Brown et al., 2014). However, without a sequenced genome for *D. pealeii*, we do not know the genomic arrangement of these genes, or if we have recovered the full complement. In addition, phylogenetic analyses suggest independent expansions of the protocadherin subfamily

in octopus and squid. Sequencing additional cephalopod genomes will clarify if these are in fact independent expansions, and whether they are present in other clades. Similar questions remain about other gene family duplications detected in the octopus genome (Albertin et al., 2015).

We identified thousands of cephalopod-specific genes in the octopus genome and transcriptomes. While their absence of similarity to sequences found in other animals makes it is difficult to speculate on their function, novel genes are exciting molecular candidates to participate in the generation of cephalopod-unique structures. Two novel gene families have been investigated in cephalopods, namely the reflectins and suckerins, which encode structural proteins involved in adaptive coloration and in the rings and teeth of squid suckers, respectively (Crookes et al., 2004; Ding et al., 2014). We were able to identify a reflectin cluster in the octopus genome, but suckerins appear to be absent. By comparing genome sequences across cephalopods, we may be able to identify other such novel genes that play a role in the evolution and development of novel structures. These genes will not only be of interest to cephalopod biologists, but also biomaterials engineers and natural products pharmacologists. Comparisons across cephalopods and other molluscs at the level of both genomes and molecular embryology can be expected to reveal taxonomically restricted regulators important in the development of these extraordinary animals.

Concluding remarks

The work presented in this dissertation has provided new molecular tools for the study of cephalopods and insights into their development. At a molecular level, organogenic stage embryos demonstrate expression of the typical invertebrate developmental regulatory gene inventory in a conserved pattern, despite the fact that the morphological arrangement of this highly derived body plan is notably different from other animals. In contrast, while early

cephalopod development demonstrates striking morphological convergence with fish, my genomic, transcriptomic, and gene expression data clearly show is that there is little or no convergence between these two early developmental programs at the molecular level.

REFERENCES

Acampora D, Postiglione MP, Avantaggiato V, Di Bonito M, Simeone A. 2000. The role of Otx and Otp genes in brain development. Int J Dev Biol **44**:669-677.

Adachi K, Ohnishi K, Kuramochi T, Yoshinaga T, Okumura S-I. 2014. Molecular cytogenetic study in Octopus (Amphioctopus) areolatus from Japan. Fisheries Science **80**:445-450.

Albertin CB, Bonnaud L, Brown CT, Crookes-Goodson WJ, da Fonseca RR, Di Cristo C, Dilkes BP, Edsinger-Gonzales E, Freeman RM, Jr., Hanlon RT, Koenig KM, Lindgren AR, Martindale MQ, Minx P, Moroz LL, Nodl MT, Nyholm SV, Ogura A, Pungor JR, Rosenthal JJ, Schwarz EM, Shigeno S, Strugnell JM, Wollesen T, Zhang G, Ragsdale CW. 2012. Cephalopod genomics: A plan of strategies and organization. Stand Genomic Sci **7**:175-188.

Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, Brenner S, Ragsdale CW, Rokhsar DS. 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature **524**:220-224.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research **25**:3389-3402.

Arendt D, Nubler-Jung K. 1994. Inversion of dorsoventral axis? Nature **371**:26.

Arendt D, Nubler-Jung K. 1999. Comparison of early nerve cord development in insects and vertebrates. Development **126**:2309-2325.

Arendt D, Technau U, Wittbrodt J. 2001. Evolution of the bilaterian larval foregut. Nature **409**:81-85.

Arnold JM. 1968. The role of the egg cortex in cephalopod development. Developmental biology **18**:180-197.

Arnold JM. 1971. Cephalopods. In: Reverberi G, editor. Experimental embryology of marine and freshwater invertebrates: Akademie Verlag, Berlin. p 265-312.

Aronowicz J, Lowe CJ. 2006. Hox gene expression in the hemichordate Saccoglossus kowalevskii and the evolution of deuterostome nervous systems. Integrative and comparative biology **46**:890-901.

Aruga J, Odaka YS, Kamiya A, Furuya H. 2007. Dicyema Pax6 and Zic: tool-kit genes in a highly simplified bilaterian. BMC evolutionary biology **7**:201.

Averof M, Patel NH. 1997. Crustacean appendage evolution associated with changes in Hox gene expression. Nature **388**:682-686.

Ballow DJ, Xin Y, Choi Y, Pangas SA, Rajkovic A. 2006. Sohlh2 is a germ cell-specific bHLH transcription factor. Gene expression patterns : GEP **6**:1014-1018.

Baratte S, Andouche A, Bonnaud L. 2007. Engrailed in cephalopods: a key gene related to the emergence of morphological novelties. Development genes and evolution **217**:353-362.

Barrallo-Gimeno A, Nieto MA. 2005. The Snail genes as inducers of cell movement and survival: implications in development and cancer. Development **132**:3151-3161.

Barucca M, Canapa A, Biscotti M. 2016. An Overview of Hox Genes in Lophotrochozoa: Evolution and Functionality. Journal of Developmental Biology **4**:12.

Barucca M, Olmo E, Canapa A. 2003. Hox and paraHox genes in bivalve molluscs. Gene **317**:97-102.

Bassaglia Y, Bekel T, Da Silva C, Poulain J, Andouche A, Navet S, Bonnaud L. 2012. ESTs library from embryonic stages reveals tubulin and reflectin diversity in Sepia officinalis (Mollusca — Cephalopoda). Gene **498**:203-211.

Begun DJ, Lindfors HA, Kern AD, Jones CD. 2007. Evidence for de novo evolution of testis-expressed genes in the Drosophila yakuba/Drosophila erecta clade. Genetics **176**:1131-1137.

Biscotti MA, Canapa A, Olmo E, Barucca M. 2007. Hox genes in the antarctic polyplacophoran Nuttallochiton mirandus. J Exp Zool B Mol Dev Evol **308**:507-513.

Boletzky SV. 2003. Biology of early life stages in cephalopod molluscs. Advances in Marine Biology, Vol 44 **44**:143-203.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**:2114-2120.

