The Essential Compendium to *Eucidaris tribuloides*: A morphological and transcriptomic characterization of embryogenesis

by

Katelyn Victoria Smith

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Biology

Auburn, Alabama
May 2, 2026

Keywords: evolutionary development, gene regulation, Cidaroidea

Approved by

Katherine M. Buckley, Chair, Associate Professor, Department of Biology
Anthony G. Moss, Associate Professor, Department of Biology
Ryan Range, Associate Professor, Department of Biology

Abstract

Evolutionary development (evo-devo) is the study of how phenotypic changes occur through evolution as associated with developmental processes. This scientific discipline serves as a way to provide causal explanations into how changes in the genotype manifest as changes in a phenotype of an organism. An especially potent mechanism for explaining evolution of organismal development are gene regulatory networks (GRNs). GRNs are a summary of all the interactions between transcription factors, *cis*-regulatory elements, and signaling molecules that control how development proceeds. They are genetically encoded developmental instructions, and mutations of these sequences can have significant impacts in how the organization and function of animal body plans.

Echinoderms have long served as morphological/genetic evo-devo study models, especially sea urchins for their unique biphasic life cycle and ease of experimental perturbation. Sea urchins, which belong to class Echinoidea, may be split into two subclasses: Euechinoidea, representing all modern sea urchins and sand dollars, and the basal sister group Cidaroidea, which contains all extant pencil sea urchins. Cidaroids act as a uniquely apt comparative model for evo-devo studies, as there is significant divergence between them and the euechinoids and clear morphological differences, yet together this clade in monophyletic. Despite these attractive characteristics, the vast majority of evo-devo studies have focused on a small handful of euechinoid sea urchins, which are all relatively closely related to one another, causing a phylogenetic gap in the current understanding of the evolution of development within Echinoidea. In this study, we sought to bridge this gap in knowledge by characterizing the

developmental morphological and transcriptomic aspects of a cidaroid sea urchin *Eucidaris tribuloides*.

In chapter 1, we developed the first standardized protocols for adult *E. tribuloides* adult husbandry and rearing conditions to bring *E. tribuloides* larvae to metamorphosis. We captured and described novel developmental stages, particularly those stages between the two-arm pluteus larvae and the metamorphosed juvenile, thus produced the first full developmental staging scheme for this cidaroid species. In an effort to understand how the differences during early development may impact the ontological trajectory of mesodermally derived tissues, we generated the first images of adult cidaroid coelomocytes. Taken together, the results of this chapter have established a morphological baseline of cidaroid development that will facilitate future comparative studies between these sea urchin species.

In chapter 2, we produced both the first developmental transcriptome assembly (from cleavage stage to the two-arm stage) and the first gene annotations of any cidaroid sea urchin species. In total, we predicted 28,331 protein sequences with both homology and functional support. Several interesting patterns arose from these efforts. First, we were able to identify with high confidence a candidate *E. tribuloides* Pmar1 transcript, which provides evidence that Pmar1 existed within echinoids before divergence occurred. Second, of the 374 regulatory genes implemented in developmental control of *S. purpuratus*, we found 371 candidate genes in our cidaroid system, suggesting that most of these genes are deeply conserved within Echinoidea. Finally, it was observed that there is a heterochronic shift of the entire skeletogenic module within cidaroids, which supports the findings of previous studies and provides an excellent example of how changes in the regulation of tissue development can have profound phenotypic

effects. The genetic resources we have generated in this chapter will serve as a powerful tool for future competitive studies investigating changes in echinoid development.

Artificial Intelligence (AI) Use Disclosure Statement

In the preparation of this thesis / dissertation, the following Artificial Intelligence (AI) tools were used: ChatGPT. These tools were used primarily to aid in writing code for transcript assembly and annotation. The author acknowledges full responsibility for the intellectual content of this work and has ensured that all AI-assisted sections have been reviewed and revised for accuracy and appropriate academic style. All AI-generated content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final document to maintain scholarly integrity of this research.

Digital Accessibility Use Disclosure Statement

In the preparation of this thesis / dissertation, the following digital accessibility tools were used to ensure this document complies with federal requirements: Microsoft Word Accessibility Checker. The author acknowledges full responsibility for the intellectual content of this work and has made a good faith effort to comply with digital accessibility requirements in publishing, wherein the nature of the content does not significantly change in order to do so. Furthermore, all content has been reviewed and revised to meet these requirements prior to final publication.

