2017-12-13

The Evolution of the Krüppel-Like Factor Gene Family and Their Function During Embryonic Development in the Ctenophore Mnemiopsis leidyi

Jason Scott Presnell
University of Miami, jspresnell@gmail.com

UNIVERSITY OF MIAMI

THE EVOLUTION OF THE KRÜPPEL-LIKE FACTOR GENE FAMILY AND THEIR FUNCTION DURING EMBRYONIC DEVELOPMENT IN THE CTENOPHORE *MNEMIOPSIS LEIDYI*

By

Jason Scott Presnell

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2017

UNIVERSITY OF MIAMI

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

THE EVOLUTION OF THE KRÜPPEL-LIKE FACTOR GENE FAMILY AND THEIR FUNCTION DURING EMBRYONIC DEVELOPMENT IN THE CTENOPHORE *MNEMIOPSIS LEIDYI*

Jason Scott Presnell

Approved:

_____	_____
William E. Browne, Ph.D. Assistant Professor of Biology	Julia Dallman, Ph.D. Associate Professor of Biology
_____	_____
Athula Wikramanayake, Ph.D. Professor of Biology	James Baker, Ph.D. Assistant Scientist of Biology
_____	_____
Mark Q. Martindale, Ph.D. Professor of Biology University of Florida	Guillermo Prado, Ph.D. Dean of the Graduate School

PRESNELL, JASON SCOTT (Ph.D., Biology)

The Evolution of the Krüppel-Like Factor Gene Family and Their Function During Embryonic Development in the Ctenophore *Mnemiopsis leidyi* (December 2017)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Assistant Professor William E. Browne.
No. of pages in text. (168)

How modifications during development influence large-scale evolutionary changes is a major emphasis in the field of evolutionary developmental biology (evo-devo). Comparing the underlying cellular and molecular basis of development in diverse organisms can provide insight into how alterations of these processes contribute to morphological diversity. Ctenophores, or comb jellies, are a phylum of marine invertebrates and, due to their unique phylogenetic position as one of the earliest branching lineages of extant animals, are an ideal system for investigating the origin and evolution of metazoan character traits. For example, the animal through-gut is thought to have originated in the metazoan stem lineage prior to diversification of Bilateria. However, as part of my dissertation, I showed that ctenophores possess a functional through-gut. These results suggest that the origin of the through-gut may have been much earlier during metazoan diversification than previously thought. Additionally, my studies of gene function in ctenophores provide insight into the molecular mechanisms that, in some part, may be driving early metazoan diversification. Members of the *Krüppel-like factor (Klf)* gene family are known to play a role in many aspects of development, including maintenance of stem cell renewal and pluripotency, and regulating the balance between

cellular proliferation and differentiation. Functional studies of these transcription factors have been restricted to only a handful of bilaterian animals, with little investigation into the evolutionary history of this gene family. To evaluate how the *Klf* gene family could influence evolutionary changes driving metazoan diversification, I examined the origins and evolutionary history of the *Klf* gene family, and characterized their function in the ctenophore, *Mnemiopsis leidyi*. Ultimately, I showed that *Klf* genes were present in the stem lineage leading to Metazoa, and they regulate cellular proliferation in putative stem cell niches during development in *M. leidyi*. My findings suggest that regulation of stem cell proliferation was an ancestral function of *Klfs* in metazoans.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Bill Browne, and my committee members for their insight, advice, and support. I would also like to thank Marine Biological Labs (Embryology Course 2014) for hosting an amazing and incredibly useful course that I attended, and applied the lessons learned into my dissertation work. Portions of my research were generously supported by the Kushlan Fund and additional funding came from the College of Arts and Sciences.

TABLE OF CONTENTS

Page

LIST OF FIGURES v

LIST OF TABLES vii

Chapter

1 Introduction 1

2 Laboratory culture and breeding of the lobate ctenophore *Mnemiopsis leidyi* 11

3 Presence of a functionally tripartite through-gut in Ctenophora has implications for metazoan character trait evolution 31

4 KLF/SP transcription factor family evolution: expansion, diversification, and innovation in eukaryotes 53

5 Zygotic function of *Krüppel-like factor* genes during embryogenesis in *Mnemiopsis leidyi* is associated with cell proliferation 96

6 Synthesis 140

References 147

LIST OF FIGURES

Page

Chapter 1

1.1 Phylogenetic tree showing the relationships of the five extant metazoan clades 9
1.2 A representative image of an adult *Mnemiopsis leidyi*. 10

Chapter 2

2.1 Body plan of *Mnemiopsis leidyi* with major axes labeled 25
2.2 Equipment used for housing, feeding, and tank maintenance 26
2.3 Mature gametes and stages of embryonic development in *Mnemiopsis leidyi* 28
2.4 Stages of juvenile development in *Mnemiopsis leidyi* 30

Chapter 3

3.1 Ctenophore digestive system anatomy 44
3.2 Waste evacuation in ctenophores 45
3.3 Aboral region of *Mnemiopsis* 46
3.4 Evolutionary scenarios for the origins of the animal through gut 47
3.5 Specialized cell types associated with the ctenophore gut 49
3.6 Alternative scenarios of animal through-gut evolution 50
3.7 Gene expression associated with the ctenophore gut 52

Chapter 4

4.1 Distribution of C2H2 zinc finger proteins, KLF-DBD containing proteins, and KLF/SP proteins in representative Eukarya taxa 84
4.2 KLF DNA binding domain 85
4.3 Combined gene tree estimates for the concatenated KLF/SP data 86
4.4 Phylogenetic distribution of transactivation/repression domains and LCRs associated with KLF/SP proteins 87
4.5 KLF/SP protein domain co-occurrence networks 89
4.6 Phylogenetic distribution of explicit domain architectures represented among KLF/SP proteins 90
4.7 Inferred relationships between key events during the evolution and expansion of the KLF/SP gene family 92

Chapter 5

5.1 Gene models for *MleKlf5a*, *MleKlf5b*, and *MleKlfX* 127
5.2 Maternally deposited *MleKlf5a* and *MleKlf5b* transcripts 128
5.3 Zygotic expression of *MleKlf5a* and *MleKlf5b* 130
5.4 Live image of a wildtype cydippid stained with vital dyes 131

5.5 Endodermal and pharyngeal phenotypes associated with knockdown of *MleKlf5a* and *MleKlf5b* expression through morpholinos and Cas9 and knockdown of *brachyury* with Cas9 132
5.6 Cellular proliferation, measured by EdU incorporation, during embryogenesis 133
5.7 *MleKlf5b* expression and EdU incorporation in the ctene row associated cells 134
5.8 Tentacle bulb phenotypes due to knockdown of *MleKlf5a* and *MleKlf5b* expression 135
5.9 Loss of lithocytes due to knockdown of *MleKlf5a* and *MleKlf5b* expression 136
5.10 Zygotic expression of *MleKlfX* 137
5.11 Percentage of embryos that had developed normally under different conditions 138

LIST OF TABLES

Page

Chapter 4

4.1 Species used in this study with genome/transcriptome reference and abbreviations from phylogenetic trees 93

Chapter 5

5.1 A list of top hits of putative off-target sites for guide RNAs (gRNAs)..... 139

Chapter 1: Introduction

Background

Multicellular animals (Metazoa) are divided into two major groups: bilaterians (animals that have bilateral symmetry), and non-bilaterians (animals that are thought to lack bilateral symmetry), see (Dunn, et al. 2014). The four non-bilaterian phyla, Cnidaria, Placozoa (but see Laumer, et al. 2017), Porifera, and Ctenophora, represent the earliest branching extant metazoan lineages (Dunn, et al. 2008; Hejnol, et al. 2009; King and Rokas 2017; Fig. 1.1). These animals display radically different body plans from bilaterian animals. Genomic studies of non-bilaterians, however, have shown that many transcription factors and signaling pathways are conserved in non-bilaterians (Putnam, et al. 2007; Srivastava, et al. 2008; Srivastava, et al. 2010; Ryan, et al. 2013; Moroz, et al. 2014). Comparative molecular and cellular studies between non-bilaterian and bilaterian animals have provided insight into the early evolution of key metazoan character traits such as the formation of mesoderm (Martindale, et al. 2004; Steinmetz, et al. 2017), muscle cells (Dayraud, et al. 2012; Steinmetz, et al. 2012; Ryan, et al. 2013), and a nervous system (Watanabe, et al. 2009; Simmons, et al. 2012; Moroz, et al. 2014; Leys 2015). Investigations of gene family evolution in non-bilaterians have provided an improved understanding of gene evolution dynamics potentially driving morphological diversity, such as transcription factor expansion (Degnan, et al. 2009), and gain/loss of regulatory domains (de Mendoza, et al. 2013). Additional studies of gene family evolution and their role in development in non-bilaterians will continue to improve our insight into the early evolution of metazoans. Here, I explore the functional anatomy of

the ctenophore gut, elucidate the evolution of the *Kruppel-like factor* (*Klf*) gene family, and examine *Klf* function during development in the ctenophore *Mnemiopsis leidyi*.

Overview of ctenophores

Ctenophores are a monophyletic group of globally distributed gelatinous, primarily pelagic, marine predators united by a suite of morphological synapomorphies (Chun 1880; Hyman 1940; Harbison, et al. 1978; Harbison 1985; Mills and Haddock 2007; Pang and Martindale 2008e). These include eight longitudinal rows of fused cilia called ctenes used for locomotion (Chun 1880; Afzelius 1961; Tamm 1973, 2014a), tentacles with colloblast cells that produce a sticky glue (Chun 1880; Abbott 1907; Bargmann, et al. 1972; Benwitz 1978; von Byern, et al. 2010), and an aboral gravity-sensing apical organ used to regulate beating of the ctene rows (Chun 1880; Horridge 1965; Tamm 2014b). Historically the phylum Ctenophora was organized into six orders based on morphology (Harbison 1985), however recent molecular analyses question the monophyly of several traditionally recognized relationships (Podar, et al. 2001; Simion, et al. 2015; Whelan, et al. 2017). Some species exhibit modified life stages, such as adult individuals of the order Platyctenida who lose their ctene rows and adopt a benthic lifestyle (Komai 1922) and *Beroe* species that lack tentacles and feed on other ctenophores (Chun 1880).

Nearly all ctenophores are considered to be self-fertile hermaphrodites (Hyman 1940; Harbison and Miller 1986) capable of producing thousands of embryos during spawning events (Baker and Reeve 1974; Reeve, et al. 1978). Availability of large numbers of transparent embryos from wild caught animals made ctenophores an early

attractive model for embryological studies (Agassiz 1874; Chun 1880; Hertwig 1880; Driesch and Morgan 1895; Fischel 1897; Yatsu 1911). These studies were among the first to recognize that ctenophore embryos undergo a unique phylum-specific cleavage program from which embryos hatch as free-swimming juveniles, or cydippids, roughly 24 hours after fertilization.

Mnemiopsis leidyi is a lobate ctenophore (Fig. 1.2), seasonally abundant in coastal areas. *M. leidyi* embryos have been used extensively to examine general aspects of ctenophore embryogenesis including: cell fate mapping, axis and symmetry formation, induction signals, organizer activity, and cleavage clocks (Freeman 1976a, b, 1977; Martindale 1986; Martindale and Henry 1995, 1996, 1997c, 1999; Henry and Martindale 2000, 2001, 2004; Fischer, et al. 2014). Recent analyses of expression patterns for many genes in *Mnemiopsis* embryos and *Pleurobrachia* adults have provided insight into the early evolution of transcription factors and signaling pathways involved in cell fate specification and axial patterning (Yamada and Martindale 2002; Derelle and Manuel 2007; Yamada, et al. 2007; Jager, et al. 2008; Pang and Martindale 2008f; Layden, et al. 2010; Pang, et al. 2010; Alié, et al. 2011; Pang, et al. 2011; Reitzel, et al. 2011; Dayraud, et al. 2012; Schnitzler, et al. 2012; Simmons, et al. 2012; Jager, et al. 2013; Schnitzler, et al. 2014; Reitzel, et al. 2016).

Recently, growing interest in ctenophore biology (Ryan, et al. 2016) has stemmed from phylogenomic analyses that have highlighted the relationship of ctenophores to other animals (Dunn, et al. 2008; Hejnol, et al. 2009; Philippe, et al. 2009; Pick, et al. 2010; Ryan, et al. 2013; Moroz, et al. 2014; Pisani, et al. 2015; Telford, et al. 2015; Whelan, Kocot and Halanych 2015; Whelan, Kocot, Moroz, et al. 2015; Shen, et al.

2017; Whelan, et al. 2017) and challenged our understanding of early metazoan evolution (Dunn, et al. 2015). While ctenophores possess many unique traits (e.g., ctene rows, colloblasts, statocyst, retention of a juvenile cydippid stage, bi-radial rotational symmetry), they also possess a suite of unique variations on features found in other animals such as muscle cells despite the absence of canonical muscle specification genes and a defined mesoderm (Hernandez-Nicaise and Amsellem 1980; Hernandez-Nicaise, et al. 1984; Mackie, et al. 1988; Ryan, et al. 2013; Moroz, et al. 2014; Vandepas, et al. 2017), a nervous system that lacks many critical neurotransmitters found in other animals (Jager, et al. 2011; Moroz, et al. 2014), a highly derived mitochondrial genome (Pett, et al. 2011; Kohn, et al. 2012), and a functional through-gut despite the lack of many hindgut specification genes (Presnell, et al. 2016). However, the cellular and molecular mechanisms that regulate the development of these variations and the unique traits specific to ctenophores (e.g., colloblasts, statocyst, complex ciliated structures) are mostly unknown (Dunn, et al. 2015). The unique phylogenetic position of ctenophores (Dunn 2017), suggests that investigating molecular mechanisms regulating aspects of their development in comparison to other animals may provide useful insight into how these processes evolved in metazoans.

Objectives

A major challenge in evo-devo has been generating molecular genetic approaches in a wide variety of animals to examine gene expression and function during development (Moczek, et al. 2015). Increasing the number of taxa for comparison is necessary for detailed understanding of how developmental processes have evolved

(Jenner and Wills 2007). As such, generating ways to rear animals in the lab and developing molecular tools to perform genetic manipulation are keys to increasing the number of taxa in important phylogenetic positions that are available for evo-devo comparisons (Moczek, et al. 2015). In *Chapter 2* of my dissertation, I detail techniques and steps for the long-term laboratory culture of *Mnemiopsis leidyi*. Their fragile nature and seasonal availability has hindered studies to gain a better understanding of the molecular, cellular, physiological, and behavioral attributes of these animals (Dunn, et al. 2015). I discuss basic husbandry of *M. leidyi* adults, spawning on controlled light regimes, collecting embryos, and rearing juveniles. I also discuss general techniques to induce spawning and subsequent preparation of embryos for gene expression knockdown analyses using both morpholino oligonucleotides and CRISPR/Cas9 via microinjection. The protocols listed in *Chapter 2* reduce the need for wild-caught animals, and make more detailed studies (e.g., routine mutational analyses) practical in ctenophores. This will allow future evolutionary studies to gain insight from comparisons with functional molecular genetic work in ctenophores.

An advantage of rearing ctenophores in the lab is the ability to observe morphological characteristics that would otherwise be difficult to document in the wild. Comparative morphological studies can provide insight into animal evolution by inferring both the origin of, and changes to, character traits in a phylogenetic context (Dunn, et al. 2014; Giribet 2015). An essential metazoan trait is the through-gut, which allows efficient uptake of food and expulsion of waste material (Dunn, et al. 2014; Hejnol and Martín-Durán 2015). Though ctenophores were known to have anal pores used for waste removal (Agassiz 1850; Chun 1880; Main 1928), recent work during the last 80 years has

suggested that ctenophores have a blind gut with one opening (Hyman 1940; Hejnol and Martín-Durán 2015). In *Chapter 3*, I present work that re-evaluates whether ctenophores possess a single opening for ingestion and egestion or possess an alimentary canal with two separate openings. The results show that the ctenophore anal pores are dedicated for the normal removal of waste material from digested food. I discuss the evolutionary consequences of the presence of the ctenophore through-gut, and whether it is homologous to the bilaterian through-gut.

Understanding gene family evolution can provide insight into how changes in gene regulation can influence morphological diversity (King, et al. 2008; Degnan, et al. 2009; de Mendoza, et al. 2013). The *Kruppel-like factor* (*Klf*) gene family is comprised of zinc finger transcription factors (McConnell and Yang 2010) that are crucial for many molecular and cellular processes during embryonic development (Bialkowska, et al. 2017). Most of the functions have been described in traditional model systems, with little to no evolutionary context. In *Chapter 4*, I present a bioinformatic pipeline that I developed to elucidate the evolutionary history of the *Klf* gene family and their associated transactivation/repression domains across a wide range of eukaryotes. I identified the *Klf* gene complement in a wide range of eukaryotes, and performed phylogenetic analyses to determine their relationships. I showed that *Klf* genes are not only present across Metazoa, but are also present in closely related unicellular organisms within the larger filozoan taxonomic group that includes the metazoans. I also determined the separate evolutionary histories of transactivation/repression domains that are associated with members of the *Klf* gene family. Overall, *Klf* gene family expansion is accompanied by a parallel diversification of transactivation/repression domains via both acquisitions of pre-

existing ancient domains as well as by the appearance of novel domains that appear to be exclusive to the *Klf* gene family (Presnell, et al. 2015).

I identified three *Klf* genes in the genome of *Mnemiopsis leidyi*, *MleKlf5a*, *MleKlf5b*, and *MleKlfX*. Although much is known about the underlying function of KLFs in a handful of species, the role of KLFs during development in other animals, particularly non-bilaterians, is poorly understood. Elucidating the function of *Klf* genes in ctenophores will provide insight into the conserved nature of KLF functions in animals, and provide better understanding of *Klf* gene family influence on early metazoan evolution. In *Chapter 5*, I characterize the expression patterns and functions of the *Klf* genes in the ctenophore *Mnemiopsis leidyi*. Expression patterns for *MleKlf5a* and *MleKlf5b* were mostly associated with embryonic tissues undergoing cellular proliferation. In contrast *MleKlfX* has a highly restricted expression pattern localized to a small number of cells associated with the apical organ. To determine the functional role of *Klf* genes in *Mnemiopsis* I used two independent gene knockdown approaches; morpholino oligonucleotides and CRISPR/Cas9. Both approaches showed that knockdown of zygotic *MleKlf5a* and *MleKlf5b* gene expression results in improper organ development in the apical organ, tentacle bulb, and gastrovascular cavity. I hypothesized that during embryogenesis these genes are required for normal organ development, in part, through the maintenance of cellular proliferation.

Overall

My dissertation seeks to elucidate aspects of *Klf* gene family evolution within metazoans using *M. leidyi* as a model for exploring *Klf* function in an early branching

metazoan lineage. I took a bioinformatic approach to examine the broad-scale mechanisms of *Klf* gene family evolution, and was able to show that many *Klf* and *Sp* genes are conserved throughout most animal phyla, with the exception of ctenophores who lack *Sp* genes and only possess three *Klf* genes. Next, I investigated the function of *Klfs* within the ctenophore *M. leidyi*. My results suggest that ctenophore *MleKlf5a* and *MleKlf5b* genes most likely play a role in regulating stem cell proliferation, as other *Klf5* genes do in other animals, suggesting this was most likely an ancestral function in the Metazoa. Subsequent studies examining downstream targets of *MleKlf5a* and *MleKlf5b* in ctenophores will provide useful insight into the evolution of *Klf* regulation of stem cell maintenance and proliferation in metazoans.

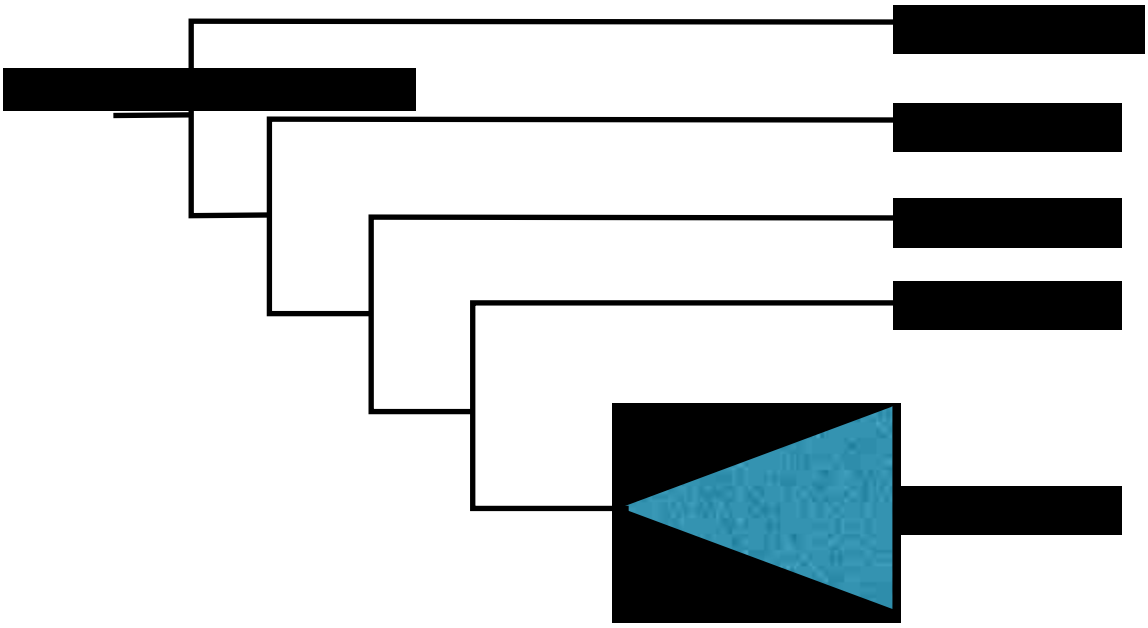


Figure 1.1. Phylogenetic tree showing the relationships of the five extant metazoan clades. The Bilaterian clade consists of animals with bilateral symmetry (e.g., mice, zebrafish, nematode worms, fruit flies). The four other phyla are Cnidaria (corals, anemones, jellyfish), Placozoa, Porifera (sponges), and Ctenophora (comb jellies). These four lineages represent the earliest branching, extant metazoan lineages. The positions of the ctenophore and sponge lineages at the base of the tree remain uncertain, see (King and Rokas 2017). The phylogeny above is based on (Dunn, et al. 2008; Hejnol, et al. 2009; Whelan, et al. 2017).

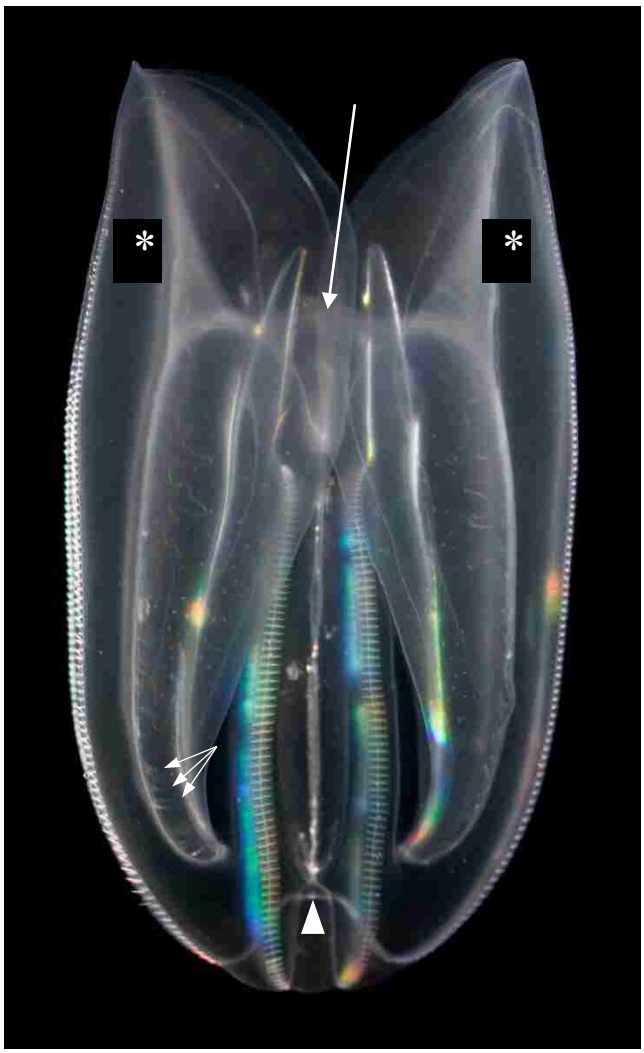


Figure 1.2. A representative image of an adult *Mnemiopsis leidyi*. The oral/aboral axis runs from the mouth (top, arrow) to the apical organ and anal pores (bottom, arrowhead). *M. leidyi* adult tentacles (small arrows) are found underneath the expandable oral lobes (asterisks) and line the feeding grooves. The ctene plates, used for locomotion, are situated in longitudinal rows along the oral/aboral axis.

Chapter 2: Laboratory culture and breeding of the lobate ctenophore *Mnemiopsis leidyi*

Background

Despite long standing interest in ctenophores due to their unique biology, ecological influence and evolutionary status, much of the previous work with ctenophores has relied on their periodic and seasonal availability. Consequently, these studies have largely been restricted in scope due to the absence of reliable methods for long-term laboratory cultures. I have developed a protocol that can be implemented in any lab to establish long-term ctenophore cultures for year-round use as models for biological investigation. The goal of this protocol is to highlight the necessary steps and resources required for culture conditions amenable to a wide range of biological studies including; anatomical, behavioral, embryological, genetics and functional genomics. In this report, I describe continuous laboratory culturing of the Atlantic lobate ctenophore *Mnemiopsis leidyi* (Fig. 2.1). I provide recommendations for general husbandry, feeding requirements, inducing spawning, embryo collection and rearing of juveniles to adults. I also discuss general techniques to prepare embryos for targeted gene knockdown and genome editing via microinjection with morpholino oligonucleotides and CRISPR/Cas9, respectively.

Advantages

Most ctenophores are oceanic and deep-water species that are particularly fragile and often have environmental requirements that are extremely difficult to replicate in the laboratory. In contrast, the coastal species *Mnemiopsis leidyi* and *Pleurobrachia bachei* are seasonally available in large numbers. *Mnemiopsis leidyi* is found in a wide swath of

tropical, subtropical and temperate coastal Atlantic environments (Mayor 1912; Seravin 1994; Bayha, et al. 2015). *M. leidyi* individuals are self-fertile hermaphrodites and well-fed adults can spawn thousands of embryos almost daily (Baker and Reeve 1974; Reeve, et al. 1978) making them amenable to laboratory research. Embryogenesis is rapid and hatching occurs less than 24 hours after fertilization. Juvenile *M. leidyi* cydippid stages can be easily reared to sexual maturity in small aquaria and *M. leidyi* egg to egg generation time can be as short as ~3 weeks (Baker and Reeve 1974; Martindale 1986). Both *M. leidyi* and *P. bachei* juveniles and adults are transparent, making observations and imaging of *in vivo* tissue and organ function straightforward (Presnell, et al. 2016). A range of methods have been developed for *M. leidyi* research including cell lineage fate mapping during embryogenesis (Martindale and Henry 1999), descriptions of gene expression patterns (Yamada and Martindale 2002; Pang and Martindale 2008d), protocols for generating primary cell culture (Vandepas, et al. 2017), use of morpholino oligonucleotides to examine molecular function (Yamada, et al. 2010; Presnell, et al. in prep), and use of CRISPR/Cas9 for genome-editing (Presnell, et al. in prep).

A growing number of genomic resources are available for both *M. leidyi* and *P. bachei* including assembled reference genomes (Ryan, et al. 2013; Moreland, et al. 2014; Moroz, et al. 2014), as well as transcriptomic data (Ryan, et al. 2013; Moreland, et al. 2014; Moroz, et al. 2014; Levin, et al. 2016; Davidson, et al. 2017; Whelan, et al. 2017). This work has provided insight into unique aspects of the ctenophore genome. Both the *M. leidyi* and *P. bachei* genomes are missing many genes typically found in other animals including determinants of neural fate and neurotransmitter biosynthesis (Ryan, et al. 2013; Moroz, et al. 2014), HOX genes (Ryan, et al. 2010; Ryan, et al. 2013; Moroz, et al.

2014), key components of the miRNA biogenesis pathway (Maxwell, et al. 2012) and mesodermal specification genes (Ryan, et al. 2013; Moroz, et al. 2014).

Importantly, detailed studies required to understand how their unique genomes contribute to the fascinating biology of ctenophores continue to be hampered by reliance on seasonal short-term availability of wild-caught animals. Reliable long-term laboratory cultures and year-round embryo availability for *M. leidyi* and *P. bachei* will facilitate comparative genetic analyses of the differences between the lobate ctenophore body plan and the putative paedomorphic cydippid body plan as well as the many strikingly unique aspects of ctenophore biology.

Procedure overview

Here I describe standardized protocols for effective and reliable long-term lab cultures of ctenophores. These protocols explain the methods I have used to successfully maintain long-term adult cultures, raise juveniles, and include feeding regimens, as well as proper water and tank setups. Although I focus on *M. leidyi*, these protocols have been used to successfully maintain several other ctenophore species (*Pleurobrachia sp.*, *Hormifera sp*., and *Bolinopsis sp*.; Personal communication, Wyatt Patry).

Water quality

M. leidyi is a coastal species and can be found in environments with variable salinity, turbidity and temperature (Baker 1973). To standardize the laboratory cultures, artificial seawater (ASW) is used exclusively (Instant Ocean Inc). The exclusive use of ASW offers several advantages: reduced contamination and parasite transmission, precise

control of multiple water quality parameters and importantly the distance of the laboratory from a natural seawater source is irrelevant. Large volumes of ASW are prepared in a mixing cistern to desired salinity, typically 1.017-1.021 specific gravity. Completely mixed ASW is then transferred to a holding cistern, ready to be used in culture tanks.

Housing

Proper housing for pelagic organisms such as ctenophores is crucial (Fig. 2.2). Ctenophores are susceptible to damage by repeated physical contact with tank walls, vigorous currents and aeration. The long-term culture tanks are based on a 'planktonkreisel' design (Greve 1968; Ward 1974; Raskoff, et al. 2003) with angled bottoms, no substrate, and very low flow rates producing a slow rotating current that minimizes contact with tank walls and keeps the ctenophores suspended in the water column (developed in consultation with Midwater Systems Inc.; Fig. 2.2a). Each main culture tank holds a volume of ~350L and is split into two chambers; a narrow rear standpipe area that maintains a constant tank volume irrespective of water flow rate and a front main chamber that houses the ctenophores (Fig. 2.2a). Water exchange between the two sections is through cutouts screened with 200μm nylon mesh. The slow circular current in the main chamber is achieved with a 'spray bar' fashioned from PVC piping with small holes positioned just below the water surface. The 'spray bar' directs water perpendicular to the nylon mesh screens dividing the main front chamber from the rear chamber (Fig. 2.2a). Filtration and waste removal is accomplished via a sump style wet-dry trickle biofilter and a venturi driven protein skimmer (Fig. 2.2b). Efficient protein

skimming is critical for reducing the organic waste load that results from continuous feeding. Debris that settles to the bottom of the culture tanks is removed periodically during partial water changes by vacuum (Fig. 2.2g). While *M. leidyi* tolerate temperatures from 18-32°C (Baker 1973), the cultures are maintained at temperatures of 20°C via an in-line seawater chiller (Fig. 2.2e). Cool water temperatures have the dual benefit of reducing physiological stress and improving long-term water quality. Each main culture tank supports ~10-15 mature adults and each healthy adult can produce hundreds to thousands of embryos on a nearly daily basis. Overcrowding adults reduces both individual feeding rates and gamete production.

Feeding

Ctenophores are voracious predators and require a constant food supply. This is particularly important for reliable production of gametes and regular breeding. There are currently no artificial diets that are suitable for long-term ctenophore maintenance. Live marine rotifers, *Brachionus plicatilis* (Reed Mariculture Inc.), are the staple food for long-term maintenance of both juvenile and adult ctenophores (Fig. 2.2c). Rotifer cultures are fed a concentrated algal paste diet (either RGcomplete or Rotifer Diet, Reed Mariculture Inc.). Rotifers are delivered directly to the ctenophore cultures in regularly spaced ~20min intervals via electronic dosing pumps (Fig. 2.2f). Extended periods of robust gamete production are achieved by supplementing rotifers with once or twice daily feedings of fish larvae (Fig. 2.2d). I have had success using freshwater zebrafish larvae (*Danio rerio*) that can be easily sourced from stock centers (e.g. UM Zebrafish Facility) or commercial aquarium stores. Freshwater fish larvae should be briefly washed in ASW

and target fed to ctenophores with a pipette. Marine larval fishes can be added directly to ctenophore culture tanks without the need for targeted feeding. Species successfully fed to ctenophores include *Coryphaena hippurus* (mahi-mahi) and *Rachycentron canadum* (Cobia). Large juvenile and adult ctenophores also receive periodic feedings of *Artemia* nauplii.

It is critical to avoid overfeeding (Presnell, et al. 2016). Ctenophores are prone to regurgitate the entire contents of their pharynx if overwhelmed with food. The transparency of ctenophores allows for rapid assessment of feeding rates. Pharyngeal occupancy of ~15-40% food is typically a sign that the feeding rate is adequate.

Parasites

A common and detrimental parasite of *M. leidyi* is the sea anemone *Edwardsiella lineata* (Reitzel, et al. 2007; Reitzel, et al. 2009). These parasitic anemones are typically observed embedded within a host ctenophore's mesoglea adjacent to the pharynx and endodermal canals. Thorough screening of wild caught ctenophores for *Edwardsiella* infection is necessary before founding long-term cultures, as heavy infestations of these parasites generally kill the host ctenophore. A culture system inadvertently inoculated with *Edwardsiella* requires disassembly and bleaching to ensure that all free-living stages of the parasite are destroyed. Some protistan epibionts (ciliates and amoebas) are known to be associated with *M. leidyi* (Moss, et al. 2001). The ciliate *Trichondina ctenophorii* can often be observed attached to the ctenes of wild caught *M. leidyi*. These ciliates and other epibionts typically detach with a few washes of clean artificial seawater and do not persist in culture. However, as with any closed loop aquaria system, a buildup of harmful

ciliates and bacteria can occur over time, thus frequent water changes and proper filtration are required to keep harmful populations in check.

Spawning, embryogenesis, juvenile staging

Spawning

Like many ctenophore species, *M. leidyi* is a self-fertile hermaphrodite (Harbison and Miller 1986), thus only a single individual is required to obtain viable embryos. Both female and male gonads are located along each ctene row (Fig. 2.3a), flanking the underlying endodermal canal (Freeman and Reynolds 1973). In gravid animals, sperm and eggs can be seen accumulating between the ctene plates at the gonopores flanking the underlying endodermal canal about an hour prior to release. During spawning, sperm compacts between the ctene rows on the male side, becomes iridescent and can be seen streaming out of the male gonopores. Mature sperm is normally released ~10 minutes prior to egg release. Mature eggs are squeezed through the female gonopores on the opposite side of the underlying endodermal canal and are fertilized almost immediately by motile sperm in the surrounding seawater. An individual spawning event can last ~1hr. Spawning can be easily entrained to artificial light cues and thus the timing of spawning can be precisely controlled. Our lab-reared southern Florida *M. leidyi* spawn 4 hours after lights off. In contrast, northern populations of *M. leidyi* near Woods Hole, MA typically spawn 8 hours after lights off (Pang and Martindale 2008e).

Embryogenesis

First cleavage occurs ~1-hour post fertilization (hpf; Fig. 2.3c, c'), and embryos undergo gastrulation around 3-5 hpf. By 9 hpf (Fig. 2.3e, f) the four pairs of ctene rows have formed on the ectodermal surface of the embryo. Additionally, this stage of development is marked by organization of different tissues in the embryo. At the oral end, ectodermal cells around the blastopore invaginate to form the pharynx. On both lateral sides of the embryo, cellular thickenings will give rise to the tentacle bulbs. At ~14 hpf (Fig. 2.3g, h) the mesoglea, made up of ECM extruded from the endoderm, hydrates as the entire embryo swells and increases in size. During this time, the pharynx elongates aborally and contacts the epithelium of the gastrovascular cavity. Ectodermal cells of the two lateral cellular thickenings invaginate and, along with the underlying endodermal cells, form the tentacle bulb. At the aboral end of the embryo, the apical organ floor thickens. Mineralized cells called lithocytes form within the epithelial floor and will eventually merge together to form the statocyst. Supporting the statocyst are four clusters of modified cilia, balancer cilia, that send signals to the ctene rows to regulate beating based on gravity-induced deflection of the statocyst (Tamm 2014b; Tamm 2015). At this stage of development, the infundibulum of the gut also becomes visible as a thin layer of endoderm enclosing a gastrovascular space that connects the pharynx, tentacles, and apical organ (Martindale and Henry 1999). Much later in development the infundibulum is remodeled to give rise to the branched endodermal canal system (Presnell, et al. 2016). Swimming juvenile cydippids hatch approximately ~18-24 hpf (Fig. 2.3i, j) and start feeding with extended tentacles about an hour after hatching.

Juvenile

After hatching, juvenile cydippid ctenophores are transferred to ~9.5L ASW rearing tanks saturated with rotifers, in which they are raised to sexual maturity, ~16 days post hatching (dph), before transferred to long-term culture tanks. Juvenile ctenophore tentacles are easily damaged and thus juveniles must be handled gently. Large transfer pipettes, with the tips cut back to enlarge the openings, are used for moving juvenile ctenophores between juvenile rearing tanks. These tanks do not require aeration or filtration. Juveniles are resistant to rapid changes in water quality, thus the still water environment of rearing tanks are generally well tolerated for several days. When the surfaces of a rearing tank become fouled, the juvenile cydippid ctenophores are moved to a clean tank with fresh ASW saturated with rotifers. A single rearing tank can initially support hundreds of juveniles if enough rotifers are provided. However, to maintain maximal growth rates for juvenile ctenophores, typically no more than ~20-30 juveniles are raised in a rearing tank. Under conditions in which food is not limiting, juvenile cydippid stage *M. leidyi* will approximately double in size each day for the first eleven days (Fig. 2.4). Post-hatching descriptions under ideal feeding conditions are briefly described.

1 to 4 days post hatching (dph)

After hatching, *Mnemiopsis* juveniles are ~0.2-0.3 mm in size. In culture, juveniles swim through the water column with extended tentacles and mouth facing forward. When actively feeding, juveniles will swim a looping pattern deploying both tentacles and tentilla that hang in a dense ‘net’ similar to adult *Pleurobrachia* (Tamm and

Moss 1985). In 1 dph juveniles (Fig. 2.4a), the infundibulum remains unbranched. At 2 dph, juveniles are morphologically similar but twice as large (Fig. 2.4b). The aboral canal visibly emerges from the infundibulum, and is typically bifurcated into two short anal canals which lead to anal pores. At 3 dph, endodermal protrusions can be seen on the oral side of the infundibulum flanking the pharynx in the tentacular plane (Fig. 2.4c). These protrusions corresponded to the developing paragastric canals. At 4 dph, the paragastric canals have extended the length of the pharynx and the gastrovascular protrusions that will become the radial and adradial canals appear. At 4 dph juvenile cydippids are approximately 1-2 mm in size (Fig. 2.4d).

5 to 8 dph

By 5 dph (Fig. 2.4e), the distance between the tentacle apparatus and the apical organ has increased and more plates have been added to the ctene rows as they extend with the meridional canals along the oral-aboral axis (Tamm 2012). By 6 dph, the stomodeum begins to enlarge (Fig. 2.4f) and by 7 dph (Fig. 2.4g) the oral lobes begin to form. Concurrently, the infundibulum and associated radial and adradial endodermal canals form and the anal canals become distinct from the aboral canal. By 8 dph, the meridional canals have extended orally beyond the ctene rows and converge around the stomodaeum, just prior to the outgrowth of the oral lobes (Fig. 2.4h). During this stage, at least some juveniles are capable of dissogony (Martindale 1987) and produce a small number of viable gametes. Progeny of these dissogony events are apparent as an additional class of ‘new’ smaller juveniles in rearing tanks. At 8 dph juvenile cydippids are approximately 5-6mm in size.

9 to 12 dph

By 9 dph transformation to the adult morphology is visible as outgrowths of the oral lobes continue (Fig. 2.4i, i'). Oral lobe outgrowth is pioneered by the extension and growth of the meridional canals that have now merged in the developing oral lobes. Oral lobe outgrowth and associated underlying endodermal canal system patterning continues through 12dph (Fig. 2.4i-l, i'-l'). The auricles first appear as protuberances flanking the developing oral lobes by 11 dph (Fig. 2.4.k, k') and begin to extend by 12 dph. Extended tentacles remain visible and are still used for prey capture. At 12 dph, juveniles are transitioning to adult morphology and are approximately 8-10mm in size (Fig. 2.4l, l').

13 to 16 dph

By 13 dph, the juvenile cydippid body plan has been extensively remodeled and closely resembles the adult lobate form (Fig. 2.4m, m'). Between 13 and 16 dph several morphogenetic patterning events associated with the final adult body plan occur (Fig. 2.4m-o, m'-o'). The characteristic, tentilla-lined, feeding grooves situated between the auricles and lobes begin to form by 14 dph (Fig. 2.4n, n'). The free hanging tentacles of the cydippid body plan begin to retract as the feeding grooves become lined with tentilla. The feeding groove tentilla are derived from the branching of the main tentacle as it emerges from the tentacle sheath and extends along the medial cleft of the feeding groove. As early as 16 dph (Fig. 2.4o,o'), mature gametes can be observed underneath the ctene rows as the animals attain sexual maturity, approximately 15-20mm in size.

Microinjection

Spawning of sexually mature *M. leidyi* can be entrained to a new user defined photoperiod within 24 hours. Thus, spawning can be induced whenever it is convenient for the researcher. Briefly, to boost gamete production and obtain single-cell embryos for microinjection, adult animals are fed fish larvae 24 hours before spawning and immediately before being placed in the dark 4 hours prior to injection. Individual gravid adults are kept in large (8 inch-diameter) glass bowl incubation dishes (Carolina Biological Supply Inc). After ~3.5 hours of darkness, animals can be re-exposed to light and screened under a dissection microscope for the presence of mature gametes. *M. leidyi* adults are broadcast spawners that produce large amounts of active sperm. When mature oocytes are presented too many sperm, high rates of polyspermy and early embryonic failure occur. To reduce polyspermy in spawning bowls, gravid *M. leidyi* are washed 2-3x during sperm release with clean ASW to lower sperm concentration during spawning. Approximately 10 minutes after sperm release, eggs are extruded through female gonopores located on the opposite side of the underlying endodermal canal. Extruded eggs are maximally competent for fertilization for a relatively short period of time. Eggs are kept in the spawning bowl for at least 10 minutes post extrusion to ensure high rates of fertilization.

To remove excess mucus, egg jelly, bacteria and soften the egg envelope, fertilized embryos are collected via pipette and successively passed through 0.5µm and 0.4µm nylon mesh filters and then rinsed three times in filtered sterile seawater (FSW) with 1% penicillin plus streptomycin (Pen/Strep) solution. Sharpened tungsten needles are then used to remove the egg envelope. Embryos must be removed from their egg shell

prior to microinjection (Martindale and Henry 1999). Dechorionated embryos are kept in gelatin-coated dishes to prevent adhesion to charged surfaces such as glass and plastic. First cleavage typically begins ~60 minutes after fertilization.