Bolognesi R, Beermann A, Farzana L, Wittkopp N, Lutz R, Balavoine G, Brown SJ, Schroder R. 2008. Tribolium Wnts: evidence for a larger repertoire in insects with overlapping expression patterns that suggest multiple redundant functions in embryogenesis. Development genes and evolution **218**:193-202.

Bonnaud L, Ozouf-Costaz C, Boucher-Rodoni R. 2004. A molecular and karyological approach to the taxonomy of Nautilus. Comptes Rendus Biologies **327**:133-138.

Boyer BC, Henry JJ, Martindale MQ. 1998. The cell lineage of a polyclad turbellarian embryo reveals close similarity to coelomate spiralians. Developmental biology **204**:111-123.

Boyer BC, Henry JQ. 1998. Evolutionary Modifications of the Spiralian Developmental Program. American Zoologist **38**:621-633.

Brown CT, Graveley BR, Rosenthal JJ. 2014. *Loligo pealeii* (squid) data dump.

Brusca RC, Brusca GJ. 2003. Invertebrates. Sunderland, Mass.: Sinauer Associates. xix, 936 p. p.

Bullock TH. 1965. Mollusca: cephalopoda. In: Bullock TH, Horridge GA, editors. Structure and Function in the Nervous Systems of Invertebrates. San Francisco: W.H. Freeman.

Bullock TH, Horridge GA. 1965. Structure and function in the nervous systems of invertebrates. San Francisco,: W. H. Freeman.

Buresi A, Andouche A, Navet S, Bassaglia Y, Bonnaud-Ponticelli L, Baratte S. 2016. Nervous system development in cephalopods: How egg yolk-richness modifies the topology of the mediolateral patterning system. Developmental biology **415**:143-156.

Buresi A, Baratte S, Da Silva C, Bonnaud L. 2012. orthodenticle/otx ortholog expression in the anterior brain and eyes of Sepia officinalis (Mollusca, Cephalopoda). Gene expression patterns : GEP **12**:109-116.

Buresi A, Canali E, Bonnaud L, Baratte S. 2013. Delayed and asynchronous ganglionic maturation during cephalopod neurogenesis as evidenced by Sof-elav1 expression in embryos of Sepia officinalis (Mollusca, Cephalopoda). J Comp Neurol **521**.

Burglin TR, Kuwabara PE. 2006. Homologs of the Hh signalling network in C. elegans. WormBook : the online review of C elegans biology:1-14.

Cai X, Zhang Y. 2014. Marine invertebrate cell culture: a decade of development. Journal of Oceanography **70**:405-414.

Callaerts P, Lee PN, Hartmann B, Farfan C, Choy DW, Ikeo K, Fischbach KF, Gehring WJ, de Couet HG. 2002. HOX genes in the sepiolid squid Euprymna scolopes: implications for the evolution of complex body plans. Proceedings of the National Academy of Sciences of the United States of America **99**:2088-2093.

Cameron RA, Rowen L, Nesbitt R, Bloom S, Rast JP, Berney K, Arenas-Mena C, Martinez P, Lucas S, Richardson PM, Davidson EH, Peterson KJ, Hood L. 2006. Unusual gene order and organization of the sea urchin hox cluster. J Exp Zool B Mol Dev Evol **306**:45-58.

Canapa A, Biscotti MA, Olmo E, Barucca M. 2005. Isolation of Hox and ParaHox genes in the bivalve Pecten maximus. Gene **348**:83-88.

Carmona-Fontaine C, Theveneau E, Tzekou A, Tada M, Woods M, Page KM, Parsons M, Lambris JD, Mayor R. 2011. Complement fragment C3a controls mutual cell attraction during collective cell migration. Dev Cell **21**:1026-1037.

Carroll SB, Grenier JK, Weatherbee SD. 2005. From DNA to diversity : molecular genetics and the evolution of animal design. Malden, MA: Blackwell Pub. ix, 258 p. p.

Chen BE, Kondo M, Garnier A, Watson FL, Puettmann-Holgado R, Lamar DR, Schmucker D. 2006. The molecular diversity of Dscam is functionally required for neuronal wiring specificity in Drosophila. Cell **125**:607-620.

Chen WV, Maniatis T. 2013. Clustered protocadherins. Development **140**:3297-3302.

Cho SJ, Valles Y, Giani VC, Jr., Seaver EC, Weisblat DA. 2010. Evolutionary dynamics of the wnt gene family: a lophotrochozoan perspective. Molecular biology and evolution **27**:1645-1658.

Choi Y, Yuan D, Rajkovic A. 2008. Germ cell-specific transcriptional regulator sohlh2 is essential for early mouse folliculogenesis and oocyte-specific gene expression. Biol Reprod **79**:1176-1182.

Chow RL, Altmann CR, Lang RA, Hemmati-Brivanlou A. 1999. Pax6 induces ectopic eyes in a vertebrate. Development **126**:4213-4222.

Cornwell CJ, Messenger JB, Williamson R. 1993. Distribution of GABA-like immunoreactivity in the octopus brain. Brain Res **621**:353-357.

Crawford K. 2001. Ooplasm segregation in the squid embryo, Loligo pealeii. Biol Bull **201**:251-252.

Croll RP, Voronezhskaya EE. 1996. Early elements in gastropod neurogenesis. Developmental biology **173**:344-347.

Crookes WJ, Ding LL, Huang QL, Kimbell JR, Horwitz J, McFall-Ngai MJ. 2004. Reflectins: the unusual proteins of squid reflective tissues. Science **303**:235-238.

De Robertis EM, Sasai Y. 1996. A common plan for dorsoventral patterning in Bilateria. Nature **380**:37-40.

Del Giacco L, Pistocchi A, Cotelli F, Fortunato AE, Sordino P. 2008. A peek inside the neurosecretory brain through Orthopedia lenses. Dev Dyn **237**:2295-2303.

Denes AS, Jekely G, Steinmetz PR, Raible F, Snyman H, Prud'homme B, Ferrier DE, Balavoine G, Arendt D. 2007. Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. Cell **129**:277-288.