Acknowledgements

I would like to thank the members of my committee for their constant support not only throughout this degree, but my academic career. Dr. Tony Moss was my first professor during my first year of undergraduate here at Auburn University. Since then, he has been an invaluable source of knowledge, wisdom, and camaraderie for topics concerning both the sciences and life. I would like to thank Dr. Kate Buckley for taking me into her lab and providing direction to my scientific career trajectory. If not for her leadership, wit, and experience, I surely would not be even half the person and scientist I am today. I thank Dr. Ryan Range for his knowledge of all things gene regulatory networks, his kind words, and his encouragement. These people have provided me with incredible education and opportunities, for which I will be forever grateful.

I would be remiss if I did not thankful my incredible labmates and friends for the love they have provided me. Dr. Megan Maloney, Tyler Smith, Amelia Williams Nick Kraieski, Bailey Steinworth, and Jake Tatum are some of the most patient, kind, knowledgeable, and incredible scientists that I have ever met. If it weren't for their guidance and help, I never would be in the position I am today. In addition, Frank Moen and Laura Oldfather have been constant sources of delight and friendship, and they have certainly made graduate school more bearable.

I thank Grant Bagley for the last ten years of love, friendship, and support. I find the English language to be too limited for me fully be able to express the gratitude I have for the unconditional love and care he has given me.

Most of all, I want to thank my parents, Jennifer Smith and Gordon Smith, and my brothers, Matthew and Ryan. These are the people who made me who I am, and without, I would

not be one *iota* the person I am today. Each member of my family has served as a role model, a friend, and a guiding hand in my life. My only hope is that I can make you proud.

Finally, I would like to show my appreciation for *Eucidaris tribuloides*. These are truly some of the most remarkable animals I have ever encountered, and words fail me in trying to describe how honored I am to have gotten to know these creatures so intimately. Such beautiful and bizarre forms serve as an incredible reminder to always stay curious.

Thank you all.

Table of Contents

List of Tables

List of Figures

Introduction

An introduction to evolutionary developmental biology

The field of evolutionary developmental biology (evo-devo) investigates how phenotypic change occurs on an evolutionary scale from a mechanistic perspective. Although the term first appeared in literature in Calow's 1983 *Evolutionary Principles* (Calow, 1983), the origin of this field can be traced further back to the study of comparative embryology in the nineteenth century. von Baer in 1828 produced what would eventually be called the "Laws of Development" in which he put forth four postulates describing how an organism grew and changed (Baer, 1999). These postulates were, in part, formed through his observations of different animals following the same general patterns of germ layer development before diverging from one another. Though these postulates have been critiqued over the last two centuries, they nonetheless became the foundation of the modern understanding of development. von Baer's work is also one of the earliest and clearest demonstrations of conserved developmental patterns occurring across disparate animal lineages. Nearly four decades later, Haeckel would further popularize the concept of evolution through development through his comparative embryology work and now well-known depictions of phylogenetic trees constructed from conserved embryologic developmental patterns (Haeckel & Royal College of Physicians of London., 1866). Although the concept that developmental ontogeny recapitulates phylogeny is no longer supported, these ideas were useful in that they generated testable hypotheses and underscored the need to understand the mechanisms underlying development to gain insight into how these features are conserved or lost across animal lineages. Lastly, Kowalevsky pioneered work at the intersection of development and evolution by demonstrating that 1) there exists

common patterns of gastrulation across invertebrates and vertebrates; 2) *Amphioxus* belongs to Cephalochordata by evidence of the similar neural development patterns between it and the rest of vertebrates; and 3) developmental comparisons between ascidian larvae vertebrates reveal tunicates belong within Chordata rather than Mollusca (Raff & Love, 2004). Kowalevsky used the presence of conserved developmental patterns observed in tunicates to (correctly) suggest that tunicates are among the most basal of chordates. The ideas and experiments of these scientists, and others, primed the field of comparative embryology to give rise to the modern concepts of evo-devo.