An injection dish containing a shallow trough is used to hold the embryos in place for microinjections. The injection trough is made by placing a glass slide at an angle of about 45° into a 30mm or 60mm petri dish and pouring molten 2% agar (dissolved in 50% FSW/50% H20) into the dish until the agar touches the top edge of the slide. Once solidified, the injection mold is rinsed with FSW. The injection mold can be re-used over multiple sessions when spiked with 1x Pen/Strep and stored at 4°C. Prior to injection, a few drops of FSW are added to the trough. Dechorionated embryos are then placed into the mold and gently lined up in a row along the back of the depression trough via mouth pipette. Due to the relatively quick early cleavage events and necessary embryo preprocessing, efficient ctenophore embryo microinjections require two people to work in tandem; one to collect, clean and dechorionate embryos, and one to prepare injection reagents and perform microinjections. These two activities are simultaneous. A pair of researchers can easily inject ~30-60 embryos in a single session. The use of reliable and uniformly sharp injection needles is key to rapid microinjection of embryos. While many options exist for making microinjection needles, we typically use aluminosilicate needles (Sutter Instrument Company) with beveled tips (Sutter Instrument Company BV-10). Injection needles are typically backfilled with an injection cocktail spiked with a trace amount of fluorescent dextran for rapid visualization of injection success. Injected embryos are then gently removed via mouth pipette from the injection dish and placed in

gelatin coated dishes with FSW. Injected embryos are kept at room temperature until the desired embryonic stage is reached.

The techniques outlined in this report can reduce and, in many cases, eliminate the need for wild-caught animals, allowing laboratories far from the ocean to perform detailed biological analyses in these remarkable animals. These protocols permit more nuanced observations of ctenophore biology than previously possible and make routine mutational analyses practical in ctenophores.

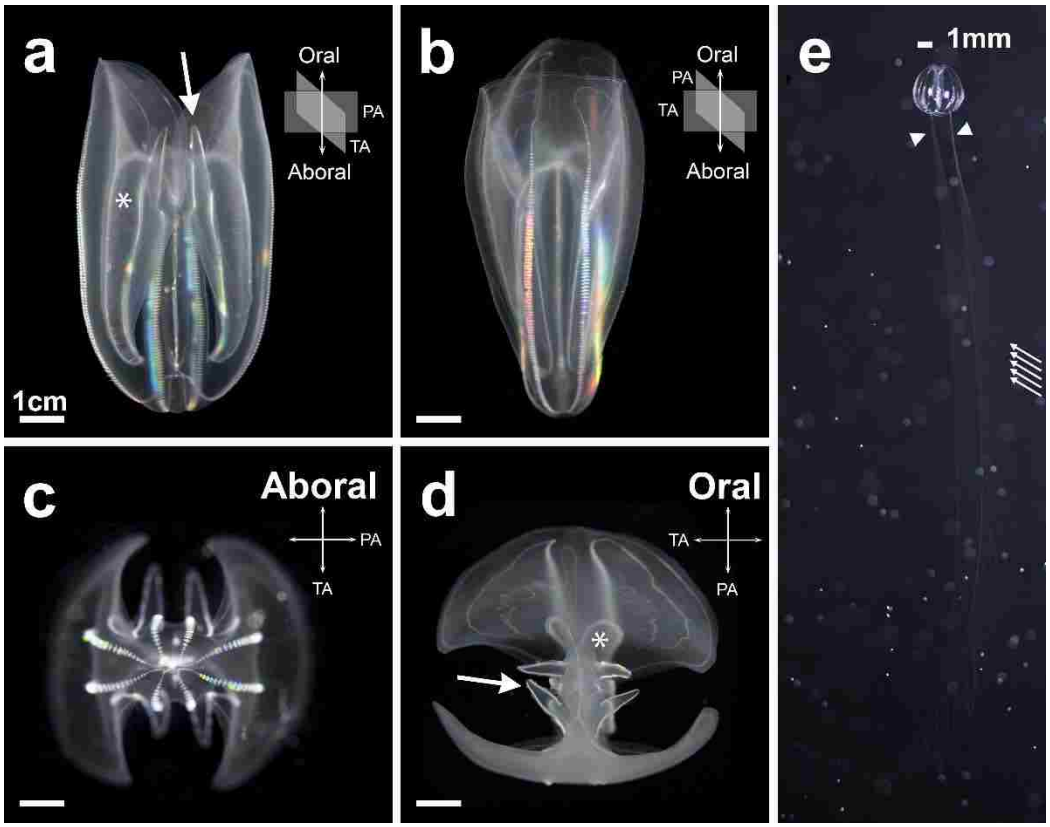


Figure 2.1. Body plan of *Mnemiopsis leidyi* with major axes labeled. a) Lateral view of an adult in the pharyngeal plane. b) Same individual as in (a), but viewed from the tentacular plane. c) Adult viewed from the aboral pole. The ctene rows can be seen oriented around the oral/aboral axis. d) Oral view, showing the mouth. e) 6 dph juvenile with extended tentacles (arrowhead) and branching tentilla (small arrows). The tentilla are covered with colloblasts, which are used for prey capture. At this stage, we feed juveniles marine rotifers (small dots in the background). The auricles (Arrows in a, d), are ciliated projections that help funnel prey into the mouth. Asterisk in a and d indicates the food grooves in the adult. The food groove is lined with the sticky tentilla, and extends underneath the lobes. TA = tentacular axis; PA = pharyngeal axis.

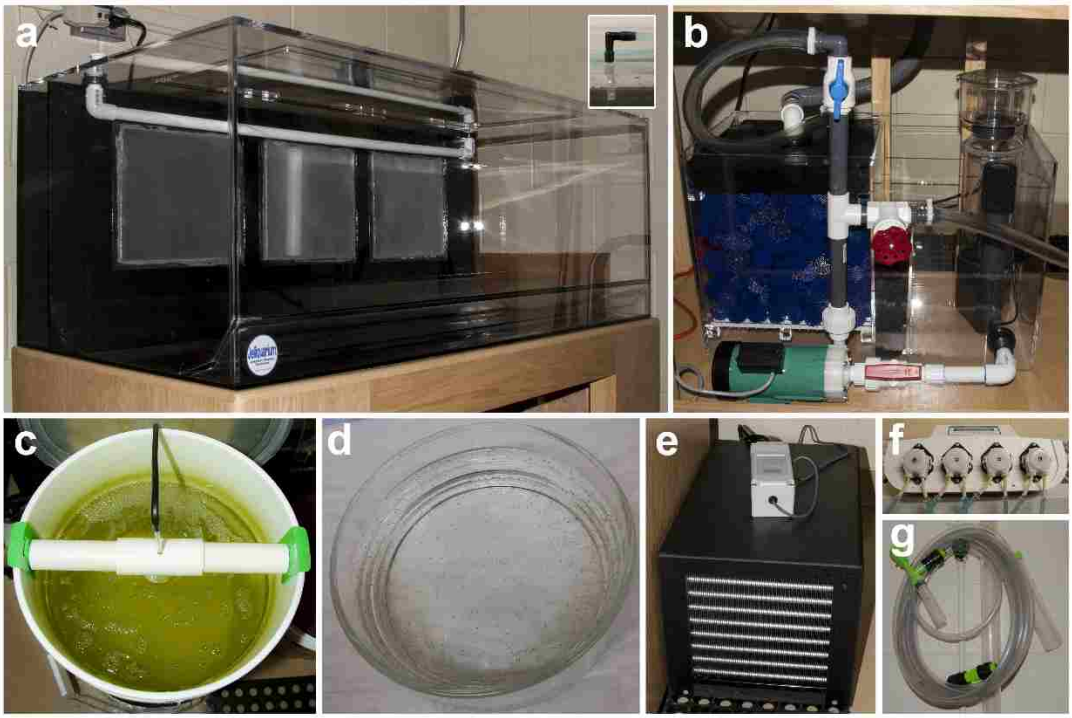


Figure 2.2. Equipment used for housing, feeding, and tank maintenance. a) Custom ~300 L pseudokreisel tank used to house adult individuals. b) The biofiltration unit of the sump is on the left. On the right is the venturi driven protein skimmer. At the bottom is the pump that is used to control the flow rate in the main tank. c) 15 L bucket filled with rotifers. d) Bowl of zebrafish larvae. e) Water chiller. f) Dosing pump used to deliver rotifers from bucket (c) through holes in the top of the main tank (a, inset). g) Vacuum used to siphon organic waste from the bottom of the tank and perform water changes.

Figure 2.3 Mature gametes and stages of embryonic development in *Mnemiopsis leidyi.* a) Mature gametes in a gravid adult are positioned underneath the comb rows along the endodermal canals. Mature sperm (left), and mature eggs (right, arrow) can typically be seen at least an hour or two before spawning. b-d, b'-d') Early development of *M. leidyi*. Differential interference contrast (DIC) images in (b-d), fluorescent images of vital dyes in (b'-d'). b) The zygote is characterized by a thin ring of cytoplasm (magenta) around a yolky center (yellow). The next two cleavages are equal and unipolar. c) 2-cell stage embryo, and d) 4 cell stage embryo. e-j, e'-j') Mid- to late- embryogenesis. DIC images of live animals in (e-j), fluorescent images of vital dyes in (e'-j'). Vital dyes can be used to label different tissues in live animals. Lysotracker (yellow) preferentially labels vacuolated cells in the endoderm, while Mitotracker (magenta) brightly stains the ctene producing cells. Tissue organization begins around 9 hpf (e, f), and organs start to develop by 13 hpf (g, h). Overall, development is rapid with hatching occurring about 18-20 hpf (i, j) in our lab reared *M. leidyi*. ao = apical organ, ct = ctene rows, gv = gastrovascular cavity, tb = tentacle bulb, ph = pharynx. Lateral view in (e, g, i, e', g', i'), with aboral end facing down. Aboral view in (f, h, j, f', h', j').

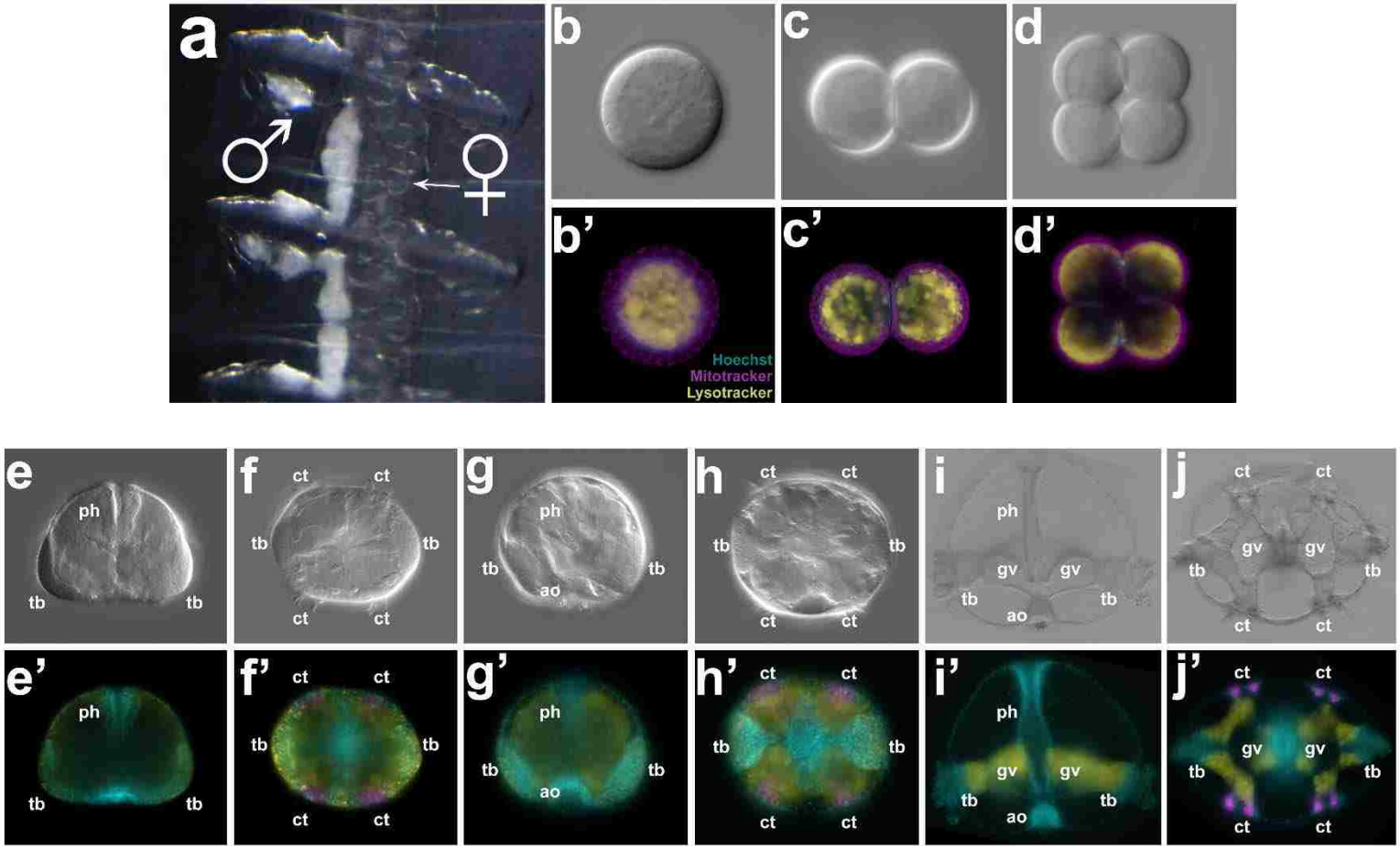
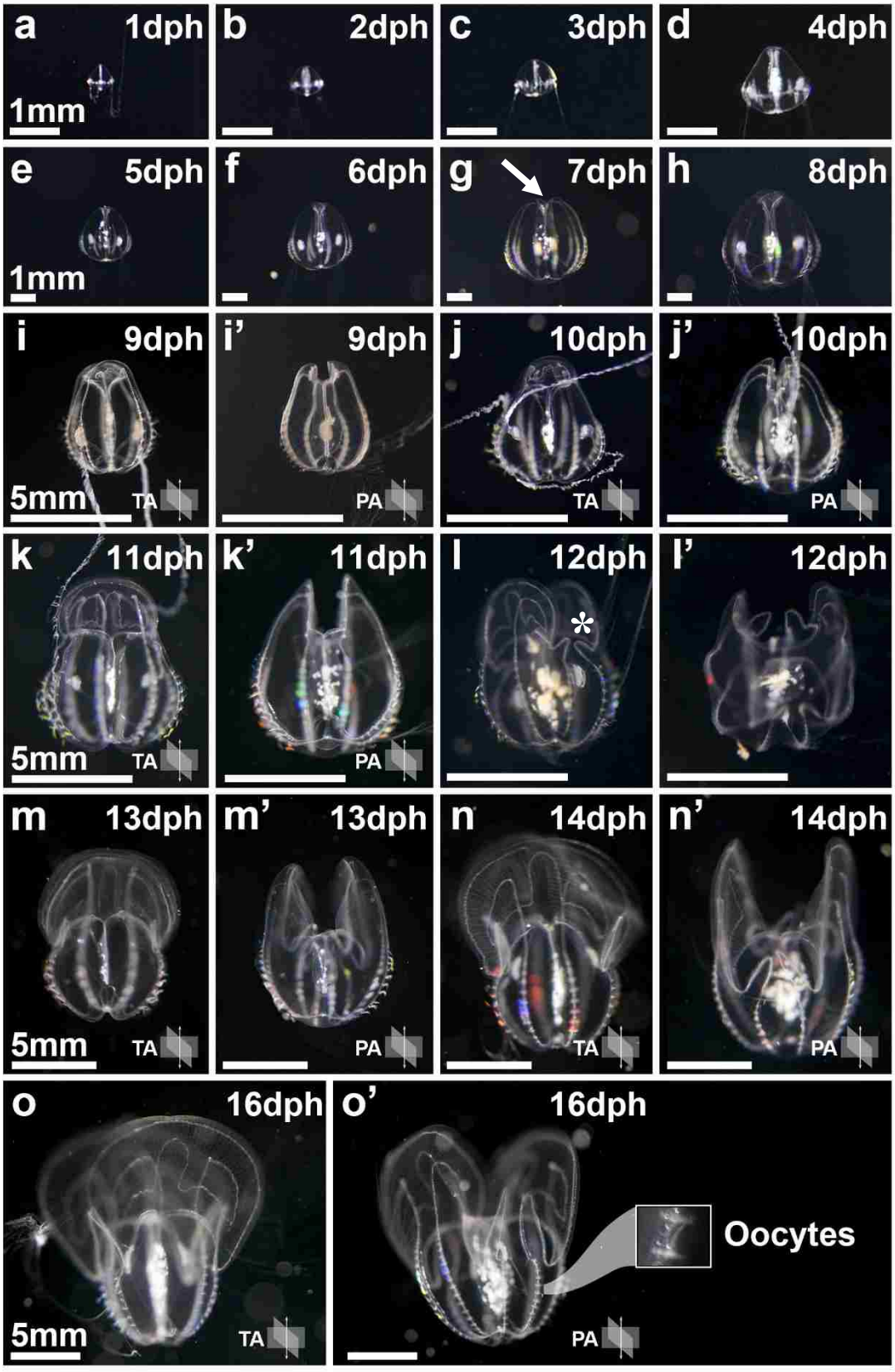

a
b
c
d
b'
c'
d'
Hoechst
Mitotracker
Lysotracker
e
f
g
h
i
j
e'
f'
g'
h'
i'
j'
ph
tb
ct
ao
gv

Figure 2.4. Stages of juvenile development in *Mnemiopsis leidyi*. a-d) 1 day post hatching (dph) through 4 dph. e-h) 5 dph through 8 dph. g) Oral lobes begin to take shape by 7 dph, indicated by the arrow. If well-fed, juveniles can double in size for the first few days after hatching (a-d). Asterisk in (l) highlights the formation of the auricles. Juveniles can reach sexual maturity as early as 16 dph (o'). In all panels, the oral end of the juvenile is facing up. Scale bars are 1mm in A-H; 5mm in i-o, and i'-o'.


a 1dph
1mm
b 2dph
c 3dph
d 4dph
e 5dph
1mm
f 6dph
g 7dph
h 8dph
i 9dph
5mm
TA
i' 9dph
PA
j 10dph
TA
j' 10dph
PA
k 11dph
5mm
TA
k' 11dph
PA
l 12dph
*
l' 12dph
m 13dph
5mm
TA
m' 13dph
PA
n 14dph
TA
n' 14dph
PA
o 16dph
5mm
TA
o' 16dph
Oocytes
PA

Chapter 3: Presence of a functionally tripartite through-gut in Ctenophora has implications for metazoan character trait evolution[1]

Background

The current paradigm of gut evolution assumes that non-bilaterian metazoan lineages either lack a gut (Porifera and Placozoa) or have a sac-like gut (Ctenophora and Cnidaria) and that a through-gut originated within Bilateria (Sedgwick 1884; Von Graff 1891; Arendt and Nübler-Jung 1997; Hejnol and Martindale 2008; Martindale and Hejnol 2009; Dunn, et al. 2014; Hejnol and Martín-Durán 2015; Telford, et al. 2015). An important group for understanding early metazoan evolution is Ctenophora (comb jellies), which diverged very early from the animal stem lineage (Dunn, et al. 2008; Hejnol, et al. 2009; Ryan, et al. 2013; Moroz, et al. 2014; Whelan, Kocot, Moroz, et al. 2015). The perception that ctenophores possess a sac-like blind gut with only one major opening remains a commonly held misconception (Buchsbaum, et al. 1987; Pang and Martindale 2008e; Martindale and Hejnol 2009; Hejnol and Martín-Durán 2015; Telford, et al. 2015). Despite descriptions of the ctenophore digestive system dating to Agassiz (Agassiz 1850) that identify two openings of the digestive system opposite of the mouth—called ‘‘excretory pores’’ by Chun (Chun 1880), referred to as an ‘‘anus’’ by Main (Main 1928), and coined ‘‘anal pores’’ by Hyman (Hyman 1940)—contradictory reports, particularly prominent in recent literature, posit that waste products are primarily expelled via the mouth (Hyman 1940; Gardiner 1972; Buchsbaum, et al. 1987; Pearse, et

[1] The results presented here were published as part of the following paper: Presnell JS, Vandepas LE, Warren KJ, Swalla BJ, Amemiya CT, Browne WE (2016) **The Presence of a Functionally Tripartite Through-Gut in Ctenophora Has Implications for Metazoan Character Trait Evolution.** *Current Biology* 26:2814-2820 doi:10.1016/j.cub.2016.08.019.

al. 1987; Bumann and Puls 1997; Martindale and Henry 1998; Martindale and Hejnol 2009; Hejnol and Martín-Durán 2015; Telford, et al. 2015). Here we demonstrate that ctenophores possess a unidirectional, functionally tripartite through-gut and provide an updated interpretation for the evolution of the metazoan through-gut. Our results resolve lingering questions regarding the functional anatomy of the ctenophore gut and long-standing misconceptions about waste removal in ctenophores. Moreover, our results present an intriguing evolutionary quandary that stands in stark contrast to the current paradigm of gut evolution: either (1) the through-gut has its origins very early in the metazoan stem lineage or (2) the ctenophore lineage has converged on an arrangement of organs functionally similar to the bilaterian through-gut.

Methods

Feeding Experiments

Lab-reared *Mnemiopsis leidyi* were kept at room temperature and starved for ~24 hours prior to feeding experiments. For lateral views, adult animals were pinned in 8-inch silicon coated dishes (SYLGARD–184, Dow Corning, Inc.) and fed zebrafish larvae. Time-lapse video was captured with a RED EPIC camera (RED, Inc.). For aboral views, 4-day post hatching juveniles were fed Brainbow fluorescent zebrafish larvae (Pan, et al. 2011), and filmed on a Zeiss SteREO Discovery.V8 with a Zeiss AxioCam MRm rev3 using Zeiss AxioVision software (Release 4.8.2). Annotations, images and video were processed with REDCINEX-PRO (RED, Inc.), Adobe Photoshop, and QuickTime Pro (Apple, Inc).

Immunohistochemical Staining

Whole *Mnemiopsis* juveniles and excised lobe tissue were relaxed in a 6.5% MgCl2 solution mixed with FSW in a 1:1 volumetric ratio and fixed in 4% paraformaldehyde at room temperature. Fixed samples were washed with PTw (0.1% Tween-20 in PBS) and incubated with phalloidin (Alexa Fluor 568 phalloidin, Molecular Probes) in PBS (1:100) for 1 hour at room temperature, washed with PBS, and incubated in SlowFade Diamond Antifade Mountant with DAPI (Molecular Probes) at room temperature. Fixed excised endodermal canals were permeabilized with 0.2% Triton X-100 in PBS, incubated in blocking solution (10% Goat serum, 0.2% Triton X-100, in PBS), and incubated with E7 antibody (Developmental Studies Hybridoma Bank) diluted 1:20 in blocking solution overnight at 4°C. Samples were incubated overnight at 4°C with secondary antibody at 1:250 in blocking solution and phalloidin-488 (Cytoskeleton, Inc.) at 100 nM in PBS. Either Alexa Fluor Goat anti-Mouse 594 or Alexa Fluor Chicken anti-Mouse 647 was used (Jackson ImmunoResearch Laboratories, Inc.). Samples were mounted in SlowFade Diamond Antifade Mountant with DAPI (Molecular Probes). Images were acquired using a Zeiss Axio Imager.Z2 with a Zeiss AxioCam MRm rev3 and Zeiss AxioVision software (Release 4.8.2), or with a Leica SP5 laser scanning microscope. Images were processed with either Adobe Photoshop or FIJI (Schindelin, et al. 2012).

Vital Dye Staining

Mnemiopsis juveniles 2 days post-hatching were incubated in FSW containing 5µg/mL Hoechst (Molecular Probes) and 100nM Lysotracker® Red DND-99 (Molecular

Probes). Tissue envelopes and cell cultures were incubated with 5µg/mL Hoechst and 66nM Lysotracker® Red DND-99 (Molecular Probes). Images were acquired using a Zeiss Axio Imager.Z2 with a Zeiss AxioCam MRm rev3 and Zeiss AxioVision software (Release 4.8.2) and edited using Adobe Photoshop or FIJI (Schindelin, et al. 2012).

Results and Discussion

The evolutionary origins of the animal through-gut are crucial for understanding the coordinated patterning of organ systems. The prevailing paradigm of gut evolution asserts that the metazoan through-gut (alimentary canal) and anus originated within Bilateria (Schmidt-Rhaesa 2007; Hejnol and Martindale 2008, 2009; Martindale and Hejnol 2009; Dunn, et al. 2014; Hejnol and Martín-Durán 2015; Telford, et al. 2015). The implication of this view of gut evolution, that non-bilaterians either lack a gut altogether or possess a blind, sac-like gut characterized by a single functional opening at the oral pole, the stomodeum, is misinformed. The idea that ctenophores possess a blind gut (Gardiner 1972; Buchsbaum, et al. 1987; Pearse, et al. 1987; Martindale and Henry 1998) appears to stem from the historical misconception that ctenophores and other non-bilaterians are simple animals lacking complex traits (Dunn, et al. 2015). Though recent evidence supports Ctenophora as the earliest-branching extant metazoan phylum (Dunn, et al. 2008; Hejnol, et al. 2009; Ryan, et al. 2013; Moroz, et al. 2014; Whelan, Kocot and Halanych 2015; Whelan, Kocot, Moroz, et al. 2015), ctenophores possess many ‘‘complex’’ traits found in other animals, such as definitive muscles and a nervous system (Jager, et al. 2011; Dayraud, et al. 2012; Ryan, et al. 2013; Moroz, et al. 2014). Thus, the revised phylogenetic position of Ctenophora has also lead to novel hypotheses

reconsidering the origin of both muscles and nerves (Ryan, et al. 2013; Moroz, et al. 2014; Ryan 2014). Here we provide new data for and discuss previous evidence of the presence of a through-gut in ctenophores and the broader implications for evolution of the metazoan through-gut and anus.

As early as the work of Louis Agassiz (Agassiz 1850), it was known that ctenophores possess digestive system openings opposite of the mouth. Despite this, contradictory reports suggest that the mouth alone is used to eliminate food waste, whereas other studies have proposed that the pores are used only occasionally to expel small waste products such as nitrogenous or intracellularly digested wastes (Hyman 1940; Pearse, et al. 1987; Bumann and Puls 1997; Martindale and Henry 1998; Hejnol and Martindale 2009; Martindale and Hejnol 2009). Collectively, these reports have misconstrued the function of the ctenophore anal pores, resulting in support for the misconception that ctenophores have a simple, blind gut (Hyman 1940; Pearse, et al. 1987; Bumann and Puls 1997; Hejnol and Martindale 2009; Hejnol and Martín-Durán 2015; Telford, et al. 2015).

The incongruence among these previous observations motivated our experiments, which were explicitly designed to explore ctenophore gut functional anatomy. Here we provide an update to previous literature on the ctenophore digestive system using high-resolution in vivo time-lapse videography combined with fluorescent labeling techniques and primary cell culture. The main longitudinal body axis in ctenophores is the oral-aboral axis, with the mouth at the oral pole and two anal pores at the aboral pole (Dunn, et al. 2015) (Figure 3.1). The unidirectional ctenophore gut is highly differentiated and organized into three functional domains (Figure 3.1A). Our results resolve conflicting

reports on the function of the ctenophore anal pores, confirming their role in relation to the overall digestive system as secondary gut openings function- ally equivalent to an anus. We unequivocally show that ctenophores possess a functional through-gut from which digestion waste products and material distributed via the endodermal canals are expelled to the exterior environment through terminal anal pores that are anatomically and physiologically specialized to control outflow from the branched endodermal canal system (Figure 3.2).

Ctenophores are highly transparent, allowing for easy in vivo observation of internal processes (Agassiz 1850; Chun 1880; Main 1928; Bumann and Puls 1997; Tamm 2014a). To visualize food passage through the ctenophore gut, we monitored the digestion of zebrafish expressing fluorescent proteins (Figure 3.2). Digestion is spatially and temporally regulated by coordinated activities throughout the ctenophore gut that include characteristic cells functioning in nutrient uptake and cells with functionally specialized cilia that control particulate and fluid movement, as well as access to the mesoglea (Figure S1) (Franc 1972; Hernandez-Nicaise 1991). Food entering the stomodeum is transported aborally in the pharynx by combined ciliary beating and muscular contractions toward the pharyngeal folds in the aboral medial third of the pharynx (Figure 3.1A). Within the pharynx, combined mechanical and enzymatic action results in the pre-digestion of food into large particles (Hernandez-Nicaise 1991; Bumann and Puls 1997) that then enter the morphologically distinct ciliary mill (Figure 3.1A). The ciliary mill is uniquely characterized by dense arrays of stiffened cilia (Tamm 2014a) that act as a physical sieve and disrupt large particles, allowing only small particles to pass through into the infundibulum (transverse canal), the first chamber of the endodermal

canal system. The filtrated material is then supplied to four radial canals, one for each radially symmetric quadrant of the ctenophore body plan (Figure 3.1). Each radial canal bifurcates into paired adradial canals that closely associate with several prominent organ systems, including ctene rows, photocytes, and gonads. In the tentacular plane, two paragastric canals extend orally and medially on either side of the pharynx, and two tentacular canals extend laterally through the mesoglea supplying the tentacle apparatus. Cilia lining the endodermal canal lumen sweep food particles through the branched canal system toward the periphery during nutrient distribution (Tamm 2014a). Collectively, the endodermal canals function in both nutrient absorption and distribution (Figure 3.5) (Chun 1880; Main 1928; Franc 1972; Hernandez-Nicaise 1991; Bumann and Puls 1997).

Both nutrient distribution and waste elimination are aided by peristaltic activity within the pharynx and the endodermal canals that direct flow of material. During waste removal, ciliary beating reverses direction, collecting waste material at the centrally located infundibulum, prior to moving into the aboral canals. The function of the aboral canals contrasts with the remainder of the endodermal canal system during both nutrient distribution and waste collection due to differentially regulated muscle and ciliary activity. The isodynamic ciliary beat cycle associated with nutrient distribution and waste elimination is a regulated process that occurs with a temporally cyclic regularity of approximately 2–2.5 hr under our continuous feeding regime. During nutrient distribution, ciliary beating within the aboral canal actively excludes particle entry. Additionally, muscles surrounding the aboral canal are relaxed, reducing the canal diameter and further limiting particle ingress. During waste elimination, these muscles contract, pulling on the canal walls and significantly enlarging canal diameter. The

muscles controlling the relative diameter of the aboral and anal canals are stereotypically localized (Figure 3.3A). In *Mnemiopsis*, muscle bundles extend in the pharyngeal plane along the oral-aboral axis on both sides of the pharynx. At the aboral end, they anastomose with aboral and anal canals; toward the oral end, muscle fibers separate from the bundle and anchor along the feeding grooves. As these muscles contract, the aboral and anal canal diameters widen, and the cilia lining the aboral canal simultaneously reverse beat direction, transporting material collected in the infundibulum into the lumen of the aboral canal immediately preceding defecation. These activities were most likely not detected in previous studies because the transparent muscles are difficult to visualize, their anastomosing ends are often lost during fixation, and the cyclical biphasic changes associated with nutrient distribution and waste elimination make detection difficult without high resolution time-lapse video.

Our results reveal that each anal canal culminates with a sphincter, consistent with a highly regulated system. The anal pores are composed of actin-rich cells in a ring configuration, consistent with a role in contraction (Figure 3.3). This actin distribution, coupled with the previous report of myosin heavy chain expression in *Pleurobrachia pilius* anal pores (Dayraud, et al. 2012), provides molecular evidence that the anal pores function as contractile, ring-like sphincters. During defecation, the sphincter of a single anal canal opens and waste is forcefully expelled (Figure 3.2). After defecation, the anal pore closes and the anal and aboral canals constrict. This temporally regulated process results in the expulsion of the entirety of material circulating in the endodermal canals (Figure 3.2). Therefore, in ctenophores, the anal pores function as the main excretory organs for defecation and serve as the terminal end of a functional through-gut.

Our functional anatomy study of the path of digestion in ctenophores confirms some of the historical observations made over the past 150 years that the anal pores open to the external environment (Agassiz 1850; Chun 1880; Main 1928; Hyman 1940; Bumann and Puls 1997). However, many previous works have differed significantly in their interpretations of ctenophore digestive system function broadly, and of the role of the anal pores in particular. While some early work referred to the anal pores as an anus that allowed for ‘‘defecation of fecal material’’ (Agassiz 1850; Main 1928), more recently the ctenophore digestive tract has been interpreted as a flow-through system, with the majority of excretion occurring through the mouth and the minor passage of intracellular waste products occurring through the anal pores (Bumann and Puls 1997). The results presented here contradict the two prevailing ideas about waste removal in ctenophores: (1) the relatively more recent dogma stating that ctenophores have only one functional opening to the gut (Hejnol and Martindale 2009; Martindale and Hejnol 2009; Dunn, et al. 2014; Hejnol and Martín-Durán 2015; Telford, et al. 2015) and (2) the historical perspective that both the mouth and anal pores are used to remove digested food waste (Hyman 1940; Bumann and Puls 1997; Tamm 2014a). Notably, authors have also disagreed as to whether the aboral pores constitute an anus (e.g., Hyman 1940).

In contrast to prior studies, we did not observe food waste expulsion or regurgitation through the mouth during our feeding experiments with *Mnemiopsis*. Ctenophore regurgitation is associated with three events: (1) over- feeding (Baker and Reeve 1974; Reeve, et al. 1978; Bumann and Puls 1997), (2) ingestion of material that cannot be predigested and/or preprocessed in the pharynx for distribution through the endodermal canals (Baker and Reeve 1974; Reeve, et al. 1978), and (3) ingestion of

unpalatable material (Baker and Reeve 1974). Unsurprisingly, most metazoans will regurgitate material through the mouth under these conditions. When captive ctenophores are over-handled or fed excessively, they will regurgitate material from the pharynx. The ctenophore pharynx can also reach a maximum concentration threshold, after which material contained within the pharynx is regurgitated (Reeve, et al. 1978). In their natural environment, the likelihood of reaching critical prey threshold within the pharynx is low. We argue that in previous studies supporting the mouth as a major site of waste removal, feeding regimes surpassed this threshold and the resulting regurgitation behavior was thus interpreted as a normal mode of waste removal. In addition, it is clear that prey with significant cuticle and/or chitinous exoskeletons can reside in the pharynx for as long as 7 hr during digestion (Granhag, et al. 2011). In our long-term lab-cultured animals, we observe extended residence time in the pharynx for rotifer and crustacean exoskeletons during digestion. As long as the pharynx is not overwhelmed, exoskeletons are typically broken down, passed into the endodermal canals, and expelled from the anal pores.

Our long-term, multigenerational ctenophore cultures coupled with our functional anatomical results and previous descriptions of ctenophore digestion clearly demonstrate that ctenophores possess a functionally tripartite through-gut. This directly refutes the historical characterization of ctenophores as possessing a sac-like gut with a single opening. The results presented here show that ingested food processed by the pharynx and distributed through the endodermal canal system is subsequently expelled *en masse* at regular intervals through the anal pores.

Comparatively, the functional anatomy of the *Mnemiopsis* digestive system is typical of Ctenophora, though some ctenophore body plans have modifications to

endodermal canal peripheral branching patterns. We have observed the same digestive system function across a range of ctenophore body plans, including representatives from the genera *Beroe*, *Bolinopsis*, *Ocyropsis*, *Cestum*, *Coeloplana*, and *Vallicula*. Among other non-bilaterians, pores connecting the gastric cavity to the external environment have been reported in Cnidaria, including both hydrozoan medusa and some sessile corals exhibiting true fluid flow-through systems (Gardiner 1972; Aria and Chan 1989; Schlichter 1991). However, cnidarians are still considered to possess a blind gut (Hyman 1940; Bumann and Puls 1997; Martindale and Henry 1998; Martindale, et al. 2002; Martindale, et al. 2004; Dunn, et al. 2014; Hejnol and Martín-Durán 2015). Thus, our results represent an additional significant morphological challenge to the historical grouping of ctenophores with cnidarians as the Coelenterata (Frey and Leuckart 1847; Haeckel 1875; Telford, et al. 2015).

The recognition of a through-gut in ctenophores has broad implications for our understanding of the evolution of the metazoan anus and the unidirectional alimentary canal (Figures 3.4 and 3.6). One scenario for the evolution of the extant ctenophore through- gut is that an anus formed independently in the ctenophore stem lineage (Figure 3.4A). This scenario does not explicitly reject current hypotheses regarding bilaterian through-gut evolution (Hejnol and Martindale 2008; Martindale and Hejnol 2009; Hejnol and Martín-Durán 2015). A second scenario is that the sophisticated morphological and cellular organization of the through-gut has very early origins in the metazoan stem lineage, suggesting the loss of a functional through-gut in several non-bilaterian lineages (Figure 4B). Secondary loss of organ systems is not atypical. Loss of the through-gut and anus has occurred multiple times within Bilateria, for example within the deuterostomes

among the echinoderms and among members of the Platyzoa (Hejnol and Martín-Durán 2015). The latter scenario would reject current hypotheses that posit that the anus first evolved after the divergence of the Acoelomorpha (Hejnol and Martindale 2008; Martindale and Hejnol 2009; Hejnol and Martín-Durán 2015). Though functionally analogous, the homology of the ctenophore anal pore to the bilaterian anus remains unclear. For the two evolutionary scenarios to be resolved, further research will be needed to distinguish between the functional and structural similarities of the ctenophore and bilaterian anus (Giribet 2015).

Our findings draw attention to the evolutionary importance of the anus. The mouth is currently considered to be homologous across most of Metazoa (Arendt, et al. 2001; Martindale and Hejnol 2009). In ctenophores, as well as many protostomes, the position of the mouth corresponds with the blastopore (Freeman 1977). Moreover, the ctenophore blastopore exhibits the same conserved gene expression patterns seen in bilaterians and cnidarians (Arendt, et al. 2001; Scholz and Technau 2003; Yamada, et al. 2007). However, the anal opening among bilaterians has a varied phylogenetic distribution, suggesting independent evolution of the through-gut and anus in many lineages (Hejnol and Martindale 2008; Hejnol and Martín-Durán 2015). Genes with clear orthology to those expressed in the bilaterian hindgut and anus (e.g., NK2.1, fox genes, and FGF8/17/18) (Hejnol and Martín-Durán 2015) are not present in either the *Mnemiopsis* or *Pleurobrachia* genomes (Ryan, et al. 2013; Moroz, et al. 2014), whereas a number of genes expressed in the bilaterian gut have clear orthologs that are expressed in ctenophores (Figure 3.7). Additional investigation of gene expression in ctenophore anal pores and more detailed functional and structural morphology analyses in these and other

non-bilaterian animals will certainly provide more insight into the evolution of the anus and the metazoan alimentary canal (Giribet 2015; Hejnol and Martín-Durán 2015). An improved understanding of the through-gut organ system is crucial for improving our understanding of body plan evolution and diversification in the metazoans.

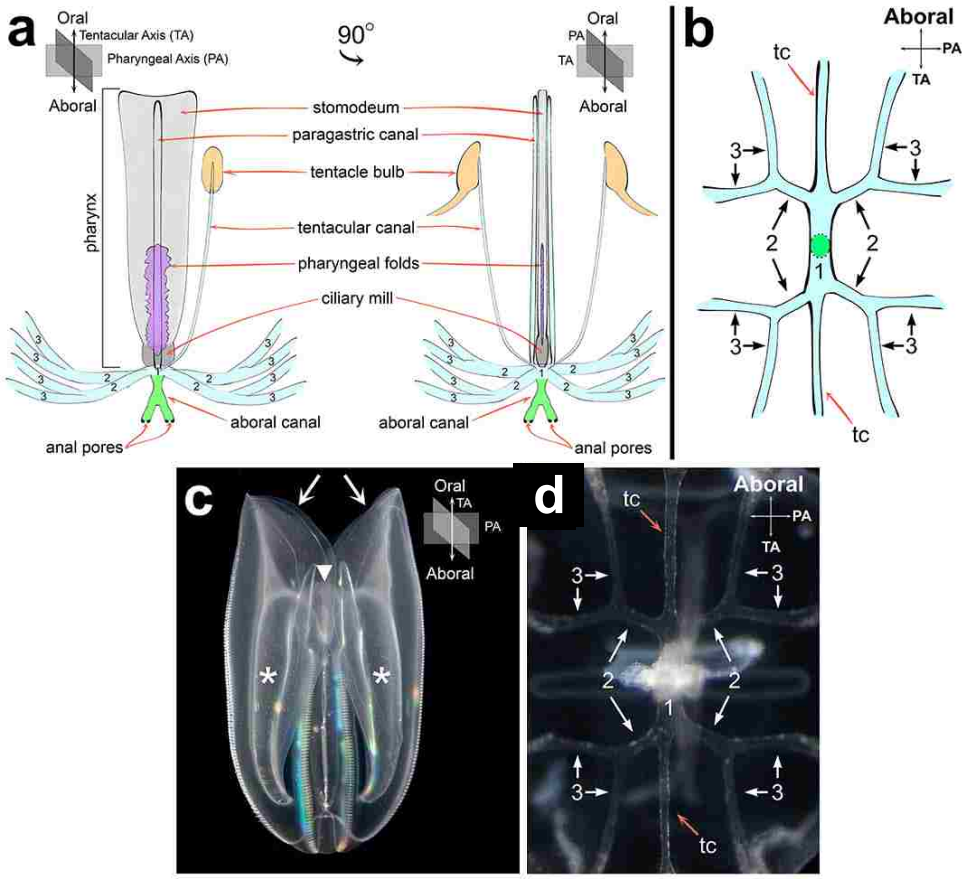


Figure 3.1. Ctenophore digestive system anatomy. (A) Schematic of the major features of the ctenophore digestive system. The pharyngeal axis (PA) is to the left, and the tentacular axis (TA) is to the right. Food enters the stomodeum and moves aborally through the pharynx (light gray), where digestive enzymes are secreted by the pharyngeal folds (purple). Large particles are then mechanically processed by a region of stiffened cilia, the ciliary mill (dark gray). Small food particles then transit into the central chamber, the infundibulum (1), of the endodermal canal system (blue), where isodynamic ciliary beating distributes particles for absorption in the radial canals (2), the paired adradial canals (3), and the paragastric canals and tentacular canals. Both solid and fluid wastes are eliminated from the branched endodermal canal system by transport through the aboral canal and anal canals (green) and evacuation through the anal pores. (B) Aboral schematic highlighting the major endodermal canal system branches. The infundibulum, or transverse canal (1), serves as the central hub from which two tentacular canals (tc) and four radial canals (2) branch. Each radial canal bifurcates into adradial canals (3). The green circle marks the aboral canal branch. (C) Adult *Mnemiopsis leidyi* are characterized by feeding grooves medial to oral lobes (arrows) lined with tentacles bearing colloblasts (asterisks) that transport prey to the stomodeum (arrowhead). (D) Aboral view of a juvenile cydippid stage *Mnemiopsis leidyi* showing the principal branches of the endodermal canal system as schematized in (B).