Dictus WJ, Damen P. 1997. Cell-lineage and clonal-contribution map of the trochophore larva of Patella vulgata (mollusca). Mech Dev **62**:213-226.

Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, Wing RA, Flavier A, Gaffney TD, Philippsen P. 2004. The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. Science **304**:304-307.

Ding D, Guerette PA, Hoon S, Kong KW, Cornvik T, Nilsson M, Kumar A, Lescar J, Miserez A. 2014. Biomimetic production of silk-like recombinant squid sucker ring teeth proteins. Biomacromolecules **15**:3278-3289.

Domazet-Loso T, Brajkovic J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. Trends in genetics : TIG **23**:533-539.

Dray N, Tessmar-Raible K, Le Gouar M, Vibert L, Christodoulou F, Schipany K, Guillou A, Zantke J, Snyman H, Behague J, Vervoort M, Arendt D, Balavoine G. 2010. Hedgehog signaling regulates segment formation in the annelid Platynereis. Science **329**:339-342.

Duboule D. 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev Suppl:135-142.

Duboule D. 2007. The rise and fall of Hox gene clusters. Development **134**:2549-2560.

Dunn CW, Giribet G, Edgecombe GD, Hejnol A. 2014. Animal Phylogeny and Its Evolutionary Implications. Annual Review of Ecology, Evolution, and Systematics **45**:371-395.

Dunn CW, Hejnol A. 2016. Inverse hourglass is due to the comb jelly.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature **452**:745-749.

Dunn CW, Ryan JF. 2015. The evolution of animal genomes. Current opinion in genetics & development **35**:25-32.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research **32**:1792-1797.

Eisenmann DM. 2005. Wnt signaling. WormBook : the online review of C elegans biology:1-17.

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

Farries MA. 2013. How 'basal' are the basal ganglia? Brain, behavior and evolution **82**:211-214.

Ferrier DE, Minguillon C, Holland PW, Garcia-Fernandez J. 2000. The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. Evolution & development **2**:284-293.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic acids research **42**:D222-230.

Fiorito G, Scotto P. 1992. Observational Learning in Octopus vulgaris. Science **256**:545-547.

Frank M, Kemler R. 2002. Protocadherins. Current opinion in cell biology **14**:557-562.

Freeman G, Lundelius JW. 1992. Evolutionary implications of the mode of D quadrant specification in coelomates with spiral cleavage. Journal of Evolutionary Biology **5**:205-247.

Freeman JL, Adeniyi A, Banerjee R, Dallaire S, Maguire SF, Chi J, Ng BL, Zepeda C, Scott CE, Humphray S, Rogers J, Zhou Y, Zon LI, Carter NP, Yang F, Lee C. 2007. Definition of the zebrafish genome using flow cytometry and cytogenetic mapping. BMC genomics **8**:195.

Fritsch M, Wollesen T, de Oliveira AL, Wanninger A. 2015. Unexpected co-linearity of Hox gene expression in an aculiferan mollusk. BMC evolutionary biology **15**:151.

Fritsch M, Wollesen T, Wanninger A. 2016. Hox and ParaHox gene expression in early body plan patterning of polyplacophoran mollusks. J Exp Zool B Mol Dev Evol **326**:89-104.

Frobius AC, Matus DQ, Seaver EC. 2008. Genomic organization and expression demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan Capitella sp. I. PloS one **3**:e4004.

Funch P, Mobjerg Kristensen R. 1995. Cycliophora is a new phylum with affinities to Entoprocta and Ectoprocta. Nature **378**:711-714.

Gabay L, Seger R, Shilo BZ. 1997. In situ activation pattern of Drosophila EGF receptor pathway during development. Science **277**:1103-1106.

Galis F, Sinervo B. 2003. Conserved Early Embryonic Stages. In: Hall BK, Olson WM, editors. Keywords and Concepts in Evolutionary Developmental Biology. Cambridge, Mass.: Harvard University Press. p xvi, 476 p.

Gehring WJ. 1998. Master control genes in development and evolution : the homeobox story. New Haven: Yale University Press. xv, 236 p. p.

Gesualdi SC, Haerry TE. 2007. Distinct signaling of Drosophila Activin/TGF-beta family members. Fly **1**:212-221.

Giribet G. 2008. Assembling the lophotrochozoan (=spiralian) tree of life. Philos Trans R Soc Lond B Biol Sci **363**:1513-1522.

Giribet G, Okusu A, Lindgren AR, Huff SW, Schrodl M, Nishiguchi MK. 2006. Evidence for a clade composed of molluscs with serially repeated structures: monoplacophorans are related to chitons. Proceedings of the National Academy of Sciences of the United States of America **103**:7723-7728.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology **29**:644-652.

Grasso FW, Basil JA. 2009. The evolution of flexible behavioral repertoires in cephalopod molluscs. Brain, behavior and evolution **74**:231-245.

Green SA, Norris RP, Terasaki M, Lowe CJ. 2013. FGF signaling induces mesoderm in the hemichordate Saccoglossus kowalevskii. Development **140**:1024-1033.

Gumienny TL, Savage-Dunn C. 2013. TGF-beta signaling in C. elegans. WormBook : the online review of C elegans biology:1-34.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols **8**:1494-1512.

Halder G, Callaerts P, Gehring WJ. 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science **267**:1788-1792.

Hallinan NM, Lindberg DR. 2011. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. Genome biology and evolution **3**:1150-1163.

Hanelt B, Van Schyndel D, Adema CM, Lewis LA, Loker ES. 1996. The phylogenetic position of Rhopalura ophiocomae (Orthonectida) based on 18S ribosomal DNA sequence analysis. Molecular biology and evolution **13**:1187-1191.

Hanlon RT, Forsythe JW. 1985a. Advances in the laboratory culture of octopuses for biomedical research. Lab Anim Sci **35**.

Hanlon RT, Messenger JB. 1996. Cephalopod behaviour. Cambridge ; New York: Cambridge University Press. xvi, 232 p. p.

Hansen EL. 1976. A cell line from embryos of Biomphalaria glabrata (Pulmonata): Establishment and characteristics. In: Maramorosch K, editor. Invertebrate tissue culture: Research applications. New York, NY: Academic Press. p 75-97.