It would, however, take several more decades before the concepts in developmental and evolutionary biology would be fully integrated. Due to technological innovations and the rediscovery of Mendel's ideas of inheritance, the study of genetics exploded. By the 1920s, evolutionary biologists began to infer phylogenetic relationships using population genetic evidence rather than developmental phenotypes and morphological differences. At the same time, the field of embryology also experienced a shift away from comparative work to focus on experimental studies to understand the mechanisms that drove developmental processes. It was primarily due to this schism that embryology was neglected from being integrated into the initial framework of the Modern Synthesis of Evolution (Hall, 2012). Evolution became defined by changes in allele frequency within a population over time. This definition, while useful, lacked the explanatory power to test how novel traits arose in a population or how homologous traits could be found across evolutionarily distant lineages with significantly different genotypes. In other words, by removing development from the Modern Synthesis, the ability to fully grasp the connection between genotype and phenotype was similarly lost.

The reintegration of development into the theory of evolution began in the 1970s and 1980s, catapulted by the discovery of homeobox genes and a genetic basis for homology of animal body plans (Goodman & Coughlin, 2000). In 1984, it was discovered that, in *Drosophila melanogaster*, a 180 bp sequence was highly similar between the genes *ANTP, FTZ,* and *UBX*, which encode proteins with significant roles in determining body orientation. This sequence, named in the homeobox, was identified in other invertebrates and vertebrates, suggesting that the way in which animals pattern their bodies is deeply conserved (McGinnis et al., 1984; Scott & Weiner, 1984). Other examples of deep genetic homologies include the anterior/posterior axis mechanisms driven by Wnt signaling (Rijsewijk et al., 1987; van 't Veer et al., 1984), and the role of TGF-B/Nodal in the formation of the dorsal/ventral axis. These, and other, studies not only rejuvenated interest in development from an evolutionary perspective but demonstrated how developmental genes can lead to observed phenotypes. Importantly, these experiments also showed how developmental genes can be regulated and deployed in different spatial and/or temporal contexts to give rise to new features.

Echinoderms as developmental model systems

The history of evolutionary developmental biology is intimately intertwined with experiments performed using echinoderms. The phylum Echinodermata is a diverse clade of marine invertebrates consisting of class Crinoidea (feather stars), Ophiuroidea (brittle stars), Asteroidea (sea stars), Concentricycloidea (sea daisies), Holothuroidea (sea cucumbers), and Echinoidea (sea urchins and sand dollars). These animals have a unique anatomy, such as water vascular system through which they manipulate the water pressure within their bodies for locomotion and food manipulation. Echinoderms undergo a significant body plan transition during development, going from a planktonic bilateral larva to a sessile pentaradial (symmetrical

on five axes) adult. This makes echinoderms particularly relevant for evo-devo studies, as these animals may be able to shed light on generally conserved developmental processes within deuterostomes.

For more than a century, these animals have served as essential model systems for investigating the function and evolution of the cellular, molecular, and genetic mechanisms that direct development. In his landmark 1892 experiment, Driesch demonstrated the totipotent nature of early sea urchin embryos by isolating two-cell and four-cell stage blastomeres (Andrews, 1892). He observed individual blastomeres developed into normal – although slightly smaller – larvae, meaning that any one of these early developmental cells had the capacity to form any part of a larva. Furthermore, these experiments showed that larval body patterning is driven by context-dependent regulation, rather than cells being hardwired for a specific fate, meaning that changes in the temporal/spatial developmental context could lead to new or altered phenotypes. This concept was further expanded upon by a series of fate mapping experiments using euechinoid *Paracentrotus lividus* embryos, which demonstrated the ways in which cells can influence the developmental fates of one another (Hörstadius, 1973). For instance, a section containing only the embryonic animal pole develops into a ball of ectodermal derived structures, whereas a section containing the animal pole and the micromeres gave rise to a mostly normal larva.

Sea urchins and other echinoderms possess several traits that have make them ideal candidates for evo-devo studies. Echinoderms generally produce thousands to millions of optically clear offspring which allow for the easy observation of the external and internal anatomy. These animals tend to be resilient to experimental manipulation techniques, such as microinjection, and are generally relatively easy to rear in a laboratory setting. From an evolution

perspective, echinoderms are relevant for study due to their life history and phylogenetic placement. This metamorphosis means that echinoderm larvae and adults are subject to different selective pressures, meaning that the two forms may evolve semi-independently. In addition, this ontogenetic transition from the ancestral to a derived body plan can be used to understand how unique morphologies arise through changes in development. Sea urchins are especially powerful evo-devo models, as several of these species, including *Strongylocentrotus purpuratus* and *Paracentrotus lividus*, have some of the best understood organismal and molecular developmental patterns of any organism to date. This means that sea urchins can act as excellent systems to compare developmental states both within the phylum and with the rest of deuterostomes.