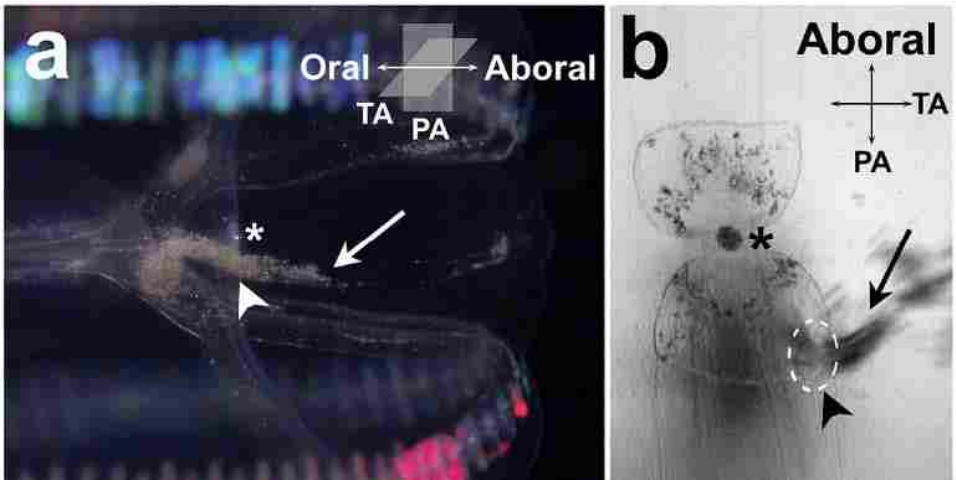


Figure 3.2. Waste evacuation in ctenophores. (A) Frame grab from time-lapse showing waste material being ejected from open anal pore of *Mnemiopsis leidyi*. (B) Frame grab showing an aboral view of a dilated anal pore during waste ejection from *Mnemiopsis leidyi*. Asterisks mark the position of the aboral apical organ, arrowheads mark the position of the open anal pore, and arrows denote waste plumes. (A) is a lateral view, with the oral end to the left; (B) is an aboral view.

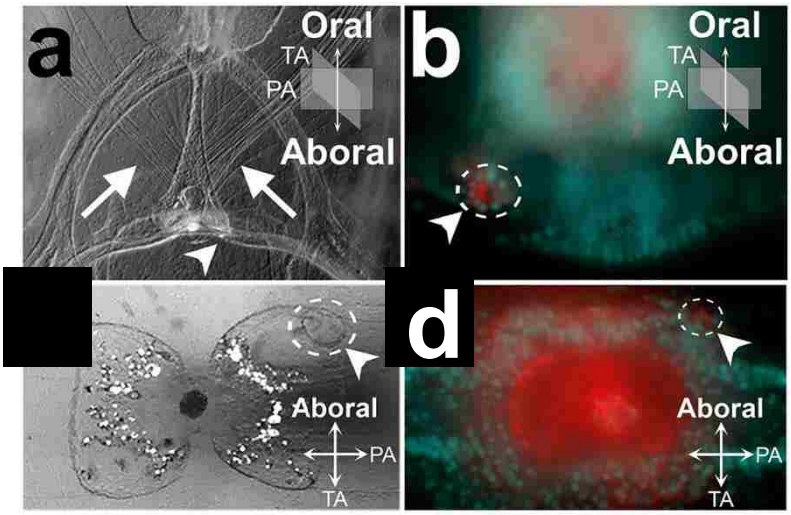


Figure 3.3. Aboral region of *Mnemiopsis*. (A) Differential interference contrast image of *Mnemiopsis*. Arrows highlight organized giant smooth muscle bundles that contact and regulate the diameter of the aboral and anal canals. (B) *Mnemiopsis* stained with phalloidin in red and DAPI in blue. The arrowhead and dashed circle mark the ring of F-actin-rich cells that contribute to the circumference of the contractile sphincter associated with the anal pore. (C) *Mnemiopsis* aboral view from merged brightfield and fluorescence microscopy. The arrowhead and dashed circle mark the position of a fully dilated anal pore. Digested fluorescent fish particles appear as bright spots in the anal canals. (D) *Mnemiopsis* stained with phalloidin in red and DAPI in blue. The arrowhead and dashed circle mark the position of F-actin associated with one of the two contractile anal pores. In all panels, arrowheads indicate the position of anal pores. Lateral views of the pharyngeal plane are shown in (A), (B), and aboral views are shown in (C), (D).

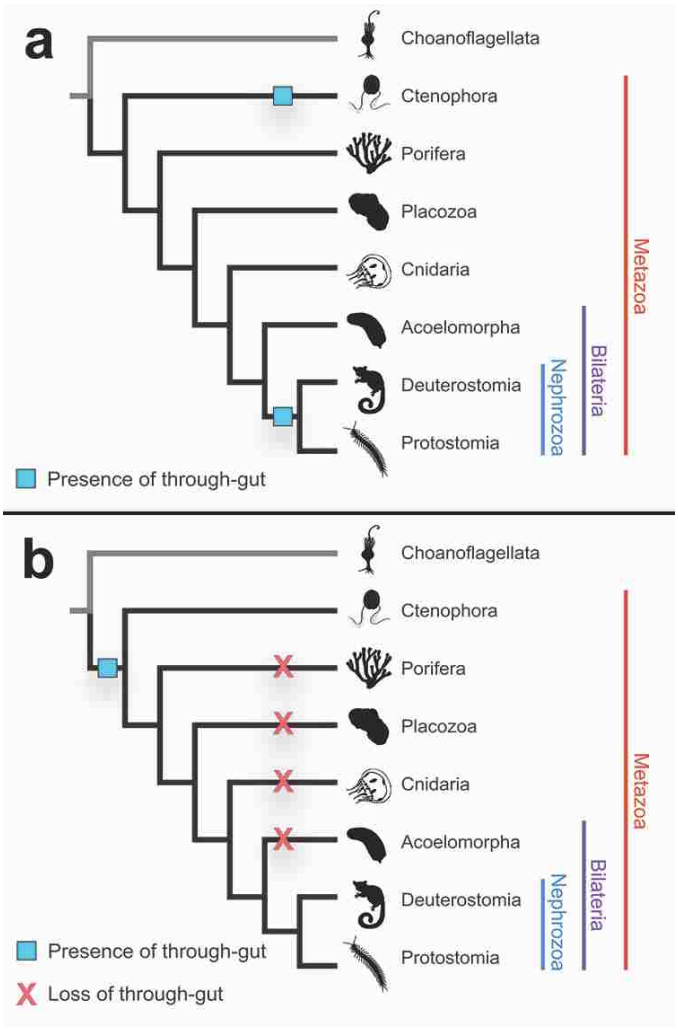
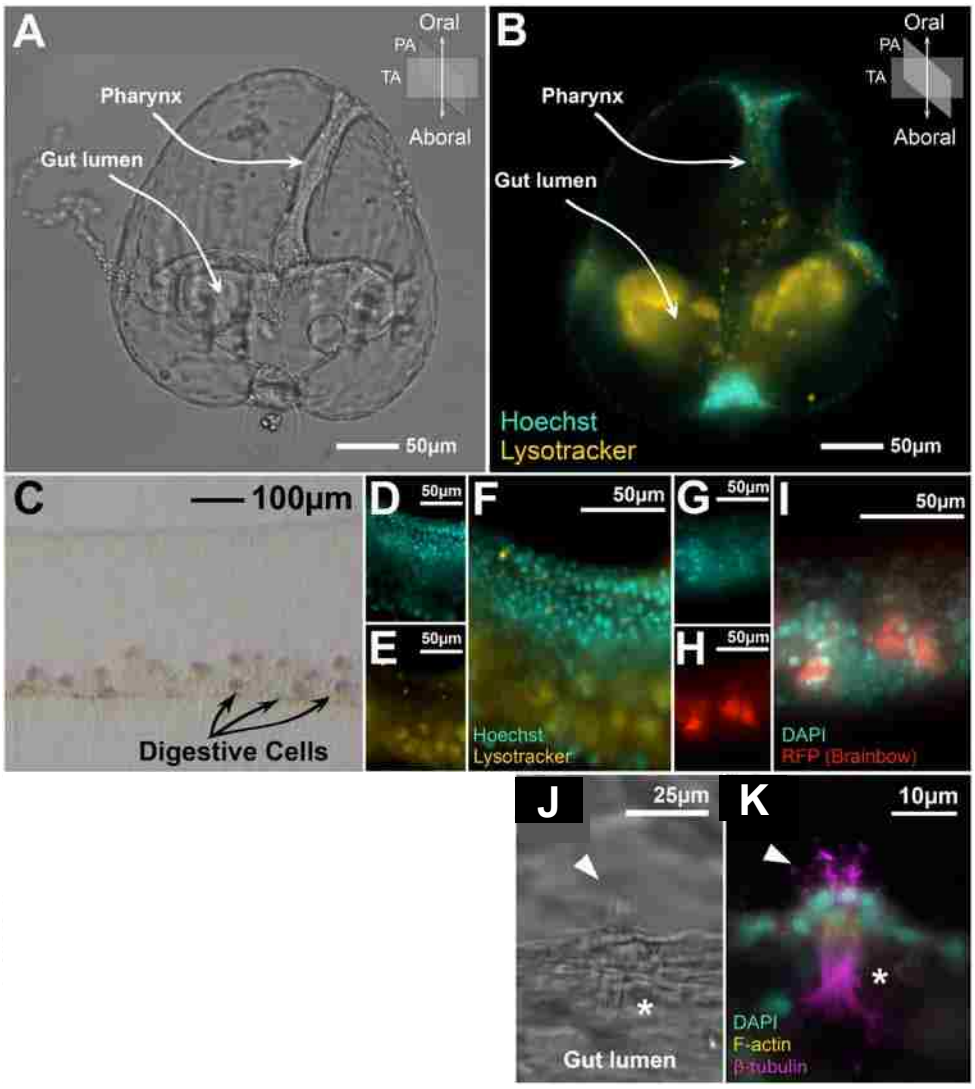


Figure 3.4. Evolutionary scenarios for the origins of the animal through-gut. (A) Scenario for the convergent acquisition of a functional through-gut in Ctenophora. After the divergence of Ctenophora, the through-gut organ system evolved independently in both the ctenophore lineage and the bilaterian lineage after the divergence of Acoelomorpha (Hejnol and Martindale 2008; Hejnol and Martín-Durán 2015). (B) Scenario for a single origin of the through-gut prior to the divergence of the ctenophore lineage. Under this scenario, several non-bilaterian lineages and the Acoelomorpha would be inferred to have subsequently lost a functional through-gut, rejecting the initial origins of a functional through-gut within Bilateria. Although phylogenetic relationships between early branching animals remain contentious, our animal phylogeny is based on recent phylogenomic inferences (Dunn, et al. 2008; Hejnol, et al. 2009; Ryan, et al. 2013; Moroz, et al. 2014; Whelan, Kocot, Moroz, et al. 2015). Organism silhouettes by Noah Schottman, Mali'o Kodis, Oliver Voigt, and Joseph Ryan are available from the PhyloPic database (http://phylopic.org) and are used here under the Creative Commons Attribution 3.0 license. See also Figure 3.6.

Figure 3.5. Specialized cell types associated with the ctenophore gut. Related to Figure 3.1. A and B are live *Mnemiopsis leidyi* 2 days post hatching. Oral end is oriented up and arrows highlight the pharynx and gut lumen. (A) DIC image. (B) Fluorescence image, nuclei are stained with Hoechst in blue and lysosomes are stained with Lysotracker in yellow. Lysosome signal strongly marks cells lining the oral region of the gut as well as cells scattered throughout the pharynx. (C) Bright-field image of endodermal canal in ex vivo tissue envelope isolated from an adult Mnemiopsis. Arrows highlight the asymmetrically distributed digestive cells that contain darkly pigmented lysosome vacuoles. (D) Adult *Mnemiopsis* endodermal canal with Hoechst stained nuclei in blue. (E) Adult *Mnemiopsis* endodermal canal with Lysotracker stained lysosomes in yellow. (F) Merged image of D and E showing the distinct asymmetric localization of digestive cells containing lysosomes in the endodermal canal wall. (G) Adult *Mnemiopsis* endodermal canal with Hoechst stained nuclei in blue. (H) Adult *Mnemiopsis* endodermal canal with RFP signal derived from digested Brainbow zebrafish. (I) Merged image of G and H showing the differential retention of RFP in digestive cells along the endodermal canal wall. (J) DIC image of cells in the endodermal canal wall organized into a ciliated rosette. The mesogleal space is at the top and the endodermal canal lumen is at the bottom. Each ciliated rosette contains a central pore that provides access between the lumen of the gut and the mesoglea. The asterisk marks specialized stiffened cilia that project into the gut lumen, the arrowhead marks long cilia that project into the mesoglea. (K) Merged fluorescent confocal image of a ciliated rosette. Nuclei are marked by DAPI in blue. F-actin is marked by phalloidin in yellow. β-tubulin is marked by E7 in purple. The ciliated rosette is formed from two concentric rings of eight cells, each having a β-tubulin rich region facing the central pore. The lower ring of cells project cilia into the gut lumen, indicated by an asterisk. The upper ring of cells project cilia into the mesoglea, indicated by an arrowhead. The two groups of specialized cilia function as ultrafilters preventing transit of, and/or blockage by, large particles between the two compartments.


A
Oral
PA
TA
Aboral
Pharynx
Gut lumen
50µm
B
Oral
PA
TA
Aboral
Pharynx
Gut lumen
Hoechst
Lysotracker
50µm
C
100µm
Digestive Cells
D
50µm
E
50µm
F
50µm
Hoechst
Lysotracker
G
50µm
H
50µm
I
50µm
DAPI
RFP (Brainbow)
J
25µm
Gut lumen
K
10µm
DAPI
F-actin
β-tubulin

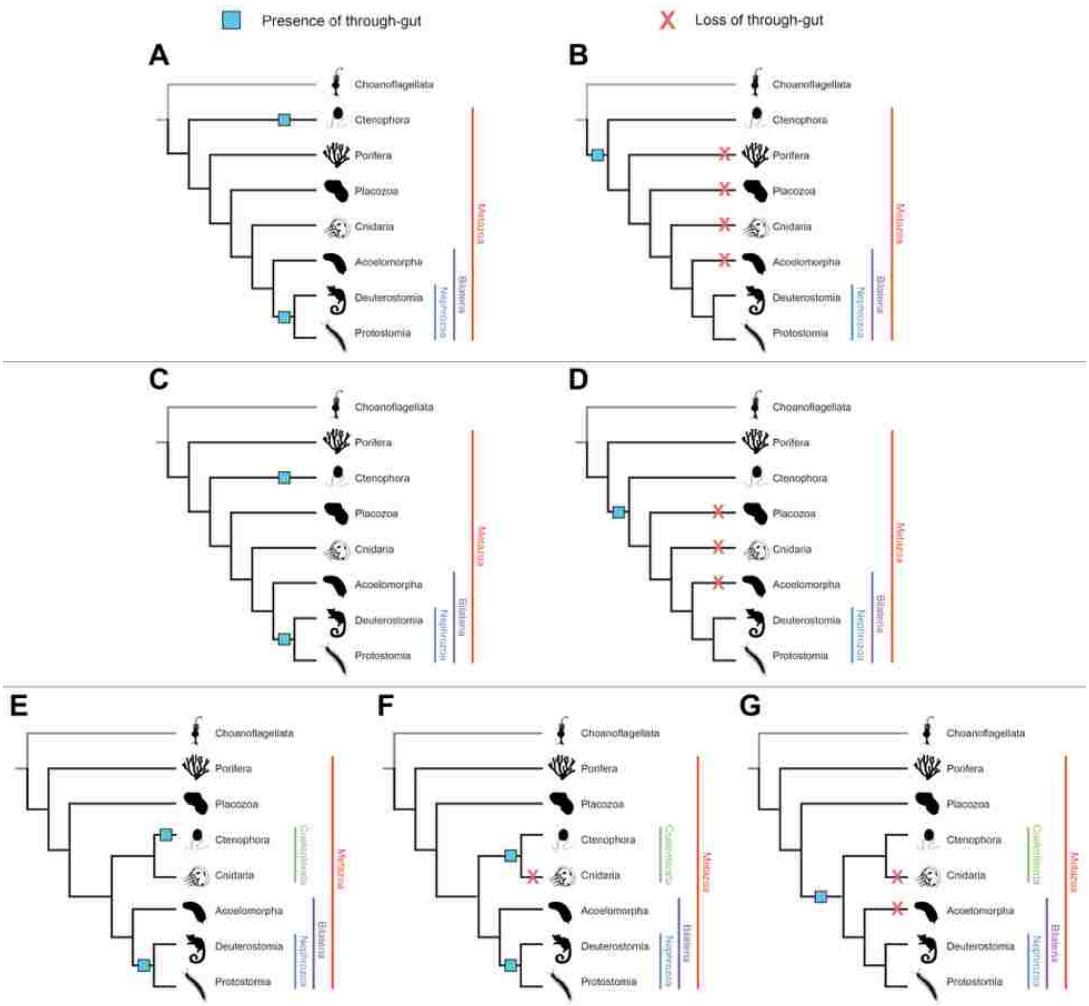
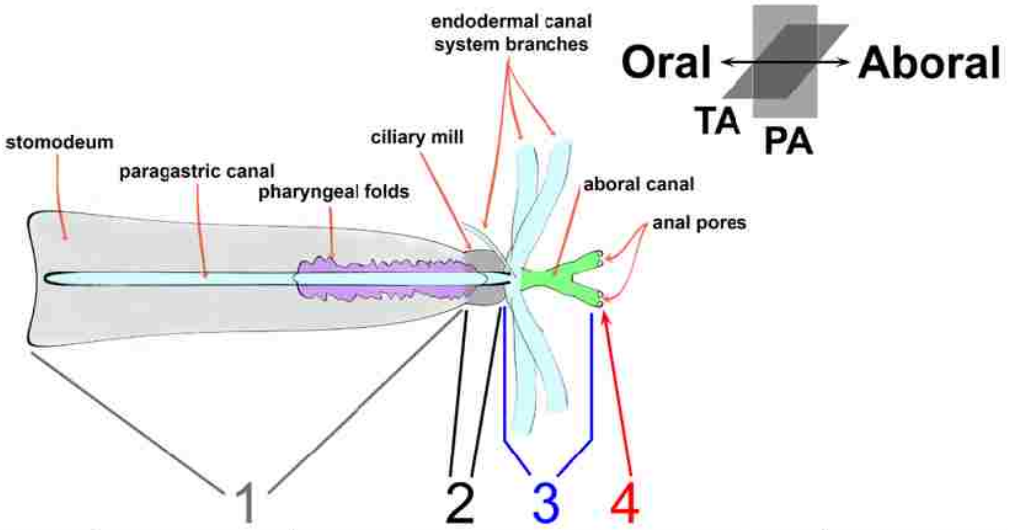


Figure 3.6. Alternative scenarios of animal through-gut evolution. Related to Figure 4. Panels represent alternative phylogenetic hypotheses for the position of Ctenophora relative to other major lineages of metazoans based on recent phylogenomic analyses. Blue squares represent the presence of a through-gut while a red X represents an inferred loss of the through-gut. Panels (A) and (B) depict Ctenophora as the earliest branching extant metazoan lineage based on data from (Dunn, et al. 2008; Hejnol, et al. 2009). Panels (C) and (D) depict the divergence of Porifera at the base of the Metazoa based on analyses from (Pick, et al. 2010). Panels (E), (F), and (G) depict the 'Coelenterata' clade inferred by (Philippe, et al. 2009; Nosenko, et al. 2013). All three scenarios allow for the independent, convergent acquisition of a through-gut within the Ctenophora lineage as presented in panels A, C, and E. Explicit scenarios for the ancestral evolution of a through- gut in the metazoan stem lineage with subsequent losses among Porifera, Placozoa, Cnidaria, and/or Acoelomorpha are presented in panels B, D, and G. Panel F represents a scenario in which the independent, convergent evolution of a through-gut occurred in the 'Coelenterata' stem lineage and was subsequently lost in Cnidarians. The phylogenetic placement of Acoelomorpha along with the absence of a through-gut is based on (Hejnol and Martindale 2008).

Figure 3.7. Gene expression associated with the ctenophore gut. Related to Figure 3.1. (A) Schematic of the ctenophore tripartite digestive system. The oral end is oriented to the left and major features are labeled as in Figure 3.1. The digestive system has been divided into four regions of gene expression based on prior gene expression studies highlighted in panel B. (B) Table of known gene expression in the ctenophore digestive system by region, including relevant citations (Yamada, et al. 2007; Pang and Martindale 2008f; Pang, et al. 2010; Pang, et al. 2011; Reitzel, et al. 2011; Dayraud, et al. 2012; Simmons, et al. 2012; Jager, et al. 2013; Schnitzler, et al. 2014).

A



B

Species	Gene (GeneID)	Digestive System Expression Pattern	Region	Citation
Mnemiopsis leidyi	*MlTbx1* (ML046417a)	aboral-most region of developing pharynx, developing canals, developing anal pores	2,3,4	Yamada *et al.*, 2007
	MlBra (ML174736a)	developing pharynx	1	
	MlNKL1 (ML058515a)	oral-most and aboral-most region of developing pharynx	1,2	Pang & Martindale 2008
	MlBar (ML03151a)	developing pharynx including aboral region	1,2	
	MlTlx-like (ML01435a)	developing pharynx	1	
	MlPrd1 (ML021120a)	developing pharynx	1	
	MlPrd2 (ML10117a)	developing pharynx	1	
	MlPrd3 (ML001520b)	developing pharynx including aboral region	1,2	
	MlFzdA (ML003224a)	developing pharynx	1	Pang *et al.*, 2010
	MlFzd3 (ML034649b)	developing pharynx	1	
	MlStrp (ML223520b)	developing pharynx	1	
	MlDsh (ML00535-7a,8a)	throughout developing digestive system	1,2,3,4	
	MlBcat (ML073715a)	throughout developing digestive system	1,2,3,4	
	MlBmp5-8 (ML218835a)	developing pharynx	1	Pang *et al.*, 2011
	MlTgf1a (ML200252a)	developing pharynx	1	
	MlTGFb3 (MLML19322b)	oral-most and aboral-most region of developing pharynx, developing canals	1,2,3	
	MlTolloid (ML016314a)	developing pharynx, developing canals	1,2,3	
	MlTgfRII (ML08593b)	throughout developing digestive system	1,2,3,4	
	MlTgfRIa (ML082117a)	oral-most and aboral-most region of developing pharynx	1,2	
	MlTgfRIb (ML131110a)	aboral-most region of developing pharynx, developing canals	2,3	
	MlTgfRIc (ML046518a)	developing pharynx	1	
	MlSmad6 (ML19701a)	aboral-most region of developing pharynx	2	
	MlSmad4 (ML02191a)	aboral-most region of developing pharynx	2	
	MlSmac1a (ML093050a)	faint signal in developing pharynx	1	
	MlSmad2 (ML011743a)	throughout developing digestive system	1,2,3,4	
	MlNR3 (ML305522a)	aboral-most region of developing pharynx, developing canals	2,3	Reitzel *et al.*, 2011
	MlLhx1/5 (ML132520a)	developing pharynx	1	Simmons *et al.*, 2012
	MleSox1 (ML047927a)	developing pharynx	1	Schnitzler *et al.*, 2014
	MleSox2 (ML234028a)	developing pharynx	1	
	MleSox3 (ML042722a)	developing pharynx including aboral region	1,2	
	MleSox4 (ML06932a)	developing pharynx	1	
	MleSoxd (ML01707a)	developing pharynx including aboral region	1,2	
Pleurobrachia pileus	*PpiMHCIIb2*	adult anal pores	4	Dayraud *et al.*, 2012
	PpiDvl1	adult anal pores	4	Jager *et al.*, 2013

Chapter 4: KLF/SP transcription factor family evolution: expansion, diversification, and innovation in eukaryotes[2]

Background

One of the most ancient and abundant classes of DNA binding domains (DBDs) is the C2H2 zinc finger class (Rubin, et al. 2000; Ravasi, et al. 2003; de Mendoza, et al. 2013). The C2H2 zinc finger domain has two cysteine and two histidine residues that coordinate a zinc ion, and typically consists of the amino acid sequence C-X(2-4)-C-C-X(12)-H-X(3-5)-H (Brown, et al. 1985; Miller, et al. 1985). C2H2 zinc finger motifs are found in many transcription factors and based on their arrangement and number can be subdivided into different families (Iuchi 2001). The Krüppel-like factor and specificity protein (KLF/SP) transcription factor gene family is characterized by a highly conserved triple-C2H2 DBD located toward the C-terminus composed of three tandem zinc fingers that are evenly spaced by conserved linker regions (Iuchi 2001) and share similarity with the *Drosophila* Krüppel gene (Rosenberg, et al. 1986). This C2H2 zinc finger DBD (KLF-DBD) binds to guanine-cytosine-rich regions and CACC elements (GT boxes) (Kadonaga, et al. 1987). The more N-terminal regions of KLF/SP transcription factors are typically highly variable and consist of different combinations of transactivation/repression domains. Historically, mammalian KLFs have been divided into 3 groups based on shared domain architecture: The KLF1, 2, 4, 5, 6, and 7 groups; the KLF3, 8, and 12 groups; and the KLF9, 10, 11, 13, 14, 16 groups (McConnell and Yang 2010), whereas SPs, which differ from KLFs by the presence of the Buttonhead

[2] This chapter presented here was published as the following paper: Presnell JS, Schnitzler CE, Browne WE (2015) **KLF/SP Transcription Factor Family Evolution: Expansion, Diversification, and Innovation in Eukaryotes.** *Genome Biology and Evolution* 7:2289-2309 doi:10.1093/gbe/evv141.

(Btd) box domain just 5' of the KLF-DBD, are typically divided into 2 groups: SP1–4 and SP5–9 (Suske, et al. 2005).

KLF/SP genes within each domain architecture group share similar functions based on the retention of explicit transactivation motif complements. A range of studies present a complex picture in which KLF/SP genes can be singly or in combination involved in temporally and spatially disparate cellular and developmental processes. For example, fly embryos mutant for *luna*, the *Drosophila* KLF6/7 ortholog, die early during development due to mitotic defects (De Graeve, et al. 2003; Weber, et al. 2014), whereas *cabut*, a KLF9/13 ortholog in *Drosophila*, is required for proper dorsal closure during gastrulation (Muñoz-Descalzo, et al. 2005). However, *cabut* also plays a role later in fly organ development by coordinating signaling for proper wing disc patterning (Rodriguez 2011). Among the vertebrates, KLF genes are often associated with balancing stem cell proliferation and differentiation, as well as regulating metabolic homeostasis. The most notable member is KLF4, one of the four pioneer transcription factors required to induce pluripotency in human and mouse fibroblasts (Takahashi and Yamanaka 2006; Soufi, et al. 2012; Soufi, et al. 2015) and a component of a core circuit of genes that maintain self-renewal in mammalian embryonic stem cells along with KLF2 and KLF5 (Jiang, et al. 2008). However, in gut epithelia, KLF4 regulates terminally differentiated cells while KLF5 is expressed in the proliferating crypt cells (McConnell, et al. 2007). In mammals, KLF2 together with KLF1 and KLF13 also regulate erythrocyte maturation and differentiation as well as globin gene activity (Miller and Bieker 1993; Basu, et al. 2005; Gordon, et al. 2008). KLF2 in zebrafish contributes to the differentiation of ectoderm derived tissues (Kotkamp, et al. 2014). In mammals, including humans, KLF11 and

KLF14 play an important role in the regulation of genes associated with diabetes and metabolic syndrome phenotypes, respectively (Small, et al. 2011; Lomberk, et al. 2013). Similarly, complex intersections with both development and metabolism exist for members of the SP subfamily. For example, in mammals, SP1, SP3, and SP7 regulate osteoblast mineralization and differentiation (Nakashima, et al. 2002; Suttamanatwong, et al. 2009). SP1 is also an important regulator of metabolic genes involved in the glycolytic pathway, fatty acid synthesis, and ribosome biogenesis (Archer 2011; Nosrati, et al. 2014). Overall, members of the KLF/SP gene family are known to function in a wide variety of biological processes (Black, et al. 2001; Zhao and Meng 2005; Wierstra 2008; McConnell and Yang 2010; Zhao, et al. 2010; Tsai, et al. 2014).

In contrast to the extensive studies highlighting the importance of the KLF/SP genes to core cellular processes, comparatively few studies have investigated the evolutionary relationship of KLF/SP genes in lineages outside of mammals (Kolell and Crawford 2002; Materna, et al. 2006; Shimeld 2008; Chen, et al. 2009; Meadows, et al. 2009; Schaeper, et al. 2010; Seetharam, et al. 2010). A KLF gene was recently identified in the choanoflagellate *Monosiga brevicollis* genome; however, that study's conclusions were restricted to examining porcine KLF paralogy (Chen, et al. 2009). A more recent study, focused on the phylogenetic distribution of C2H2 zinc finger families in eukaryotes, also showed that KLFs were present in *Monosiga* but absent in the fungal taxa surveyed (Seetharam and Stuart 2013). No study had examined the phylogenetic context of the different transactivation/repression domains associated with the KLF/SP gene family. Pinpointing the origin and evolutionary history of this gene family and associated domains can help determine possible relationships of the KLF/SP repertoire

expansion to key innovations in the evolution of metazoan cellular diversity. Hypotheses of metazoan gene evolution are greatly aided by sampling a wide range of taxa that include non-metazoan representatives. Here, we infer the evolutionary history of the KLF/SP gene family and their associated transactivation/repression domains across a wide range of eukaryotes. Species were chosen to represent well established groups across the Eukarya including the following: Bikonta, Amorphea (Amoebozoa + Apusozoa + Opisthokonta), Opisthokonta (Holomycota + Holozoa), and Holozoa (Ichthyosporea + Filasterea + Choanoflagellata + Metazoa) (fig. 4.1) (Adl, et al. 2012; Derelle and Lang 2012; Paps, et al. 2013). Several transactivation/repression domains found in KLF/SPs are also present in the genomes of ancient unicellular lineages that lack KLF/SPs, lending support for domain shuffling playing a major role in the acquisition of transactivation/repression domains during the expansion and diversification of the KLF/SP gene family. We show that domain connectivity and resulting unique domain architectures among these transcription factors have become increasingly complex in metazoans. Thus, a pattern of gene duplication along with domain shuffling and the rare emergence of *de novo* domains have collectively played a vital role in the evolution of the KLF/SP gene family.

Methods

KLF Identification Pipeline

To broadly identify C2H2 zinc finger and KLF/SP proteins associated with a diverse range of Eukaryote lineages, publicly available sequenced genomes listed in table 1 were comprehensively searched including 26 metazoans, 4 unicellular holozoans, 7

holomycotans, 1 apusozoan, 4 amoebozoans, and 6 bikonts. We used the HMMER 3.0 program (Eddy 1998), to identify proteins that contained C2H2 zinc fingers. The Hidden Markov Model (hmm) of the C2H2 zinc finger domain PF00096 (Punta, et al. 2012) was downloaded from the Pfam database. The *hmmsearch* command was then used to search protein models of the representative 48 Eukaryote species using the PF00096 hmm as a query. HMMER identified all protein sequences (using default settings) that contained at least one C2H2 zinc finger corresponding to the hmm. This output was used for subsequent analyses. Raw outputs for this and subsequent steps of the pipeline can be found online at https://goo.gl/dbdBil.

We then used a perl script modified from Zeng et al. (Zeng, et al. 2011) to search the protein sequences of the HMMER output for the 81-amino acid triple-C2H2 zinc finger DBD conserved in KLF/SPs (KLF-DBD; fig. 4.2). The amino acid sequence is as follows: C-X4-C-X12-H-X3-H-X7-C-X4-C-X12-H-X3-H-X7-C-X2-C-X12-H-X3-H, where X can be any amino acid. Sequences meeting the following criteria were initially considered putative KLF/SPs: 1) Presence of the KLF-DBD and 2) presence of only three zinc fingers (i.e., no additional zinc fingers other than the KLF-DBD). Sequences fitting these initial criteria were only found within the Filozoa (online at https://goo.gl/dbdBil). We additionally found 13 non-filozoan amorphean sequences possessing a KLF-DBD; however, in all cases these sequences also contained additional zinc fingers (online at https://goo.gl/dbdBil). To determine the relationship between the non-filozoan amorphean sequences and the putative filozoan KLF/SPs, all sequences were aligned and included in our phylogenetic analyses.

Where possible, partial KLF/SP gene models were manually extended by obtaining genomic scaffold regions spanning the location of incompletely annotated gene models. Briefly, the CLCBio software package was used to map partial KLF/SP nucleotide sequences to genomic scaffolds and identify associated open reading frames. Extended exonic nucleotide sequences were then translated to obtain enhanced annotation of KLF/SP amino acid sequences. Protein schematics were created using the PROSITE MyDomians Image Creator (http://prosite.expasy.org/cgi-bin/prosite/mydomains/, last accessed April 30, 2014). Scripts for the KLF/SP identification pipeline are publicly available online at https://goo.gl/dbdBil.

Transactivation/Repression Domain Identification and Characterization

Custom perl scripts were used to search for KLF/SP-associated transactivation/repression domains. The motifs used for the perl scripts followed these amino acid sequences: Btd box, C-X-C-P-X-C (Wimmer, et al. 1993); SP box, S/T-P-L-X-ϕ-L-X-X-X-C-X-R/K-ϕ (Harrison, et al. 2000); SID, D-X1-4-X-A-ϕ-X-X-L-M/V/L/A-X-F/M/L/I (Zhang, et al. 2001); PVDLS, P-V-D-L-S/T (Crossley, et al. 1996); R2, S-V-I-R-H-T-X-D/E (Cook, et al. 1999); and R3, ϕ-X-X-G-X-ϕ-ϕ-ϕ-ϕ-ϕ-ϕ-P/S-Q/P (Cook, et al. 1999), where X can be any amino acid and ϕ represents one of V, I, L, M, F, W, G, A, or P. To confirm the accuracy of the automated identification of transactivation/repression domains, we first searched the completely annotated human KLF/SP amino acid sequences. In all cases, our searches identified the human KLF/SP sequences that had previously been characterized and annotated as containing the specified domains. For example, the search with the Btd box model identified all 9 human SP sequences. Thus,

our models are both specific and sensitive enough to capture the intended domains. To identify the nine-amino-acid transactivation domain (9aaTAD), we utilized a prediction tool from http://www.med.muni.cz/9aaTAD/index.php (last accessed April 30, 2014), using the "moderately stringent pattern" option (Piskacek, et al. 2007). Compositionally biased, low complexity regions (LCRs) associated with KLF/SPs were identified using four independent assessments: The CAST algorithm, http://athina.biol.uoa.gr/cgi-bin/CAST/cast_cgi (Promponas, et al. 2000); ScanProsite, http://prosite.expasy.org/scanprosite/ (last accessed April 30, 2014) (de Castro, et al. 2006); the SEG algorithm, http://mendel.imp.ac.at/METHODS/seg.server.html (last accessed April 30, 2014) (Wootton 1994); and manual curation based on criteria from (Sim and Creamer 2004). A putative LCR was assigned if it was identified in at least two of the four methods (supplementary table S2, online at https://goo.gl/dbdBil).

Additionally, we searched the entire genomes from our representative 48 eukaryotes for the following transactivation/repression domains in proteins excluding identified KLF/SP genes: Btd box, SP box, SID, PVDLS, R2, and R3 domain motifs. Perl scripts for each transactivation/repression domain model were run against the complete set of protein models from each genome (Supplementary Material online at https://goo.gl/dbdBil).

Domain Co-occurrence Networks

To visually represent the different domain architectures of the identified KLF/SP proteins, we created co-occurrence network maps. For each species network map, all unique domain combination pairs were identified, summed, and divided by the total

number of proteins (KLFs only, SPs only, or both KLF/SPs) to obtain a percentage of co-occurrence for a particular domain pair. Composite network maps reflecting an organismal clade or grade were generated by the same procedure as for individual species maps. The size of the circles (domains) reflects how often specific domains appear relative to the KLF-DBD, which has 100% representation. Lines connecting two domains represent that unique combination pair. Line weights represent the relative frequency of that specific pair combination. A given domain pair combination was only counted once. Network maps were visualized using Microsoft PowerPoint.

Phylogenetic Analysis

Putative filozoan KLF/SP and non-filozoan amorphean KLF-DBD amino acid sequences were aligned using default settings of the MUSCLE alignment package in CLCBio (Edgar 2004). To improve statistical support in our analyses, we concatenated the transactivation/repression domains with the KLF-DBD. The highly variable 9aaTAD and LCR sequences were excluded from the alignment. Individual aligned sequences ranged from 81 to 112 amino acids in length. Each domain, including the KLF-DBD, can be distinguished as separate blocks within the alignment (Supplementary Material online at https://goo.gl/dbdBil). For the non-filozoan amorphean sequences, only the core three zinc fingers corresponding to the KLF-DBD model were included in the alignment. Duplicate sequences were removed and are listed in supplementary table S5, Supplementary Material online at https://goo.gl/dbdBil. ProtTest v2.4 was used to determine the LG+I+G model as the best-fit model for protein evolution (Abascal, et al. 2005).

Maximum likelihood (ML) analyses were performed using the MPI version of RAxMLv7.2.8 (Stamatakis 2006). We executed 300 independent ML searches on randomized maximum parsimony starting trees using the standardized RAxML search algorithm, followed by comparison of likelihood values among all 300 resulting ML trees. The final log-likelihood score of the best ML tree was −12974.002279. One hundred bootstrap replicates were computed and applied to the best scoring ML tree. ML bootstrap values are indicated on the ML tree (supplementary fig. S2, online at https://goo.gl/dbdBil). Bayesian analyses were performed with MrBayes3.2.5 (Ronquist and Huelsenbeck 2003). We ran two independent 5 million generation runs of five chains each with default heating and with the "LG+I+G" amino acid model option. The "average standard deviation of split frequencies" between the two runs was 0.048635. This diagnostic is an indicator of how well the two runs converge. A value below 0.01 is a strong indication of convergence, while a value between 0.01 and 0.05 is typically acceptable for convergence. Additional convergence diagnostics were examined with AWTY (Nylander et al. 2008), which was used to determine if a sufficient number of generations had been completed for posterior probabilities to stabilize, and to determine the amount of burn-in. From 60,001 trees, 45,001 were sampled (25% burn-in was confirmed as adequate with Tracer v1.6) (Rambaut and Drummond 2007) and used to create a consensus tree. The runs reached stationarity, and adjusting the burn-in did not affect the topology of the tree. Bayesian posterior probabilities (BPP) were calculated and are shown on the Bayesian consensus tree (supplementary fig. S3, online at https://goo.gl/dbdBil). FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/) was used for tree visualization.

Results

C2H2 Zinc Finger and KLF Identification

To better understand the evolution of the KLF/SP gene family and to gain insight into the genetic repertoire of transactivation/repression domains known to regulate disparate aspects of metabolism, and growth and development, we comprehensively searched for, identified, and characterized KLF/SP gene family complements from 48 eukaryotic genomes using a combination of hmms and custom perl models (Table 1) with the assumption that the amorphean bikont split occurred at or near the origin of the Eukarya (Derelle and Lang 2012). We found C2H2 zinc finger proteins highly represented in all 48 eukaryotic genomes. The 81 amino acid KLF-DBD contains 3 highly conserved C2H2 zinc fingers separated by 2 highly conserved linker sequences (fig. 4.2). We found this domain architecture to be restricted to the opisthokont lineage (fig. 4.1). Among a small number of non-filozoans, represented by Sphaeroforma and several holomycotans, sequences containing the KLF-DBD motif were also found to possess additional zinc fingers. Further analyses revealed the presence of a conserved aspartic acid residue in the second zinc finger of the KLF-DBD at position 44 (D44) among 397 putative KLF/SP filozoan sequences, the single exception being a ctenophore gene. The D44 residue is critical for stabilizing proper KLF/SP DNA binding (Feng, et al. 1994; Schuetz, et al. 2011) and is notably absent from all non-filozoan KLF-DBD motifs and the ctenophore sequence MleKLFX. Our phylogenetic analyses consistently recovered a moderately supported clade that represents the 397 filozoan genes possessing the canonical D44 aspartic acid residue in the second zinc finger of the KLF-DBD to the exclusion of the remaining 13 non-filozoan sequences, along with MleKLFX (fig. 4.3).

We then searched the human genome with a refined KLF-DBD model including the canonical D44 residue. We recovered the 26 human KLF/SPs plus Wilms Tumor 1 (WT1). The WT1 gene contains an extra zinc finger and is phylogenetically distinct (Shimeld 2008). Therefore the high stringency D44 KLF-DBD model is specific and sensitive. When the D44 KLF-DBD model is run against the eukaryotic genomes, we recover only putative KLF/SP protein sequences along with a small number of WT1 orthologs and WT1-like sequences that contain additional zinc fingers (Supplementary Material online at https://goo.gl/dbdBil). Therefore, we operationally define bona fide KLF/SP orthologs as containing only 3 zinc fingers that conform to an 81 amino acid, 3 zinc finger KLF-DBD motif containing a canonical aspartic acid residue at position 44, D44 (fig. 4.2). Thus in our analyses KLF genes first appear in the filasterean *Capsaspora* while the SP subfamily first appears near the base of the metazoan lineage in sponges and is absent from ctenophores (fig. 4.1). The full set of KLF/SPs identified in this study with their corresponding accession numbers or protein/transcript identifiers are available online in supplementary table S1, at https://goo.gl/dbdBil. Gene names given to each KLF/SP gene from this study can be found in supplementary table S4, online at https://goo.gl/dbdBil.

C2H2 zinc finger proteins were identified in both unicellular and multicellular bikont species. However, the KLF-DBD was not present in any of the bikont genomes searched (fig. 4.1). A phylogenetically diverse range of unicellular species, including amoebas and fungi, were used in this study. Although C2H2 zinc finger proteins were identified in all unicellular amorpheans, the KLF-DBD was only found in *Fonticula* (sister to fungi), fungi (with the exceptions of *Mortierella* and *Encephalitozoon*), the

ichthyosporean *Sphaeroforma*, the filasterean *Capsaspora*, and choanoflagellates (fig. 4.1). Among unicellular taxa bona fide KLF genes were only found in *Capsaspora* and choanoflagellates, two sister groups to the Metazoa (fig. 4.1). Within the metazoans C2H2 zinc finger proteins along with the KLF-DBD were present in all species. However, neither representative ctenophore, *Mnemiopsis* or *Pleurobrachia*, show evidence for SP genes and both have fewer KLF genes as compared with other non-bilaterian metazoans. Sponge, placozoan, and cnidarian KLF/SP complements are similar to bilaterian protostome lineages and early branching deuterostome lineages. The total complement of KLF/SP genes substantially increases within the jawed vertebrates (fig. 4.1) and teleosts have an average of 33 KLF/SPs, which is slightly higher than cartilaginous fishes (22) and tetrapods (25).

Transactivation/Repression Domains

The defining characteristic of the KLF/SP family, the highly conserved KLF-DBD, is located near the C-terminal region of nearly all identified KLF/SP proteins (supplementary figs. S4, online at https://goo.gl/dbdBil). In contrast, N-terminal regions of KLF/SP proteins are generally less conserved. These more variable N-terminal regions encode a variety of transactivation/repression domains, some of which have been well characterized in mammalian model systems (Suske, et al. 2005; McConnell and Yang 2010). We generated custom perl script models for the following transactivation/repression domains to facilitate their identification across filozoan KLF/SP proteins (Supplementary Material online at https://goo.gl/dbdBil): SID repressor domain, PVDLS repressor domain, R2/R3 repressor domains, the Btd box, and the SP

box. We screened for the presence of a conserved 9aaTAD (Piskacek, et al. 2007). We also identified several nonrandom LCRs implicated in transactivation/repression that are highly biased for a particular amino acid residue (Wootton and Federhen 1993). After determining the distribution of transactivation/repression domains within KLF/SPs (fig. 4.4), we extended our search to include the full set of 48 eukaryotic genomes (Supplementary Material online at https://goo.gl/dbdBil) used in the study to better understand the representation of these domains in other proteins and to determine the extent domain shuffling may have played in the expansion of the KLF/SP gene family.