Hartenstein V, Stollewerk A. 2015. The evolution of early neurogenesis. Dev Cell **32**.

Harvey RP. 1996. NK-2 homeobox genes and heart development. Developmental biology **178**:203-216.

Haszprunar G. 2000. Is the Aplacophora monophyletic? A cladistic point of view. American Malacological Bulletin **15**:115-130.

Hejnol A. 2010. A twist in time--the evolution of spiral cleavage in the light of animal phylogeny. Integrative and comparative biology **50**:695-706.

Henry JJ, Martindale MQ. 1999. Conservation and innovation in spiralian development. In: Dorresteijn AWC, Westheide W, editors. Reproductive Strategies and Developmental Patterns in Annelids. Dordrecht: Springer Netherlands. p 255-265.

Henry JJ, Perry KJ, Fukui L, Alvi N. 2010. Differential localization of mRNAs during early development in the mollusc, Crepidula fornicata. Integrative and comparative biology **50**:720-733.

Henry JQ. 2014. Spiralian model systems. Int J Dev Biol **58**:389-401.

Herculano-Houzel S, Mota B, Lent R. 2006. Cellular scaling rules for rodent brains. Proceedings of the National Academy of Sciences of the United States of America **103**:12138-12143.

Hinman VF, O'Brien EK, Richards GS, Degnan BM. 2003. Expression of anterior Hox genes during larval development of the gastropod Haliotis asinina. Evolution & development **5**:508-521.

Holland LZ, Carvalho JE, Escriva H, Laudet V, Schubert M, Shimeld SM, Yu JK. 2013. Evolution of bilaterian central nervous systems: a single origin? EvoDevo **4**:27.

Holland PW. 2013. Evolution of homeobox genes. Wiley Interdiscip Rev Dev Biol **2**:31-45.

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. 1994. Gene duplications and the origins of vertebrate development. Dev Suppl:125-133.

Holley SA, Jackson PD, Sasai Y, Lu B, De Robertis EM, Hoffmann FM, Ferguson EL. 1995. A conserved system for dorsal-ventral patterning in insects and vertebrates involving sog and chordin. Nature **376**:249-253.

Hooper JE, Scott MP. 2005. Communicating with Hedgehogs. Nature reviews Molecular cell biology **6**:306-317.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assuncao JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero

R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Urun Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberlander M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SM, Enright A, Geisler R, Plasterk RH, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nusslein-Volhard C, Hubbard TJ, Roest Crollius H, Rogers J, Stemple DL. 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature **496**:498-503.

Hulpiau P, van Roy F. 2009. Molecular evolution of the cadherin superfamily. The international journal of biochemistry & cell biology **41**:349-369.

Hulpiau P, van Roy F. 2011. New insights into the evolution of metazoan cadherins. Molecular biology and evolution **28**:647-657.

Hvorecny LM, Grudowski JL, Blakeslee CJ, Simmons TL, Roy PR, Brooks JA, Hanner RM, Beigel ME, Karson MA, Nichols RH, Holm JB, Boal JG. 2007. Octopuses (Octopus bimaculoides) and cuttlefishes (Sepia pharaonis, S. officinalis) can conditionally discriminate. Animal cognition **10**:449-459.

Iijima M. 2006. Evolution of Hox genes in molluscs: a comparison among seven morphologically diverse classes. Journal of Molluscan Studies **72**:259-266.

Irie N, Kuratani S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nature communications **2**:248.

Iuchi S. 2001. Three classes of C2H2 zinc finger proteins. Cellular and molecular life sciences : CMLS **58**:625-635.

Jacob MH. 1984. Neurogenesis in Aplysia californica resembles nervous system formation in vertebrates. The Journal of neuroscience : the official journal of the Society for Neuroscience **4**:1225-1239.

Jacobs DK, Wray CG, Wedeen CJ, Kostriken R, DeSalle R, Staton JL, Gates RD, Lindberg DR. 2000. Molluscan engrailed expression, serial organization, and shell evolution. Evolution & development **2**:340-347.

Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. BMC genomics **12**:164.

Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P. 2010. Gene expression divergence recapitulates the developmental hourglass model. Nature **468**:811-814.

Kang D. 2003. A hedgehog homolog regulates gut formation in leech (Helobdella). Development **130**:1645-1657.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature **428**:617-624.

Kim CH, Ueshima E, Muraoka O, Tanaka H, Yeo SY, Huh TL, Miki N. 1996. Zebrafish elav/HuC homologue as a very early neuronal marker. Neuroscience letters **216**:109-112.

King PH, Levine TD, Fremeau RT, Jr., Keene JD. 1994. Mammalian homologs of Drosophila ELAV localized to a neuronal subset can bind in vitro to the 3' UTR of mRNA encoding the Id transcriptional repressor. The Journal of neuroscience : the official journal of the Society for Neuroscience **14**:1943-1952.

Klug A. 2010. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. Quarterly reviews of biophysics **43**:1-21.

Koch BJ, Ryan JF, Baxevanis AD. 2012. The diversification of the LIM superclass at the base of the metazoa increased subcellular complexity and promoted multicellular specialization. PloS one **7**:e33261.

Kocot KM. 2013. Recent advances and unanswered questions in deep molluscan phylogenetics. American Malacological Bulletin **31**:195-208.

Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B, Halanych KM. 2011. Phylogenomics reveals deep molluscan relationships. Nature **477**:452-456.

Koenig K. 2015. Complex eye formation in the squid Doryteuthis pealeii and its evolutionary implications.

Kristof A, de Oliveira AL, Kolbin KG, Wanninger A. 2016. Neuromuscular development in Patellogastropoda (Mollusca: Gastropoda) and its importance for reconstructing ancestral gastropod bodyplan features. **54**:22-39.

Kroger B, Vinther J, Fuchs D. 2011. Cephalopod origin and evolution. BioEssays : news and reviews in molecular, cellular and developmental biology **33**:602-613.

Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ, Holstein TW. 2005. Unexpected complexity of the Wnt gene family in a sea anemone. Nature **433**:156-160.