General Sea Urchin Development

As a result of these previous studies, a well understood general model of sea urchin development has been produced, though it is heavily biased towards what is known of euechinoid developmental patterns. Broadly speaking, sea urchins exhibit a biphasic life history in which they are swimming pluteus larvae as juveniles and then metamorphose into benthic adults. There are some species of direct developing sea urchins, most notably *Heliocidaris erythrogramma,* however this strategy is outside the scope of this study.

Upon fertilization, the sea urchin zygote undergoes radial cell cleavage, initially producing eight equally sized cells and establishing the animal and vegetal poles. After a fourth round of cell division, the 16-cell stage, the vegetal pole undergoes unequal cell division to produce micromere and macromere cells. The micromeres will undergo another cycle of unequal cell division, producing the large micromeres and small micromeres, which become part of the skeletal mesenchyme and the germ cell layer respectively. The macromere cells divide to give

rise to the *veg1* and *veg2* cells, the latter of which eventually develop into the non-skeletal mesenchyme and part of the gut. The cells of the animal pole generally become the ectoderm tissues of the animal.

As cell cleavage proceeds, blastula is eventually formed, which is a ciliated hollow ball just one cell layer thick. Eventually, the large micromeres specify into primary mesenchyme cells, which ingress into the blastocoel before gastrulation occurs. During gastrulation, cells at the vegetal pole invaginate, forming the archenteron, and move inward toward the center of the blastocoel. During this time, non-skeletal mesenchyme cells will delaminate off the tip of the archenteron to populate various spaces of the blastocoel. Gastrulation finished once the archenteron has made contact with the wall of the blastocoel and the mouth of the embryo opens up. The animal will go through stages of early skeletal development until reaching the pluteus stage. From here, the larvae will continue to develop, growing several pairs of arms, and increase in size. Eventually, juvenile structures, such as tube feet and pedicellaria, will begin to grow from the left/right coelomic pouches. The juvenile body forms within the left coelomic pouch, from which it will erupt from, completing metamorphosis and living the rest of its life as a benthic adult sea urchin.

Eucidaris living things

Phylogeny & Fossils.

Class Echinoidea has traditionally been separated into two subclasses: Euechinoidea and Cidaroidea. Euechinoidea contains all modern sea urchins, irregular sea urchins (sea biscuits and heart urchins) and sand dollars. Cidaroidea is the basal sister group to this clade (Kroh & Smith, 2010), containing all pencil urchin species, including *E. tribuloides*. Evidence of this phylogenetic relationship can be found from examining the rich cidaroid and euechinoid fossil

records. These records have shown that cidaroids diverged within Echinoidea sometime during the Guadalupian era (~272 – 259 mya) of the mid Triassic and diversified through the Cretaceous (~145 – 66 mya) and Tertiary (~66 – 2.6 mya) periods. This means that the developmental processes of this ancient sea urchin lineage have been evolving independently from euechinoids for at least 268 million years (J. R. Thompson et al., 2015).

It appears that there were several drivers in the differentiation of the cidaroid and euechinoid subclasses, particularly in relation to test morphology. First, the supportive structures that hold up the Aristotle's lantern and act as muscular attachment points the associated muscular systems protrude from different plates of the test between these two sea urchin clades(Gao et al., 2015). The plates of the test also take on different forms. The plate structure of cidaroids is generally simpler than that of euechinoids, with the attachment points of the primary spines being generally more robust. There are differences in the spine structure itself as well; the core of the cidaroid primary spine is bare whereas that of euechinoids is covered in a layer of epithelium, and the inner layers of the primary spines differ in respect to density and stereom (Grossmann & Nebelsick, 2013). It is believed that this is what allows for the conspicuous settling and crustacean of algae and small marine invertebrates on cidaroids spines(Grossmann & Nebelsick, 2013). While the exact purpose of allowing for epibionts to settle on the animal in unknown, it is possible that such action incurs a level of camouflage for cidaroid species. Finally, cidaroids and euechinoids undergo relatively drastic different development patterns, the details of which will be expanded upon in the coming pages.