Low Complexity Regions (LCRs) are nonrandom regions of protein sequences that are highly biased for a particular amino acid residue (Wootton and Federhen 1993). LCRs are commonly found in transcription factors (Faux, et al. 2005), have been shown to influence transcriptional regulation (Gerber, et al. 1994), and are typically c-LCRs located centrally within the protein (Coletta, et al. 2010). The composition of LCRs typically found in KLF/SP proteins includes serine/threonine (S/T)-rich, glutamine (Q)-rich, and proline (P)-rich regions. S/T-rich regions are generally associated with enhanced transcriptional activation, whereas P-rich regions have been associated with transcriptional repression (Hanna-Rose and Hansen 1996). Q-rich regions, which are more frequently found among members of the SP subgroup, are known to interact with TAFII110 to activate transcription (Hoey, et al. 1993; Gill, et al. 1994). We used four different algorithms (see Materials and Methods) to identify putative LCRs present in filozoan KLF/SP protein sequences (supplementary table S2, Supplementary Material online at https://goo.gl/dbdBil). We further required a putative LCR to be detected by a minimum of two methods for annotation as an LCR. S/T-rich LCR regions occur most

frequently and are found in at least one KLF/SP family member in all filozoan taxa except the poriferan *Oscarella* (fig. 4.4). P-rich LCR regions have lower representation among KLF/SPs, and are found in all filozoans except for *Ciona*, *Daphnia*, *Tribolium*, *Lottia*, *Trichoplax*, *Oscarella*, and *Monosiga* (fig. 4.4). KLF/SPs with Q-rich LCR regions were present in all filozoans except *Ciona*, *Tribolium*, *Capitella*, *Lottia*, cnidarians, *Oscarella*, ctenophores, and *Monosiga* (fig. 4.4).

The 9aaTAD, first identified in yeast transcription factors (Piskacek, et al. 2007), is a short motif highly conserved throughout eukaryotes. The 9aaTAD has been shown to interact with TAF9 of the RNA polymerase II holoenzyme and this domain motif has been identified in many transcription factors (Piskacek 2009). Using a 9aaTAD prediction tool (Piskacek, et al. 2007), we found corresponding motifs in all filozoan KLF/SPs except *Capitella*, *Lottia*, and *Salpingoeca* (fig. 4.4).

The Btd box and SP box are conserved domains that, in combination with the KLF-DBD, characterize the SP subfamily. The Btd box was first identified in the Drosophila gene *buttonhead* (Wimmer, et al. 1993) and the associated SP box was subsequently discovered in SP5 (Harrison, et al. 2000). The Btd box motif is typically found just N-terminal of the KLF-DBD, while the SP box motif is located proximal to the N-terminus of the protein. Although not definitive, there is evidence suggesting that the Btd box is involved in transactivation (Athanikar, et al. 1997). We examined filozoan KLF/SP sequences for the presence of the Btd and SP boxes. Both Btd and SP box motifs are absent in unicellular KLFs, absent in both ctenophore KLFs, but present in all other metazoan phyla (fig. 4.4). Uniquely within the poriferans, the SP box was not found in any *Amphimedon* KLF/SP, but was identified in *Oscarella* KLF/SPs.

Using the same domain models to screen the complete set of 48 eukaryotic whole genomes, we identified Btd box motifs in a number of genes other than the KLF/SP family with the notable exception of the fungi, *Encephalitozoon* and *Saccharomyces* (Supplementary Material online at https://goo.gl/dbdBil). In contrast, the distribution of the SP box motif is highly restricted. This domain motif first appears coincident with the SP subfamily in poriferans. The SP box is present in one out of two SP genes in *Oscarella* but is not detected in the two SP genes identified in the *Amphimedon* genome. Excluding the ctenophores, metazoan genomes outside the poriferans have SP box domain motifs among a small set of genes to the exclusion of the SPs.

The Sin3a protein acts as a transcriptional repressor, and is able to recruit and bind histone deacetylases (Laherty, et al. 1997; Silverstein and Ekwall 2005). Mammalian KLF9, 10, 11, 13, 14, and 16 are known to interact with the Sin3a protein through a Sin3a interacting domain (SID) which binds the PAH domain of the Sin3a protein (Imataka, et al. 1992; Blok, et al. 1995; Cook, et al. 1998; Song, et al. 1999; Kaczynski, et al. 2002). No SID motif was detected in unicellular KLF/SP sequences. Within metazoans a SID-containing KLF was identified in all taxa except ctenophores and the protostomes *Capitella*, *Lottia*, and *Drosophila* (fig. 4.4). To gain further insight into the evolutionary history of the SID motif, we searched the complete set of 48 whole genomes and identified a conserved SID in a number of genes in all representative eukaryote genomes (Supplementary Material online at https://goo.gl/dbdBil).

KLF10 and KLF11 proteins contain R2 and R3 repressor domains in combination with an SID motif (Cook, et al. 1999). We searched filozoan KLF/SP complements and were only able to identify R2 and R3 domains in representative vertebrate taxa (fig. 4.4).

We then searched the complete set of 48 whole genomes for the presence of the R2 and R3 domains. Our search revealed the presence of an R3 domain in all representative eukaryotic genomes, while the R2 domain was restricted to vertebrate KLF10/11 genes (Supplementary Material online at https://goo.gl/dbdBil).

KLF3, KLF8, and KLF12 can co-repress transcription through interaction with the C-terminal binding protein mediated by the PVDLS domain (Crossley, et al. 1996; Turner and Crossley 1998; Imhof, et al. 1999; van Vliet, et al. 2000). A PVDLS domain was identified in at least one KLF/SP in all jawed vertebrates and in *Drosophila* (fig. 4.4). To help resolve this relationship between the fly PVDLS containing KLF and the vertebrate KLF3/8/12 genes, we searched additional protostome genomes for the presence of KLF genes containing the PVDLS motif. Curiously, the association of this motif with KLF genes was not detected in other *Drosophilid* species, but was identified in two hymenopterans *Apis millifera* and *Nasonia vitripennis* (data not shown). Analysis of the complete set of 48 whole genomes revealed the PVDLS domain in multiple genes in all representative eukaryote species (Supplementary Material online at https://goo.gl/dbdBil).

Co-occurrence Networks

Co-occurrence networks are visual representations of domain pair occurrences within a given protein or protein family. Typically, domains are represented as circles, and lines connect domains that appear together within a protein. These maps can be useful for visualizing the frequencies of certain domains occurring with each other, that is, domain architectures. To explore differences in domain architecture complexity, we

generated domain co-occurrence maps for different KLF/SP domain networks (fig. 4.5). These maps can also show the general N-terminal to C-terminal relationship between different domains (fig. 4.5A). Our co-occurrence maps indicate the frequency of unique domain pairs as well as how often an individual domain appears within a given network in extant species (fig. 4.5B–I). Composite networks of larger taxonomic groupings represent the consensus map of all included species. The composite networks are additive and thus do not compensate for missing data due to poor genome annotations. Importantly, these co-occurrence maps are not intended to be ancestral reconstructions. They are used to show observed domain relationships in, or among groupings of, extant species. Nonetheless, evolutionary inferences can be drawn from these network maps when combined with inferences of ancestral presence or absence of discrete transactivation/repression domains as represented in figure 7.

The unicellular KLF/SP network (*Capsaspora* + choanoflagellates) is one of the least complex with only LCR domains linked to the KLF-DBD (fig. 4.5C). Within metazoans, there appears to be a gradient of network complexity; the non-bilaterian network (fig. 4.5D) is less complex than the invertebrate bilaterian network (fig. 4.5E), and both of these are less complex than the vertebrate network (fig. 4.5F). A small number of lineage-specific networks showed a significant departure from larger composite networks. For example, the Ctenophora network (fig. 4.5G), in contrast to both the Poriferan network (fig. 4.5H) and larger non-bilaterian composite network (fig. 4.5D), is composed of only S/T-rich and P-rich LCRs, with the P-rich LCRs + KLF-DBD domain pair occurring at greater frequency. The Poriferan network (fig. 4.5H) bears substantially greater similarity to the more inclusive non-bilaterian composite network

(fig. 4.5D). The urochordate *Ciona* also presents an interesting departure from composite networks (fig. 4.5I), diverging dramatically from the invertebrate bilaterians in having S/T-rich LCRs as the most prevalent domain linked to the KLF-DBD. The relationship of Ctenophora and Porifera to other metazoans shown in figure 5J and 5K are based on recent hypotheses of metazoan phylogeny (Dunn, et al. 2008; Ryan, et al. 2013).

KLF Domain Architecture and Phylogenetics

All KLF/SP gene family members share homology at the conserved KLF-DBD, typically located toward the C-terminus. The N-terminal regions share very little similarity across the entire family. Distinct KLF/SP subgroups can, however, be defined based on comparable structure and function (McConnell and Yang 2010). These subgroups share domain architectures defined by unique combinations of transactivation/repression domains occurring with the KLF-DBD (fig. 4.6; supplementary table S3, Supplementary Material online at https://goo.gl/dbdBil). Domain architectures have been described for mammalian KLF/SP genes (Kaczynski, et al. 2003; Suske, et al. 2005; McConnell and Yang 2010; Archer 2011). Historically, KLFs have been divided into five domain architectures and named according to the genes in each class: KLF1/2/4, KLF6/7, KLF3/8/12, KLF9/13, and KLF10/11. Similarly, the SPs have been divided into two architecture classes: SP1–4 and SP5–9. Using this existing classification scheme as a foundation, we determined the distribution of domain architectures for all filozoan KLF/SP genes. The 9aaTAD was not used in determining domain architecture as it does not contribute to any unique domain combination.

We performed ML and Bayesian phylogenetic analyses on an alignment of non-filozoan amorphean and filozoan concatenated sequences that included the KLF-DBD and transactivation domains (supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil). The overall topologies of the ML and Bayesian trees share significant overlap, albeit with only moderate support for deeper nodes, for example, BPP = 0.84 (fig. 4.3). A number of informative clades with congruence across both ML and Bayesian analyses were recovered. We primarily focused on overlapping nodes with BPP $\geq$ 0.90. Most of these nodes are consistent with explicit domain architectures and well-accepted inferences of metazoan phylogeny. Many of the well-supported orthologous sequence clades are composed of vertebrates with corresponding nonvertebrate putative orthologs diverging prior to the highly supported vertebrate nodes (fig. 4.3; supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil).

Among the unicellular KLF representatives, the *Capsaspora* and *Salpingoeca* domain architectures consist of alternating S/T- and Q-rich LCRs (fig. 4.6) with scattered P-rich LCRs at low frequency (fig. 4.5C). Although the S/T- and Q-rich LCR organization superficially resembles the architecture found among SP1–4 genes, these unicellular KLFs lack the characteristic Btd box and SP box motifs that define the SPs. Therefore, we created a unicellular specific KLF architecture class (fig. 4.6). Although a single unicellular KLF clade composed of *CowKLFa*, *SroKLFa*, and *MbrKLFa* branch sister to a highly supported cnidarian + bilaterian KLF15 clade (BPP = 0.99), the remaining unicellular KLF sequences are divergent and characterized by long branches

(fig. 4.3; supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil).

The KLF6/7 and KLF1/2/4 classes are defined by the presence of mostly S/T-rich or P-rich LCRs, respectively. KLF6/7 domain architectures were identified in *Monosiga* and all metazoans except for *Oscarella* and *Lottia* (fig. 4.6). The single *Monosiga* KLF gene superficially consists of a KLF6/7 domain architecture; however, phylogenetically this sequence groups with other unicellular KLFs. *Amphimedon* sequences with KLF6/7 architecture grouped more closely with KLF1/2/4 and KLF5, albeit with low statistical support. Our phylogenetic analysis recovered a well-supported clade (BPP = 0.91) that includes bilaterian KLF6/7 along with representative cnidarian and *Trichoplax* KLF6/7. The KLF1/2/4 class was identified in all metazoans except *Oscarella*, *Trichoplax*, *Lottia*, *Drosophila*, *Daphnia*, and *Ciona* (fig. 4.6). Our phylogenetic analyses recovered a KLF 1/2/4 clade including invertebrate bilaterians (BPP = 0.96), a teleost-specific KLF1 clade (BPP = 1), and a mammalian KLF1 clade (BPP = 1; fig. 4.3).

The KLF9/13 class possesses a SID near the N-terminal region and typically contains one or more c-LCRs located between the SID and KLF-DBD. The related KLF10/11 class shares the SID with KLF9/13, but uniquely contains the R2 and R3 repressor domains situated between the SID and the KLF-DBD. The R2 and R3 repressor domains are generally flanked by one or more P-rich or S/T rich c-LCRs. Based on phylogenetic analyses, all metazoans, with the notable exception of ctenophores, have at least one KLF9/13 gene (supplementary figs. S2–S4, Supplementary Material online at https://goo.gl/dbdBil). Mammalian KLF14 and KLF16 form a unique, highly supported (BPP = 1) clade within the 9/13 group supporting the assignment of these orthologs to a

mammalian-specific lineage (fig. 4.3; supplementary fig. S5, Supplementary Material online at https://goo.gl/dbdBil). The KLF10/11 class has a more restricted distribution and is exclusively found in vertebrates (fig. 4.6). All SID containing KLF sequences form a clade with low support (BPP = 0.75; supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil).

Finally, the KLF3/8/12 class is characterized by a PVDLS domain and often has additional c-LCRs between the PVDLS and the KLF-DBD. According to domain architecture, this class is found in all the jawed vertebrates and in the Endopterygota as represented in this study by *Drosophila* (fig. 4.6). Our phylogenetic analyses recovered a KLF3/8/12 clade with high support (BPP = 0.95; fig. 4.3). Notably, the *Drosophila* KLF sequence that contains a PVDLS domain (CG42741) falls outside of the KLF3/8/12 clade which consists exclusively of vertebrate sequences. Within this clade the KLF8 orthologs are recovered with BPP = 1. Despite poor annotation in the lamprey genome, we identified a lamprey sequence within the vertebrate KLF3/8/12 clade (supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil).

The SP subfamily is defined by the presence of a Btd box 5′ proximal to the KLF-DBD and generally possesses an additional SP box located near the N-terminus. The SP factors can be further separated into two domain architecture classes, SP1–4 and SP5–9 (Bouwman and Philipsen 2002; Archer 2011). The SP1–4 class typically has S/T-rich c-LCRs adjacent to Q-rich c-LCRs. The SP5–9 class is characterized by having either predominately S/T, or with lesser frequency, P-rich c-LCRs. Notably, the SP subfamily is absent in both ctenophore genomes. The SP1–4 domain architecture is found in all metazoans except *Oscarella*, cnidarians, *Capitella*, *Lottia*, *Tribolium*, and *Ciona*. The

SP5–9 domain architecture is found in all metazoan genomes excluding ctenophores, *Drosophila*, and *Daphnia* (fig. 4.6). Phylogenetic analyses recovered moderate support (BPP = 0.84) for a metazoan SP clade (fig. 4.3; supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil). Well-supported SP clades (BPP ≥ 0.9) include vertebrate SP2, SP6, and SP7.

Discussion

KLF/SP Gene Family Origins in Eukarya

The evolution of the KLF/SP family and associated transactivation/repression domains is represented in figure 7. Although C2H2 zinc finger domains are ubiquitous across a wide range of eukaryotes (de Mendoza, et al. 2013), the KLF-DBD first appears in the Holomycota (figs. 1 and 7). Historically, the single criteria for defining the KLF/SP gene family has been the presence of a highly conserved KLF-DBD composed of three C2H2 zinc fingers each separated by a seven amino acid linker region (McConnell and Yang 2010). The first 2 zinc fingers are 23 amino acids long (from C to H), while the third zinc finger is only 21 amino acids long. Our analyses reveal that all filozoan KLF/SP sequences recovered except one, the highly divergent MleKLFX ctenophore sequence, contains a canonical aspartic acid residue in the second zinc finger, D44, which is critical for proper DNA binding (fig. 4.2; Feng, et al. 1994; Schuetz, et al. 2011). All non-filozoan KLF-DBD containing genes lacked this diagnostic residue and possess additional zinc fingers. Furthermore, none of the transactivation/repression domains considered in this study were found in the non-filozoan KLF-DBD containing sequences, despite several relevant domains being well represented in all eukaryotic genomes

examined (fig. 4.7, Supplementary Material online at https://goo.gl/dbdBil). Thus the non-filozoan KLF-DBD sequences were not classified as bona fide KLF/SP orthologs. Moreover, our phylogenetic analyses recovered a moderately supported filozoan clade that possesses the canonical D44 residue to the exclusion of all non-filozoan sequences and the highly divergent MleKLFX ctenophore sequence (fig. 4.3). We operationally define a bona fide KLF/SP gene as containing only 3 zinc fingers conforming to 81 amino acid, 3 zinc finger KLF-DBD motif typically found in the C-terminal region of the parent sequence and containing a canonical aspartic acid residue at position 44, D44, following the general consensus sequence C-X4-C-X12-H-X3-H-X7-C-X4-C-X7-D-X4-H-X3-H-X7-C-X2-C-X12-H-X3-H.

Therefore, the KLF-DBD likely has its origins in the opisthokont stem lineage prior to the divergence of the Holomycota (figs. 1 and 7). In our study, KLF genes first appear in the filasterian *Capsaspora*, while the SP subfamily is restricted to metazoans, notably excluding the ctenophores. Thus we infer the origin of KLF genes in the filozoan stem lineage and a later origin of the SP genes after the divergence of the metazoan lineage (figs. 1 and 7). Our phylogeny assumes an early divergence of the ctenophore lineage from the metazoan stem (Dunn et al. 2008; Ryan et al. 2013). An alternate view of early animal phylogeny (Philippe et al. 2009; Nosenko et al. 2013) would infer the origin of SP genes prior to the divergence of the poriferans and assume a subsequent loss of SP orthologs in the ctenophore lineage. Under either scenario of early metazoan lineage divergence, the appearance of KLF genes precedes the origin of the SP gene subfamily.

Apart from the KLF-DBD, the most common sequence features are the presence of one or more S/T-, P-, or Q-rich c-LCRs (figs. 4-6; supplementary fig. S4, Supplementary Material online at https://goo.gl/dbdBil). Our results highlight promiscuous variation in LCR composition and length between KLF/SP gene family members (supplementary fig. S4, Supplementary Material online at https://goo.gl/dbdBil). In contrast to earlier studies with more restricted sampling, we are unable to discriminate phylogenetic relationships between KLF/SP architectural classes composed of explicit c-LCR compositional types linked to the KLF-DBD. For example, many representative genes nested within the KLF1/2/4 and KLF6/7 clades include all three predominant c-LCR compositional types (figs. 3 and 6; supplementary fig. S4, Supplementary Material online at https://goo.gl/dbdBil). Variability, expansion, and extinction of LCRs have been associated with gene conversion due to mismatch repair of DNA heteroduplexes (Radding 1982) and higher rates of recombination relative to flanking sequences due to unequal crossing over and replication slippage (DePristo, et al. 2006). Both of these core genetic mechanisms would contribute to our observations of significant LCR homoplasy within and between orthologous domain architectures.

Our analyses of the frequency of appearance of explicit compositionally biased c-LCRs and their frequency of co-occurrence with flanking transactivation domains suggest functional consequences. For example, S/T- and Q-rich LCRs are typically acidic and associated with transcriptional activation, whereas P-rich regions, with their bulky side chains, are typically neutral and associated with context-dependent repression. The high frequency of S/T-rich c-LCRs suggests an ancestral KLF that was likely involved in transcriptional activation. Despite high rates of gene conversion and unequal crossover

associated with LCRs, we observe P-rich c-LCRs at low frequency. Conversely, among the KLF/SP genes that retain P-rich c-LCRs, our analyses also show an increase in the frequency of connectivity with other repressor motifs in metazoans (fig. 4.5). This suggests the evolution of coordinated multidomain repression during the expansion of the KLF/SP gene family. For example, known repressor domains such as the SID and PVDLS motifs are typically located 5′ of c-LCRs and show a similar pattern of increased frequency during metazoan lineage diversification (figs. 5 and 7). This pattern suggests that within the KLF/SP gene family, despite a selective preference for acidic c-LCRs as evidenced by their high frequency, the evolution of a repertoire of transcriptional repressor combinations through the differential pairing of c-LCRs with repressor motifs occurred. Our analyses further suggest that these shuffling events may be associated with cell type diversification in metazoans. For example, chordates possess >100 distinct cell types, whereas earlier diverging metazoan lineages range from 4 to 59 distinct cell types (Chen, et al. 2014). This pattern mirrors the increased frequency of repressor domain connectivity observed across members of the KLF/SP gene family in metazoans (figs. 5-7).

The total complement of KLF/SP genes across non-bilaterian metazoans, protostomes, and invertebrate deuterostomes shows only a ~2-fold variance of 6–12 genes. Vertebrates have an average number of 24 KLF/SPs, likely due to the 2 rounds of whole genome duplication (WGD) in the vertebrate stem lineage (supplementary fig. S5, Supplementary Material online at https://goo.gl/dbdBil). Despite the additional round of WGD in the teleosts, their KLF/SP gene complement is not substantially greater than other jawed vertebrates. It seems that KLF/SP gene duplicates were not necessarily

preferentially retained, which has been shown to be the case with other transcription factors (de Mendoza, et al. 2013). The lamprey *Petromyzon marinus* has an atypically depauperate KLF/SP complement as compared with other vertebrates. It has been shown that during embryogenesis lampreys can lose ~20% of germline DNA in somatic tissues due to genomic rearrangements (Smith, et al. 2009; Smith, et al. 2012). The lamprey genome was derived completely from somatic tissue, thus the true *P. marinus* KLF/SP complement may be underrepresented in our analyses (Smith, et al. 2013).

Ctenophores have a reduced KLF/SP complement compared with other nonbilaterian metazoans and lack members of the SP subfamily altogether (fig. 4.1). This pattern of gene underrepresentation in ctenophores relative to other metazoans has been observed in a number of studies (Ryan, et al. 2010; Maxwell, et al. 2012; Sebé-Pedrós, et al. 2013; Moroz, et al. 2014). Among the filozoan KLF orthologs, a single member, the highly divergent ctenophore gene MleKLFX, possesses atypical domain architecture in which a KLF-DBD lacking the canonical D44 residue is located in the N-terminal region of the protein instead of the C-terminal region (supplementary fig. S4, Supplementary Material online at https://goo.gl/dbdBil). This gene consistently groups with the non-filozoan amorphean KLF-DBD sequences presumably due to the effects of long-branch attraction (fig. 4.3; supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil).

The Btd box is found in SPs and along with the KLF-DBD represents the defining characteristic of the SP subfamily. The Btd box was initially described in the Drosophila *btd* gene which has high similarity to mouse SP1 and SP3 (Wimmer et al. 1993). Most of the amino acid sequence conservation resides in the zinc fingers, and *btd* is often

classified as an SP homolog in flies. However, the second zinc finger of the btd gene is slightly different than the corresponding zinc finger in the KLF-DBD. The second zinc finger of *btd* is only 21 amino acids long, lacking 2 residues between the 2 C residues. According to this structural difference, *btd* would be classified in the Zif268/EGR-1 family of zinc finger transcription factors (Iuchi 2001). However, phylogenetic analysis of *btd* shows that it is more similar to SPs than to EGR-1 family genes and btd-like orthologs were not found in other organisms (data not shown). These suggest that at some point along the stem lineage leading to *Drosophilia*, there was a lineage-specific deletion of 2 amino acid residues in the second zinc finger of *btd*. Notably, the loss of these two residues does not hinder *btd* from binding transcription factor binding sites similar to SP factors (Wimmer et al. 1993).

Recently, three SPs in the cnidarian *Nematostella* and the placozoan *Trichoplax* were identified and grouped into three separate clades: SP1–4, SP5/btd, and SP6-9 (Schaeper, et al. 2010) leading to a proposed ancestral complement of three SPs in the Metazoa. In our study, we recovered two SPs in sponges, two SP genes in *Trichoplax*, and four SP genes in *Nematostella*. The *Trichoplax* btd gene and SP6–9 gene from Schaeper et al. (2010) match with two *Trichoplax* SPs from our study (supplementary table S1 and supplementary fig. S4, Supplementary Material online at https://goo.gl/dbdBil). However, the *Trichoplax* SP1–4 gene from Schaeper et al. (2010) matches with the *Trichoplax* KLF9/13 gene in our study. We were able to extend this *Trichoplax* sequence and identified a SID motif in the previously unannotated 5′ end of this gene model thus confirming the placement of this sequence within the KLF9/13 group (supplementary fig. S4, Supplementary Material online at https://goo.gl/dbdBil).

Therefore, our analyses suggest an ancestral complement of a single SP1–4 ortholog and a single SP5–9 ortholog in the early metazoan stem lineage followed by an expansion after the divergence of the Placozoa (fig. 4.1).

Our analyses also suggest that two KLF9/13 genes were present early in Metazoa (figs. 6 and 7; supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil). In mammals, two additional genes with similar domain architecture, KLF14 and KLF16, were present and a recent study provided evidence that KLF14 evolved from KLF16 through a retrotransposon event (RTE) within the mammalian stem lineage (Parker-Katiraee, et al. 2007). Thus, we infer that KLF16 arose by a tandem duplication of one of the two ancestral KLF9/13 genes in the mammalian stem lineage followed by KLF14 evolution by RTE of KLF16 (supplementary fig. S5, Supplementary Material online at https://goo.gl/dbdBil).

Transactivation/Repression Domains Show Unique Evolutionary History

Our results highlight that individual transactivation/repression domains associated with KLF/SP transcription factors have unique evolutionary histories (fig. 4.7). The Btd box, PVDLS, R3, and SID are ancient domain motifs present in other genes in all representative eukaryote genomes in this study. The notable exceptions being the *Saccharomyces* and *Encephalitozoon* fungal genomes in which no Btd box domains were detected. These four ancient domains first appear in differential combinations with the KLF-DBD in KLF/SP genes after the divergence of the metazoans and correlate strongly with the observed expansion of domain architecture repertoires (fig. 4.6). Given the phylogenetic distribution of domain architectures associated with the putative acquisition

of ancient motifs by domain shuffling, we infer that particular domain architectures stem from unique independent ancestral shuffling events along the metazoan stem lineage (figs. 4.6 and 4.7).

Our analyses also infer two instances of *de novo* domain origin within the KLF/SP gene family (fig. 4.7). The SP subfamily appears early in metazoan evolution in the Poriferan lineage (fig. 4.1) coincident with the appearance of the SP box motif (fig. 4.4). However, in contrast to the Btd box motif, the SP box motif has a very limited genomic and phylogenetic distribution. In sponges, the SP box motif is uniquely associated with the SP genes. In later diverging lineages of metazoans, the SP box is associated with a small number of genes in addition to the KLF/SP family. Thus the SP genes provide an example of a metazoan-specific multidomain protein that consists of both ancient domains, including the Btd box and the KLF-DBD, coupled with the *de novo* origin of a metazoan-specific domain motif, the SP box. Another example of *de novo* domain motif origin within the KLF/SP family is the vertebrate-specific KLF10/11 genes (fig. 4.7; supplementary fig. S5, Supplementary Material online at https://goo.gl/dbdBil). This architectural class is composed of ancient SID, R3, and KLF-DBD domains combined with a vertebrate-specific R2 domain. Our exhaustive search uncovered no R2 domains in any invertebrate genome or outside of the KLF10/11 genes within the vertebrate genomes in this study. Interestingly, the R3 domain is present in all eukaryote genomes but there is no evidence for it being shuffled into KLFs until much later than the SID motif (fig. 4.7).

The PVDLS domain represents an intriguing case of putative convergence. In our analyses, the PVDLS domain appears in vertebrate KLFs defining the 3/8/12 architecture

group and in a lone *Drosophila* KLF (CG42741). Our phylogenetic analyses suggest, however, that the fly sequence is not nested within the highly supported KLF3/8/12 clade (supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil). To help elucidate the incongruous relationship between the representative fly gene containing a PVDLS motif and the PVDLS containing vertebrate KLF3/8/12 class, we searched a number of additional protostome genomes for the presence of KLF genes containing the PVDLS motif (data not shown). Our search yielded only two other instances, both within the hexapods, of a KLF gene also containing a PVDLS motif: the hymenopterans *A. mellifera* and *N. vitripennis*. Exhaustive searches of 12 other *Drosophilid* species did not uncover any PVDLS containing KLFs. Based on these additional results, we infer that a putative ancestral KLF3/8/12 gene most likely evolved early in the vertebrate stem lineage, whereas the PVDLS motif was likely also convergently acquired in the hexapod lineage leading to the Endopterygota.

Our analysis across 48 eukaryotic genomes illuminates the origin and evolutionary history of the KLF/SP gene family. We also identify and characterize associated transactivation/repression domains, including LCRs, enabling us to develop models of KLF/SP domain co-occurrence evolution. By extending our domain search to include entire proteomes, we find evidence for a complex intersection of domain shuffling, gene duplication, and *de novo* domain evolution as the primary mechanisms for the diversification of the KLF/SP gene family across the Metazoa. Our results uncover a pattern of an increased frequency of repressive domain connectivity repertoires (P-rich LCRs, SID, R2/R3, and PVDLS domains) in the KLF/SP gene family among metazoans suggesting a role in mediating diverse transcriptional repression activity. Our

phylogenetic results further suggest that the expansion of the KLF/SP gene family mirrors increased cell type diversity during metazoan lineage diversification. The expansion and diversification of the KLF/SP gene family within the Metazoa may thus reflect the accumulation of differential transcriptional repression strategies associated with the development of extensive repertoires of cell types required to support complex tissues.

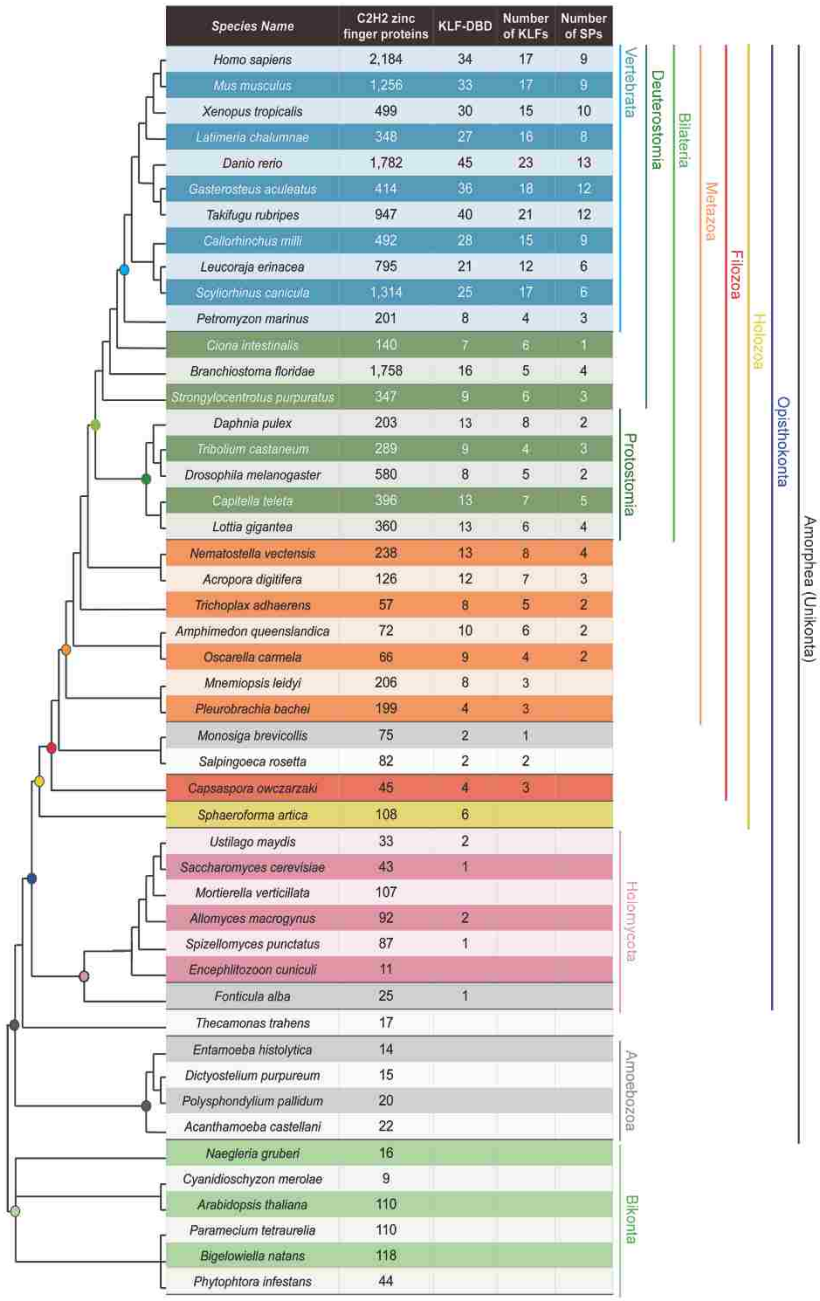
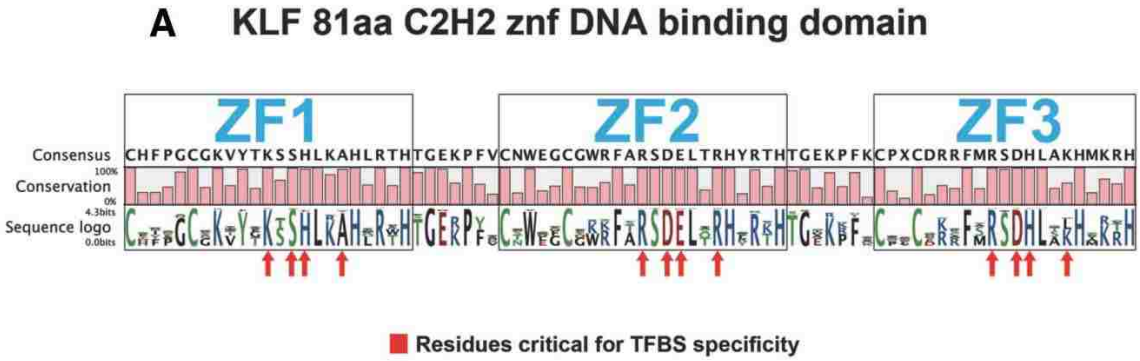


Figure 4.1. Distribution of C2H2 zinc finger proteins, KLF-DBD containing proteins, and KLF/SP proteins in representative Eukarya taxa. Rows indicate representative genomes searched. Columns indicate the total number of protein sequences that contain at least one C2H2 zinc finger using the Pfam PF00096 HMM model, the total number of protein sequences that contain the archetypical KLF-DBD, the total number of bona fide KLF sequences recovered, and the total number of SP sequences recovered. Phylogeny is based on Adl et al. (2012), Derelle and Lang (2012), Dunn et al. (2008), Ryan et al. (2013), and Sebé-Pedrós et al. (2013).



B Stringent KLF-DBD model

C-X_4-C-X_{12}-H-X_3-H-X_7-C-X_4-C-X_7-**D**-X_4-H-X_3-H-X_7-C-X_2-C-X_{12}-H-X_3-H

Figure 4.2. KLF DNA binding domain (KLF-DBD). A) Sequence logo from alignment of filozoan KLF-DBDs. Schematic represents the consensus sequence of the highly conserved 81 amino acid KLF-DBD from the identified filozoan KLF/SP sequences in this study. The height of the red bars indicates percent conservation of each residue, with the tallest bar representing 100%. The height of individual lettered residues in the sequence logo reflects their conservation. Each zinc finger is indicated with blue text and bounded with a black box. The red arrows point to amino acid residues critical for mediating sequence specific DNA binding (Feng, et al. 1994; Schuetz, et al. 2011). B) Stringent KLF-DBD consensus sequence model which includes the canonical D44 aspartic acid.

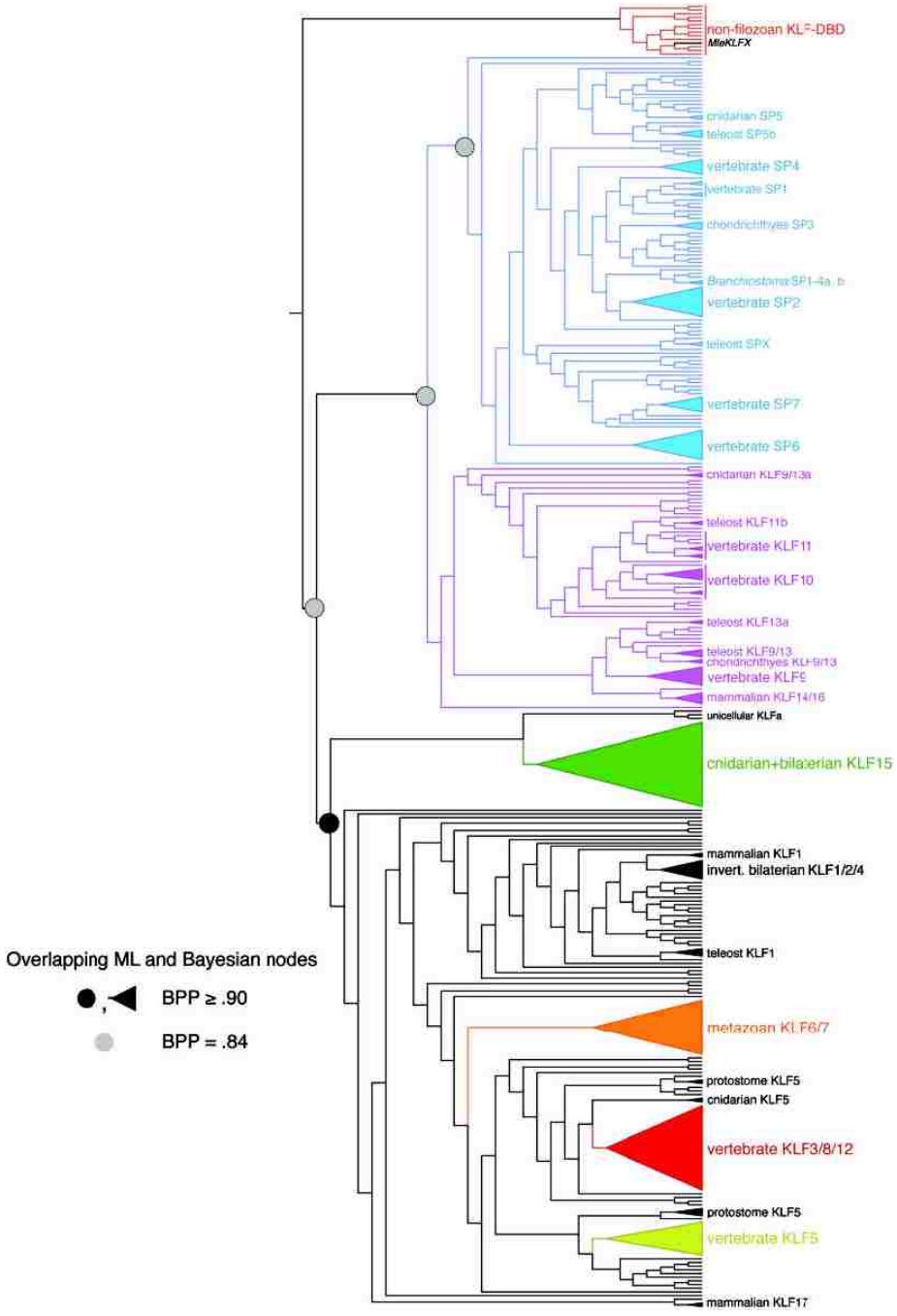


Figure 4.3. Combined gene tree estimates for the concatenated KLF/SP data set using Bayesian criterion (MrBayes) and ML criterion (RAxML). Gray node labels indicate congruent topology with BPP support = 84%. Black node labels indicate congruent topology with BPP support ≥90%. Clades collapsed to triangles indicate congruent topologies with BPP support ≥90%. The single highly divergent ctenophore MleKLFX sequence clusters with non-filozoan KLF-DBD presumably due to long-branch attraction. Bayesian and ML trees with support values and branch lengths are available in supplementary figs. S2 and S3, Supplementary Material online at https://goo.gl/dbdBil.

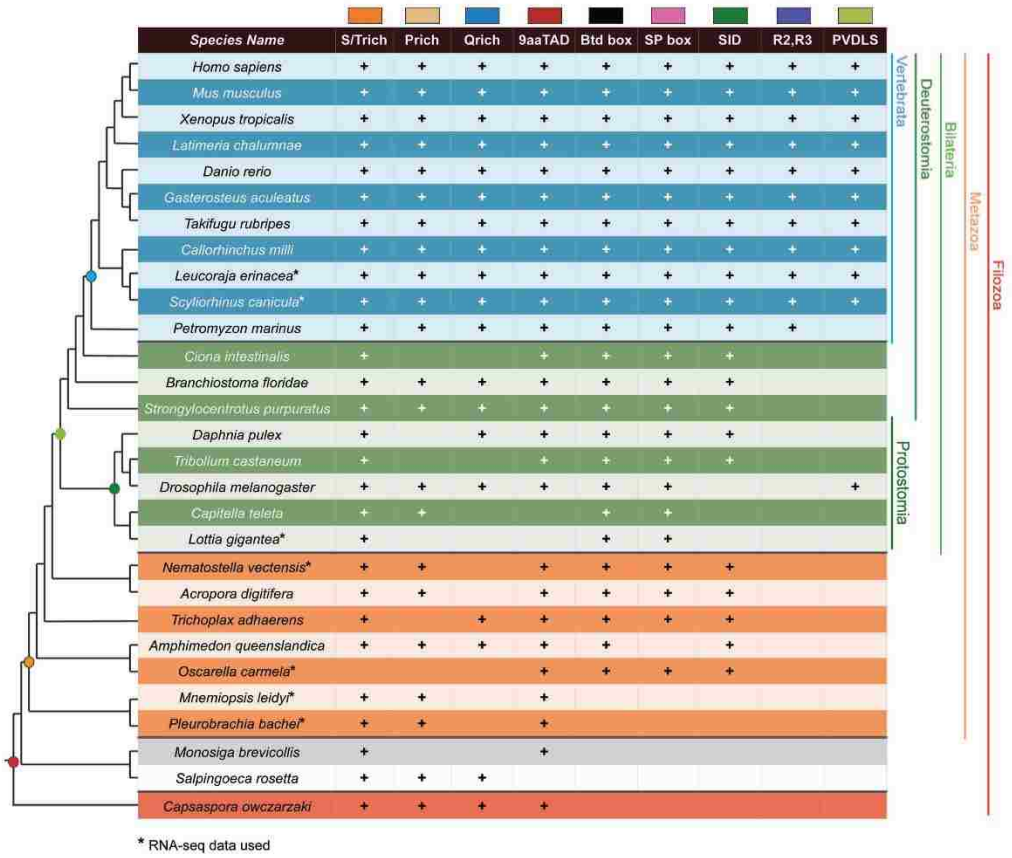
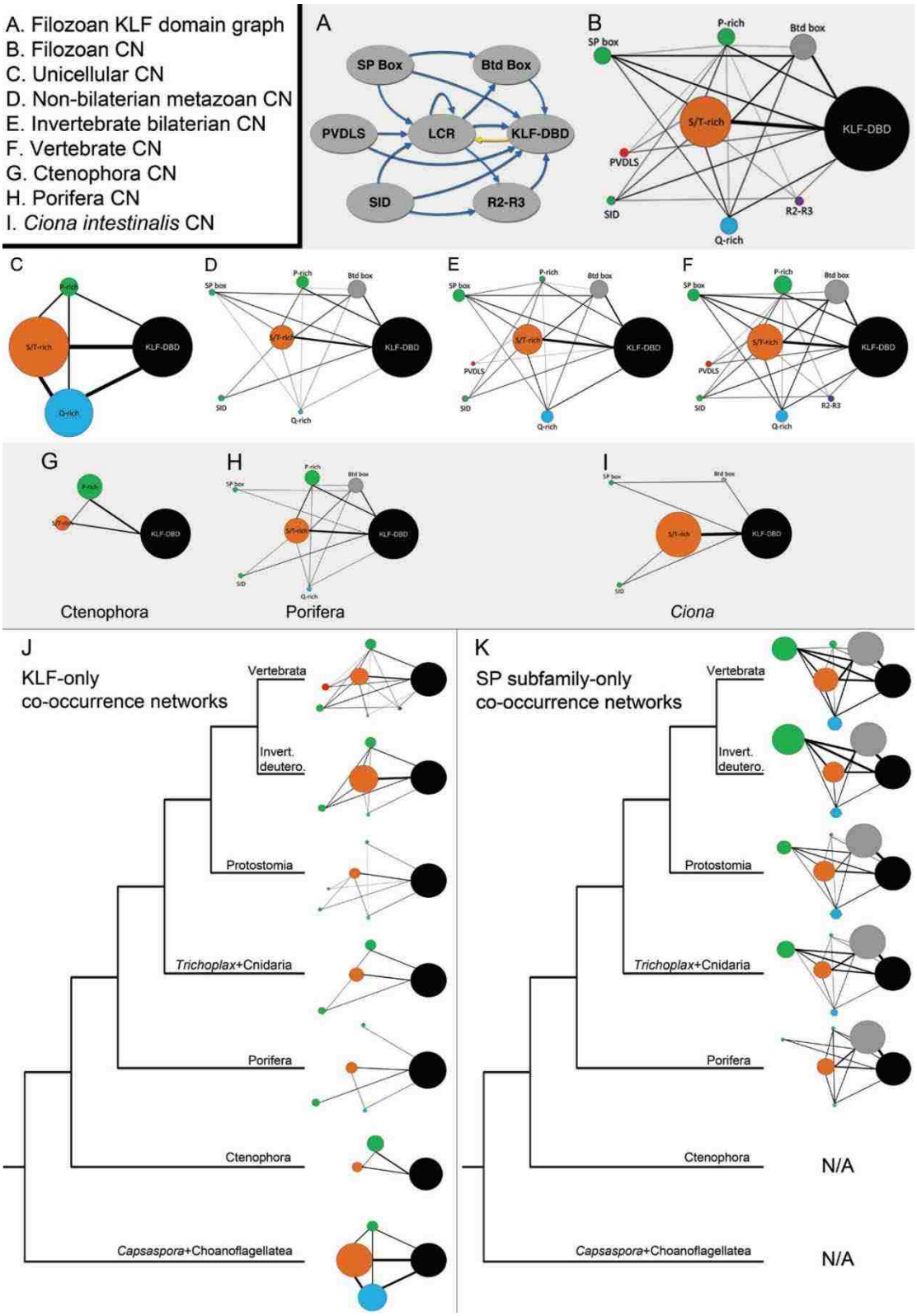


Figure 4.4. Phylogenetic distribution of transactivation/repression domains and LCRs associated with KLF/SP proteins. The + indicates the presence of the corresponding domain or LCR in at least one KLF/SP protein in the indicated taxa. Only filozoan lineages containing bona fide KLF/SP proteins are shown. An asterix indicates that RNA-seq data were used for that species. Phylogeny is based on Dunn et al. (2008), Ryan et al. (2013), and Sebé-Pedrós et al. (2013).