Lambert JD, Nagy LM. 2001. MAPK signaling by the D quadrant embryonic organizer of the mollusc Ilyanassa obsoleta. Development **128**:45-56.

Laumer CE, Bekkouche N, Kerbl A, Goetz F, Neves RC, Sorensen MV, Kristensen RM, Hejnol A, Dunn CW, Giribet G, Worsaae K. 2015. Spiralian phylogeny informs the evolution of microscopic lineages. Current biology : CB **25**:2000-2006.

Lauri A, Brunet T, Handberg-Thorsager M, Fischer AH, Simakov O, Steinmetz PR, Tomer R, Keller PJ, Arendt D. 2014. Development of the annelid axochord: insights into notochord evolution. Science **345**:1365-1368.

Layden MJ, Meyer NP, Pang K, Seaver EC, Martindale MQ. 2010. Expression and phylogenetic analysis of the zic gene family in the evolution and development of metazoans. EvoDevo **1**:12.

Lee JJ, von Kessler DP, Parks S, Beachy PA. 1992. Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. Cell **71**:33-50.

Lee PN, Callaerts P, de Couet HG. 2009. Culture of Hawaiian bobtail squid (Euprymna scolopes) embryos and observation of normal development. Cold Spring Harb Protoc **2009**:pdb prot5323.

Lee PN, Callaerts P, De Couet HG, Martindale MQ. 2003. Cephalopod Hox genes and the origin of morphological novelties. Nature **424**:1061-1065.

Lefebvre JL, Kostadinov D, Chen WV, Maniatis T, Sanes JR. 2012. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. Nature **488**:517-521.

Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, Fernandez-Valverde SL, Nakanishi N, Simmons D, Simakov O, Larsson T, Liu SY, Jerafi-Vider A, Yaniv K, Ryan JF, Martindale MQ, Rink JC, Arendt D, Degnan SM, Degnan BM, Hashimshony T, Yanai I. 2016. The mid-developmental transition and the evolution of animal body plans. Nature **531**:637-641.

Levin M, Hashimshony T, Wagner F, Yanai I. 2012. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev Cell **22**:1101-1108.

Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. Proceedings of the National Academy of Sciences of the United States of America **103**:9935-9939.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics **12**:323.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754-1760.

Lindgren AR, Pankey MS, Hochberg FG, Oakley TH. 2012. A multi-gene phylogeny of Cephalopoda supports convergent morphological evolution in association with multiple habitat shifts in the marine environment. BMC evolutionary biology **12**:129.

Liu H, Chang LH, Sun Y, Lu X, Stubbs L. 2014. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. Genome biology and evolution **6**:510-525.

Lumsden A, Krumlauf R. 1996. Patterning the vertebrate neuraxis. Science **274**:1109-1115.

Lyons DC, Henry JQ. 2014. Ins and outs of Spiralian gastrulation. Int J Dev Biol **58**:413-428.

Lyons DC, Perry KJ, Henry JQ. 2015. Spiralian gastrulation: germ layer formation, morphogenesis, and fate of the blastopore in the slipper snail Crepidula fornicata. EvoDevo **6**:24.

Marquis F. 1989. Die Embryonalentwicklung des Nervensysem von Octopus vulgaris Lam. (Cephalopoda, Octopoda), eine histologische Analyse. Verhandl Naturf Ges (Basel) **99**.

Marthy H-J. 1987. Ontogenesis of the Nervous System in Cephalopods. In: Ali MA, editor. Nervous Systems in Invertebrates. Boston, MA: Springer US. p 443-459.

Maslakova SA, Hiebert TC. 2014. From trochophore to pilidium and back again - a larva's journey. Int J Dev Biol **58**:585-591.

McGinnis W, Garber RL, Wirz J, Kuroiwa A, Gehring WJ. 1984. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell **37**:403-408.

McGlinn E, Tabin CJ. 2006. Mechanistic insight into how Shh patterns the vertebrate limb. Current opinion in genetics & development **16**:426-432.

Meyer NP, Seaver EC. 2009. Neurogenesis in an annelid: characterization of brain neural precursors in the polychaete Capitella sp. I. Developmental biology **335**:237-252.

Mi H, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function analysis with the PANTHER classification system. Nature protocols **8**:1551-1566.

Morishita H, Yagi T. 2007. Protocadherin family: diversity, structure, and function. Current opinion in cell biology **19**:584-592.

Moroz LL. 2009. On the independent origins of complex brains and neurons. Brain, behavior and evolution **74**:177-190.

Moroz LL, Edwards JR, Puthanveettil SV, Kohn AB, Ha T, Heyland A, Knudsen B, Sahni A, Yu F, Liu L, Jezzini S, Lovell P, Iannucculli W, Chen M, Nguyen T, Sheng H, Shaw R, Kalachikov S, Panchin YV, Farmerie W, Russo JJ, Ju J, Kandel ER. 2006. Neuronal transcriptome of Aplysia: neuronal compartments and circuitry. Cell **127**:1453-1467.

Naef A. 1928. Die Cephalopoden (Embryologie). Washington: Fauna Flora Golfo Napoli: Smithsonian Institution.

Navet S, Andouche A, Baratte S, Bonnaud L. 2009. Shh and Pax6 have unconventional expression patterns in embryonic morphogenesis in Sepia officinalis (Cephalopoda). Gene expression patterns : GEP **9**:461-467.

Navet S, Bassaglia Y, Baratte S, Martin M, Bonnaud L. 2008. Somatic muscle development in Sepia officinalis (cephalopoda - mollusca): a new role for NK4. Dev Dyn **237**:1944-1951.

Nederbragt AJ, te Welscher P, van den Driesche S, van Loon AE, Dictus WJ. 2002. Novel and conserved roles for orthodenticle/ otx and orthopedia/ otp orthologs in the gastropod mollusc Patella vulgata. Development genes and evolution **212**:330-337.

Nithianantharajah J, Komiyama NH, McKechanie A, Johnstone M, Blackwood DH, St Clair D, Emes RD, van de Lagemaat LN, Saksida LM, Bussey TJ, Grant SG. 2013. Synaptic scaffold evolution generated components of vertebrate cognitive complexity. Nature neuroscience **16**:16-24.