Natural history & ecology

Eucidaris tribuloides (rock-boring urchin or pencil urchin) is a sea urchin species native to the Atlantic coast of the Americas, historically distributed from North Carolina down to Brazil.

These animals inhabit rocky reef habitats where they use their relatively large Aristotle's lantern to graze on rock and other similar substrates. Stomach dissections have indicated that the bulk of *E. tribuloides*' diet is comprised of macro algae, encrusting sponges, and tunicates. Previous studies have found that these animals spawn during the summer and late fall months (June to November(McClintock & Watts, 1990). Like most other sea urchin species, these animals are broadcast spawners whose reproductive cycle may be influenced by the lunar cycle, where these animals reach peak gravidity during the full moon (Lessios, 1991). It is possible that they use their robust spines (see figure 0.1) to brace themselves within rock or shelter as an antipredation mechanism.

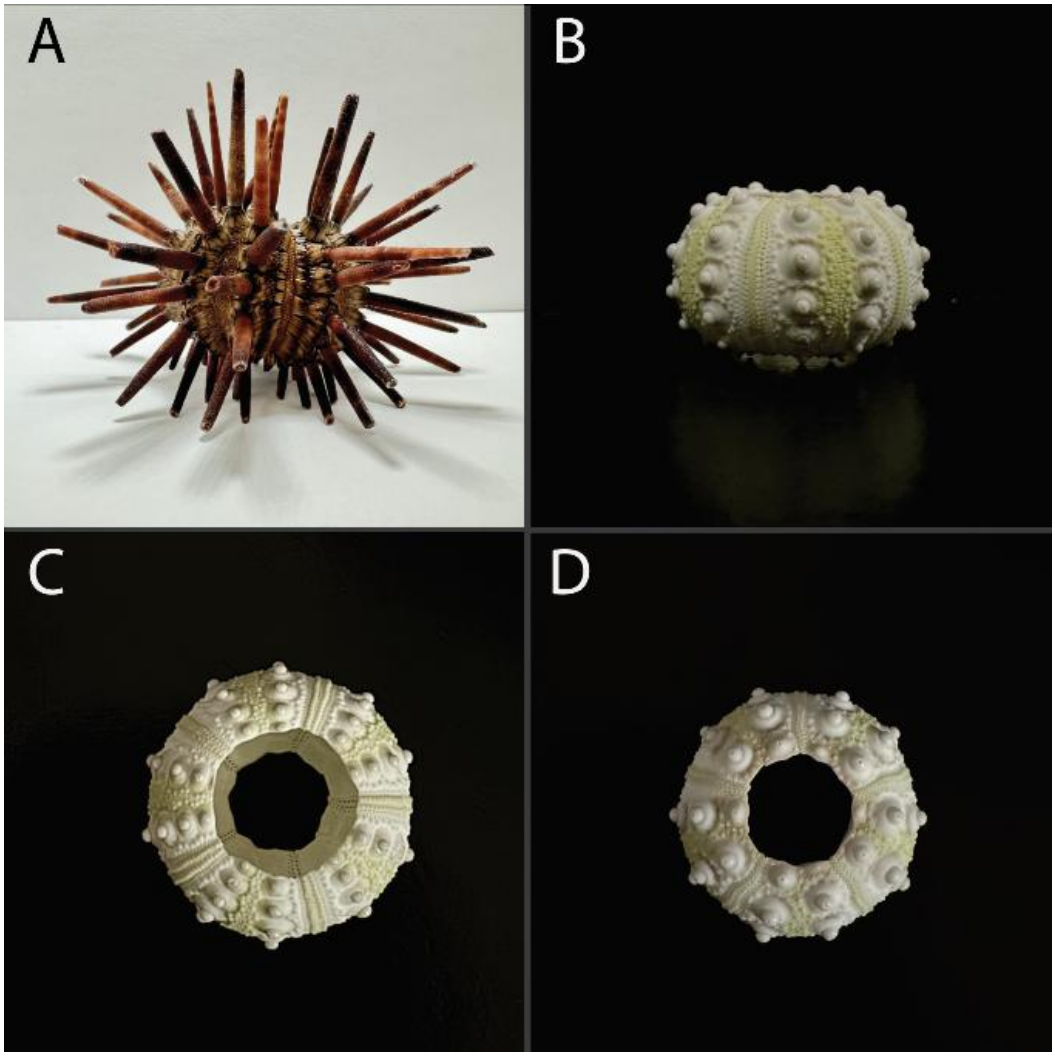

Figure 0.1: General morphology of adult *E. tribuloides*. **(A)** Living adult specimen of unknown sex. **(B)** Side view of adult bleached test. **(C)** Oral view of adult bleached test. **(D)** Aboral view of adult bleached test.