Figure 4.5. KLF/SP protein domain co-occurrence networks. In all networks, each circle represents a transactivation/repression domain or an LCR. A line connecting two domains indicates a co-occurrence of those two domains. Domains are arranged in approximately the same 5′–3′ spatial orientation as they appear encoded in KLF/SP sequences. (A) General network diagram showing connectivity and unidirectional spatial relationships between transactivation domains among filozoan KLF/SPs. Blue arrows represent connectivity upstream of the KLF-DBD; the gold arrow represents connectivity downstream of the KLF-DBD. (B–I) KLF/SP co-occurrence networks from different taxonomic groups. Circle size indicates the relative frequency of occurrence in the network, with the KLF-DBD always representing 100%. Circle color follows the same convention as seen in figure 4.4. Repeated domains were counted as occurring only once. Lines connecting circles indicate the presence of that specific domain pair co-occurrence in at least one KLF/SP. Line width indicates the frequency of domain pair co-occurrence. Only LCR domains which are found N-terminal of the KLF-DBD are represented in these networks (supplementary fig. S4, Supplementary Material online at https://goo.gl/dbdBil). (B) Complete filozoan KLF/SP network. (C) Representative unicellular KLF/SP network. (D) KLF/SP network from nonbilaterian metazoans. (E) Invertebrate bilaterian KLF/SP network. (F) Vertebrate KLF/SP network. (G-H) Representative Ctenophora and poriferan KLF/SP networks for comparison with each other and with the network in D. (I) *Ciona* KLF/SP network for comparison with the networks in E and F. (J, K) Co-occurrence network maps for the KLF subfamily and SP subfamily mapped onto the filozoan phylogeny (Dunn et al. 2008; Ryan et al. 2013) for evolutionary comparison. Each network represents a composite for the taxonomic group indicated. (J) Co-occurrence maps for domains found in the KLF subfamily. (K) Co-occurrence maps for domains found in the SP subfamily. The unicellular filozoan genomes and ctenophore genomes do not contain SP genes.


A. Filozoan KLF domain graph
B. Filozoan CN
C. Unicellular CN
D. Non-bilaterian metazoan CN
E. Invertebrate bilaterian CN
F. Vertebrate CN
G. Ctenophora CN
H. Porifera CN
I. Ciona intestinalis CN
A
SP Box
Btd Box
PVDLS
LCR
KLF-DBD
SID
R2-R3
B
P-rich
Btd box
SP box
S/T-rich
KLF-DBD
PVDLS
SID
R2-R3
Q-rich
C
D
E
F
G
H
I
Ctenophora
Porifera
Ciona
J
KLF-only
co-occurrence networks
Vertebrata
Invert. deutero.
Protostomia
Trichoplax+Cnidaria
Porifera
Ctenophora
Capsaspora+Choanoflagellatea
K
SP subfamily-only
co-occurrence networks
Vertebrata
Invert. deutero.
Protostomia
Trichoplax+Cnidaria
Porifera
Ctenophora
N/A
Capsaspora+Choanoflagellatea
N/A

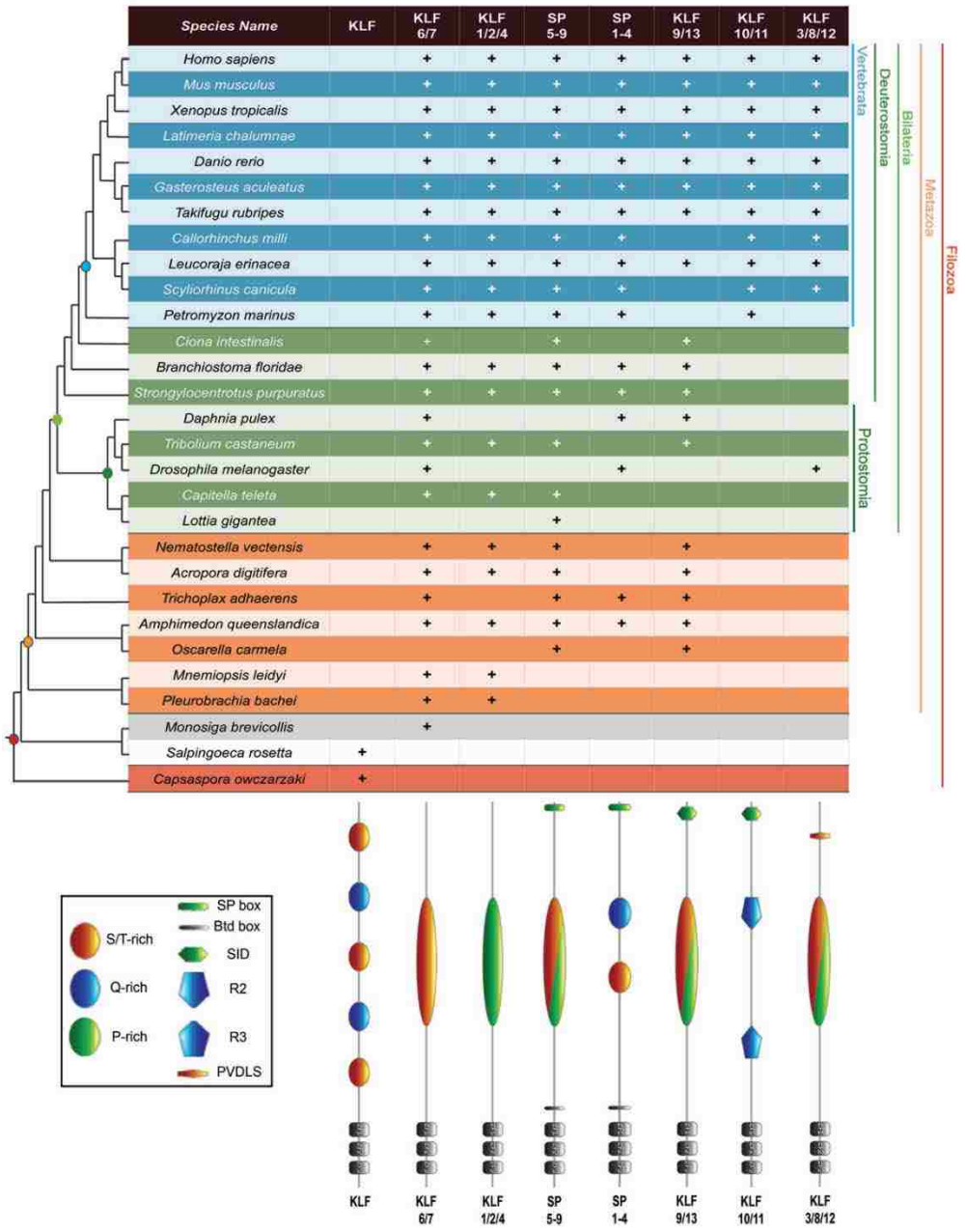


Figure 4.6. Phylogenetic distribution of explicit domain architectures represented among KLF/SP proteins. The key at lower left identifies LCRs and transactivation/repression domains used to determine domain architectures. The protein schematics along lower right represent the particular combinations of domains and LCRs with the KLF-DBD that define each specific KLF/SP protein architecture. All groups, except for the ancient unicellular KLF architecture recovered, are named according to established human KLF/SP paralogy groups that conform to each specific architecture. The three C-terminal zinc fingers of the KLF-DBD are indicated with grey boxes labeled zf1, zf2, and zf3. Architecture schematics are not to scale.

Figure 4.7. Inferred relationships between key events during the evolution and expansion of the KLF/SP gene family. Symbol key is at upper left. Colored rectangles represent the origin of particular transactivation/repression domains or LCRs co-occurring with the KLF-DBD (fig. 4.5). Yellow hexagons represent the origin of specific KLF/SP domain architectures (fig. 4.6). A black X over a hexagon represents the loss of specific domain architecture. Colored triangles represent the presence of specific transactivation domain motifs within whole eukaryote genomes to the exclusion of the KLF/SP gene family. (A) We infer the origin of the KLF-DBD in the opisthokont stem lineage prior to the divergence of the Holomycota. However, bona fide KLF gene architectures do not appear until the divergence of the filozoan lineage (KLF origin). The ancient unicellular KLF domain architecture is not recovered in metazoan lineages. The ancient PVDLS, SID, Btd box, and R3 domains were recovered, to the exclusion of KLF/SPs, in all eukaryote genomes searched. Notably, the Btd box was not recovered in *Saccharomyces* and *Encephalitozoon* fungal genomes. Our analysis suggests that the origin of the SP subfamily is in the metazoan stem lineage prior to the divergence of the poriferans; it is not present in Ctenophora. The SP box motif only appears in SP genes in poriferans and is not found in additional genes until the divergence of *Trichoplax*. The R2 repressor domain appears to be a de novo innovation restricted to KLF genes in the vertebrate stem lineage, contributing to the KLF10/11 architecture class. Composite domain co-occurrence maps for each taxonomic group are shown to the right of the tree. Representative examples of putative domain shuffling events during the evolution and expansion of the KLF/SP gene family. (B) An ancient Btd box and a metazoan SP gene may have contributed to the origin of the SP gene subfamily early in metazoan evolution. (C) An ancient SID likely combined with a pre-existing ancestral KLF gene to form the KLF9/13 group, also early in metazoan evolution. (D) An ancient PVDLS domain combined with a pre-existing ancestral KLF gene to form the KLF3/8/12 group. We infer an independent convergent acquisition of the PVDLS domain within a KLF gene in the Protostomia lineage (see Discussion). Domain icon colors are the same as figure 4.4.

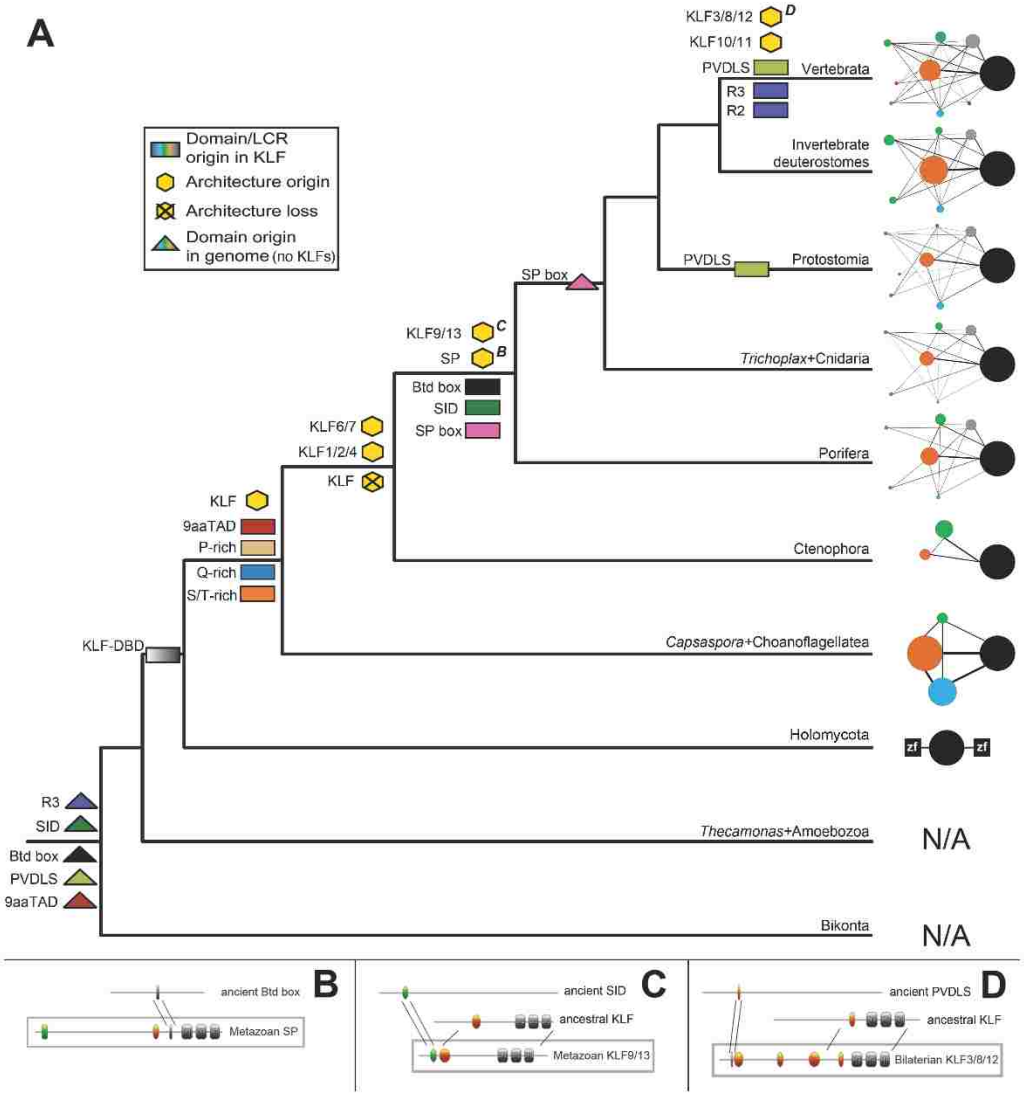

A
Domain/LCR origin in KLF
Architecture origin
Architecture loss
Domain origin in genome (no KLFs)
KLF3/8/12 D
KLF10/11
PVDLS
R3
R2
Vertebrata
Invertebrate deuterostomes
PVDLS
Protostomia
SP box
KLF9/13 C
SP B
Btd box
SID
SP box
Trichoplax+Cnidaria
KLF6/7
KLF1/2/4
KLF
Porifera
KLF
9aaTAD
P-rich
Q-rich
S/T-rich
Ctenophora
KLF-DBD
Capsaspora+Choanoflagellatea
Holomycota
zf
zf
R3
SID
Btd box
PVDLS
9aaTAD
Thecamonas+Amoebozoa
N/A
Bikonta
N/A
B
ancient Btd box
Metazoan SP
C
ancient SID
ancestral KLF
Metazoan KLF9/13
D
ancient PVDLS
ancestral KLF
Bilaterian KLF3/8/12

Table 4.1. Species used in this study with genome/transcriptome reference and abbreviations from phylogenetic trees

Species	Reference	Species Abbreviation from phylogenetic trees
Homo sapiens	Lander et al. (2001)	Hsa
Mus musculus	Chinwalla et al. (2002)	Mus
Xenopus tropicalis	Hellsten et al. (2010)	Xtr
Latimeria chalumnae	Amemiya et al. (2013)	Lch
Danio rerio	Howe et al. (2013)	Dre
Takifugu rubripes	Aparicio et al. (2002)	Tru
Gasterosteus aculeatus	Jones et al. (2012)	Gac
Callorhinchus milii	Venkatesh et al. (2014)	Cmi
Leucoraja erinacea	http://skatebase.org/downloads	Ler
Scyliorhinus canicula	http://skatebase.org/downloads	Sca
Petromyzon marinus	Smith et al. (2013)	Pma
Ciona intestinalis	Dehal et al. (2002)	Cin
Branchiostoma floridae	Putnam et al. (2008)	Bfl
Strongylocentrotus purpuratus	Sodergren et al. (2006)	Spu
Daphnia pulex	Colbourne et al. (2011)	Dpu
Drosophila melanogaster	Adams et al. (2000)	Dme
Tribolium castaneum	Richards et al. (2008)	Tca
Capitella teleta	Simakov et al. (2013)	Cte
Lottia gigantea	Simakov et al. (2013)	Lgi
Nematostella vectensis	Putnam et al. (2007)	Nve
Acropora digitifera	Shinzato et al. (2011)	Adi
Trichoplax adhaerens	Srivastava et al. (2008)	Tad
Amphimiedon queenslandica	Srivastava et al. (2010)	Aqu

Oscarella carmela	http://www.compagen.org/datasets.html	Oca
Mnemiopsis leidyi	Ryan et al. (2013)	Mle
Pleurobrachia bachei	Moroz et al. (2014)	Pba
Monosiga brevicollis	King et al. (2008)	Mbr
Salpingoeca rosetta	Fairclough et al. (2013)	Sro
Capsaspora owczarzaki	Suga et al. (2013)	Cow
Sphaeroforma arctica	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html	
Ustilago maydis	Kamper et al. (2006)	
Saccharomyces cerevisiae	Goffeau et al. (1996)	
Mortierella verticillata	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html	
Allomyces macrogynus	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html	
Spizellomyces punctatus	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html	
Encephalitozoon cuniculi	Katinka et al. (2001)	
Fonticula alba	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html	
Thecamonas trahens	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html	
Entamoeba histolytica	Loftus et al. (2005)	
Dictyostelium purpureum	Sucgang et al. (2011)	
Polysphondylium pallidum	Heidel et al. (2011)	
Acanthamoeba castellani	Clarke et al. (2013)	
Naegleria gruberi	Fritz-Laylin et al. (2010)	

Cyanidioschyzon merolae	Matsuzaki et al. (2004)
Arabidopsis thaliana	Arabidopsis Genome Initiative (2000)
Paramecium tetraurelia	Aury et al. (2006)
Bigelowiella natans	Curtis et al. (2012)
Phytophtora infestans	Haas et al. (2009)

Chapter 5: Zygotic function of *Krüppel-like factor* genes during embryogenesis in *Mnemiopsis leidyi* is associated with cell proliferation

Background

Krüppel-like factor (Klf) genes belong to the C2H2 zinc finger family. KLF transcription factors are defined by a highly-conserved zinc finger DNA binding domain, with variable N-terminal domains that mediate various protein-protein interactions (Presnell, et al. 2015). Some KLFs can bind directly to condensed chromatin, and through the recruitment of various chromatin remodeling complexes, can alter the chromatin state (Soufi, et al. 2015). For example, KLFs bound to condensed chromatin enable the recruitment of protein machinery necessary for transcription, such as the activation of the zygotic genome (Iwafuchi-Doi and Zaret 2014). The *Klf* gene family is found in all metazoan lineages, and is found in closely related unicellular organisms including filozoans and choanoflagellates (Presnell, et al. 2015). However, most of what is known about the function of *Klf* genes comes from work done in vertebrates and cell lines (see Nagai, et al. 2009). Within mammals, 17 *Klf* genes have been identified and functions associated with many of these genes have been characterized, ranging from metabolism and homeostasis, to apoptosis and tumorigenesis, and balancing cell proliferation and differentiation (reviewed in McConnell and Yang 2010; Bialkowska, et al. 2017; Kim, et al. 2017). In vertebrates, the role of *Klfs* in stem cell renewal and pluripotency has been thoroughly investigated (Takahashi and Yamanaka 2006; Takahashi, et al. 2007; Jiang, et al. 2008; Parisi, et al. 2008; Nandan and Yang 2009; McConnell and Yang 2010; Parisi, et al. 2010). For example, overexpression of the four transcription factors *Oct4*, *Sox2*, *Klf4*, and *c-myc* (OSKM) can induce pluripotency in differentiated mammalian cells

(Takahashi and Yamanaka 2006; Takahashi, et al. 2007). Additionally, in mouse embryonic stem cells (mESCs) *Klf2*, *Klf4*, and *Klf5* participate in an autoregulatory activation loop and together participate in maintaining a transcriptional circuit with *Oct4*, *Sox2*, *c-myc*, and *Nanog*, that regulates stem cell self-renewal (Jiang, et al. 2008).

Although much is known about the underlying function of *Klf* genes associated with embryonic stem cell maintenance and cell proliferation/differentiation in a handful of vertebrate species, the role of *Klfs* during development and their potential to regulate stem cell proliferation and differentiation in other animals is poorly understood. I was interested in how *Klf* function has evolved during metazoan diversification, and whether its role in stem cells has been conserved in metazoans. My recent phylogenetic analysis identified three *Klf* genes in the lobate ctenophore *Mnemiopsis leidyi* (*Chapter 4*; Presnell, et al. 2015). Ctenophores are one of the earliest branching extant metazoan lineages, making them fundamental to understanding early animal evolution (Dunn, et al. 2015). Therefore, understanding the role of *Klf* genes in ctenophores will provide insight into the relationship between *Klfs* and stem cell proliferation and differentiation during the course of animal evolution.

Ctenophores are a phylum of marine, invertebrate predators that are characterized by unique morphological features such as rows of plates of fused giant cilia (ctenes) used for locomotion, an aboral and neuronal-rich gravity-sensing (apical) organ which controls ctene row beating, and muscular tentacles that are lined with sticky cells (colloblasts) used for prey capture (Chun 1880; von Byern, et al. 2010; Ryan, et al. 2013; Moroz, et al. 2014; Tamm 2014a). Adults have remarkable regenerative capabilities (Coonfield 1936; Freeman 1967; Martindale 1986) and possess tissues that undergo constant cell-turnover

to replace old or damaged cells (e.g., tentacles; von Byern, et al. 2010). Along with the ability to rear individuals in the lab, the recent development of primary cell culture techniques (Vandepas, et al. 2017), and an emerging molecular genetic toolkit (Pang and Martindale 2008a, b, c, d; Yamada, et al. 2010), make ctenophores a valuable system for investigating the molecular mechanisms underlying stem cell regulation.

Three axes define the ctenophore body plan: the oral/aboral axis which runs longitudinally, and two orthogonal axes, the pharyngeal axis bisecting the pharynx, and the tentacular axis bisecting the tentacles and perpendicular to the pharyngeal axis (see *Chapter 3*, Fig. 3.1). *Mnemiopsis leidyi* embryos undergo a stereotypical, ctenophore-specific cleavage program (Freeman 1976b). The first cleavage typically begins at ~1 hour post fertilization (hpf), with subsequent cell divisions occurring every 20 minutes (Fischer, et al. 2014). The pharyngeal and tentacular axes are established during the first and second cleavages, respectively (Freeman 1977). During gastrulation, which takes place between 3 and 5 hpf, ectodermal, aboral micromeres migrate via epiboly to the oral side of the embryo and then invaginate, forming the blastopore (Martindale and Henry 1997a). After gastrulation, ~9 hpf, the *M. leidyi* embryo has begun to establish different tissues. Four pairs of ctene rows, one pair in each quadrant towards the aboral end of the embryo, are typically one of the first structures to form (Fischer, et al. 2014). Each ctene plate is made up of thickened ectodermal cells which produce motile cilia that are fused, forming paddle-like structures that beat and propel the animal through the water column (Tamm 1973). These cilia-bearing cells are called polster cells (Tamm 1973). Ctene plates are situated in rows in an oral/aboral fashion, and each quadrant of the animal possesses two ctene rows. At the oral pole, ectodermal cells surrounding the blastopore

invaginate forming the stomodeum and pharynx. Along the tentacular axis, on either side of the embryo, are two cellular thickenings that will ultimately give rise to the tentacle bulbs and tentacles.

By 14 hpf, organs have started to take shape. As the embryo develops, ectodermal cells associated with the two lateral thickenings invaginate and, along with the medial internal most cells derived from the endoderm, form the tentacle bulb (Martindale and Henry 1997b, 1999). The newly formed tentacles are rooted within this protective sheath. Blastomere fate maps from *M. leidyi* embryos demonstrated that multiple cell lineages contribute to both the tentacle and tentacle bulb (Martindale and Henry 1997b, 1999). For example, the core tentacle muscles are derived from endodermal cells, whereas the colloblasts and other epithelial cells of the tentacle are derived from ectodermal cells (Martindale and Henry 1997b, 1999). As the tentacle develops, it emerges from the medial-most aboral endodermal region of the tentacle bulb.

During this period of development, the pharynx also begins to elongate towards the aboral end of the embryo. Aborally, the pharynx connects with the endoderm of the developing gastrovascular cavity. At this stage, ECM is extruded from the endoderm into the mesogleal space and the embryo increases dramatically in size. The gastrovascular cavity is clearly visible by 14 hpf and gives rise to the infundibulum and endodermal canal system (Presnell, et al. 2016).

The ctenophore apical organ, located at the aboral pole, is composed of an epithelial floor that supports four clusters of specialized balancer cilia. Each tightly clustered set of balancer cilia is situated within a quadrant of the animal and connected to the ctene rows by a ciliated groove. The balancer cilia play an integral role in supporting

the statocyst. The statocyst is a small spheroid structure composed of mineralized cells called lithocytes (Horridge 1964). Deflection of the statocyst by gravity on the balancer cilia propagates a signal along the ciliated grooves that influences the rate of ctene row beating (Tamm 2014b). The lithocytes of the statocyst initially develop within two discrete territories of the apical organ epithelial floor along the tentacular axis (Tamm 2014b). Newly born lithocytes emerge from these two territories and migrate aborally along balancer cilia to ultimately merge with other lithocytes to form the mature statocyst (Tamm 2014b). Also embedded within the apical organ epithelial floor are putative photoreceptor cells (Horridge 1964; Schnitzler, et al. 2012) and other types of neurons (Jager, et al. 2011).

Using EdU pulse/chase assays, discrete territories of rapidly proliferating cells expressing genes associated with somatic stem cells such as paralogs of *Piwi*, *Nanos*, *Vasa*, and *Sox* have been identified in adult *Pleurobrachia pileus* in the ctene rows, apical organ, and tentacle apparatus (Jager, et al. 2008; Alié, et al. 2011). In the ctene rows, putative stem cells were identified at the lateral edges of each ctene plate which give rise to the polster cells (Alié, et al. 2011). Within the tentacle apparatus, putative somatic stem cells were in the basal portion (tentacle root) along the median and lateral ridges which harbor tentacle muscle cell progenitors and colloblast progenitors, respectively (Alié, et al. 2011).

By comparing the expression patterns of these genes from adult *P. pileus* to *M. leidyi* embryos, it was shown that *Sox1*, *Sox3*, *Piwi1*, and *Vasa1* expression during the embryonic development of *M. leidyi* corresponded to similar territories in these developing organs (Schnitzler, et al. 2014; Reitzel, et al. 2016). EdU staining of recently

hatched juvenile cydippids also demonstrated that the expression of these genes was found in areas of cell proliferation, including the apical organ and the tentacle bulbs (Schnitzler, et al. 2014; Reitzel, et al. 2016). These data are consistent with the hypothesis that *Sox1* and *Sox3* genes in ctenophores are expressed in embryonic stem cells and thus could be playing a role in stem cell maintenance and/or cell proliferation and renewal.

The major transcription factors that regulate self-renewal of stem cells and induce pluripotency in mammals include the *Oct4*, *Sox2*, *Klf4, and c-myc* (OSKM) genes (Takahashi and Yamanaka 2006; Takahashi, et al. 2007). While the expression patterns of *Sox* paralog genes are associated with putative stem cells (Jager, et al. 2008; Schnitzler, et al. 2014), *M. leidyi* lacks the *Oct4* ortholog but contains other *Pou* family genes (Ryan, et al. 2010). Nothing is known about *Pou*, *Klf*, or *c-myc* expression and function in ctenophores. Here I examine the expression patterns of *Klf* genes and use both morpholino oligonucleotides (MOs) and CRISPR/Cas9 to disrupt zygotic *Klf* expression during the embryonic development of *M. leidyi*. Determining the role of *Klf* genes will help elucidate the transcriptional regulatory network controlling stem cell fate in ctenophores and provide insight into the mechanisms underlying the evolution of metazoan stem cells.

Results

My previous phylogenetic analysis of *Klf* genes in eukaryotes identified three *Klf*s in *M. leidyi*: *MleKlf5a, MleKlf5b,* and *MleKlfX* (*Chapter 4*; Presnell, et al. 2015). *MleKlf5a* and *MleKlf5b* genes are closely related to other *Klf5* genes. These two genes

appear to be the result of a lineage-specific duplication. *MleKlfX*, however, is highly derived with no clear orthology to any specific *Klf* clade (Presnell, et al. 2015). Gene models generated from the reference genome (Ryan, et al. 2013) and transcriptomic data (Davidson, et al. 2017) show that *MleKlf5a* (*ML00922a)*, *MleKlf5b* (*ML25776a)*, and *MleKlfX* (*ML20061a*) span roughly 23.8, 26.5, and 4.2 kb and contain 4, 9, and 5 exons, respectively (Fig. 5.1A). I performed *in situ* hybridization to examine the expression patterns of these genes during embryogenesis, followed by assessment of gene function using both splice-blocking morpholino oligonucleotides (sbMOs) and CRISPR/Cas9 (Jinek, et al. 2012).

The sbMOs were designed to target the junctions between intron 2 and exon 3 (#1), and intron 3 and exon 4 (#2) for *MleKlf5a*; and exon 6 and intron 6 (#1), and exon 7 and intron 7 (#2) for *MleKlf5b* (Fig. 5.1B). Mechanistically, sbMOs can cause retention of introns or cause exons to be skipped during mRNA processing (Draper, et al. 2001), and RT-PCR can be used to assess mis-splicing events caused by the sbMOs. The percentage of embryos injected with the control MO that developed normally (79.6%; n=49) was similar to both un-manipulated embryos (85.1%, p=0.28) and dechorionated, uninjected embryos (80.6%, p=0.70; Fig. 5.11). Closely related KLFs bind to shared downstream targets which can result in complex functional outcomes. For example, KLF2, KLF4, and KLF5 have redundant roles in regulation of *Nanog* (Jiang, et al. 2008), while KLF1 and KLF3 compete for binding sites to exert different functional outcomes (Ilsley, et al. 2017). For this reason, to maximize the efficiency of generating an observable phenotype I injected #1 and #2 MOs for both *MleKlf5a* and *MleKlf5b*, simultaneously (KLF-MO).

Guide RNAs (gRNAs) were designed to target sites spanning the entire genes for all three *MleKlf* genes. Similar to the sbMO injections, I simultaneously injected solutions that contained recombinant Cas9 protein combined with multiple gRNAs for both *MleKlf5a* and *MleKlf5b* (KLF-Cas9). To assess which gRNAs were effective in shuttling the Cas9 nuclease to its target site, I sequenced short regions spanning each putative cut site from individual embryos (Fig. 5.1C). Uninjected embryos from the same batches as injected embryos were also sequenced for comparison. Sequence traces from KLF-Cas9 injected embryos showed a noticeable degradation of signal caused by INDELS, compared to control traces (Fig. 5.1C). This indicated that the gRNAs were successful in shuttling the Cas9 nuclease to its target site followed by efficient nuclease activity. For instance, out of 10 KLF-Cas9 injected embryos sequenced, only 1 embryo had sequence traces that showed no evidence of Cas9 activity for both *MleKlf5a* or *MleKlf5b*. The remaining KLF-Cas9 injected embryos had sequence traces that showed INDELS around the cut sites, indicating successful Cas9 activity. To test the possibility of nonspecific binding of the gRNAs, I sequenced top putative off-target sites and observed no evidence of Cas9 activity (Table 5.2). These lines of evidence, coupled with similar phenotypes generated from KLF-MO injections (see below), suggest that the KLF-Cas9 phenotypes observed in injected embryos were due to the successful disruption of *MleKlf5a* and *MleKlf5b* gene expression.

***MleKlf5a* and *MleKlf5b* transcripts are maternally deposited**

Transcriptional profiling in *M. leidyi* during early embryonic cleavage stages found high levels of *MleKlf5a* and *MleKlf5b* transcripts and showed that zygotic genome

activation most likely occurs ~1.5 hpf (Davidson, et al. 2017). I detected ubiquitous expression of *MleKlf5a* and *MleKlf5b* transcripts in the zygote, early cleavage stages, and during gastrulation (Fig. 5.2). KLF-MO and KLF-Cas9 injected embryos developed normally through gastrulation, suggesting that zygotic, not maternal, expression of *MleKlf5a* and *MleKlf5b* was disrupted.

Zygotic *MleKlf5a* and *MleKlf5b* regulate cellular proliferation during development

I first observed spatially localized expression patterns of both *MleKlf5a* and *MleKlf5b* just after gastrulation at mid-embryogenesis, ~9 hpf, and persisting through hatching, ~20 hpf (Fig. 5.3). In general, expression patterns of *MleKlf5a* and *MleKlf5b* were found in the pharynx, gastrovascular cavity, tentacle bulbs, and apical organ (Fig. 5.3). Additional *MleKlf5b* expression was observed in cells surrounding the ctene rows (Fig. 5.7A).

Pharynx and Gastrovascular cavity

At ~9 hpf, the ectodermal cells surrounding the blastopore invaginate to form the stomodeum and pharynx. As development continues, by ~14 hpf, the mesoglea (situated between the ectoderm and endoderm) is hydrated and the embryo dramatically increases in size. This is accompanied by growth of the pharynx in the aboral direction where it connects with the endoderm of the newly formed gastrovascular cavity. Initially the endoderm is made up of large, vacuolated cells that are preferentially labeled with the vital dye Lysotracker. As extracellular matrix (ECM) is extruded from these cells during development, Lysotracker signal becomes restricted to the epithelial cells of the

gastrovascular cavity (Fig. 5.4B, D). Spatially distinct expression patterns of *MleKlf5a and MleKlf5b* were observed along the oral/aboral axis of the pharynx beginning at 9 hpf. Throughout development, transcript signals of *MleKlf5a* and *MleKlf5b* were localized to the inner cells of the aboral most portion of the pharynx (Fig. 3B, D, F, H J, L), and was absent in the stomodeum. I also observed *MleKlf5a* and *MleKlf5b* expression in the epithelial cells of the gastrovascular cavity (Fig. 5.3D, F, J, L). This suggests that zygotic *MleKlf5a* and *MleKlf5b* expression most likely plays a role in the development of the pharynx and gastrovascular cavity. To test this, I knocked down *MleKlf5a* and *MleKlf5b* expression using both sbMOs (KLF-MO) and Cas9 (KLF-Cas9).

Upon knockdown of *MleKlf5a* and *MleKlf5b* expression, I observed disrupted endodermal patterning associated with the gastrovascular cavity 52% and 77% of the time in KLF-MO and KLF-Cas9 injected embryos, respectively. The two types of endodermal phenotypes were characterized as mild and severe (Fig. 5.5). Embryos with mild phenotypes typically underwent mesogleal hydration accompanied by the increase in embryo size and pharynx elongation, however patterning of the gut endoderm was disorganized (Fig. 5.5B, C.). Embryos that failed to undergo mesogleal hydration and overall size increase were classified as exhibiting a severe phenotype (Fig. 5.5.F, G). In these embryos, Lysotracker positive cells remained throughout the endoderm due to a lack of ECM release into the mesoglea. I observed a higher percentage of KLF-Cas9 injected embryos exhibit the severe phenotype (71% Cas9 vs 26% sbMOs), while more KLF-MO injected embryos presented the mild phenotype (6% Cas9 vs 26% sbMOs). Overall, KLF-MO injected embryos presented an allelic series in which phenotypes ranged from mild to severe. The observed broad range of phenotypes was due to the

varying penetrance caused by KLF-MO activity. This is due to the lower efficiency of sbMOs, which must target many mRNA transcripts. Compared to sbMOs, however, the efficiency of Cas9 activity is higher due to the relatively fewer genomic DNA sites it must target, thus producing more consistent, severe phenotypes, as seen in the higher percentage of severe endodermal phenotypes in KLF-Cas9 injected embryos.

Embryos with the mild endodermal phenotypes also had defects in pharyngeal patterning, mostly in the aboral region of the pharynx (Fig. 5.5D). For example, the junction between the gastrovascular cavity endoderm and the pharynx was disrupted and in some cases I observed instances in which the aboral end of the pharynx was bifurcated (Fig. 5.5E). *MleKlf5a* and *MleKlf5b* were expressed in this part of the pharynx, and could play a role in the development of specialized structures found in the aboral portion of the pharynx, such as the ciliary mill and pharyngeal folds (Tamm 2014a; Presnell, et al. 2016). In *M. leidyi*, the pharynx is derived from the ectoderm while the gastrovascular cavity is derived from the endoderm, and thus may be analogous to the foregut/midgut junction boundary. In the severe embryos, pharynx formation was also severely disrupted (Fig. 5.5F, G). This suggests that *MleKlf5a* and *MleKlf5b* may be playing a role in patterning the junction between the gastrovascular cavity and the pharynx.

The failure of mesogleal hydration accompanied by an increase in embryo size, and disruption of pharynx formation in the severe embryos was similar to the phenotypes seen in *M. leidyi* embryos in which *Brachyury* (*MlBra*) translation was blocked with MOs (Yamada et al. 2010). *Brachyury*, a member of the T-box transcription factor family, is known for its roles in mesoderm specification and regulation of morphogenesis during development (Papaioannou 2014). Characterization of this gene in non-bilaterians

suggests that in these animals *Brachyury* regulates morphogenetic movements, such as invagination of the stomodeum and pharynx (Technau 2001; Yamada, et al. 2010; Yasuoka, et al. 2016; Servetnick, et al. 2017). Therefore, I decided to knockdown *Brachyury* expression with Cas9 (Bra-Cas9) to compare the pharyngeal phenotype to what was seen in KLF-Cas9 injected embryos. Bra-Cas9 injected embryos had a cluster of stomodeal cells which had initially invaginated, but additional pharyngeal tissue was not present (Fig. 5.5H, K), like that seen in the Bra-MO injected embryos from (Yamada, et al. 2010). In KLF-Cas9 injected embryos, ectodermal cells that initially invaginated and gave rise to the stomodeum were present, as seen in Bra-Cas9 injected embryos. However, in KLF-Cas9 injected embryos, additional pharyngeal tissue was observed (Fig. 5.5I, J). This pharyngeal phenotype observed in the KLF-Cas9 injected embryos was slightly different than the phenotype generated by a loss of *Bra* expression. This suggests that the pharyngeal phenotype in KLF-Cas9 injected embryos was not caused by Cas9 artifacts, but by loss of *MleKlf5a* and *MleKlf5b* expression. One possible cause of pharyngeal growth failure in KLF-Cas9 embryos was the loss of the proliferation in the aboral portion of the pharynx, where both *MleKlf5a* and *MleKlf5b* are expressed.

Next, I investigated the distribution of EdU incorporation in the pharynx to test the idea that *MleKlf5a* and *MleKlf5b* are regulating pharyngeal growth through cellular proliferation. EdU incorporation was observed in two spatially distinct regions of the pharynx, one in the stomodeum and oral portion of the pharynx, and the other in the aboral most region near the junction with the gastrovascular cavity (Fig. 5.6A, B). High levels of cellular proliferation in the aboral portion of the pharynx was most likely due to the formation of the pharyngeal folds and numerous cilia that will, during feeding, propel

food from the mouth to the infundibulum (Tamm 2014a; Presnell, et al. 2016). EdU assays on KLF-Cas9 injected embryos, showed that proliferation was unaffected in the oral-most portion of the pharynx however, proliferation was lost in the aboral-most portion (Fig. 5.6G). Genes such as *Piwi1*, *Vasa1*, *Sox1*, and *Sox3* are also expressed in the developing pharynx in *M. leidyi* (Schnitzler, et al. 2014; Reitzel, et al. 2016), consistent with the presence of an active stem cell niche in this organ. My results suggest that *MleKlf5a* and *MleKlf5b* are required for normal endodermal patterning and mesoglea expansion, and, additionally, regulate pharyngeal morphogenesis through the maintenance of proliferative cells in the aboral most portion of the pharynx.

Ctene rows

Ctene rows take shape around 9 hpf, as a pair of ectodermal thickenings in each quadrant of the embryo. The ctenes are produced from a set of cells called polster cells, which have a high abundance of mitochondria, as seen with Mitotracker staining (Fig. 5.4A, C). From 9-20 hpf, *MleKlf5b* expression was observed in a set of ectodermal cells that surrounded each ctene row pair (Fig. 3G-L; 5.7A). These cells meet in a central location on the oral side of each ctene row pair, and extend aborally on both sides of each ctene row pair (Fig. 5.7A).

Injection with KLF-MO and KLF-Cas9 produced no effect on polster cell or ctene formation, as indicated by normal Mitotracker signal (Fig. 5.7C, D). This is not surprising given the evidence that ctene row fate is specified by maternal factors (Fischer, et al. 2014). Next, I examined the proliferative nature of these ctene row associated cells and observed EdU incorporation, suggesting these may be polster cell progenitors similar to

those identified in *Pleurobrachia* (Alié, et al. 2011). However, new ctene plates and polster cells do not form during embryogenesis, only after hatching (Tamm 2012). One possibility for observing EdU incorporation in these cells is that they may play a role in overall growth and maintenance of the ectodermal epithelia around the ctene rows. Once the embryo hatches, these cells could give rise to new polster cells. Examining the post-embryonic expression and function of *MleKlf5b* will provide insight into its role in ctene row growth in juveniles.

Tentacle Bulbs

Ctenophore tentacles are comprised of a muscular core surrounded by epithelial cells which include colloblasts, specialized cells unique to ctenophores that contain a glue-like substance for prey capture. In adult *Pleurobrachia*, both muscle and colloblast progenitor cells are in the tentacle root, the basal portion of the tentacle bulb (Alié et al 2011). Muscle stem cells are found within the median ridge, a column of cells along the basal-most portion of the tentacle root, while colloblast stem cells are in the lateral ridges which extend outward on either side of the median ridge. At 9 hpf, and through 14 hpf, I observed expression of *MleKlf5a* and *MleKlf5b* in the endodermal, medial extensions of the initial tentacle bulb thickenings (Fig. 5.3). Once ectoderm invagination took place, and the tentacle bulbs started to take shape, the expression patterns of *MleKlf5a* and *MleKlf5b* were localized to the median ridges (Fig. 5.3).