Nixon M, Young JZ. 2003. The brains and lives of cephalopods. Oxford ; New York: Oxford University Press. xiv, 392 p. p.

Nomaksteinsky M, Rottinger E, Dufour HD, Chettouh Z, Lowe CJ, Martindale MQ, Brunet JF. 2009. Centralization of the deuterostome nervous system predates chordates. Current biology : CB **19**:1264-1269.

Noonan JP, Grimwood J, Schmutz J, Dickson M, Myers RM. 2004. Gene conversion and the evolution of protocadherin gene cluster diversity. Genome research **14**:354-366.

Northcutt RG. 2012. Evolution of centralized nervous systems: two schools of evolutionary thought. Proceedings of the National Academy of Sciences of the United States of America **109 Suppl 1**:10626-10633.

Nowick K, Hamilton AT, Zhang H, Stubbs L. 2010. Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. Molecular biology and evolution **27**:2606-2617.

Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag. xv, 160 p. p.

Packard A. 1972. Cephalopods and fish: the limits of convergence. Biological Reviews **47**:241-307.

Page LR. 2006. Early differentiating neuron in larval abalone (Haliotis kamtschatkana) reveals the relationship between ontogenetic torsion and crossing of the pleurovisceral nerve cords. Evolution & development **8**:458-467.

Palmieri N, Kosiol C, Schlotterer C. 2014. The life cycle of Drosophila orphan genes. eLife **3**:e01311.

Pang K, Ryan JF, Program NCS, Mullikin JC, Baxevanis AD, Martindale MQ. 2010. Genomic insights into Wnt signaling in an early diverging metazoan, the ctenophore Mnemiopsis leidyi. EvoDevo **1**:10.

Pani AM, Mullarkey EE, Aronowicz J, Assimacopoulos S, Grove EA, Lowe CJ. 2012. Ancient deuterostome origins of vertebrate brain signalling centres. Nature **483**:289-294.

Paps J, Xu F, Zhang G, Holland PW. 2015. Reinforcing the egg-timer: recruitment of novel lophotrochozoa homeobox genes to early and late development in the pacific oyster. Genome biology and evolution **7**:677-688.

Paridaen JT, Huttner WB. 2014. Neurogenesis during development of the vertebrate central nervous system. EMBO Rep **15**:351-364.

Perea-Atienza E, Gavilan B, Chiodin M, Abril JF, Hoff KJ, Poustka AJ, Martinez P. 2015. The nervous system of Xenacoelomorpha: a genomic perspective. J Exp Biol **218**:618-628.

Perez-Paralle ML, Carpintero P, Pazos AJ, Abad M, Sanchez JL. 2005. The HOX gene cluster in the bivalve mollusc Mytilus galloprovincialis. Biochem Genet **43**:417-424.

Pernice M, Deutsch JS, Andouche A, Boucher-Rodoni R, Bonnaud L. 2006. Unexpected variation of Hox genes' homeodomains in cephalopods. Molecular phylogenetics and evolution **40**:872-879.

Perry KJ, Lyons DC, Truchado-Garcia M, Fischer AH, Helfrich LW, Johansson KB, Diamond JC, Grande C, Henry JQ. 2015. Deployment of regulatory genes during gastrulation and germ layer specification in a model spiralian mollusc Crepidula. Dev Dyn **244**:1215-1248.

Pickford GE, McConnaughey BH. 1949. The *Octopus bimaculatus* problem: a study in sibling species. Bulletin of the Bingham Oceanographic Collection **12**:1-66.

Pires-daSilva A, Sommer RJ. 2003. The evolution of signalling pathways in animal development. Nature reviews Genetics **4**:39-49.

Ponder WF, Lindberg DR. 2008. Phylogeny and evolution of the Mollusca. Berkeley: University of California Press. xi, 469 p. p.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS one **5**:e9490.

Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research **33**:D501-504.

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature **453**:1064-1071.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science **317**:86-94.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**:841-842.

Raff RA. 1996. The shape of life : genes, development, and the evolution of animal form. Chicago: University of Chicago Press. xxiii, 520 p. p.

Reichert H. 2005. A tripartite organization of the urbilaterian brain: developmental genetic evidence from Drosophila. Brain Res Bull **66**:491-494.

Reiner A, Medina L, Veenman CL. 1998. Structural and functional evolution of the basal ganglia in vertebrates. Brain Res Brain Res Rev **28**:235-285.

Render J. 1997. Cell fate maps in the Ilyanassa obsoleta embryo beyond the third division. Developmental biology **189**:301-310.

Richter S, Loesel R, Purschke G, Schmidt-Rhaesa A, Scholtz G, Stach T, Vogt L, Wanninger A, Brenneis G, Doring C, Faller S, Fritsch M, Grobe P, Heuer CM, Kaul S, Moller OS, Muller CH, Rieger V, Rothe BH, Stegner ME, Harzsch S. 2010. Invertebrate neurophylogeny: suggested terms and definitions for a neuroanatomical glossary. Front Zool **7**:29.

Rinkevich B. 2011. Cell cultures from marine invertebrates: new insights for capturing endless stemness. Mar Biotechnol (NY) **13**:345-354.

Robinow S, White K. 1991. Characterization and spatial distribution of the ELAV protein during Drosophila melanogaster development. J Neurobiol **22**:443-461.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**:139-140.

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LS, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJ, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vosshall LB, Zhang J, Zhao Q, Zheng XH, Lewis S. 2000. Comparative genomics of the eukaryotes. Science **287**:2204-2215.

Ryan JF, Pang K, Schnitzler CE, Nguyen AD, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, Program NCS, Smith SA, Putnam NH, Haddock SH, Dunn CW, Wolfsberg TG, Mullikin JC, Martindale MQ, Baxevanis AD. 2013. The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. Science **342**:1242592.

Ryan K, Garrett N, Mitchell A, Gurdon JB. 1996. Eomesodermin, a key early gene in Xenopus mesoderm differentiation. Cell **87**:989-1000.