Current knowledge gaps and thesis goals

We believe that *E. tribuloides* provides a unique and excellent system in which to study the developmental mechanisms that lead to diversification and divergence in animal body plan. As aforementioned, these animals occupy an interesting branch of the animal tree as invertebrate deuterostomes. They are morphologically distinct and phylogenetically distant enough from a

well understood system, the euechinoids and in particular *Strongylocentrotus purpuratus*, to have clear evolutionary divergence of character and gene expression yet are still the sister group to the comparative model, all while maintaining a monophyletic relationship. What's more, despite these characteristics, *E. tribuloides* is still a relatively unknown organism to science, meaning this system is ripe with opportunity for scientific discovery.

For an animal to become a model research system in understanding transcriptomic regulation as a mechanism for developmental evolution, there must at minimum be a basic understanding of the organism develops and what genes are being expressed during these different stages. *E. tribuloides* has scant resources available in regard to both of these elements. Our goal in chapter 1 is to bridge this gap in our knowledge of the morphology of cidaroid development by tracking the ontological trajectory of *E. tribuloides* from fertilization through metamorphosis. We also aim to outline and standardize husbandry protocols for maintaining adults and rearing *E. tribuloides* offspring in an artificial, reproducible laboratory environment. In chapter 2, our goal is to increase the genetic resources available for *E. tribuloides*, particularly as they pertain to early development. We seek to do this by producing the first developmental transcriptome of *E. tribuloides* and identify general trends of gene expression, especially in regards to the mesenchyme tissues. While there have been other studies that have investigated isolated components of *E. tribuloides* development, this is, the best of our knowledge, the first attempt to understand embryogenesis of this animal with such a complete and holistic approach.

Chapter 1

Introduction

Developmental staging as a tool

While the use of genetic and molecular tools have been useful in the discovery of many developmental processes, there still exists a need for descriptive developmental staging in order to gain a full appreciation of a species' ontogenetic trajectory; this is particularly true for non-model organisms that have limited genetic resources. Staging as an evo-devo tool allows for defined development events that can be used to standardize experimental results(Hopwood, 2007), define successful experimental replication(Kimmel et al., 1995), and facilitate communication across laboratories(Kimmel et al., 1995). The identification of key developmental features has been essential for the resolution of the phylogenetic placement for several species, such the tunicate's larval notochord grouping them with Chordates (Kowalevsky, 1871), the temporal patterning of organogenesis in turtles supporting their placement as basal reptiles(Werneburg & Sánchez-Villagra, 2009), and nauplius larval form of barnacles being used to identify them as belonging to Crustacea rather than Mollusca(J. V. Thompson, 1829). Finally, having a well-defined timeline is essential for performing developmental perturbation experiments, which have been used for uncovering and/comparing conserved and diverged developmental regulatory states.

Existing echinoid models and limitations

Euechinoid sea urchins dominate the current echinoid development/evo-devo landscape, and while these animals have been instrumental to our current understanding of ontogeny, the

dependence on these systems has left evolutionary breadth of echinoid developmental biology under sampled. Of the modern sea urchin species, *Strongylocentrotus purpuratus*, *Paracentrotus lividus,* and *Lytechinus variegatus* have some of the most well-characterized developmental frameworks. *S. purpuratus* has served as a foundational model to study the ways in which genetic regulatory interactions organize and drive development(Davidson et al., 2002), especially those systems that control the specification and growth of the larval skeletal mesenchyme(Oliveri et al., 2008; Rafiq et al., 2012). *P. lividus* has historically been used to understand echinoid morphogenesis and developmental patterns(Horstadius, 1939), developmental regulation(Saudemont et al., 2010), and even the effects of ocean acidification and anthropogenically sourced pollution on the growth and health of echinoid larvae(Martin et al., 2011; Moulin et al., 2011). Finally, *L. variegatus* has been used to make significant contributions towards standardizing euechinoid developmental staging and larval anatomy(McEdward & Herrera, 1999).