I examined the effects of knockdown of *MleKlf5a* and *MleKlf5b* expression on tentacle bulb formation, and observed a decrease in the size of the tentacle bulbs in both KLF-MO and KLF-Cas9 embryos. To quantify these results, I measured the length of the

median ridge (Fig. 5.8H), and found significant differences in this parameter between the control and injected embryos (Mann Whitney U test, $p<0.05$; Fig. 5.8G). In control embryos the median ridge was, on average, 23 μm in length (Fig. 5.8A). In embryos injected with KLF-MO and KLF-Cas9, the average lengths of the median ridge were 18.29 μm (Fig. 5.8C, D) and 9.88 μm (Fig. 5.8E, F), respectively. In some instances, I observed embryos that lacked tentacle bulb tissue altogether, about 15% of the KLF-MO embryos and 29% of the KLF-Cas9 embryos. These results suggest that *MleKlf5a* and *MleKlf5b* play a role in tentacle bulb development through regulating proliferation of the median ridge.

To test the idea that *MleKlf5a* and *MleKlf5b* are regulating maintenance of proliferative cells in the median ridge, I examined EdU incorporation in the tentacle bulb in both control and KLF-Cas9 embryos. EdU signal was localized in the endodermal cells of the cellular thickenings in control embryos, where I observed *MleKlf5a* and *MleKlf5b* expression (Fig. 5.6A-D). Additionally, in the fully developed tentacle bulb, I observed proliferative cells in both the median ridge and in the outer epithelium of the tentacle bulb (Fig. 5.6B, D). In KLF-Cas9 embryos, proliferation in the tentacle bulb epithelium was unaffected, but I observed a loss of EdU incorporation in the median ridge (Fig. 5.6G). Stem cell genes, *Piwi1, Vasa1, Sox1*, and *Sox3* are expressed in the median ridge in both *M. leidyi* and *Pleurobrachia* (Alié, et al. 2011; Schnitzler, et al. 2014; Reitzel, et al. 2016), and the median ridge harbors stem cells for tentacle muscle cells (Alié, et al. 2011). These results suggest that *MleKlf5a* and *MleKlf5b* are required for normal development of the tentacle bulb, through maintenance of the proliferative cells in the

median ridge stem cell niche. The absence of tentacle bulbs could be due to the loss of the stem cell niche that initially gives rise to cells fated to become tentacle bulbs.

Apical Organ

At 14 hpf, the ectodermal cells that give rise to the epithelial floor of the apical organ invaginate. This is shortly followed by extension of the balancer and dome cilia. Mature lithocytes travel along the balancer cilia where they merge and form the statocyst. Prior to invagination, I did not observe expression of *MleKlf5a* or *MleKlf5b* in the aboral ectodermal cells (Fig. 5.3B, H). At 14 hpf, I observed expression of both *MleKlf5a* and *MleKlf5b* in the epithelial floor (Fig. 5.3D, J), and by cydippid stage, expression of *MleKlf5a* and *MleKlf5b* was restricted to two territories in the apical organ floor where new lithocytes form before leaving the epithelial floor (Fig. 5.3F, L).

Knockdown of *MleKlf5a* and *MleKlf5b* resulted in fewer number of lithocytes compared to control embryos (Mann Whitney U, $p<0.05$; Fig. 5.9). I observed, on average, 6.5 lithocytes in control embryos (Fig. 5.9A, B). KLF-MO injected embryos had, on average 4.36 lithocytes, and three of these embryos had lost lithocytes entirely (Fig. 5.9C, D). KLF-Cas9 injected embryos also had fewer lithocytes, average of 2.25, and five embryos completely lacked lithocytes (Fig. 5.9E, F). In embryos which lacked lithocytes, balancer cilia and dome cilia were still present (Fig. 5.9C, E). *MleKlf5a* and *MleKlf5b* expression was not observed in the cells that give rise to these ciliated structures (Fig. 5.6H). Lithocytes and the cilia are derived from separate cell lineages, endodermal and ectodermal respectively (Martindale and Henry 1999).

Cell proliferation in the apical organ was observed in the developing polar fields, ciliated cells that extend from the aboral pole in the pharyngeal axis, and in a small subset of cells located in the tentacular axis deep within the epithelial floor (Fig. 5.6E, F). These deeper proliferative cells may be lithocyte progenitor cells, due to their proximity to where new lithocytes are formed. *MleKlf5a* and *MleKlf5b* expression was found in the apical organ where lithocytes are born (Fig. 5.6H). My results suggest that lithocyte production is regulated by both *MleKlf5a* and *MleKlf5b* possibly through regulation of cell proliferation in lithocyte progenitor cells found in the apical organ.

Zygotic *MleKlfX* expression is restricted to a small subset of cells in the apical organ

MleKlfX is unique due to its highly derived sequence, in which the DNA binding domain is located on the opposite end of the protein compared to what is found in other KLFs. According to my phylogenetic analyses (see *Chapter 4*), *MleKlfX* is not closely related to other filozoan *Klfs* indicating rapid evolution of this gene (Presnell, et al. 2015). This gene is also divergent from the other ctenophore *KlfX* that was identified in *Pleurobrachia* (Presnell, et al. 2015), which suggests that the significant divergence of *MleKlfX* could be lineage-specific and most likely indicative of some function that may be restricted to a subset of ctenophores.

MleKlfX expression was only found in late stage embryos, beginning around ~16 hpf. Expression was observed in two layers of cells, deep and superficial, within the apical organ floor (Fig. 5.10). The cells found on the surface were located within each quadrant of the animal, just medial of the ciliated groove. The deeper cells were located in the center of the epithelial floor. Two cell clusters were along the tentacular plane

perpendicularly situated to the other two cell clusters, forming a cross-shaped pattern. Knockdown of *MleKlfX* failed to produce any clear morphological phenotype. I did not observe any EdU incorporation in these territories, and thus, these were most likely differentiated cells. The ctenophore apical organ floor contains numerous neuronal cell types (Jager, et al. 2011; Moroz, et al. 2014; Tamm 2014a) suggesting that *MleKlfX* is expressed in a set of differentiated neurons. The cells in which *MleKlfX* was expressed are in a region of the apical organ that is similar to the position of the lamellate bodies, putative photoreceptor cells (Horridge 1964; Schnitzler, et al. 2012). These results suggest that *MleKlfX* plays a role in putative neuronal cell fates of the apical organ in *M. leidyi*.

Discussion

In this chapter, I examined the expression and function of *Klf* genes in the ctenophore *Mnemiopsis leidyi*. Using *in situ* hybridization I showed that transcripts of *MleKlf5a* and *MleKlf5b* were initially found throughout the embryo during early cleavage stages. These transcripts were maternally deposited, since zygotic genome activation occurs roughly 1.5 hours after fertilization (Davidson, et al. 2017). *Klf* genes in other animals, including *Drosophila* (De Graeve, et al. 2003; Weber, et al. 2014), sea urchin (Materna, et al. 2006), zebrafish (Li, et al. 2011; Kotkamp, et al. 2014), *Xenopus* (Gao, et al. 2015), and humans (Blakeley, et al. 2015) are also known to be maternally expressed within the embryo.

Spatially restricted expression patterns of zygotic *MleKlf5a* and *MleKlf5b* showed significant overlap in embryos after gastrulation. Prominent areas where these genes were

expressed were the pharynx, the gastrovascular cavity, the median ridge of the tentacle bulbs, and cells in the epithelial floor of the apical organ. *MleKlf5b* had a unique expression pattern located in ctene row associated-cells, which may be polster progenitor cells. Both *MleKlf5a* and *MleKlf5b* expression patterns were found in areas that exhibited high cell proliferation, as indicated by EdU incorporation. These proliferation zones in *M. leidyi* embryos are similar to the zones of proliferation, and putative stem cell niches, found in adult *Pleurobrachia* (Alié, et al. 2011).

In contrast, *MleKlfX* expression was found in a small subset of apical organ cells, which lacked any EdU signal. These cells, the lamellate bodies, are thought to be photoreceptors, and have been shown to express *MlOpsin2* (Schnitzler, et al. 2012), *Mlgli*, a gene associated with neuronal specification (Layden, et al. 2010), and *MlLhx1/5* part of the LIM homeobox family that regulates neuronal specification (Simmons, et al. 2012). Opsins are expressed in the photocytes along the endodermal canals and in the apical organ epithelial floor, and the apical organ expression of *MlOpsin2* is thought to be associated with light-sensing neuronal cells (Horridge 1964; Schnitzler, et al. 2012). Ctenophores use these light-sensing cells for initiating spawning due to light cues, specifically the onset of darkness (*Chapter 2*; Pang and Martindale 2008g). Knockdown of the *MleKlfX* gene using Cas9 showed no obvious morphological phenotype. If *MleKlfX* is playing a role in the opsin-expressing cells, then a loss of *MleKlfX* could possibly result in a behavioral phenotype manifested by a decrease in the ability for the apical organ to respond to light cues, possibly perturbing spawning timing. Opsins are also thought to be involved in the bioluminescent pathway in ctenophores (Schnitzler, et al. 2012), and *MleKlfX* may regulate some aspect of the neuronal pathway associated

with bioluminescence. Comparison of *Klf* gene expression patterns and functions in other ctenophore species will be valuable in determining the evolutionary significance of *MleKlfX*.

Using both sbMOs and CRISPR/Cas9 to disrupt *MleKlf5a* and *MleKlf5b* expression, I demonstrated that zygotic *MleKlf5a* and *MleKlf5b* play a role in regulating cell proliferation associated with normal organ growth during early development. For example, in KLF-Cas9 injected embryos, I observed a loss of EdU incorporation in the aboral portion of the pharynx. This loss of cell proliferation was associated with abnormal pharyngeal morphogenesis. I also observed a loss of EdU signal in the median ridge, and in addition, KLF-MO and KLF-Cas9 injected embryos had decreased median ridge lengths. In some cases, the median ridge or tentacle bulb was completely missing. In the apical organ, I observed a reduction in the number of lithocytes in both KLF-MO and KLF-Cas9 injected embryos.

Normally during development, ectodermal cells invaginate to form the stomodeum and pharynx, and the pharynx elongates in the aboral direction and connects with the gastrovascular cavity. This is accompanied by hydration of the mesoglea and an increase in embryo size. Whether mesogleal hydration and overall growth of the embryo is causally associated with pharyngeal elongation is unclear, as the mechanisms for pharyngeal and mesogleal expansion are unknown. In Bra-Cas9 injected embryos, I also observed a loss of pharyngeal elongation coupled with a lack of mesoglea hydration and embryo growth. This corroborated results from *Brachyury* MO knockdown (Bra-MO) in *M. leidyi* (Yamada, et al. 2010). Additionally, most of the Bra-MO injected embryos lacked tentacle bulbs and an apical organ (Yamada, et al. 2010), similar to the more

severe phenotypes I observed in KLF-MO and KLF-Cas9 injected embryos. These results hint at possible interactions between *Klf* and *Brachyury* in ctenophore development. Brachyury and KLF transcription factors have been shown to bind similar promoter sites in human embryonic stem cells (Faial, et al. 2015), and both KLF4 and KLF5 putatively bind to the *Brachyury* promoter in mouse embryonic stem cells (Aksoy, et al. 2014).

The median ridge, lithocytes, and gastrovascular cavity are derived from endodermal lineages (Martindale and Henry 1999). Regulation of endoderm development by *Klf5* has been shown in other animals (Ohnishi, et al. 2000; McConnell and Yang 2010; Gao, et al. 2015). My results suggest that *Klf5* function in endoderm specification and differentiation is conserved across metazoans.

I observed a lack of an effect associated with KLF-MO and KLF-Cas9 on ctene row development. Polster cells, which produce the motile cilia of each ctene, are specified early in development (Fischer, et al. 2014), and thus the initial formation of ctene plates would not be affected by a loss of zygotic *MleKlf5a* and *MleKlf5b*. Additionally, only *MleKlf5b* expression is associated with the ctene rows, in cells that surround each ctene row pair. These cells are most likely polster progenitor cells, which are the stem cells that generate new polster cells in juveniles and adults. If *MleKlf5b* is responsible for maintaining polster cell stem cells, then I would expect to see a loss of new polster cell production in juveniles that have lost *MleKlf5b* expression.

My results show that the association of *Klfs* with cell proliferation is conserved throughout metazoans and likely represents an ancestral function of this gene family. Although expression patterns for *M. leidyi Pou* and *c-myc* genes are currently lacking, some *Sox* paralogs are also expressed in similar embryonic territories as *MleKlf5a* and

MleKlf5b (Schnitzler, et al. 2014). This hints at the retention of a transcriptional circuit known to regulate pluripotency (OSKM) in ctenophores, and thus would be conserved in metazoans (Rosselló, et al. 2013). My phylogenetic analysis of the *Klf* gene family, see *Chapter 4*, identified *Klf* genes in the genomes of closely related unicellular organisms (Presnell, et al. 2015). Some of these organisms can undergo transient multicellularity (Brunet and King 2017), requiring some balance between cell proliferation and differentiation. Specifically, the choanoflagellate *Salpingoeca rosetta* forms multicellular colonies through cellular divisions (Fairclough, et al. 2013). Solitary swimmer cells produce daughter cells that form the multicellular colonies, and transcriptome profiling of the swimmer cell type found upregulated expression of *Sro_06628* (Fairclough, et al. 2013), one of the *S. rosetta Klf* genes identified in my phylogenetic analysis (Presnell, et al. 2015). *Klfs*, therefore, are good candidates for genes that, at least in the filozoan lineage, have played a role in enabling multicellularity through the differential regulation of cell fate specification, and regulation of stem cell renewal and proliferation.

Methods

Microinjection setup

Lab reared *Mnemiopsis leidyi* were spawned in glass bowls and eggs were allowed to be fertilized for approximately 10 minutes before being washed in 0.2 µm filtered artificial sea water (FSW). Fertilized eggs were passed through 0.5 µm and 0.4 µm nylon mesh filters to remove excess debris and mucus. Embryos were then transferred to gelatin-coated dishes and were dechorionated using tungsten needles.

Injection needles were pulled from either borosilicate or aluminosilicate glass capillary tubes (World Precision Instruments, Inc. 1B150F-4 or Sutter Instrument) using a micropipette puller (Sutter Instrument) with the following parameters: *Heat = 764, Pull = 60, Velocity = 90, Time = 200*. Needles were backfilled with different injection cocktails (see below). Injected embryos were gently transferred to glass dishes with FSW spiked with a 1% solution of penicillin and streptomycin (Pen/Strep) and kept at room temperature until they reached the developmental stage of interest.

Gene expression, in situ hybridization

RNA was extracted from embryos at different developmental stages using TRIzol and reversed transcribed to generate cDNA (SMARTer kit, Clontech). *MleKlf5a*, *MleKlf5b*, and *MleKlfX* sequences were amplified (see primers used below) from the cDNA template and cloned using the pGEM Easy T kit (Promega).

MleKLf5a (*ML00922a*)	Fwd: 5'-ATGAGTGCTATGACATG-3'
	Rev: 5'-AAACGTGTTCAAATGCCTCTT-3'
MleKlf5b (*ML25776a*)	Fwd: 5'-ATGGACGTTTCCACGC-3'
	Rev: 5'-AGACGAGCTAGGGGGAACG-3'
MleKlfX (*ML20061a*)	Fwd: 5'-GGCAGTTTAGTTCGATCGG-3'
	Rev: 5'-TGCAGTGAGTGGTAGGTT-3'

Digoxigenin labeled antisense RNA probes were *in vitro* transcribed (MEGAscript™ T7 and SP6 kit, ThermoFisher Scientific) and stored in hybridization buffer (Hyb) at -20°C at a stock concentration of 50 ng/µL. I followed the *in-situ* hybridization protocol from (Pang and Martindale 2008d) with a few modifications: For

later stages (14-20 hpf), embryos were relaxed in a 1:1 FSW:6.5% $MgCl_2$ (in dH_2O) for 20 minutes prior to fixation. Embryos were fixed with 4% paraformaldehyde (PFA) overnight at 4°C and used immediately or stored at -20°C in 100% methanol. I omitted the glutaraldehyde fixation step. Incubation time for the hybridization step was 24 hours. Images were acquired using a Zeiss Axio Imager.Z2 with a Zeiss AxioCam MRm Rev3 camera running through Zeiss AxioVision (Release 4.8.2) or Zeiss Zen Blue software. Images were processed in Adobe Photoshop. Brightfield images were inverted, pseudocolored, and overlaid onto fluorescent images of nuclei labeled with either 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI, Molecular Probes) or Hoechst 33342 (Molecular Probes).

Vital Dye staining

Live embryos were incubated in FSW containing a final concentration of 100 nM MitoTracker® (Deep Red FM, Molecular Probes), 100 nM LysoTracker® (Red DND-99, Molecular Probes), and 10 ng/µl Hoechst 33342 for one hour at room temperature. Live embryos were placed on glass slides and were relaxed by a drop of 6.5% $MgCl_2$ (in dH_2O) situated on the coverslip. Images were acquired using a Zeiss Axio Imager.Z2 with a Zeiss AxioCam MRm Rev3 camera and Zeiss AxioVision software (Release 4.8.2) or Zeiss Zen Blue. Excitation and emission maximum for MitoTracker® channel was 644/665 nm, for LysoTracker® was 577/590 nm, and Hoechst was 350/461. Z-stacks were acquired for each fluorescent channel plus the differential interference contrast (DIC) channel. Single planes for each channel were processed for brightness and contrast, and were merged together using Adobe Photoshop.

Cell proliferation staining

For measuring proliferating cells, I used the Click-iT® EdU Alexa Fluor® 647 Imaging Kit (ThermoFisher Scientific). EdU (5-ethynyl-2'-deoxyuridine) is a thymidine analog which is incorporated into DNA during DNA synthesis. Embryos were collected at different developmental stages and incubated (pulse) with 100 µM EdU in a 1:1 volumetric ratio of FSW to 6.5% $MgCl_2$ (in dH_2O) at room temperature. Pulse time for all stages was 25 minutes. The EdU solution was washed out (chase) and embryos were either fixed immediately (no chase) or allowed to further develop and then fixed (24-hour chase). Embryos were fixed with 4% PFA (in FSW) for 30 minutes at room temperature, washed with 3% BSA (in PBS), and incubated with 0.5% Triton X-100 (in 1x PBS) for 20 minutes at room temperature. Fixed embryos were washed with 3% BSA (in 1x PBS) and stored at 4°C until used for the following EdU detection steps. I followed the manufacturer's protocol for EdU detection. Briefly, I incubated the fixed embryos in the Click-iT® reaction cocktail for 30 minutes at room temperature in the dark, washed the samples with 3% BSA (in 1x PBS) and with 1x PBS, then incubated the embryos with 5 µg/mL Hoechst 33342 (in 1x PBS) for 30 minutes at room temperature in the dark. Embryos were finally washed with 1x PBS and mounted on glass slides and images were acquired using either a Zeiss Axio Imager.Z2 with a Zeiss AxioCam MRm Rev3 camera running through Zeiss AxioVision (Release 4.8.2) or Zeiss Zen Blue software, or a Leica SP5 confocal microscope. Images were processed with either Adobe Photoshop or FIJI (Schindelin, et al. 2012).

Morpholino oligonucleotides

I obtained splice-blocking morpholino oligonucleotides (MO; Gene Tools) that were complimentary to intron/exon boundaries for both *MleKlf5a* (*ML00922a*) and *MleKlf5b* (*ML25776a*). For *MleKlf5a* the sequence of MO #1 was 5'-TCTCGTGTCTGAAACAATTTTAAGT-3' and of MO #2 was 5'-GTCTACCACCTGCAAGATTTTAAGT-3'. For *MleKlf5b* the sequence of MO#1 was 5'-CAGTTGATTTCTCACCTGCCAAGAA-3' and of MO #2 was 5'-CAAACAGACTTACCTTCAAATGTGA-3'. A standard control MO (Gene Tools, 5'-CCTCTTACCTCAgTTACAATTTATA-3') was used as a negative control. Stock solutions of MO were resuspended in dH_2O, and stored at room temperature. MO injection cocktail solutions consisted of final MO concentration of ~333 nM for each MO along and 35% glycerol with ~0.5 mg/ml fluorescent dextran (rhodamine or Alexa-Fluor 488, 10,000 MW, ThermoFisher Scientific). Injected embryos were gently transferred to glass dishes filled with FSW and kept at room temperature until they reached the developmental stage of interest. After imaging, RNA was extracted from single embryos. The OneTaq® One-Step RT-PCR Kit (New England Biolabs) was then used to amplify specific regions of cDNA to test for MO activity.

CRISPR/Cas9

Recombinant Cas9 protein was obtained through PNA Bio. Guide RNAs (gRNAs) targeting my genes of interest, as well as potential off-target binding sites, were designed using the CasOT program (Xiao, et al. 2014). The *Mnemiopsis leidyi* genome downloaded from the genome portal website (Moreland, et al. 2014), and then was

uploaded into the CasOT program. I used the default settings for identifying the presence of protospacer-adjacent motifs (PAMs) with the sequence NGG. I designed the gRNA sequences to contain a 5' guanine to enable *in vitro* transcription using T7 polymerase. For each gene, gRNA sequences that contained the fewest mismatches in the seed region (mostly 0) were initially selected. I then sorted these gRNA sequences by fewest potential off-target sites that have one mismatch in the seed region, and ultimately chose between four and five gRNA sequences to transcribe for each gene. I followed the methods from (Bassett, et al. 2013; Kistler, et al. 2015) to generate gRNAs using a template-free PCR. A universal CRISPR primer (5'-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTG ATAACGGA CTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3') was annealed to a gene specific primer (see below).

MleKLf5a (*ML00922a*)

gRNA #1: 5'-GAAATTAATACGACTCACTATAGGAGCAACGGGTCCGTCCGTGTTTTAGAGC TAGAAATAGC-3'

gRNA #2: 5'-GAAATTAATACGACTCACTATAGGTTGAGGGACGCGGGAGCAAGTTTTAGAG CTAGAAATAGC-3'

gRNA #3: 5'-GAAATTAATACGACTCACTATAGGACGGAGGGAATCGGCGATGTTTTAGAGC TAGAAATAGC-3'

gRNA #4: 5’-

GAAATTAATACGACTCACTATAGGTTAGACTCTGTCGCGGGGGTTTTAGAGCT

AGAAATAGC-3’

gRNA #5: 5’-

GAAATTAATACGACTCACTATAGGAGCTGACGCGGCACTATGTTTTAGAGCT

AGAAATAGC-3’

MleKlf5b (*ML25776a*)

gRNA #1: 5’-

GAAATTAATACGACTCACTATAGGTGGTGATATACCAGGCGGTTTTAGAGCT

AGAAATAGC

gRNA #2: 5’-

GAAATTAATACGACTCACTATAGGATCTTTCACGCTTAGGGGCGTTTTAGAGC

TAGAAATAGC

gRNA #3: 5’-

GAAATTAATACGACTCACTATAGGCGCTTGGAGGGAACCTAAGTTTTAGAGC

TAGAAATAGC

gRNA #4: 5’-

GAAATTAATACGACTCACTATAGGCTGAAACACCGGTCGCAGGTTTTAGAGC

TAGAAATAGC

MleKlfX (*ML20061a*)

gRNA #1: 5’-

GAAATTAATACGACTCACTATAGGAGATGTGTTCATGGCCCGGTTTTAGAGCT

AGAAATAGC

gRNA #2: 5’-

GAAATTAATACGACTCACTATAGGACACTTTCGTATGCATACGTTTTAGAGCT

AGAAATAGC

gRNA #3: 5’-

GAAATTAATACGACTCACTATAGGAGAAAATGCGGCTCCTTTGTTTTAGAGCT

AGAAATAGC

MleBra (*ML174736a*)

gRNA #1: 5’-

GAAATTAATACGACTCACTATAGGAGGCCGTGCTGGTCGGCAAGTTTTAGAG

CTAGAAATAGC

gRNA #2: 5’-

GAAATTAATACGACTCACTATAGGACCAGCACGGCCTCTATCAGTTTTAGAG

CTAGAAATAGC

gRNA #3: 5’-

GAAATTAATACGACTCACTATAGGTTAACGGCGAGTGGGTACCGTTTTAGAG

CTAGAAATAGC

gRNA #4: 5’-

GAAATTAATACGACTCACTATAGGTGACGTTGGAAGGCGTAGCGTTTTAGAG

CTAGAAATAGC

gRNA #5: 5’-

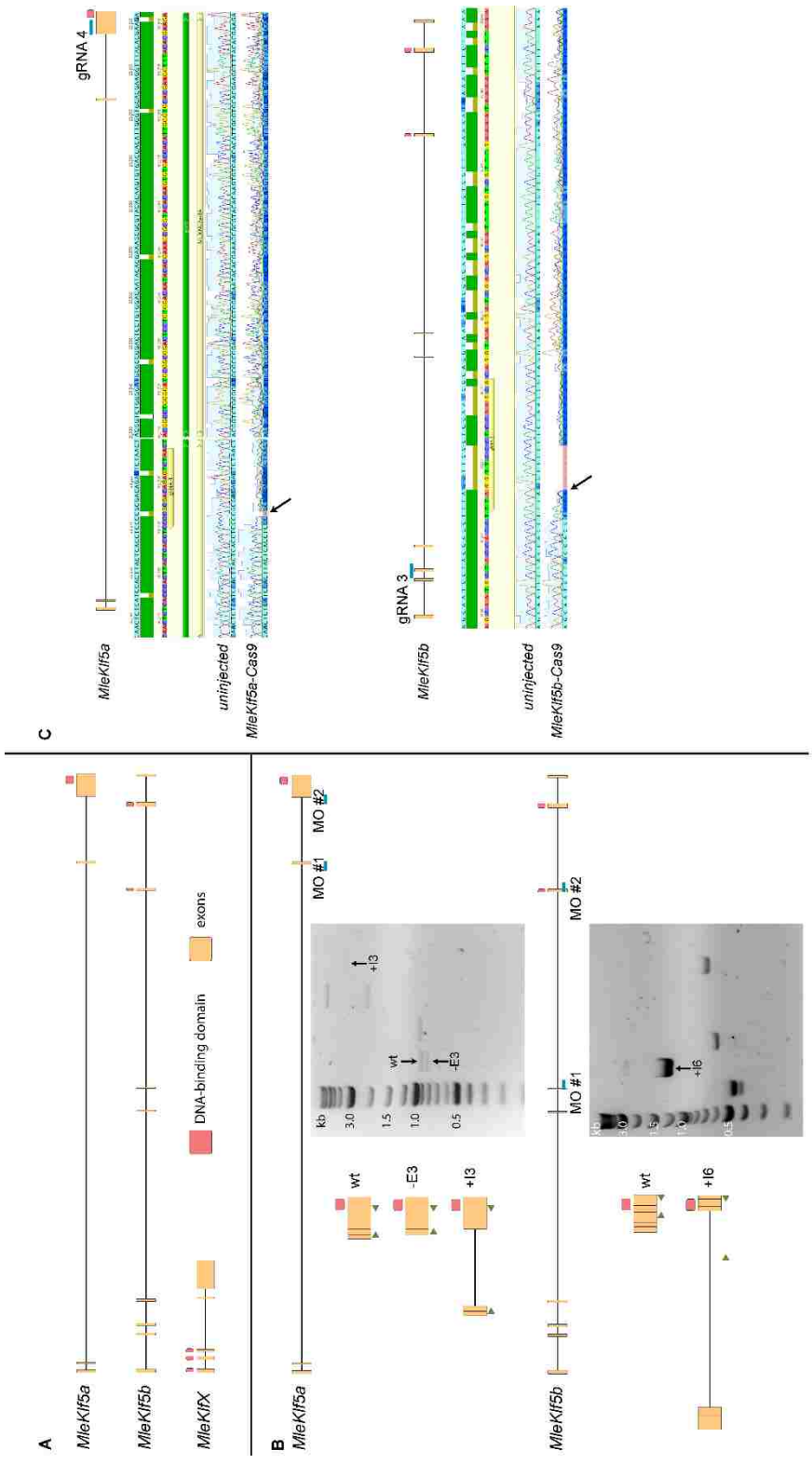
GAAATTAATACGACTCACTATAGGTTCCTCCTTCACGCGAGGAGTTTTAGAGC

TAGAAATAGC

The PCR product was then purified using a PCR purification kit (Qiagen). I then followed the manufacturer's protocol for T7 mediated *in vitro* transcription of gRNAs (MEGAscript™ T7 kit, ThermoFisher Scientific). Cas9 injection cocktail solutions consisted of a final concentration of ~400 ng/µl Cas9 protein, ~100 ng/µl for each gRNA, 35% glycerol, and ~0.5 mg/ml fluorescent dextran (Alexa-Fluor 488, 10,000 MW, ThermoFisher Scientific). Injected embryos were gently transferred to glass dishes filled with FSW and kept at room temperature until they reached the developmental stage of interest. After imaging, genomic DNA was extracted from single embryos using the (QIAamp DNA Micro Kit). Genomic regions flanking putative cut sites, either for *MleKlf5a* and *MleKlf5b,* or for off-target sites, were amplified and Sanger sequenced.

Both sbMO and gRNA/Cas9 solutions were delivered via microinjections into zygotes that had been removed from their chorion and egg jelly (shucked). Embryos must be shucked to be injected (Martindale and Henry 1997c), and for every injection experiment a pool of shucked, uninjected embryos was used as a control for phenotypes caused by shucking. I scored the viability of un-shucked (wildtype) embryos, and found that about 85% of hatchlings counted over multiple spawns developed normally (n=161). Approximately 80% of shucked, uninjected embryos developed normally (n=217), which was not statistically different than the viability rate of un-shucked embryos (Mann Whitney U test; p=0.28).

Figure 5.1 A) Gene models for *MleKlf5a*, *MleKlf5b*, and *MleKlfX*. The zinc finger DNA binding domain is marked with a red square. Exons are yellow squares, and introns are represented as lines. B) Location of MOs for *MleKlf5a* and *MleKlf5b*. Schematics of the different mis-splicing events (on the left) that were detected through RT-PCR (labeled with arrows on the gel). C) Location of a gRNA site for *MleKlf5a* and *MleKlf5b*, and the corresponding sequence trace files showing degradation of the signal at the cut site (arrow). This is due to the presence of INDELS that were created by Cas9 activity.


A
MleKlf5a
MleKlf5b
MleKlfX
DNA-binding domain
exons
B
MleKlf5a
MO #1
MO #2
wt
-E3
+I3
kb
3.0
1.5
1.0
0.5
MleKlf5b
MO #2
MO #1
wt
+I6
C
gRNA 4
MleKlf5a
uninjected
MleKlf5a-Cas9
gRNA 3
MleKlf5b
uninjected
MleKlf5b-Cas9

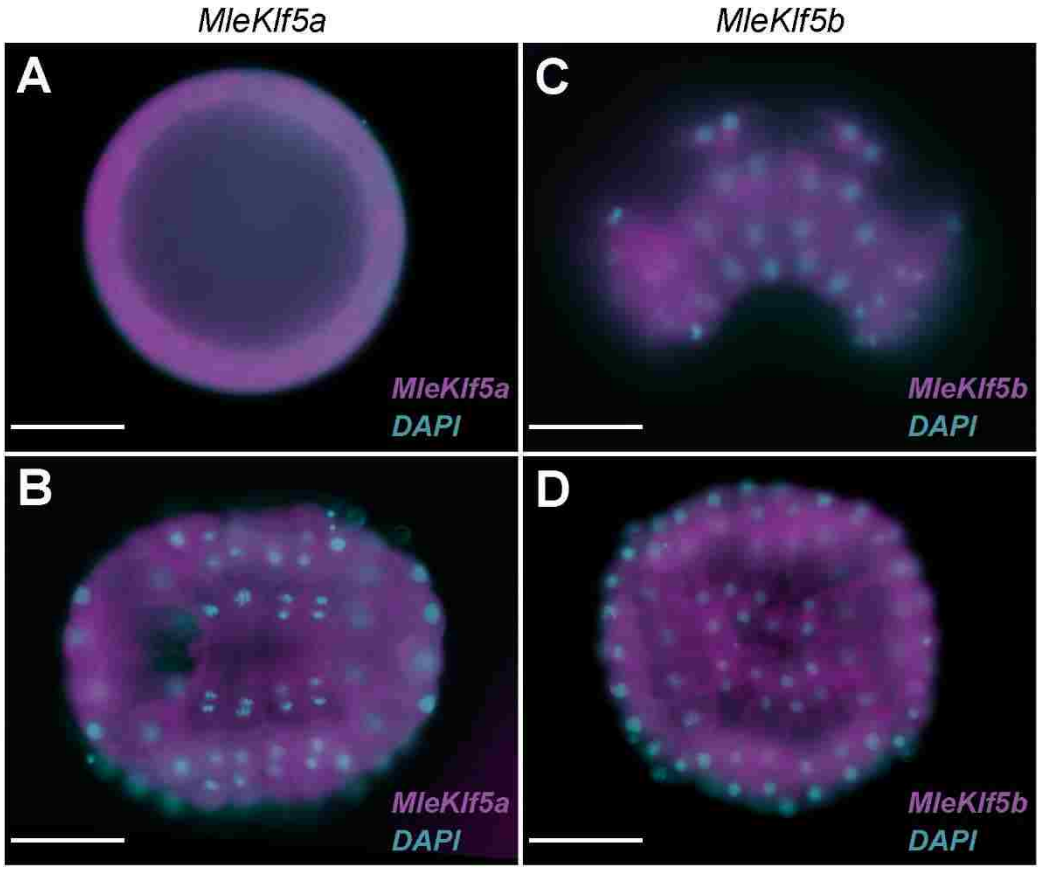
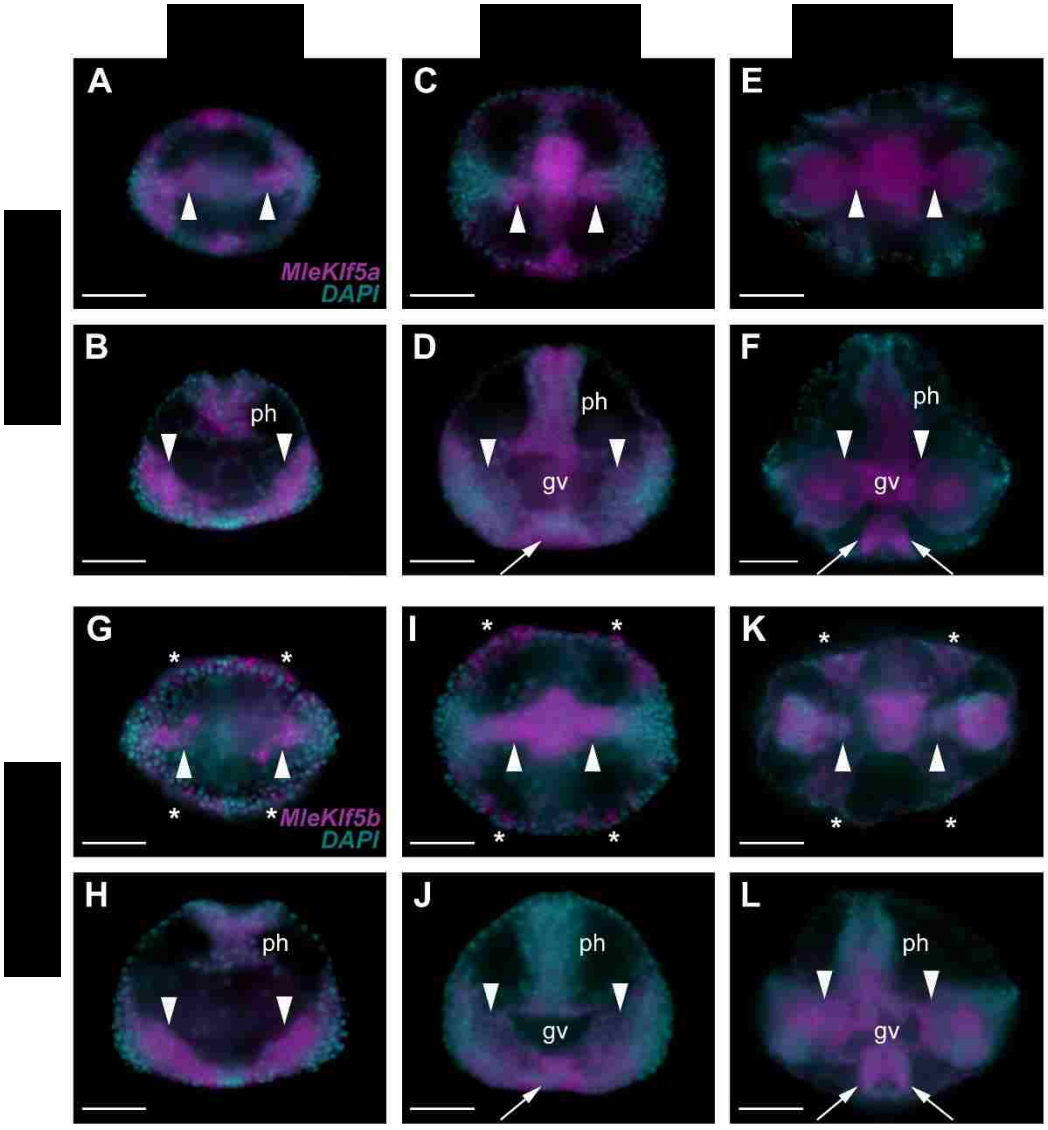


Figure 5.2. Maternally deposited *MleKlf5a* and *MleKlf5b* transcripts. A) Expression of *MleKlf5a* in the zygote stage. Transcript signal is found in the thin ring of cytoplasm. B) *MleKlf5a* expression during gastrulation was found in all cells. C) *MleKlf5b* expression was found in all the blastomeres during all cleavage stages. D) Expression of *MleKlf5b* was found in all cells during gastrulation. In C) the embryo is laterally oriented, oral pole at the top. B) and D) Oral view. Scale bars = 50 μm.

Figure 5.3. Zygotic expression of *MleKlf5a* (A-F) and *MleKlf5b* (G-P). Aboral view in panels (A, C, E, G, I, K), lateral view in panels (B, D, F, H, J, P). During 9 hpf, *MleKlf5a* expression (A, B) was found in the endodermal cells of the developing tentacle bulb (arrowheads), throughout the cells in the pharynx (ph), and in two patches of ectodermal cells along the pharyngeal axis (A). *MleKlf5b* expression (G, H) was also seen in the endodermal cells of the developing tentacle bulb (arrowheads), and in the pharynx (ph). Unlike *MleKlf5a*, *MleKlf5b* was also expressed in the cells surrounding the ctene rows (asterisks). During 14 hpf, expression of *MleKlf5a* (C, D) remained in the pharynx (ph), and was found in the median ridges of the tentacle bulbs (arrowheads). New expression in the developing apical organ was also observed (arrow, D). *MleKlf5b* (I, J) expression remained in the pharynx (ph) and in cells around the ctene rows (asterisks). Median ridge expression (arrowheads) was also seen, along with expression in the apical organ (arrow). These same patterns continued into hatching stages. *MleKlf5a* expression (E, F) was seen in the pharynx (ph), median ridges (arrowheads), lateral ridges, gastrovascular cavity (gv), and two territories in the apical organ floor (arrows). Similar expression patterns were observed for *MleKlf5b*, in addition to the expression patterns seen in the cells around the ctene rows (asterisks). Scale bars = 50 μm.


A
C
E
MleKlf5a
DAPI
B
D
F
ph
gv
G
I
K
MleKlf5b
DAPI
H
J
L

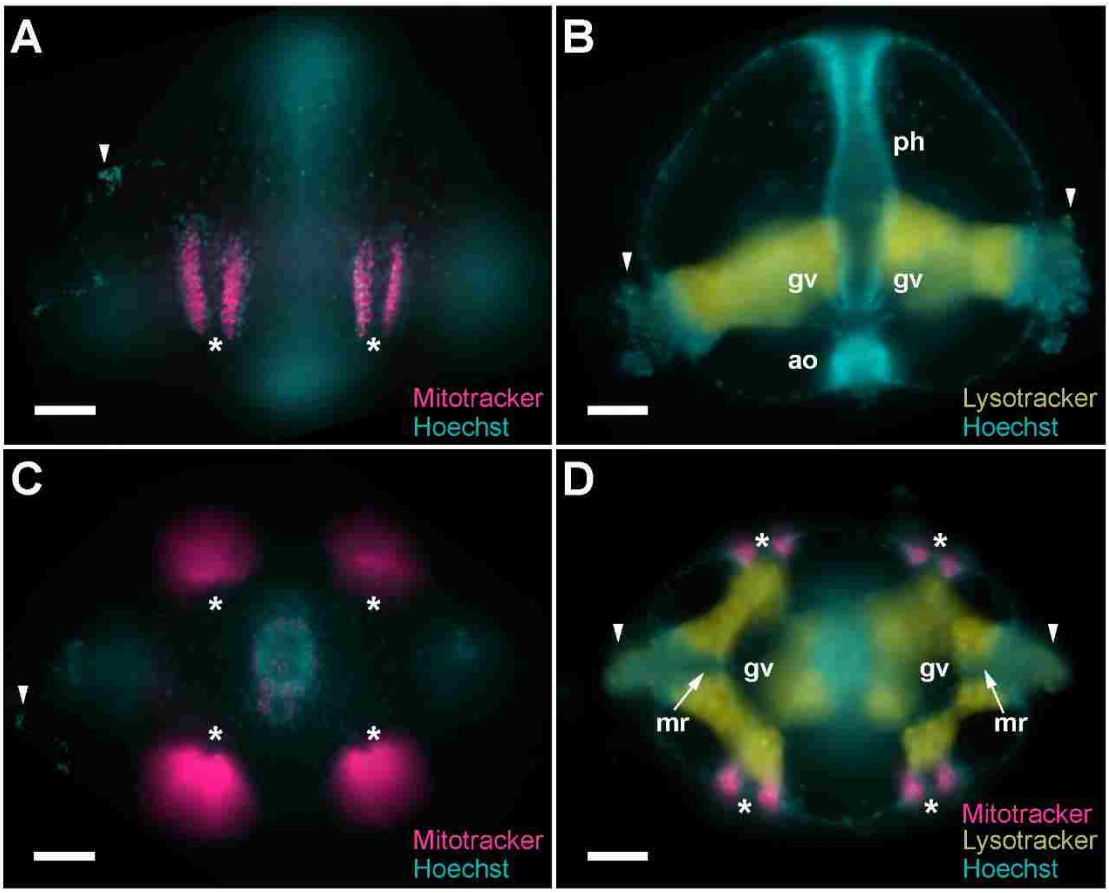


Figure 5.4. Live image of a wildtype cydippid stained with vital dyes. A) Outer lateral epithelium, with the tentacular axis running left to right. The polster cells, which bear the ctene cilia are preferentially stained with Mitotracker (asterisks). Nuclei in the tentacle can be seen (arrowhead). B) Medial plane of lateral view, showing elongated pharynx (ph) connected to the gastrovascular cavity (gv) epithelium which is preferentially stained with Lysotracker. Two tentacle bulbs with some exposed tentacle (arrowheads) can be seen on either side of the of the gastrovascular cavity. Nuclei of the lithocytes can be seen just aboral to the thick apical organ (ao) floor. C) Surface aboral view in which the Mitotracker-positive pairs of ctene rows (asterisks) can be seen in each quadrant. Extended tentacle can be seen (asterisk). The apical organ Mitotracker signal is associated with the polar fields (along the pharyngeal axis) and cells that bear the dome cilia (ring of cells around the apical organ floor). D) Medial plane of aboral view. The four quadrants of the embryo are clearly visible, with a quarter of the of gastrovascular cavity (gv) extending towards each pair of ctene rows (asterisks). The two tentacle bulbs and tentacles (arrowheads) are visible on either side of the embryo, with the median ridge (mr) situated in between the gastrovascular cavity epithelium. In panels (A, B), oral is up, aboral is down. Scale bars = 50 μm.