Sadamoto H, Takahashi H, Okada T, Kenmoku H, Toyota M, Asakawa Y. 2012. De novo sequencing and transcriptome analysis of the central nervous system of mollusc Lymnaea stagnalis by deep RNA sequencing. PloS one **7**:e42546.

Salvini-Plawen Lv, Steiner G. 1996. Synapomorphies and plesiomorphies in higher classification of the Mollusca. Origin and evolutionary radiation of the Mollusca. Oxford ; New York: Oxford University Press. p xiv, 392 p.

Samadi L, Steiner G. 2009. Involvement of Hox genes in shell morphogenesis in the encapsulated development of a top shell gastropod (Gibbula varia L.). Development genes and evolution **219**:523-530.

Samadi L, Steiner G. 2010. Expression of Hox genes during the larval development of the snail, Gibbula varia (L.)-further evidence of non-colinearity in molluscs. Development genes and evolution **220**:161-172.

Santagata S, Resh C, Hejnol A, Martindale MQ, Passamaneck YJ. 2012. Development of the larval anterior neurogenic domains of Terebratalia transversa (Brachiopoda) provides insights into the diversification of larval apical organs and the spiralian nervous system. EvoDevo **3**:3.

Schmitz JF, Zimmer F, Bornberg-Bauer E. 2016. Mechanisms of transcription factor evolution in Metazoa. Nucleic acids research.

Schnitzler CE, Simmons DK, Pang K, Martindale MQ, Baxevanis AD. 2014. Expression of multiple Sox genes through embryonic development in the ctenophore Mnemiopsis leidyi is spatially restricted to zones of cell proliferation. EvoDevo **5**:15.

Schreiner D, Weiner JA. 2010. Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. Proceedings of the National Academy of Sciences of the United States of America **107**:14893-14898.

Sebe-Pedros A, de Mendoza A. 2015. Transcription factors and the origin of animal multicellularity. In: Ruiz-Trillo I, Nedelcu A, editors. Evolutionary transitions to multicellular life : principles and mechanisms. New York, NY: Springer Berlin Heidelberg. p 379-394.

Semmler H, Chiodin M, Bailly X, Martinez P, Wanninger A. 2010. Steps towards a centralized nervous system in basal bilaterians: insights from neurogenesis of the acoel Symsagittifera roscoffensis. Development, growth & differentiation **52**:701-713.

Sen S, Reichert H, VijayRaghavan K. 2013. Conserved roles of ems/Emx and otd/Otx genes in olfactory and visual system development in Drosophila and mouse. Open Biol **3**:120177.

Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D. 2004. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. Nature **431**:67-71.

Seo S, Lim JW, Yellajoshyula D, Chang LW, Kroll KL. 2007. Neurogenin and NeuroD direct transcriptional targets and their regulatory enhancers. EMBO J **26**:5093-5108.

Sever R, Glass CK. 2013. Signaling by nuclear receptors. Cold Spring Harbor perspectives in biology **5**:a016709.

Shannon M, Hamilton AT, Gordon L, Branscomb E, Stubbs L. 2003. Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters. Genome research **13**:1097-1110.

Shigeno S, Kidokoro H, Tsuchiya K, Segawa S, Yamamoto M. 2001a. Development of the brain in the oegopsid squid, Todarodes pacificus: an atlas up to the hatching stage. Zool Sci **18**.

Shigeno S, Parnaik R, Albertin CB, Ragsdale CW. 2015. Evidence for a cordal, not ganglionic, pattern of cephalopod brain neurogenesis. Zoological Letters **1**:1-13.

Shigeno S, Ragsdale CW. 2015. The gyri of the octopus vertical lobe have distinct neurochemical identities. J Comp Neurol **523**:1297-1317.

Shigeno S, Sasaki T, Moritaki T, Kasugai T, Vecchione M, Agata K. 2008. Evolution of the cephalopod head complex by assembly of multiple molluscan body parts: Evidence from Nautilus embryonic development. J Morphol **269**.

Shigeno S, Tsuchiya K, Segawa S. 2001b. Embryonic and paralarval development of the central nervous system of the loliginid squid Sepioteuthis lessoniana. J Comp Neurol **437**:449-475.

Shomrat T, Graindorge N, Bellanger C, Fiorito G, Loewenstein Y, Hochner B. 2011. Alternative sites of synaptic plasticity in two homologous "fan-out fan-in" learning and memory networks. Current biology : CB **21**:1773-1782.

Showell C, Binder O, Conlon FL. 2004. T-box genes in early embryogenesis. Dev Dyn **229**:201-218.

Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty. Nature **457**:818-823.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology **7**:539.

Sigwart JD, Lindberg DR. 2015. Consensus and Confusion in Molluscan Trees: Evaluating Morphological and Molecular Phylogenies. Systematic biology **64**:384-395.

Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo DH, Larsson T, Lv J, Arendt D, Savage R, Osoegawa K, de Jong P, Grimwood J, Chapman JA, Shapiro H, Aerts A, Otillar RP, Terry AY, Boore JL, Grigoriev IV, Lindberg DR, Seaver EC, Weisblat DA, Putnam NH, Rokhsar DS. 2013. Insights into bilaterian evolution from three spiralian genomes. Nature **493**:526-531.

Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M. 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC evolutionary biology **7**:33.

Slack JM, Holland PW, Graham CF. 1993. The zootype and the phylotypic stage. Nature **361**:490-492.

Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SC, Rouse GW, Giribet G, Dunn CW. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature **480**:364-367.

Söller R, Warnke K, Saint-Paul U, Blohm D. 2000. Sequence divergence of mitochondrial DNA indicates cryptic biodiversity in Octopus vulgaris and supports the taxonomic distinctiveness of Octopus mimus (Cephalopoda: Octopodidae). Marine Biology **136**:29-35.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF,

Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS. 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature **466**:720-726.

Stathopoulos A, Tam B, Ronshaugen M, Frasch M, Levine M. 2004. pyramus and thisbe: FGF genes that pattern the mesoderm of Drosophila embryos. Genes & development **18**:687-699.

Stoger I, Sigwart JD, Kano Y, Knebelsberger T, Marshall BA, Schwabe E, Schrodl M. 2013. The continuing debate on deep molluscan phylogeny: evidence for Serialia (Mollusca, Monoplacophora + Polyplacophora). Biomed Res Int **2013**:407072.