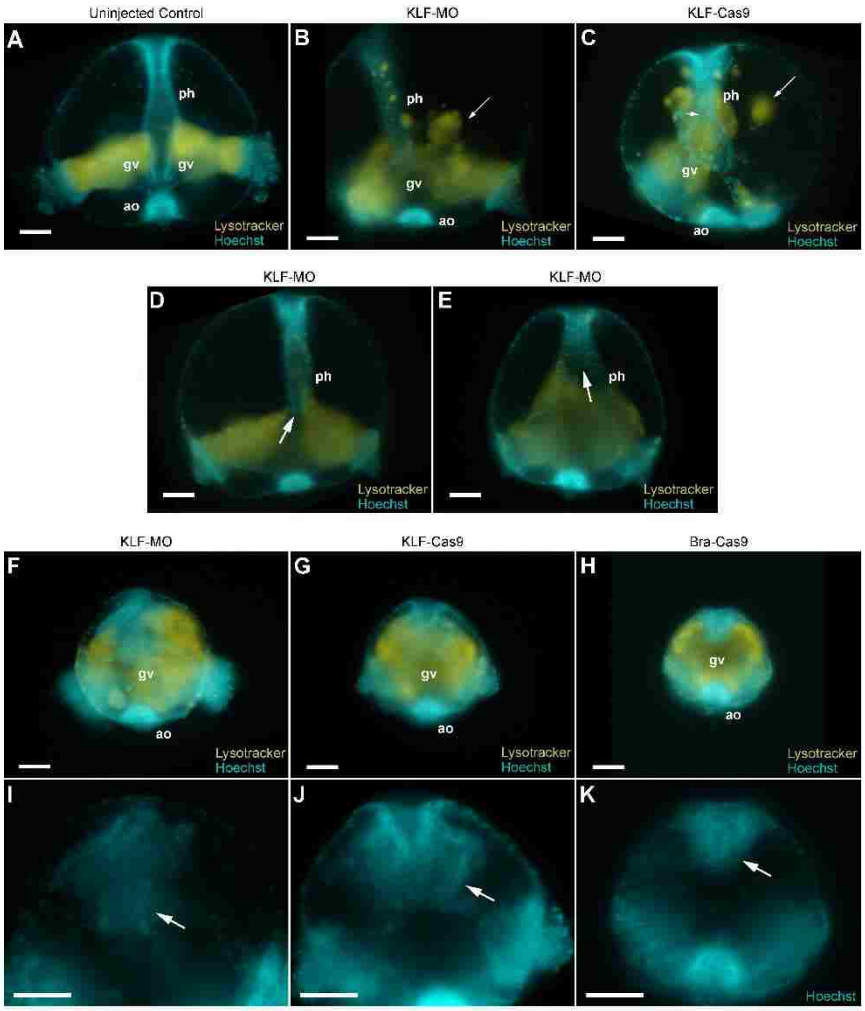


Figure 5.5. Endodermal and pharyngeal phenotypes associated with knockdown of *MleKlf5a* + *MleKlf5b* expression through morpholinos (KLF-MO) and Cas9 (KLF-Cas9), and knockdown of *brachyury* expression with Cas9 (Bra-Cas9). A) Live image of a wildtype embryo from Fig. 5.4. B, C) Live images of embryos with disorganized gastrovascular cavity (gv) epithelium (mild phenotype). These embryos still underwent mesogleal hydration, and pharynx (ph) elongation. Epithelial tissue invading mesogleal space (arrow). D, E) Live images of embryos with defects in the aboral portion of the pharynx. D) The junction between the pharynx and gastrovascular cavity is disrupted (arrow). E) Pharynx with bifurcation at the aboral end (arrow). F, G) Embryos with severe endodermal phenotypes, that failed to undergo mesogleal hydration and embryo growth. The gastrovascular cavity (gv) fills the entire embryo, and the pharynx fails to elongate. H) Bra-Cas9 embryos have a similar phenotype to the severe KLF-Cas9 injected embryos. These embryos also lack pharyngeal elongation, and mesogleal expansion. I-K) Close-up view of embryos in (F-H) showing the difference in KLF-MO/Cas9 and Bra-Cas9 at the oral pole. I, J) Pharyngeal tissue is formed in KLF knockdown embryos (arrow), which fails to proliferate towards the aboral end. K) In Bra-Cas9 embryos, no additional pharyngeal tissue was observed, only stomodeal cells (arrow). Aboral end is down. ao – apical organ. Scale bars = 50 μm.

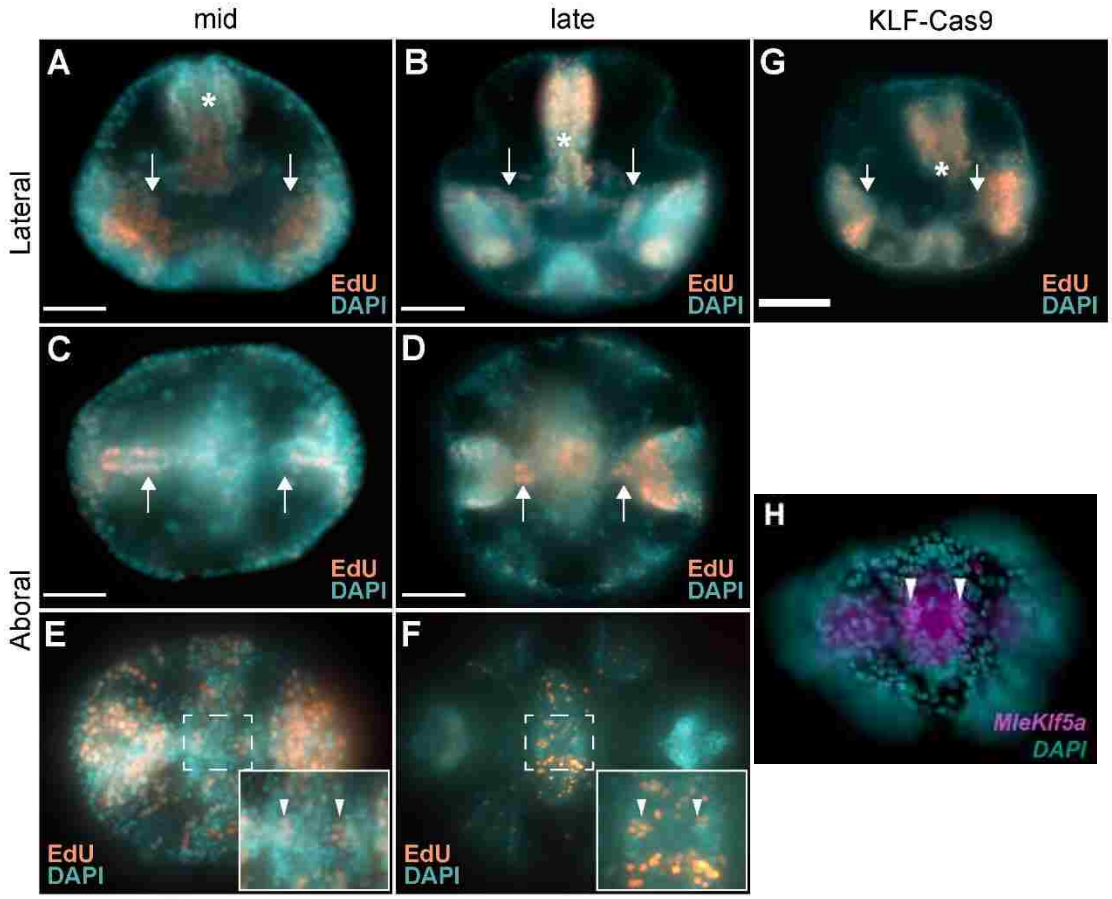


Figure 5.6. Cellular proliferation, measured by EdU incorporation, during embryogenesis. A, B) Lateral view. A) A 14 hpf embryo with high cellular proliferation in the pharynx and endodermal portions of the developing tentacle bulbs (arrows). B) Proliferation occurs in the median ridge (arrows) of the tentacle bulbs. Proliferation in the pharynx occurs in two territories; an oral territory (above the asterisk), and an aboral territory (below the asterisk). C, D) Median plane, aboral view in which the median ridge proliferation can be seen (arrows). E, F) Surface view of the apical organ. The developing polar fields are proliferative, as are two small clusters of cells deeper in the floor of the apical organ (inset). These cells most likely give rise to the lithocytes. G) Knockdown of *MleKlf5a* + *MleKlf5b* expression with Cas9 (KLF-Cas9) results in a loss of proliferation in the median ridge (arrows), as well as the proliferation in the oral portion of the pharynx. H) Aboral view of 20 hpf embryo showing *MleKlf5a* expression in territories of the apical organ where lithocytes most likely are formed (arrowheads). Scale bars = 50 µm.

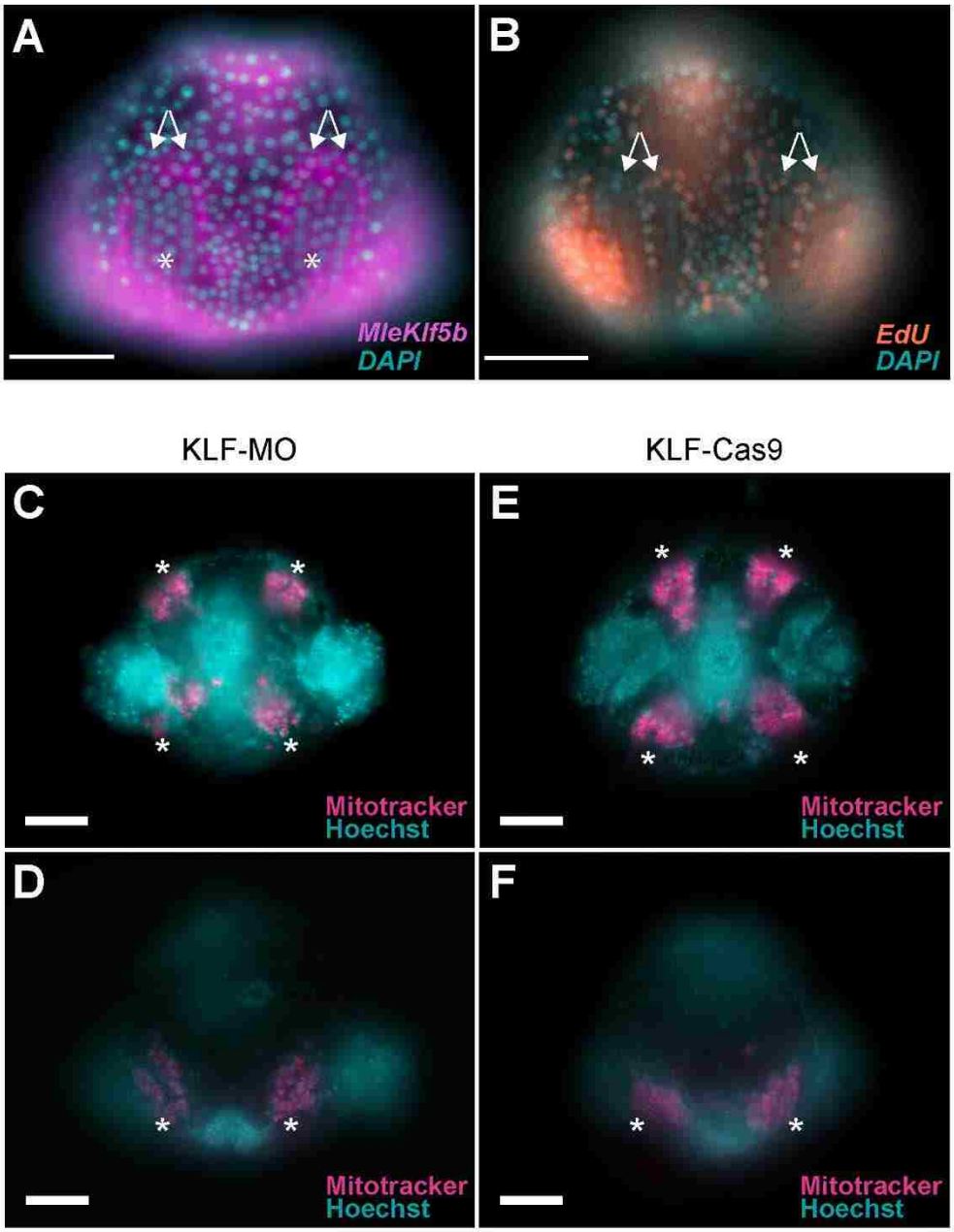


Figure 5.7. *MleKlf5b* expression and EdU incorporation in the ctene row associated cells. A) Image of *in situ* hybridization in 10 hpf embryo showing *MleKlf5b* expression in the cells (arrows) surrounding the polster cells (asterisk). B) These *MleKlf5b*-postive cells are proliferative, as indicated by EdU incorporation. C-F) Knockdown of *MleKlf5a* + *MleKlf5b* expression with MO (KLF-MO) or Cas9 (KLF-Cas9) does not alter polster cell specification, and they retain the high Mitotracker signal (asterisks). C, E) Aboral view. D, F) Lateral view, aboral end is facing down. Scale bars = 50 µm.

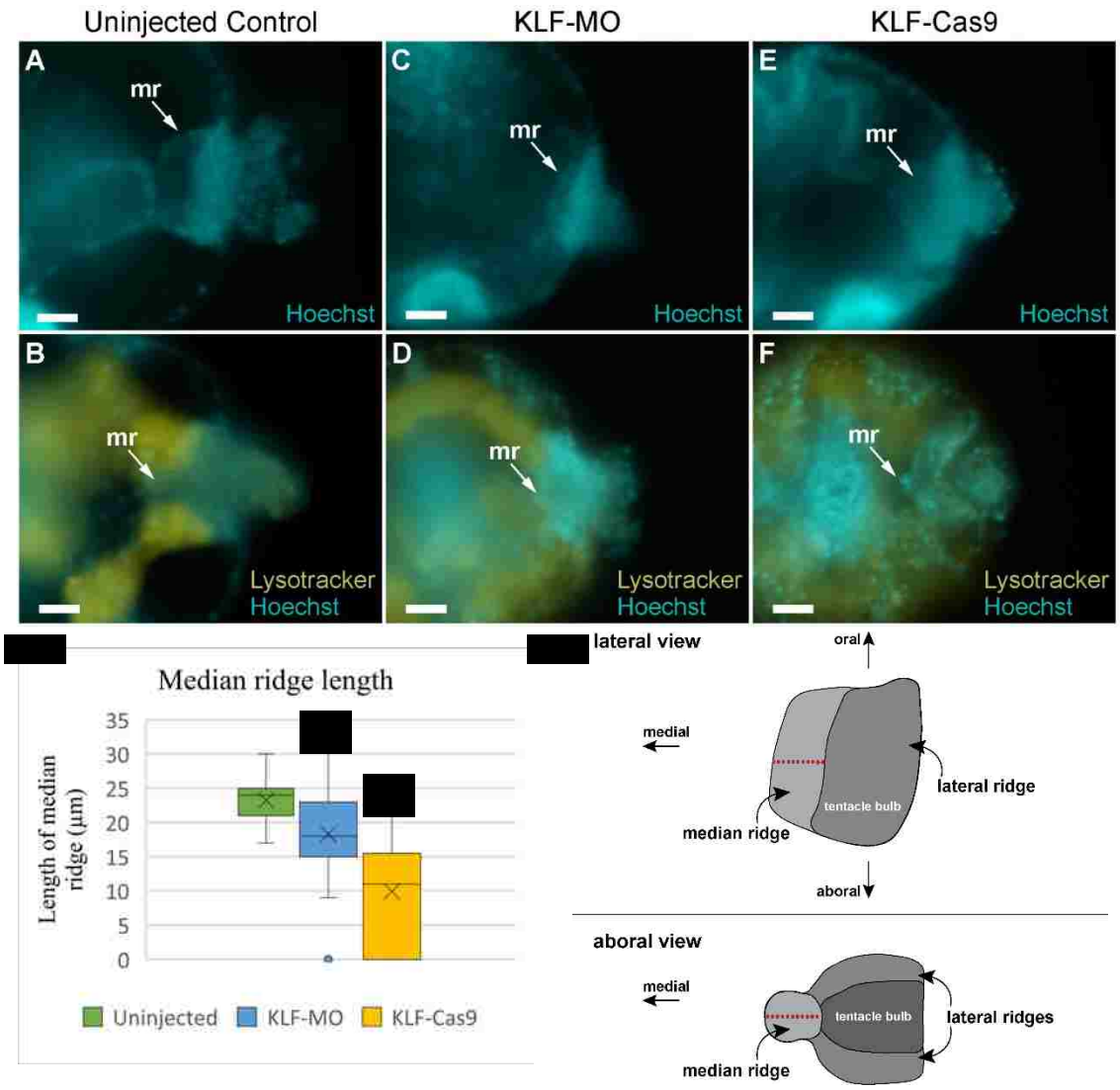


Figure 5.8. Tentacle bulb phenotypes due to knockdown of *MleKlf5a* and *MleKlf5b* expression. A, B) Magnified live image of a tentacle bulb from the wildtype embryo seen in Fig. 5.4. Median ridge (mr). C, D) Knockdown of *MleKlf5a* + *MleKlf5b* with MO (KLF-MO) causes a decrease in the overall tentacle bulb with a pronounced effect on the median ridge (mr). E, F) A similar phenotype was observed in knockdown of *MleKlf5a* + *MleKlf5b* with Cas9 (KLF-Cas9). A, C, E) Lateral view. aboral is down. B, D, F) Aboral view. Scale bars = 25 µm. G) A plot showing the distributions of median ridge lengths in uninjected control embryos (mean 23 µm), KLF-MO embryos (mean 18.29 µm), and KLF-Cas9 embryos (mean 9.88 µm). * denotes significantly different values compared to uninjected controls (Mann Whitney U test; $p<0.05$). H) Schematic of a tentacle bulb in a lateral view (top) and an aboral view (bottom), with features labeled. The tentacle (not shown) is rooted and housed in the space labeled tentacle bulb. Length of the median ridge is denoted with a red, dashed line.

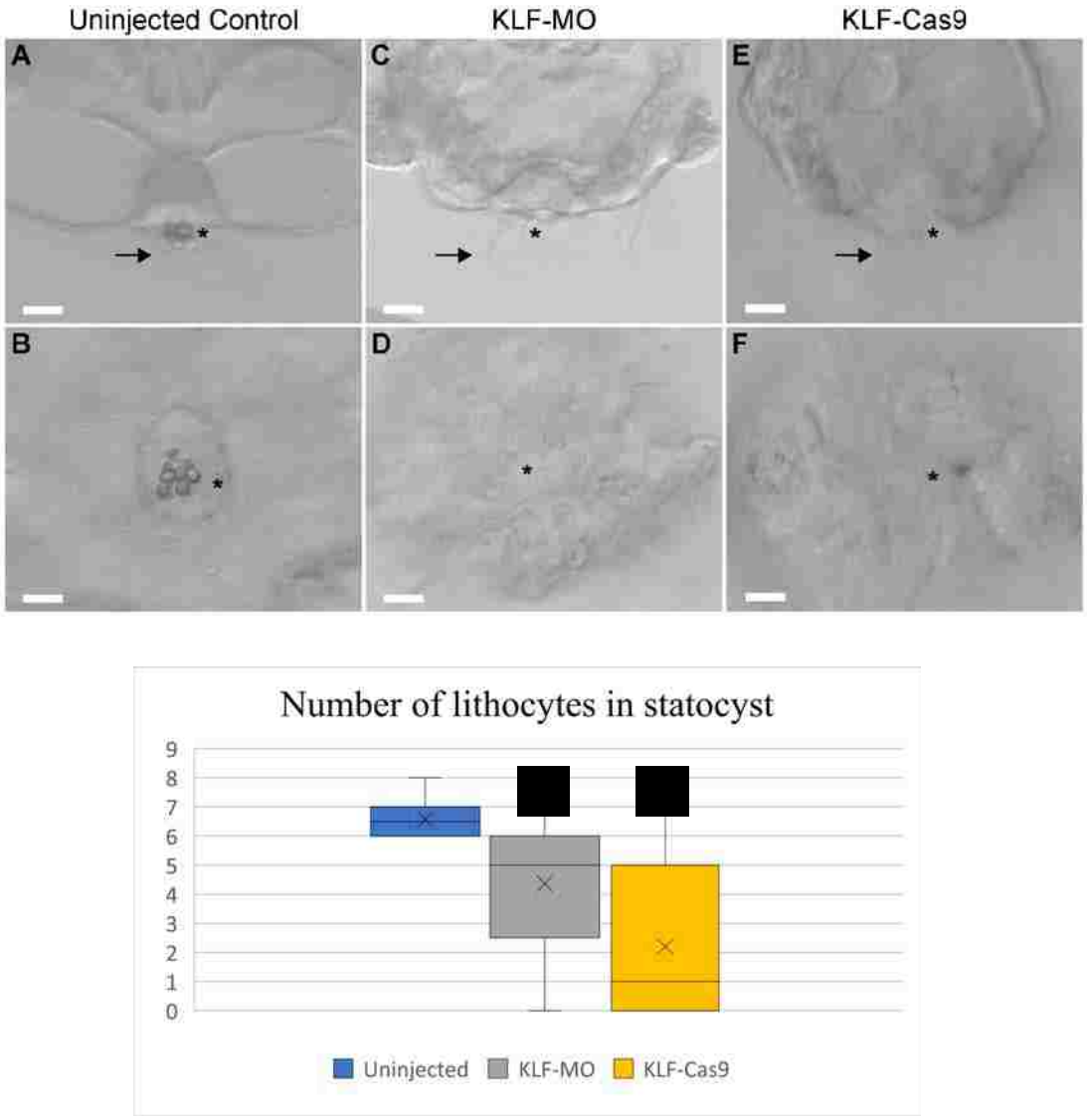


Figure 5.9. Loss of lithocytes due to knockdown of *MleKlf5a* and *MleKlf5b* expression. A) Magnified live image of the aboral, apical organ from the wildtype embryo from Fig. 5.4. The lithocytes (asterisk), between 6-8, are housed in the dome cilia (arrow). B) Knockdown of *MleKlf5a* + *MleKlf5b* with MO (KLF-MO) results in a reduction of lithocyte production, sometimes resulting in a loss of lithocytes. Balancer cilia (asterisk) and dome cilia (arrow) still form. C) A similar phenotype was observed in Cas9-mediated knockdown of *MleKlf5a* + *MleKlf5b* (KLF-Cas9). Scale bars = 25 µm. G) A plot showing the distributions of lithocyte number in uninjected controls (mean of 6.5), KLF-MO embryos (mean of 4.36), and KLF-Cas9 embryos (mean of 2.25). * denotes significantly different values compared to uninjected controls (Mann Whitney U test; $p<0.05$).

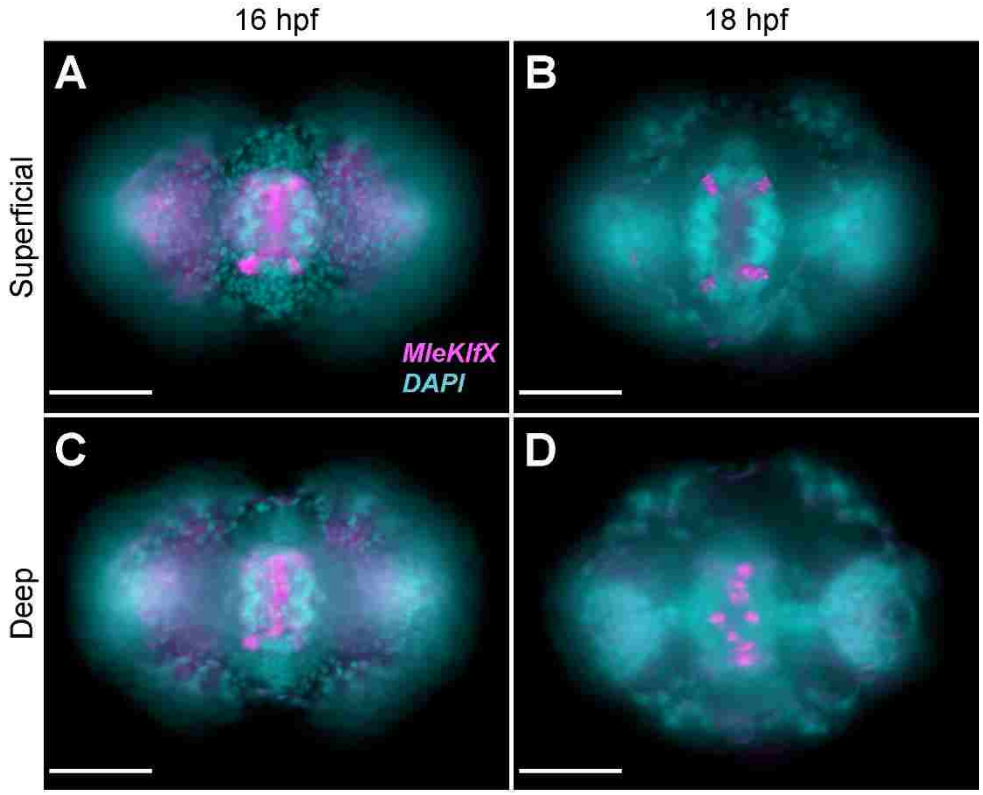


Figure 5.10. Zygotic expression of *MleKlfX*. A, B) On the surface of the apical organ floor, *MleKlfX* is expressed in four clusters of cells, one in each quadrant and medial to the ciliated grooves. C, D) Expression of *MleKlfX* is additionally found deeper in the apical organ floor, in four clusters of cells: two clusters along the pharyngeal axis, and two perpendicular along the tentacular axis. Scale bars = 50 μm.

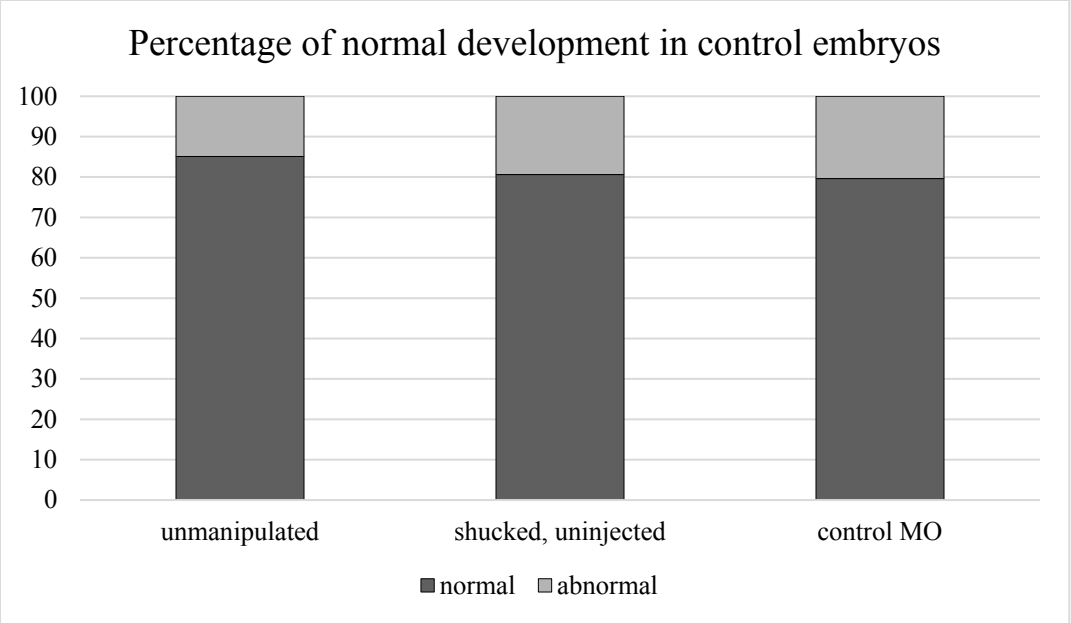


Figure 5.11. Percentage of embryos that had developed normally under different conditions. Unmanipulated embryos were control embryos that were washed and filtered, but not removed from the egg shell. 85.1% of these embryos developed normally (number counted, n=161). Shucked, uninjected embryos were control embryos that were washed, filtered, removed from the egg shell, but not injected. 80.6% of these embryos developed normally (n=217). 79.6% of embryos injected with the control MO developed normally. These results were not statistically different, p=0.28.

Table 5.1. A list of top hits of putative off-target sites for guide RNAs (gRNAs). Genes that contain putative off-target sites identified for each gRNA are shown. KLF-Cas9 injected embryos were sequenced to determine if these genomic sites showed any evidence of off-target Cas9 activity.

Guide RNA (gRNA)	Off target site	Evidence of cut	# embryos sequenced
KLF5a gRNA.4	ML021138a	No	3
	ML090223a	No	3
KLF5a gRNA.5	ML096813a	No	2
	ML233111a	No	1
	ML015610a	No	1
KLF5b.gRNA.3	ML02979a	No	2
	ML200217a	No	2
	ML00332a	No	1
	ML00363a	No	2
KLF5b gRNA.4	ML077212a	No	2

Chapter 6: Synthesis

One of the main goals of evo-devo is to elucidate how changes in developmental processes over time have contributed to the overall evolution of different body plans (Moczek, et al. 2015). A thorough understanding of the cellular and molecular processes underlying development within a phylogenetic context are needed to make reasonable comparisons and inferences of the developmental influence on evolution (Raff 2000; Moczek, et al. 2015). Gene family evolution studies can provide insight into what genes are conserved in multiple lineages, how these genes changed over time, and how those changes contributed to morphological diversity (King, et al. 2008; Degnan, et al. 2009; de Mendoza, et al. 2013; Moczek, et al. 2015). In my dissertation, I elucidated the evolutionary history of the *Klf* gene family and examined their role in regulating development in the ctenophore *Mnemiopsis leidyi*. I also took an in-depth look at ctenophore gut morphology to re-examine the function of the anal pores, and discussed the evolutionary implications of the ctenophore through gut.

Ctenophores are considered one of the earliest branching extant metazoan lineages (Dunn, et al. 2008; Philippe, et al. 2009; Pick, et al. 2010; Ryan, et al. 2013; Moroz, et al. 2014; Hejnol and Martín-Durán 2015; Telford, et al. 2015; Whelan, Kocot and Halanych 2015; Shen, et al. 2017; Whelan, et al. 2017), and are characterized by unique synapomorphic morphologies (Chun 1880; Harbison, et al. 1978; Harbison 1985; Pang and Martindale 2008e). They also exhibit interesting animal trait variations, for example, presence of muscle cells (in some cases striated muscle, Mackie, et al. 1988) but no evidence of many canonical muscle specification genes (Ryan, et al. 2013; Moroz, et al. 2014), and an extensive set of neurons that lack many canonical animal

neurotransmitters and neural fate genes (Jager, et al. 2011; Moroz, et al. 2014). Understanding the genetic regulation of development in ctenophores, and comparing that with what is found in other animals will provide insight into the underlying molecular mechanisms of the evolutionary changes driving metazoan diversification.

Since comparative analyses are a major component of evo-devo studies, one of the major challenges to overcome is the limited number of organisms that can be subjected to detailed functional assessments (Moczek, et al. 2015). Although protocols have been published for standard molecular work in the ctenophore *M. leidyi* (Pang and Martindale 2008g), and plenty of gene expression patterns have been investigated (Yamada and Martindale 2002; Yamada, et al. 2007; Pang and Martindale 2008f; Layden, et al. 2010; Pang, et al. 2010; Pang, et al. 2011; Reitzel, et al. 2011; Schnitzler, et al. 2012; Simmons, et al. 2012; Schnitzler, et al. 2014; Reitzel, et al. 2016), most molecular work in ctenophores is still reliant on the seasonal availability of these animals. Establishment of long-term laboratory cultures and the ability for lab-induced spawning eliminates the reliance on wild caught animals. Therefore, my detailed protocols and guidelines for the establishment of permanent laboratory cultures of *M. leidyi* in *Chapter 2* should make functional, and subsequent comparative, analyses, like that of (Yamada, et al. 2010) and my work presented in *Chapter 5*, more routine.

One advantage of maintaining lab-reared cultures of *M. leidyi* is the ability for more nuanced observations of general morphology and behaviors that would otherwise be missed in the wild. Due to the transparent nature of *M. leidyi* adults, I was able to observe the full cycle of digestion, from ingestion of prey to the release of waste from the anal pores. I showed that *M. leidyi* adults have a through-gut, and discussed in *Chapter 3*

the evolutionary implications of this. The possession of a through gut in ctenophores suggests that the ctenophore lineage independently evolved a through gut or that a loss of the anal opening occurred in other non-bilaterian lineages (Presnell, et al. 2016). Either way, this study shows why more careful morphological, behavioral, and physiological studies are needed for ctenophores (and other non-bilaterians; Dunn, et al. 2008) to aid inferences made about metazoan character trait evolution (Giribet 2015).

Although the functions of the *Krüppel-like factor* (*Klf*) gene family in regulating aspects of development such as stem cell renewal, cell proliferation and differentiation, and fate specification have been characterized in a handful of animals (McConnell and Yang 2010; Bialkowska, et al. 2017), very little has been done to examine the evolutionary history of this gene family. Because of their diverse roles in regulating many cellular processes, studies of *Klf* gene family evolution can provide inferences about how changes in these genes contributed to modifications found during metazoan diversification. I was able to show in *Chapter 4* that *Klf* genes are present in metazoans and closely related unicellular organisms. Overall the mechanisms for *Klf* gene family diversification included domain shuffling, gene duplication, and *de novo* domain evolution. *Klf* gene family expansion mirrored the increased cell type diversity found during metazoan diversification. This may reflect the association of *Klf* gene family variation with the development of cell type repertoires needed to support complex tissues within the Metazoa (Presnell, et al. 2015).

To elucidate the evolution of *Klf* function in metazoans, I examined the expression patterns and molecular function of *Klf* genes during development in the ctenophore *M. leidyi*. I characterized the expression patterns of the three *M. leidyi Klf*

genes identified in *Chapter 4*, *MleKlf5a*, *MleKlf5b*, and *MleKlfX*, during embryogenesis, and found that *MleKlf5a* and *MleKlf5b* were expressed in putative stem cell niches undergoing cellular proliferation. *MleKlfX*, a highly divergent *Klf* gene, was expressed in a spatially restricted territory in the apical organ. The expression of *MleKlfX* was most likely found in differentiated neuronal cells that are thought to be photoreceptors involved in the bioluminescence pathway (Horridge 1964; Schnitzler, et al. 2012). To further characterize these genes during development, I used two independent methods, morpholino oligonucleotides and CRISPR/Cas9, to knockdown gene expression. When *MleKlf5a* and *MleKlf5b* expression was knocked-down, I observed a loss of normal organ development coupled with a loss of cellular proliferation in putative stem cell niches. For example, in the apical organ, the number of lithocytes was reduced, and in the tentacle bulb, the median ridge decreased in size. The results from *Chapter 5* support the idea that regulation of stem cell proliferation and differentiation by *Klf* genes is conserved among metazoans, and represents an ancestral function of this gene family.

Future Directions

Although my dissertation provides useful insight into *Klf* function during ctenophore development, future studies can build off this work to further elucidate the molecular mechanisms regulated by *Klf* genes during *M. leidyi* development.

What genetic mechanisms regulate the development of the unique traits in ctenophores is still an open question (Dunn, et al. 2015). Based on my findings in *Chapter 5*, *MleKlf5a* and *MleKlf5b* genes seem to play a role in the normal development of the lithocytes and the tentacle bulb. To begin to generate gene regulatory networks for

the development of these novel traits, expression analyses of candidate genes via *in situ* hybridization can be assessed in KLF-Cas9 embryos. For example, members of the Wnt and TGF-beta signaling pathways, and genes of the homeobox family exhibit spatially distinct expression patterns in areas of the apical organ and tentacle bulb (Pang and Martindale 2008f; Pang, et al. 2010; Pang, et al. 2011). These genes would be useful in marking specific groups of cells associated with these organs, and to help determine which cell types are specifically affected in KLF-Cas9 embryos. This would help refine the functional role of *MleKlf5a* and *MleKlf5b* during development of the apical organ and tentacle bulbs.

Additionally, RNA sequencing data can be used to generate a broader set of genes that exhibit differential expression upon knockdown of *MleKlf5a* and *MleKlf5b* expression. This would provide insight into what genes are regulated by these *Klf*s in ctenophores. This can be combined with chromatin immunoprecipitation coupled with sequencing (ChIP-seq) to elucidate the downstream targets of both *MleKlf5a* and *MleKlf5b* by identifying where MleKLF5a and MleKLF5b directly bind along the genome. However, a robust antibody is needed for this procedure, and currently, ctenophore-specific antibodies do not exist. One possible workaround to making a KLF-specific antibody, which is difficult due to the highly conserved DNA-binding domain, is to use CRISPR/Cas9 to knock-in tags onto the genes and use antibodies against those tags. For example, a knock-in of GFP onto the end of *MleKlf5a* would enable an antibody against GFP to be used during ChIP-seq to generate information on what genes MleKLF5a is directly binding. Ultimately, these experiments would provide a better

understanding of how an ancient gene family has been co-opted for the development of novel features, such as those found in ctenophores.

In order to make comparisons with other ctenophore species, *Klf* gene expression patterns can also be examined during juvenile stages and in adult tissues in *Mnemiopsis*. *MleKlf5a* and *MleKlf5b* are expressed in putative stem cells in *M. leidyi* embryos (Schnitzler, et al. 2014; Reitzel, et al. 2016), and these areas are similar to what has been identified in adult *Pleurobrachia* (Alié, et al. 2011). Furthermore, adult *Mnemiopsis* have the remarkable ability to regenerate large portions of missing tissue (Coonfield 1936; Martindale 1986). Regeneration requires reprogramming of differentiated cells (Jopling, et al. 2011), and *Klf4* is one of the known transcription factors involved in reprogramming (Takahashi and Yamanaka 2006; Takahashi, et al. 2007). *Klf5* and *Klf4* are closely related, see *Chapter 4*, and thus it is likely that *MleKlf5a* and/or *MleKlf5b* could potentially regulate regeneration in *M. leidyi*.

Dissertation summary

Overall, my dissertation encompassed multiple approaches in order to gain insight into the early evolution of metazoans, including investigation of the functional morphology of the ctenophore gut, as well as, examining the evolutionary history of the *Klf* gene family. I was able to show that *Klf* genes, in ctenophores, regulate cellular proliferation associated with proper organogenesis during embryonic development. My results suggest that regulation of stem cell proliferation was most likely an ancestral function of *Klf* genes in the Metazoa. The results presented here provide a framework for future studies elucidating the molecular genetic regulatory networks underlying

ctenophore development. Comparison to other animals will provide better insight into the role *Klf* genes had on the origins and evolution of important metazoan character traits such as multicellularity, stem cell maintenance, and cell fate specification.

References

Abascal F, Zardoya R, Posada D. 2005. Prottest: Selection of best-fit models of protein evolution. Bioinformatics 21:2104-2105.

Abbott JF. 1907. The morphology of coeloplana: University of Chicago.

Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V. 2012. The revised classification of eukaryotes. Journal of Eukaryotic Microbiology 59:429-514.

Afzelius BA. 1961. The fine structure of the cilia from ctenophore swimming-plates. J Biophys Biochem Cytol 9:383-394.

Agassiz A. 1874. Embryology of the ctenophorae. Memoirs of the American Academy of Arts and Sciences 10:357-398.

Agassiz L. 1850. Contributions to the natural history of the acalephæ of north america. Part ii: On the beroid medusæ of the shores of massachusetts, in their perfect state of development. Memoirs of the American Academy of Arts and Sciences 4:313-374.

Aksoy I, Giudice V, Delahaye E, Wianny F, Aubry M, Mure M, Chen J, Jauch R, Bogu GK, Nolden T. 2014. Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. Nature Communications 5:3719.

Alié A, Leclere L, Jager M, Dayraud C, Chang P, Le Guyader H, Queinnec E, Manuel M. 2011. Somatic stem cells express *Piwi* and *Vasa* genes in an adult ctenophore: Ancient association of "germline genes" with stemness. Dev Biol 350:183-197.

Archer MC. 2011. Role of sp transcription factors in the regulation of cancer cell metabolism. Genes & Cancer 2:712-719.

Arendt D, Nübler-Jung K. 1997. Dorsal or ventral: Similarities in fate maps and gastrulation patterns in annelids, arthropods and chrodates. Mechanisms of Development 61:7-21.

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

Aria MN, Chan IM. 1989. Two types of excretory pores in the hydrozoan medusa *Aequorea victoria* (Murbach and Shearer, 1902). Journal of Plankton Research 11:609-614.

Athanikar JN, Sanchez HB, Osborne TF. 1997. Promoter selective transcriptional synergy mediated by sterol regulatory element binding protein and sp1: A critical role for the btd domain of sp1. Molecular and Cellular Biology 17:5193-5200.

Baker LD. 1973. The ecology of the ctenophore *Mnemiopsis mccradyi* Mayer in Biscayne Bay, Florida. University of Miami technical report, University of Miami-Rosenstiel School of Marine and Atmospheric Science. 131 pp.

Baker LD, Reeve MR. 1974. Laboratory culture of the lobate ctenophore *Mnemiopsis mccradyi* with notes on feeding and fecundity. Marine Biology 26:57-62.

Bargmann W, Jacob K, Rast A. 1972. [Tentacles and colloblasts of the ctenophore *Pleurobrachia pileus*]. Z Zellforsch Mikrosk Anat 123:121-152.

Bassett Andrew R, Tibbit C, Ponting Chris P, Liu J-L. 2013. Highly efficient targeted mutagenesis of *Drosophila* with the Crispr/Cas9 system. Cell Reports 4:220-228.

Basu P, Morris PE, Haar JL, Wani MA, Lingrel JB, Gaensler KM, Lloyd JA. 2005. Klf2 is essential for primitive erythropoiesis and regulates the human and murine embryonic beta-like globin genes in vivo. Blood 106:2566-2571.

Bayha KM, Chang MH, Mariani CL, Richardson JL, Edwards DL, DeBoer TS, Moseley C, Aksoy E, Decker MB, Gaffney PM. 2015. Worldwide phylogeography of the invasive ctenophore *Mnemiopsis leidyi* (Ctenophora) based on nuclear and mitochondrial DNA data. Biological Invasions 17:827.

Benwitz G. 1978. Elektronenmikroskopische untersuchung der colloblasten-entwicklung bei der ctenophore *Pleurobrachia pileus* (Tentaculifera, Cydippea). Zoomorphologie 89:257-278.

Bialkowska AB, Yang VW, Mallipattu SK. 2017. Krüppel-like factors in mammalian stem cells and development. Development 144:737-754.

Black AR, Black JD, Azizkhan-Clifford J. 2001. Sp1 and Krüppel-like factor family of transcription factors in cell growth regulation and cancer. Journal of Cellular Physiology 188:143-160.

Blakeley P, Fogarty NM, Del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P, Niakan KK. 2015. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. Development 142:3151-3165.

Blok LJ, Grossmann ME, Perry JE, Tindall DJ. 1995. Characterization of an early growth response gene, which encodes a zinc finger transcription factor, potentially involved in cell cycle regulation. Molecular Endocrinology (Baltimore, Md.) 9:1610-1620.

Bouwman P, Philipsen S. 2002. Regulation of the activity of sp1-related transcription factors. Molecular and Cellular Endocrinology 195:27-38.