Strausfeld NJ, Hirth F. 2013. Deep homology of arthropod central complex and vertebrate basal ganglia. Science **340**:157-161.

Strugnell J, Norman M, Jackson J, Drummond AJ, Cooper A. 2005. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework. Molecular phylogenetics and evolution **37**:426-441.

Sweeney AM, Haddock SH, Johnsen S. 2007. Comparative visual acuity of coleoid cephalopods. Integrative and comparative biology **47**:808-814.

Sweet HC. 1998. Specification of first quartet micromeres in Ilyanassa involves inherited factors and position with respect to the inducing D macromere. Development **125**:4033-4044.

Szabo A, Cobo I, Omara S, McLachlan S, Keller R, Mayor R. 2016. The Molecular Basis of Radial Intercalation during Tissue Spreading in Early Development. Dev Cell **37**:213-225.

Tabata T, Kornberg TB. 1994. Hedgehog is a signaling protein with a key role in patterning Drosophila imaginal discs. Cell **76**:89-102.

Takeichi M. 2007. The cadherin superfamily in neuronal connections and interactions. Nature reviews Neuroscience **8**:11-20.

Tarazona OA, Slota LA, Lopez DH, Zhang G, Cohn MJ. 2016. The genetic program for cartilage development has deep homology within Bilateria. Nature **533**:86-89.

Tautz D, Domazet-Loso T. 2011. The evolutionary origin of orphan genes. Nature reviews Genetics **12**:692-702.

Tessmar-Raible K, Raible F, Christodoulou F, Guy K, Rembold M, Hausen H, Arendt D. 2007. Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. Cell **129**:1389-1400.

Tomer R, Denes AS, Tessmar-Raible K, Arendt D. 2010. Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. Cell **142**:800-809.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25**:1105-1111.

UniProt C. 2015. UniProt: a hub for protein information. Nucleic acids research **43**:D204-212.

Urbach R. 2007. A procephalic territory in Drosophila exhibiting similarities and dissimilarities compared to the vertebrate midbrain/hindbrain boundary region. Neural Dev **2**:23.

Vanhalst K, Kools P, Staes K, van Roy F, Redies C. 2005. delta-Protocadherins: a gene family expressed differentially in the mouse brain. Cellular and molecular life sciences : CMLS **62**:1247-1259.

Vellutini BC, Hejnol A. 2016. Expression of segment polarity genes in brachiopods supports a nonsegmental ancestral role of engrailed for bilaterians. bioRxiv.

Vidal EA, Villanueva R, Andrade JP, Gleadall IG, Iglesias J, Koueta N, Rosas C, Segawa S, Grasse B, Franco-Santos RM, Albertin CB, Caamal-Monsreal C, Chimal ME, Edsinger-Gonzales E, Gallardo P, Le Pabic C, Pascual C, Roumbedakis K, Wood J. 2014. Cephalopod culture: current status of main biological models and research priorities. Adv Mar Biol **67**:1-98.

Vinogradov AE. 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. Cytometry **31**:100-109.

Vinogradov AE. 2013. Density peaks of paralog pairs in human and mouse genomes. Gene **527**:55-61.

Wada H, Saiga H, Satoh N, Holland PW. 1998. Tripartite organization of the ancestral chordate brain and the antiquity of placodes: insights from ascidian Pax-2/5/8, Hox and Otx genes. Development **125**:1113-1122.

Wadeson PH, Crawford K. 2003. Formation of the blastoderm and yolk syncytial layer in early squid development. Biol Bull **205**:179-180.

Wakamatsu Y, Weston JA. 1997. Sequential expression and role of Hu RNA-binding proteins during neurogenesis. Development **124**:3449-3460.

Wanninger A, Wollesen T. 2015. Mollusca. In: Wanninger A, editor. Evolutionary developmental biology of invertebrates. Vienna: Springer Verlag. p 103-153.

Wells MJ. 1978. Octopus : physiology and behaviour of an advanced invertebrate. London New York: Chapman and Hall. xiii, 417 p., 413 leaves of plates p.

Wharton K, Derynck R. 2009. TGFbeta family signaling: novel insights in development and disease. Development **136**:3691-3697.

Wissler L, Gadau J, Simola DF, Helmkampf M, Bornberg-Bauer E. 2013. Mechanisms and dynamics of orphan gene emergence in insect genomes. Genome biology and evolution **5**:439-455.

Wojtowicz WM, Wu W, Andre I, Qian B, Baker D, Zipursky SL. 2007. A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. Cell **130**:1134-1145.

Wollesen T, Rodriguez Monje SV, Todt C, Degnan BM, Wanninger A. 2015. Ancestral role of Pax2/5/8 in molluscan brain and multimodal sensory system development. BMC evolutionary biology **15**:231.

Yagi T. 2008. Clustered protocadherin family. Development, growth & differentiation **50 Suppl 1**:S131-140.

Yamamoto M, Shimazaki Y, Shigeno S. 2003. Atlas of the embryonic brain in the pygmy squid, Idiosepius paradoxus. Zoolog Sci **20**:163-179.

Yochelson EL, Flower RH, Webers GF. 1973. The bearing of the new late Cambrian monoplacophoran Knightconus upon the origin of the Cephalopoda. Lethaia **6**.

Yoshida MA, Ishikura Y, Moritaki T, Shoguchi E, Shimizu KK, Sese J, Ogura A. 2011. Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. Gene **483**:63-71.

Young JZ. 1971. The anatomy of the nervous system of Octopus vulgaris. Oxford,: Clarendon Press. xxxi, 690 p. p.

Yu WP, Rajasegaran V, Yew K, Loh WL, Tay BH, Amemiya CT, Brenner S, Venkatesh B. 2008. Elephant shark sequence reveals unique insights into the evolutionary history of vertebrate genes: A comparative analysis of the protocadherin cluster. Proceedings of the National Academy of Sciences of the United States of America **105**:3819-3824.

Zipursky SL, Sanes JR. 2010. Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. Cell **143**:343-353.