Brown RS, Sander C, Argos P. 1985. The primary structure of transcription factor TFIIIa has 12 consecutive repeats. FEBS letters 186:271-274.

Brunet T, King N. 2017. The origin of animal multicellularity and cell differentiation. Developmental Cell 43:124-140.

Buchsbaum R, Buchsbaum M, Pearse J, Pearse V. 1987. Animals without backbones: An introduction to the invertebrates: University of Chicago Press.

Bumann D, Puls G. 1997. The ctenophore *Mnemiopsis leidyi* has a flow-through system for digestion with three consecutive phases of extracellular digestion. Physiological Zoology 70:1-6.

Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. 2014. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. Molecular Biology and Evolution 31:1402-1413.

Chen Z, Lei T, Chen X, Zhang J. 2009. Porcine klf gene family: Structure, mapping, and phylogenetic analysis. Genomics 95:111 - 119.

Chun C. 1880. Die Ctenophoren des golfes von Neapel und der angrenzenden meeres-abschnitte. Leipzig: W. Engelmann.

Coletta A, Pinney J, Solis D, Marsh J, Pettifer S, Attwood T. 2010. Low-complexity regions within protein sequences have position-dependent roles. BMC Systems Biology 4:43.

Cook T, Gebelein B, Belal M, Mesa K, Urrutia R. 1999. Three conserved transcriptional repressor domains are a defining feature of the TIEG subfamily of sp1-like zinc finger proteins. Journal of Biological Chemistry 274:29500-29504.

Cook T, Gebelein B, Mesa K, Mladek A, Urrutia R. 1998. Molecular cloning and characterization of TIEG2 reveals a new subfamily of transforming growth factor-β-inducible sp1-like zinc finger-encoding genes involved in the regulation of cell growth. Journal of Biological Chemistry 273:25929-25936.

Coonfield BR. 1936. Regeneration in *Mnemiopsis leidyi*, Agassiz. The Biological Bulletin 71:421-428.

Crossley M, Whitelaw E, Perkins A, Williams G, Fujiwara Y, Orkin SH. 1996. Isolation and characterization of the cDNA encoding bKlf/Tef-2, a major CACC-box-binding protein in erythroid cells and selected other cells. Molecular and Cellular Biology 16:1695-1705.

Davidson PL, Koch BJ, Schnitzler CE, Henry JQ, Martindale MQ, Baxevanis AD, Browne WE. 2017. The maternal-zygotic transition and zygotic activation of the *Mnemiopsis leidyi* genome occurs within the first three cleavage cycles. Molecular Reproduction and Development.

Dayraud C, Alié A, Jager M, Chang P, Le Guyader H, Manuel M, Queinnec E. 2012. Independent specialisation of myosin II paralogues in muscle vs. non-muscle functions during early animal evolution: A ctenophore perspective. BMC Evolutionary Biology 12:107.

de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. 2006. Scanprosite: Detection of prosite signature matches and prorule-associated functional and structural residues in proteins. Nucleic Acids Research 34:W362-W365.

De Graeve F, Smaldone S, Laub F, Mlodzik M, Bhat M, Ramirez F. 2003. Identification of the *Drosophila* progenitor of mammalian Krüppel-like factors 6 and 7 and a determinant of fly development. Gene 314:55-62.

de Mendoza A, Sebé-Pedrós A, Šestak MS, Matejčić M, Torruella G, Domazet-Lošo T, Ruiz-Trillo I. 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proceedings of the National Academy of Sciences 110:E4858-E4866.

Degnan BM, Vervoort M, Larroux C, Richards GS. 2009. Early evolution of metazoan transcription factors. Current Opinion in Genetics & Development 19:591-599.

DePristo MA, Zilversmit MM, Hartl DL. 2006. On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. Gene 378:19-30.

Derelle R, Lang BF. 2012. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. Mol Biol Evol 29:1277-1289.

Derelle R, Manuel M. 2007. Ancient connection between nkl genes and the mesoderm? Insights from tlx expression in a ctenophore. Dev Genes Evol 217:253-261.

Draper BW, Morcos PA, Kimmel CB. 2001. Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: A quantifiable method for gene knockdown. Genesis 30:154-156.

Driesch H, Morgan TH. (Driesch1895 co-authors). 1895. Zur analysis der ersten entwickelungsstadien des ctenophoreneies. I. Von der entwickelung einzelner ctenophorenblastomeren. Arch. Enwicklungsmech. Organ. 2:204-215.

Dunn CW. 2017. Ctenophore trees. Nature Ecology & Evolution:1, 1600–1601.

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, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745-749.

Dunn CW, Leys SP, Haddock SHD. 2015. The hidden biology of sponges and ctenophores. Trends in Ecology & Evolution 30:282-291.

Eddy SR. 1998. Profile hidden markov models. Bioinformatics 14:755-763.

Edgar RC. 2004. Muscle: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.

Faial T, Bernardo AS, Mendjan S, Diamanti E, Ortmann D, Gentsch GE, Mascetti VL, Trotter MW, Smith JC, Pedersen RA. 2015. Brachyury and Smad signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. Development 142:2121-2135.

Fairclough S, Chen Z, Kramer E, Zeng Q, Young S, Robertson H, Begovic E, Richter D, Russ C, Westbrook MJ, et al. 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. Genome Biol 14:R15.

Faux NG, Bottomley SP, Lesk AM, Irving JA, Morrison JR, de la Banda MC, Whisstock JC. 2005. Functional insights from the distribution and role of homopeptide repeat-containing proteins. Genome Research 15:537-551.

Feng WC, Southwood CM, Bieker JJ. 1994. Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (eKlf), an erythroid cell-specific transcription factor. J Biol Chem 269:1493-1500.

Fischel A. 1897. Experimentelle untersuchungen am ctenophorenei. Archiv für Entwicklungsmechanik der Organismen 6:109-130.

Fischer AH, Pang K, Henry JQ, Martindale MQ. 2014. A cleavage clock regulates features of lineage-specific differentiation in the development of a basal branching metazoan, the ctenophore *Mnemiopsis leidyi*. EvoDevo 5:4.

Franc JM. 1972. Activites des rosettes ciliees et leurs supports ultrastructuraux chez les ctenaires. Zeitschrift fur Zellforschung und mikroskopische Anatomie 130:527-544.

Freeman G. 1976a. The effects of altering the position of cleavage planes on the process of localization of developmental potential in ctenophores. Developmental Biology 51:332-337.

Freeman G. 1977. The establishment of the oral-aboral axis in the ctenophore embryo. Journal of Embryology and Experimental Morphology 42:237-260.

Freeman G. 1976b. The role of cleavage in the localization of developmental potential in the ctenophore *Mnemiopsis leidyi*. Dev Biol 49:143-177.

Freeman G. 1967. Studies on regeneration in the creeping ctenophore, *Vallicula multiformis*. J Morphol 123:71-83.

Freeman G, Reynolds GT. 1973. The development of bioluminescence in the ctenophore *Mnemiopsis leidyi*. Dev Biol 31:61-100.

Frey H, Leuckart R. 1847. Beiträge zur kenntnis wirbelloser theire: Mit besondere berücksichtigung der fauna des norddeutschen meeres. Braunschweig: Friedrich Vieweg und sohn.

Gao Y, Cao Q, Lu L, Zhang X, Zhang Z, Dong X, Jia W, Cao Y. 2015. Kruppel-like factor family genes are expressed during *Xenopus* embryogenesis and involved in germ layer formation and body axis patterning. Developmental Dynamics 244:1328-1346.

Gardiner MS. 1972. The biology of invertebrates. New York,: McGraw-Hill.

Gerber HP, Seipel K, Georgiev O, Hofferer M, Hug M, Rusconi S, Schaffner W. 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 263:808-811.

Gill G, Pascal E, Tseng ZH, Tjian R. 1994. A glutamine-rich hydrophobic patch in transcription factor sp1 contacts the dTAFII110 component of the Drosophila TFIId complex and mediates transcriptional activation. Proceedings of the National Academy of Sciences 91:192-196.

Giribet G. 2015. Morphology should not be forgotten in the era of genomics–a phylogenetic perspective. Zoologischer Anzeiger 256:96-103.

Gordon AR, Outram SV, Keramatipour M, Goddard CA, Colledge WH, Metcalfe JC, Hager-Theodorides AL, Crompton T, Kemp PR. 2008. Splenomegaly and modified erythropoiesis in Klf13$^{-/-}$ mice. Journal of Biological Chemistry 283:11897-11904.

Granhag L, Møller LF, Hansson LJ. 2011. Size-specific clearance rates of the ctenophore *Mnemiopsis leidyi* based on *in situ* gut content analyses. Journal of Plankton Research 33:1043-1052.

Greve W. 1968. The “planktonkreisel”, a new device for culturing zooplankton. Marine Biology 1:201-203.

Haeckel E. 1875. Die gastraea-theorie, die phylogenetische classification des thierreichs und die homologie der keimblätter. Jenn Z Naturwiss 8:1-55.

Hanna-Rose W, Hansen U. 1996. Active repression mechanisms of eukaryotic transcription repressors. Trends in Genetics 12:229-234.

Harbison GR. 1985. On the classification and evolution of the Ctenophora. In: Morris SC, George JD, Gibson R, Platt HM, editors. The origins and relationships of lower invertebrates. Oxford: Clarendon Press. p. 645-652.

Harbison GR, Madin LP, Swanberg NR. 1978. On the natural history and distribution of oceanic ctenophores. Deep Sea Research 25:233-256.

Harbison GR, Miller RL. (Harbison1986 co-authors). 1986. Not all ctenophores are hermaphrodites. Studies on the systematics, distribution, sexuality and development of two species of *Ocyropsis*. Marine Biology 90:413-424.

Harrison SM, Houzelstein D, Dunwoodie SL, Beddington RSP. 2000. Sp5, a new member of the sp1 family, is dynamically expressed during development and genetically interacts with brachyury. Developmental Biology 227:358-372.

Hejnol A, Martín-Durán JM. 2015. Getting to the bottom of anal evolution. Zoologischer Anzeiger 256:61-74.

Hejnol A, Martindale MQ. 2008. Acoel development indicates the independent evolution of the bilaterian mouth and anus. Nature 456:382-386.

Hejnol A, Martindale MQ. 2009. The mouth, the anus, and the blastopore—open questions about questionable openings. In: Telford MJ, Littlewood DTJ, editors. Animal evolution: Genomes, fossils, and trees: Oxford University Press. p. 33-40.

Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, Martinez P, Baguñà J, Bailly X, Jondelius U, et al. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proceedings of the Royal Society B: Biological Sciences 276:4261-4270.

Henry JQ, Martindale MQ. 2004. Inductive interactions and embryonic equivalence groups in a basal metazoan, the ctenophore *Mnemiopsis leidyi*. Evol Dev 6:17-24.

Henry JQ, Martindale MQ. 2001. Multiple inductive signals are involved in the development of the ctenophore *Mnemiopsis leidyi*. Dev Biol 238:40-46.

Henry JQ, Martindale MQ. 2000. Regulation and regeneration in the ctenophore *Mnemiopsis leidyi*. Dev Biol 227:720-733.

Hernandez-Nicaise M-L. 1991. Ctenophora. In. Microscopic anatomy of invertebrates: Wiley-Liss. p. 359-418.

Hernandez-Nicaise M-L, Amsellem J. 1980. Ultrastructure of the giant smooth muscle fiber of the ctenophore *Beroe ovata*. J Ultrastruct Res 72:151-168.

Hernandez-Nicaise M-L, Nicaise G, Malaval L. 1984. Giant smooth muscle fibers of the ctenophore *Mnemiopsis leydii*: Ultrastructural study of in situ and isolated cells. Biological Bulletin 167:210-228.

Hertwig R. 1880. Über den bau der ctenophoren: Jena: G. Fischer.

Hoey T, Weinzierl ROJ, Gill G, Chen J-L, Dynlacht BD, Tjian R. 1993. Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. Cell 72:247-260.

Horridge GA. 1964. Presumed photoreceptive cilia in a ctenophore. Journal of Cell Science 3:311-317.

Horridge GA. 1965. Relations between nerves and cilia in ctenophores. American Zoologist 5:357-375.

Hyman LH. 1940. Protozoa through Ctenophora. In The invertebrates. New York: McGraw-Hill. p. 662–695.

Ilsley MD, Gillinder KR, Magor GW, Huang S, Bailey TL, Crossley M, Perkins AC. 2017. Krüppel-like factors compete for promoters and enhancers to fine-tune transcription. Nucleic Acids Research 45(11): 6572-7588.

Imataka H, Sogawa K, Yasumoto K, Kikuchi Y, Sasano K, Kobayashi A, Hayami M, Fujiikuriyama Y. 1992. 2 regulatory proteins that bind to the basic transcription element (bte), a GC-box sequence in the promoter region of the rat p-4501a1 gene. Embo Journal 11:3663-3671.

Imhof A, Schuierer M, Werner O, Moser M, Roth C, Bauer R, Buettner R. 1999. Transcriptional regulation of the AP-2α promoter by bteb-1 and ap-2rep, a novel wt-1/egr-related zinc finger repressor. Molecular and Cellular Biology 19:194-204.

Iuchi S. 2001. Three classes of C2H2 zinc finger proteins. Cellular and Molecular Life Sciences CMLS 58:625-635.

Iwafuchi-Doi M, Zaret KS. 2014. Pioneer transcription factors in cell reprogramming. Genes & Development 28:2679-2692.

Jager M, Chiori R, Alie A, Dayraud C, Queinnec E, Manuel M. 2011. New insights on ctenophore neural anatomy: Immunofluorescence study in *Pleurobrachia pileus* (Muller, 1776). Journal of experimental zoology. Part B, Molecular and Developmental Evolution 316B:171-187.

Jager M, Dayraud C, Mialot A, Queinnec E, le Guyader H, Manuel M. 2013. Evidence for involvement of wnt signalling in body polarities, cell proliferation, and the neuro-sensory system in an adult ctenophore. PLoS ONE 8:e84363.

Jager M, Queinnec E, Chiori R, Le Guyader H, Manuel M. 2008. Insights into the early evolution of sox genes from expression analyses in a ctenophore. J Exp Zool B Mol Dev Evol 310:650-667.

Jenner RA, Wills MA. 2007. The choice of model organisms in evo–devo. Nature Reviews Genetics 8:311-314.

Jiang J, Chan Y-S, Loh Y-H, Cai J, Tong G-Q, Lim C-A, Robson P, Zhong S, Ng H-H. 2008. A core klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol 10:353-360.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337:816-821.

Jopling C, Boue S, Belmonte JCI. 2011. Dedifferentiation, transdifferentiation and reprogramming: Three routes to regeneration. Nature Reviews Molecular Cell Biology 12:79-89.

Kaczynski JA, Conley AA, Zapico MF, Delgado SM, Zhang JS, Urrutia R. 2002. Functional analysis of basic transcription element (bte)-binding protein (bteb) 3 and bteb4, a novel sp1-like protein, reveals a subfamily of transcriptional repressors for the bte site of the cytochrome p4501a1 gene promoter. Biochemical Journal 366:873-882.

Kaczynski JA, Cook T, Urrutia R. 2003. Sp1-and Kruppel-like transcription factors. Genome Biol 4.

Kadonaga JT, Carner KR, Masiarz FR, Tijian R. 1987. Isolation of cDNA-encoding transcription factor sp1 and functional-analysis of the DNA-binding domain. Cell 51:1079-1090.

Kim C-K, He P, Bialkowska AB, Yang VW. 2017. Sp and klf transcription factors in digestive physiology and diseases. Gastroenterology 152:1845-1875.

King N, Rokas A. 2017. Embracing uncertainty in reconstructing early animal evolution. Current Biology 27:R1081-R1088.

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, et al. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451:783-788.

Kistler KE, Vosshall LB, Matthews BJ. 2015. Genome engineering with CRISPR-Cas9 in the mosquito *Aedes aegypti*. Cell Reports 11:51-60.

Kohn AB, Citarella MR, Kocot KM, Bobkova YV, Halanych KM, Moroz LL. 2012. Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore, *Pleurobrachia bachei*. Mol Phylogenet Evol 63:203-207.

Kolell KJ, Crawford DL. 2002. Evolution of sp transcription factors. Molecular Biology and Evolution 19:216-222.

Komai T. 1922. Studies on two aberrant ctenophores: Coeloplana and Gastrodes: Kyoto, Japan.

Kotkamp K, Mössner R, Allen A, Onichtchouk D, Driever W. 2014. A Pou5f1/Oct4 dependent Klf2a, Klf2b, and Klf17 regulatory sub-network contributes to evl and ectoderm development during zebrafish embryogenesis. Developmental Biology 385:433-447.

Laherty CD, Yang W-M, Sun J-M, Davie JR, Seto E, Eisenman RN. 1997. Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell 89:349-356.

Laumer CE, Gruber-Vodicka H, Hadfield MG, Pearse VB, Riesgo A, Marioni JC, Giribet G. 2017. Placozoans are eumetazoans related to Cnidaria. bioRxiv 200972; doi: https://doi.org/10.1101/200972.

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 1:1-16.

Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, et al. 2016. The mid-developmental transition and the evolution of animal body plans. Nature 531:637.

Leys SP. 2015. Elements of a 'nervous system'in sponges. Journal of Experimental Biology 218:581-591.

Li IC, Chan C-T, Lu Y-F, Wu Y-T, Chen Y-C, Li G-B, Lin C-Y, Hwang S-PL. 2011. Zebrafish Krüppel-like factor 4a represses intestinal cell proliferation and promotes differentiation of intestinal cell lineages. PLoS ONE 6:e20974.

Lomberk G, Grzenda A, Mathison A, Escande C, Zhang J-S, Calvo E, Miller LJ, Iovanna J, Chini EN, Fernandez-Zapico ME. 2013. Krüppel-like factor 11 regulates the expression of metabolic genes via an evolutionarily conserved protein interaction domain functionally disrupted in maturity onset diabetes of the young. Journal of Biological Chemistry 288:17745-17758.

Mackie GO, Mills CE, Singla CL. 1988. Structure and function of the prehensile tentilla of *Euplokamis* (Ctenophora, Cydippida). Zoomorphology 107:319-337.

Main RJ. 1928. Observations of the feeding mechanism of a ctenophore, *Mnemiopsis leidyi*. The Biological Bulletin 55:69-78.

Martindale MQ. 1987. Larval reproduction in the ctenophore *Mnemiopsis mccradyi* (order Lobata). Marine Biology 94:409-414.

Martindale MQ. 1986. The ontogeny and maintenance of adult symmetry properties in the ctenophore, *Mnemiopsis mccradyi*. Dev Biol 118:556-576.

Martindale MQ, Finnerty JR, Henry JQ. 2002. The radiata and the evolutionary origins of the bilaterian body plan. Molecular Phylogenetics and Evolution 24:358-365.

Martindale MQ, Hejnol A. 2009. A developmental perspective: Changes in the position of the blastopore during bilaterian evolution. Developmental Cell 17:162-174.

Martindale MQ, Henry JQ. 1995. Diagonal development: Establishment of the anal axis in the ctenophore *Mnemiopsis leidyi*. The Biological Bulletin 189:190-192.

Martindale MQ, Henry JQ. 1996. Development and regeneration of comb plates in the ctenophore *Mnemiopsis leidyi*. The Biological Bulletin 191:290-292.

Martindale MQ, Henry JQ. 1997a. The Ctenophora. Gilbert SF, Raunio AM, editors. Embryology: Constructing the organism: Sinauer Press. p. 87-111.

Martindale MQ, Henry JQ. 1997b. Experimental analysis of tentacle formation in the ctenophore *Mnemiopsis leidyi*. The Biological Bulletin 193:245.

Martindale MQ, Henry JQ. 1997c. Reassessing embryogenesis in the ctenophora: The inductive role of e1 micromeres in organizing ctene row formation in the 'mosaic' embryo, *Mnemiopsis leidyi*. Development 124:1999-2006.

Martindale MQ, Henry JQ. 1998. The development of radial and biradial symmetry: The evolution of bilaterality. American Zoologist 38:672-684.

Martindale MQ, Henry JQ. 1999. Intracellular fate mapping in a basal metazoan, the ctenophore *Mnemiopsis leidyi*, reveals the origins of mesoderm and the existence of indeterminate cell lineages. Dev Biol 214:243-257.

Martindale MQ, Pang K, Finnerty JR. 2004. Investigating the origins of triploblasty: `mesodermal' gene expression in a diploblastic animal, the sea anemone *Nematostella vectensis* (phylum, Cnidaria; class, Anthozoa). Development 131:2463-2474.

Materna SC, Howard-Ashby M, Gray RF, Davidson EH. 2006. The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. Developmental Biology 300:108-120.

Maxwell EK, Ryan JF, Schnitzler CE, Browne WE, Baxevanis AD. 2012. MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore *Mnemiopsis leidyi*. BMC Genomics 13:714.

Mayor AG. 1912. Ctenophores of the Atlantic coast of North America: Carnegie Institution of Washington.

McConnell BB, Ghaleb AM, Nandan MO, Yang VW. 2007. The diverse functions of Krüppel-like factors 4 and 5 in epithelial biology and pathobiology. BioEssays 29:549-557.

McConnell BB, Yang VW. 2010. Mammalian Krüppel-like factors in health and diseases. Physiological Reviews 90:1337-1381.

Meadows SM, Salanga MC, Krieg PA. 2009. Kruppel-like factor 2 cooperates with the ets family protein erg to activate flk1 expression during vascular development. Development 136:1115-1125.

Miller I, Bieker J. 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Mol Cell Biol 13:2776 - 2786.

Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIa from *Xenopus* oocytes. Embo Journal 4:1609-1614.

Mills CE, Haddock SHD. 2007. Key to the Ctenophora. Carlton JT, editor. Light and Smith's manual: Intertidal invertebrates of the central California coast: University of California Press. p. 189-199.

Moczek AP, Sears KE, Stollewerk A, Wittkopp PJ, Diggle P, Dworkin I, Ledon-Rettig C, Matus DQ, Roth S, Abouheif E. 2015. The significance and scope of evolutionary developmental biology: A vision for the 21st century. Evolution & Development 17:198-219.

Moreland RT, Nguyen AD, Ryan JF, Schnitzler CE, Koch BJ, Siewert K, Wolfsberg TG, Baxevanis AD. 2014. A customized web portal for the genome of the ctenophore *Mnemiopsis leidyi*. BMC Genomics 15:316.

Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM, et al. 2014. The ctenophore genome and the evolutionary origins of neural systems. Nature 510:109-114.

Moss AG, Estes AM, Muellner LA, Morgan DD. 2001. Protistan epibionts of the ctenophore *Mnemiopsis mccradyi* Mayer. Hydrobiologia 451:295-304.

Muñoz-Descalzo S, Terol J, Paricio N. 2005. Cabut, a C2H2 zinc finger transcription factor, is required during *Drosophila* dorsal closure downstream of jnk signaling. Developmental Biology 287:168-179.

Nagai R, Friedman SL, Kasuga M. 2009. The biology of Krüppel-like factors: Springer.

Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B. 2002. The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. Cell 108:17-29.

Nandan MO, Yang VW. 2009. The role of krüppel-like factors in the reprogramming of somatic cells to induced pluripotent stem cells. Histology and Histopathology 24:1343.

Nosenko T, Schreiber F, Adamska M, Adamski M, Eitel M, Hammel J, Maldonado M, Müller WEG, Nickel M, Schierwater B, et al. 2013. Deep metazoan phylogeny: When different genes tell different stories. Molecular Phylogenetics and Evolution 67:223-233.

Nosrati N, Kapoor NR, Kumar V. 2014. Combinatorial action of transcription factors orchestrates cell cycle-dependent expression of the ribosomal protein genes and ribosome biogenesis. FEBS Journal 281:2339-2352.

Ohnishi S, Laub F, Matsumoto N, Asaka M, Ramirez F, Yoshida T, Terada M. 2000. Developmental expression of the mouse gene coding for the Krüppel-like transcription factor Klf5. Developmental Dynamics 217:421-429.

Pan YA, Livet J, Sanes JR, Lichtman JW, Schier AF. 2011. Multicolor brainbow imaging in zebrafish. Cold Spring Harbor Protocols 2011:pdb.prot5546.

Pang K, Martindale MQ. 2008a. Ctenophore tissue preparation and extraction of DNA. CSH Protoc 2008:pdb prot5088.

Pang K, Martindale MQ. 2008b. Ctenophore tissue preparation and extraction of rna. CSH Protoc 2008:pdb prot5089.

Pang K, Martindale MQ. 2008c. Ctenophore whole-mount antibody staining. CSH Protoc 2008:pdb prot5086.

Pang K, Martindale MQ. 2008d. Ctenophore whole-mount in situ hybridization. CSH Protoc 2008:pdb prot5087.

Pang K, Martindale MQ. 2008e. Ctenophores. Current Biology 18:R1119-R1120.

Pang K, Martindale MQ. 2008f. Developmental expression of homeobox genes in the ctenophore *Mnemiopsis leidyi.* Dev Genes Evol 218:307-319.

Pang K, Martindale MQ. 2008g. *Mnemiopsis leidyi* spawning and embryo collection. CSH Protoc 2008:pdb prot5085.

Pang K, Ryan JF, Baxevanis AD, Martindale MQ. 2011. Evolution of the Tgf-beta signaling pathway and its potential role in the ctenophore, *Mnemiopsis leidyi*. PLoS ONE 6:e24152.

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

Papaioannou VE. 2014. The t-box gene family: Emerging roles in development, stem cells and cancer. Development 141:3819-3833.

Paps J, Medina-Chacón LA, Marshall W, Suga H, Ruiz-Trillo I. 2013. Molecular phylogeny of unikonts: New insights into the position of Apusomonads and Ancyromonads and the internal relationships of Opisthokonts. Protist 164:2-12.

Parisi S, Cozzuto L, Tarantino C, Passaro F, Ciriello S, Aloia L, Antonini D, De Simone V, Pastore L, Russo T. 2010. Direct targets of Klf5 transcription factor contribute to the maintenance of mouse embryonic stem cell undifferentiated state. BMC Biol 8:128.

Parisi S, Passaro F, Aloia L, Manabe I, Nagai R, Pastore L, Russo T. 2008. Klf5 is involved in self-renewal of mouse embryonic stem cells. Journal of Cell Science 121:2629-2634.

Parker-Katiraee L, Carson AR, Yamada T, Arnaud P, Feil R, Abu-Amero SN, Moore GE, Kaneda M, Perry GH, Stone AC, et al. 2007. Identification of the imprinted *Klf14* transcription factor undergoing human-specific accelerated evolution. PLoS Genet 3:e65.

Pearse V, Pearse J, Buchsbaum M, Buchsbaum R. 1987. Living invertebrates: Blackwell/Boxwood.

Pett W, Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxevanis AD, Lavrov DV. 2011. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: Insight from mtDNA and the nuclear genome. Mitochondrial DNA 22:130-142.

Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Quéinnec E. 2009. Phylogenomics revives traditional views on deep animal relationships. Current Biology 19:706-712.

Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, Wiens M, Alié A, Morgenstern B, Manuel M, et al. 2010. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. Molecular Biology and Evolution 27:1983-1987.

Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, Lartillot N, Wörheide G. (Pisani2015 co-authors). 2015. Genomic data do not support comb jellies as the sister group to all other animals. Proc Natl Acad Sci 112.

Piskacek M. 2009. Common transactivation motif 9aatad recruits multiple general co-activators taf9, med15, cbp and p300. Nature Precedings.

Piskacek S, Gregor M, Nemethova M, Grabner M, Kovarik P, Piskacek M. 2007. Nine-amino-acid transactivation domain: Establishment and prediction utilities. Genomics 89:756-768.

Podar M, Haddock SHD, Sogin ML, Harbison GR. 2001. A molecular phylogenetic framework for the phylum Ctenophora using 18s rrna genes. Molecular Phylogenetics and Evolution 21:218-230.

Presnell JS, Schnitzler CE, Browne WE. 2015. Klf/sp transcription factor family evolution: expansion, diversification, and innovation in eukaryotes. Genome Biology and Evolution 7:2289-2309.

Presnell JS, Vandepas Lauren E, Warren Kaitlyn J, Swalla Billie J, Amemiya Chris T, Browne William E. 2016. The presence of a functionally tripartite through-gut in Ctenophora has implications for metazoan character trait evolution. Current Biology 26:2814-2820.

Promponas VJ, Enright AJ, Tsoka S, Kreil DP, Leroy C, Hamodrakas S, Sander C, Ouzounis CA. 2000. Cast: An iterative algorithm for the complexity analysis of sequence tracts. Bioinformatics 16:915-922.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. 2012. The pfam protein families database. Nucleic Acids Research 40:D290-D301.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86-94.

Radding CM. 1982. Homologous pairing and strand exchange in genetic recombination. Annual Review of Genetics 16:405-437.

Raff RA. 2000. Evo-devo: The evolution of a new discipline. Nature Reviews Genetics 1:74-79.

Raskoff KA, Sommer FA, Hamner WM, Cross KM. 2003. Collection and culture techniques for gelatinous zooplankton. The Biological Bulletin 204:68-80.

Ravasi T, Huber T, Zavolan M, Forrest A, Gaasterland T, Grimmond S, Group RG, Members G, Hume DA. 2003. Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome. Genome Research 13:1430-1442.

Reeve MR, Walter MA, Ikeda T. 1978. Laboratory studies of ingestion and food utilization in lobate and tentaculate ctenophores. Limnology and Oceanography 23:740-751.

Reitzel AM, Daly M, Sullivan JC, Finnerty JR. 2009. Comparative anatomy and histology of developmental and parasitic stages in the life cycle of the lined sea anemone *Edwardsiella lineata*. Journal of Parasitology 95:100-112.

Reitzel AM, Pang K, Martindale MQ. (Reitzel2016 co-authors). 2016. Developmental expression of “germline”- and “sex determination”-related genes in the ctenophore *Mnemiopsis leidyi*. EvoDevo 7:1-16.

Reitzel AM, Pang K, Ryan JF, Mullikin JC, Martindale MQ, Baxevanis AD, Tarrant AM. 2011. Nuclear receptors from the ctenophore *Mnemiopsis leidyi* lack a zinc-finger DNA-binding domain: Lineage-specific loss or ancestral condition in the emergence of the nuclear receptor superfamily? EvoDevo 2:3.

Reitzel AM, Sullivan JC, Brown BK, Chin DW, Cira EK, Edquist SK, Genco BM, Joseph OC, Kaufman CA, Kovitvongsa K. 2007. Ecological and developmental dynamics of a host-parasite system involving a sea anemone and two ctenophores. Journal of Parasitology 93:1392-1402.

Rodriguez I. 2011. Drosophila tieg is a modulator of different signalling pathways involved in wing patterning and cell proliferation. PLoS ONE 6:e18418.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rosenberg UB, Schroder C, Preiss A, Kienlin A, Cote S, Riede I, Jackle H. 1986. Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIa. Nature 319:336-339.

Rosselló RA, Chen C-C, Dai R, Howard JT, Hochgeschwender U, Jarvis ED. 2013. Mammalian genes induce partially reprogrammed pluripotent stem cells in non-mammalian vertebrate and invertebrate species. eLife 2:e00036.

Rubin GM, Yandell MD, Wortman JR, Miklos GLG, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, et al. 2000. Comparative genomics of the eukaryotes. Science 287:2204-2215.

Ryan JF. 2014. Did the ctenophore nervous system evolve independently? Zoology (Jena) 117:225-226.

Ryan JF, Pang K, Program NCS, Mullikin JC, Martindale MQ, Baxevanis AD. 2010. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the parahoxozoa. EvoDevo 1:9.

Ryan JF, Pang K, Schnitzler CE, Nguyen A, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, Smith SA, et al. 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science 342:1242592.

Ryan JF, Schnitzler CE, Tamm SL. (Ryan2016 co-authors). 2016. Meeting report of ctenopalooza: The first international meeting of ctenophorologists. EvoDevo 7:19.

Schaeper ND, Prpic N-M, Wimmer EA. 2010. A clustered set of three sp-family genes is ancestral in the Metazoa: Evidence from sequence analysis, protein domain structure, developmental expression patterns and chromosomal location. BMC Evolutionary Biology 10:88.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: An open-source platform for biological-image analysis. Nature Methods 9:676-682.

Schlichter D. 1991. A perforated gastrovascular cavity in the symbiotic deep-water coral *Leptoseris fragilis*: A new strategy to optimize heterotrophic nutrition. Helgoländer Meeresuntersuchungen 45:423-443.

Schmidt-Rhaesa A. 2007. Evolution of organ systems: Oxford University Press.

Schnitzler CE, Pang K, Powers ML, Reitzel AM, Ryan JF, Simmons D, Tada T, Park M, Gupta J, Brooks SY, et al. 2012. Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidyi*: A new view of ctenophore photocytes. BMC Biol 10:107.

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.

Scholz CB, Technau U. 2003. The ancestral role of *brachyury*: Expression of *nembra1* in the basal cnidarian *Nematostella vectensis* (Anthozoa). Development Genes and Evolution 212:563-570.

Schuetz A, Nana D, Rose C, Zocher G, Milanovic M, Koenigsmann J, Blasig R, Heinemann U, Carstanjen D. 2011. The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. Cellular and Molecular Life Sciences 68:3121-3131.

Sebé-Pedrós A, Ariza-Cosano A, Weirauch MT, Leininger S, Yang A, Torruella G, Adamski M, Adamska M, Hughes TR, Gómez-Skarmeta JL, et al. 2013. Early evolution of the t-box transcription factor family. Proceedings of the National Academy of Sciences 110:16050-16055.

Sedgwick A. 1884. Memoirs: On the origin metameric segmentation and some other morphological question. Quarterly Journal of Microscopical Science 2:43-82.

Seetharam A, Bai Y, Stuart G. 2010. A survey of well conserved families of C2H2 zinc-finger genes in *Daphnia*. BMC Genomics 11:276.

Seetharam A, Stuart G. 2013. A study on the distribution of 37 well conserved families of C2H2 zinc finger genes in eukaryotes. BMC Genomics 14:420.

Seravin LN. 1994. The systematic revision of the genus *Mnemiopsis* (Ctenophora, Lobata). 2. Species attribution on *Mnemiopsis* from the black sea and the species composition of the genus *Mnemiopsis*. Zoologichesky Zhurnal 73:19-34.

Servetnick MD, Steinworth B, Babonis LS, Simmons D, Salinas-Saavedra M, Martindale MQ. 2017. Cas9-mediated excision of *Nematostella brachyury* disrupts endoderm development, pharynx formation and oral-aboral patterning. Development 144:2951-2960.

Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nature Ecology & Evolution 1:0126.

Shimeld SM. 2008. C2h2 zinc finger genes of the gli, zic, Klf, sp, wilms' tumour, huckebein, snail, ovo, spalt, odd, blimp-1, fez and related gene families from *Branchiostoma floridae*. Development Genes and Evolution 218:639-649.

Silverstein R, Ekwall K. 2005. Sin3: A flexible regulator of global gene expression and genome stability. Current Genetics 47:1-17.

Sim KL, Creamer TP. 2004. Protein simple sequence conservation. Proteins: Structure, Function, and Bioinformatics 54:629-638.

Simion P, Bekkouche N, Jager M, Quéinnec E, Manuel M. 2015. Exploring the potential of small RNA subunit and its sequences for resolving phylogenetic relationships within the phylum Ctenophora. Zoology (Jena) 118:102-114.

Simmons DK, Pang K, Martindale MQ. 2012. Lim homeobox genes in the ctenophore *Mnemiopsis leidyi*: The evolution of neural cell type specification. EvoDevo 3:2.

Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, Kong A, Thorsteindottir U, Shin SY, Richards HB, Soranzo N, et al. 2011. Identification of an imprinted master trans regulator at the Klf14 locus related to multiple metabolic phenotypes. Nat Genet 43:561-564.

Smith JJ, Antonacci F, Eichler EE, Amemiya CT. 2009. Programmed loss of millions of base pairs from a vertebrate genome. Proceedings of the National Academy of Sciences 106:11212-11217.

Smith JJ, Baker C, Eichler EE, Amemiya CT. 2012. Genetic consequences of programmed genome rearrangement. Curr Biol 22:1524-1529.

Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat Genet 45:415-421, 421e411-412.

Song A, Chen Y-F, Thamatrakoln K, Storm TA, Krensky AM. 1999. Rflat-1: A new zinc finger transcription factor that activates rantes gene expression in T lymphocytes. Immunity 10:93-103.

Soufi A, Donahue G, Zaret KS. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151:994-1004.

Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 161:555-568.

Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, et al. 2008. The *Trichoplax* genome and the nature of placozoans. Nature 454:955-U919.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, et al. 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466:720-U723.

Stamatakis A. 2006. Raxml-vi-hpc: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Steinmetz PR, Aman A, Kraus JE, Technau U. 2017. Gut-like ectodermal tissue in a sea anemone challenges germ layer homology. Nature Ecology & Evolution 1:1535.

Steinmetz PR, Kraus JE, Larroux C, Hammel JU, Amon-Hassenzahl A, Houliston E, Wörheide G, Nickel M, Degnan BM, Technau U. 2012. Independent evolution of striated muscles in cnidarians and bilaterians. Nature 487:231-234.

Suske G, Bruford E, Philipsen S. 2005. Mammalian sp/Klf transcription factors: Bring in the family. Genomics 85:551-556.

Suttamanatwong S, Jensen ED, Schilling J, Franceschi RT, Carlson AE, Mansky KC, Gopalakrishnan R. 2009. Sp proteins and runx2 mediate regulation of matrix gla protein (mgp) expression by parathyroid hormone. Journal of Cellular Biochemistry 107:284-292.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131:861-872.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663-676.

Tamm SL. 2014a. Cilia and the life of ctenophores. Invertebrate Biology 133:1-46.

Tamm SL. 2014b. Formation of the statolith in the ctenophore *Mnemiopsis leidyi*. Biol Bull 227:7-18.

Tamm SL. 2015. Functional consequences of the asymmetric architecture of the ctenophore statocyst. The Biological Bulletin 229:173-184.

Tamm SL. 1973. Mechanisms of ciliary co-ordination in ctenophores. Journal of Experimental Biology 59:231-245.

Tamm SL. 2012. Patterns of comb row development in young and adult stages of the ctenophores *Mnemiopsis leidyi* and *Pleurobrachia pileus*. J Morphol 273:1050-1063.

Tamm SL, Moss AG. 1985. Unilateral ciliary reversal and motor responses during prey capture by the ctenophore *Pleurobrachia*. J Exp Biol 114:443-461.

Technau U. 2001. Brachyury, the blastopore and the evolution of the mesoderm. BioEssays 23:788-794.

Telford MJ, Budd GE, Philippe H. 2015. Phylogenomic insights into animal evolution. Current Biology 25:R876-R887.

Tsai M-Y, Lu Y-F, Liu Y-H, Lien H-W, Huang C-J, Wu J-L, Hwang S-PL. 2014. Modulation of p53 and met expression by krüppel-like factor 8 regulates zebrafish cerebellar development. Developmental Neurobiology 75(9):908-26.

Turner J, Crossley M. 1998. Cloning and characterization of mctbp2, a co-repressor that associates with basic Krüppel-like factor and other mammalian transcriptional regulators. The EMBO Journal 17:5129-5140.

van Vliet J, Turner J, Crossley M. 2000. Human Krüppel-like factor 8: A CACCC-box binding protein that associates with ctbp and represses transcription. Nucleic Acids Research 28:1955-1962.

Vandepas LE, Warren KJ, Amemiya CT, Browne WE. 2017. Establishing and maintaining primary cell cultures derived from the ctenophore *Mnemiopsis leidyi*. J Exp Biol 220:1197-1201.

von Byern J, Mills CE, Flammang P. 2010. Bonding tactics in ctenophores — morphology and function of the colloblast system. von Byern J, Grunwald I, editors. Biological adhesive systems: From nature to technical and medical application. Vienna: Springer Vienna. p. 29-40.

Von Graff L. 1891. Die organisation der turbellaria acoela. Leipzig: von Wilhelm Engelmann.

Ward WW. 1974. Aquarium systems for the maintenance of ctenophores and jellyfish and for the hatching and harvesting of brine shrimp (artemia salina) larvae. Chesapeake Science 15:116-118.

Watanabe H, Fujisawa T, Holstein TW. 2009. Cnidarians and the evolutionary origin of the nervous system. Development, Growth & Differentiation 51:167-183.

Weber U, Rodriguez E, Martignetti J, Mlodzik M. 2014. *Luna*, a *Drosophila* Klf6/Klf7, is maternally required for synchronized nuclear and centrosome cycles in the preblastoderm embryo. PLoS ONE 9:e96933.

Whelan NV, Kocot KM, Halanych KM. 2015. Employing phylogenomics to resolve the relationships among cnidarians, ctenophores, sponges, placozoans, and bilaterians. Integrative and Comparative Biology 55:1084-1095.

Whelan NV, Kocot KM, Moroz LL, Halanych KM. 2015. Error, signal, and the placement of Ctenophora sister to all other animals. Proceedings of the National Academy of Sciences 112:5773-5778.

Whelan NV, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, Moroz LL, Halanych KM. 2017. Ctenophore relationships and their placement as the sister group to all other animals. Nature Ecology & Evolution:1.

Wierstra I. 2008. Sp1: Emerging roles—beyond constitutive activation of tata-less housekeeping genes. Biochemical and Biophysical Research Communications 372:1-13.

Wimmer EA, Jackle H, Pfeifle C, Cohen SM. 1993. A *Drosophila* homolog of human sp1 is a head-specific segmentation gene. Nature 366:690-694.

Wootton JC. 1994. Non-globular domains in protein sequences: Automated segmentation using complexity measures. Comput Chem 18:269-285.

Wootton JC, Federhen S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. Comput Chem 17:149-163.

Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, Gao G, Zhang B. 2014. CasOT: A genome-wide Cas9/gRNA off-target searching tool. Bioinformatics 30:1180-1182.

Yamada A, Martindale MQ. 2002. Expression of the ctenophore brain factor 1 forkhead gene ortholog (ctenobf-1) mrna is restricted to the presumptive mouth and feeding apparatus: Implications for axial organization in the metazoa. Dev Genes Evol 212:338-348.

Yamada A, Martindale MQ, Fukui A, Tochinai S. 2010. Highly conserved functions of the brachyury gene on morphogenetic movements: Insight from the early-diverging phylum Ctenophora. Dev Biol 339:212-222.

Yamada A, Pang K, Martindale MQ, Tochinai S. 2007. Surprisingly complex t-box gene complement in diploblastic metazoans. Evolution & Development 9:220-230.

Yasuoka Y, Shinzato C, Satoh N. 2016. The mesoderm-forming gene brachyury regulates ectoderm-endoderm demarcation in the coral *Acropora digitifera*. Current Biology 26:2885-2892.

Yatsu N. 1911. Observations and experiments on the ctenophore egg. II. Notes on the early cleavage stages and experiments on cleavage. Annot. Zool. Japan 7:333-346.

Zeng V, Villanueva KE, Ewen-Campen BS, Alwes F, Browne WE, Extavour CG. 2011. De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean *Parhyale hawaiensis*. BMC Genomics 12.

Zhang J-S, Moncrieffe MC, Kaczynski J, Ellenrieder V, Prendergast FG, Urrutia R. 2001. A conserved α-helical motif mediates the interaction of sp1-like transcriptional repressors with the corepressor mSin3a. Molecular and Cellular Biology 21:5041-5049.

Zhao CT, Meng AM. 2005. Sp1-like transcription factors are regulators of embryonic development in vertebrates. Development Growth & Differentiation 47:201-211.

Zhao X, Monson C, Gao C, Gouon-Evans V, Matsumoto N, Sadler KC, Friedman SL. 2010. Klf6/copeb is required for hepatic outgrowth in zebrafish and for hepatocyte specification in mouse ES cells. Developmental Biology 344:79-93.