These echinoids have been preferentially used for embryogenic and comparative developmental studies for several reasons, least of all that considerable groundwork has already been done to define developmental stages, standardize husbandry protocols, and develop molecular tools for these systems. While cidaroid urchins generally share many of the characteristics that make euechinoids attractive study organisms (clear larvae, large number of progeny, relatively simple larval anatomy), they lack these resources and so have been remained underrepresented in developmental studies.

This small number of study organisms(McClay, 2011) matters because all three of these urchins species are relatively closely related to one another, having diverged from one another about 35 – 50 million years ago(Lee, 2003). The consequence of these close phylogenetic

relationships is that only limited evolutionary information may be ascertained by comparative morphological and/or molecular developmental studies. By focusing attention on this narrow slice of the echinoid clade, the ability to infer lineage specific developmental innovations, ancestral developmental states, and how these mechanisms have changed is constrained. Addressing these limitations in echinoderm evo-devo would require expanding the current understanding of development of a distantly related lineage, for which Cidaroidea would be very appropriate.

What is known about E. tribuloides development

The first description of the development of *E. tribuloides* was published in 1981, in which embryonic anatomy and behavior was observed from egg fertilization to the two-arm stage(Schroeder, 1981). It was observed that these animals produced clear eggs and had either an extremely reduced or entirely absent hyaline layer. During the 16-cell stage, animals had a variable number and morphology of micromeres and mesomeres. After the blastula hatched, it was noted that while the embryo could swim, it was very slow and lacked an apical tuft of cilia that is common in other sea urchins. Gastrulation began before the ingression of primary mesenchyme cells and other mesenchyme cells populated the blastocoel before the first skeletal spicules began to form. The skeletal rods grew to form the post-oral arms, but during the 7-day period of this study, the left and right elements of the larval skeleton never joined. It was noted that, even though algal cells were available, *E. tribuloides* larvae were never seen eating. Finally, the larvae were shown to continue to swim extremely slowly even into the two-arm stage.

Since 1981, there have been two other major studies that have focused on the morphological development of specific embryonic organ systems in *E. tribuloides*, one of which being the nervous system(Bishop et al., 2013). Briefly, the *E. tribuloides* blastula lacked a thickened

animal plate or apical ciliary tuft, and they had cilia that were associated with the expression of Hnf6, a significant transcription factor used in the process of ciliogenesis. As the animal transitions through the prism and early two-arm stages, different types of neurons with distinct gene expression patterns and morphology arise within the ciliary band that circles the perimeter of the animal. Strikingly, it appeared that *E. tribuloides* lacked an apical organ at the eight-arm stage, a feature that is broadly shared among euechinoids.

The scope of the second study primarily focused on the morphogenesis of the muscle systems of *E. tribuloides* and how these system incur larval movement(MacNeil et al., 2017). During the two-arm phase, the first muscles form in association with the esophagus, and as development continues, these muscles lengthen and branch out to the left and right sides. During the four- and six-arm stages, the lateral filaments, cardiac sphincter muscles, and medial ring muscles appear. During this time, a second muscular ring, termed the pyramidal muscle, develops posterior to the medial ring, a feature that is not observed in *L. variegatus* or other euechinoids. Other unique musculature features were seen in *E. tribuloides* as development continued, such as connections between the esophageal and medial ring muscles and the muscles of *E. tribuloides* being generally thicker and more robust in comparison to *L. variegatus*. These differences in musculature appeared to have functional consequences, as during late larval stages *E. tribuloides* larvae demonstrated distinct arm flexing movements. Such behavior has not been observed in *L. variegatus* or other euechinoid larvae.

Finally, the test of post-metamorphosed *E. tribuloides* and *S. purpuratus* juveniles have been compared to understand similarities and differences in skeletal structure that have may have arisen due to their phylogenetic history(Gao et al., 2015). Several significant differences were between the two lineage's test structure. First, ambulacral plates of *E. tribuloides* appear to be

much simpler than that of *S. purpuratus*, being made from simple plates as opposed to being made of demiplates (series of smaller, interconnecting plates). The Aristotle's lantern, while developing around the same time, is supported by different structures in these two species. In *E. tribuloides*, the perignathic girdle (the general support structure of the Aristotle's lantern) is made of apophyses, which are projections of the test connected to the interambulacral plates. The perignathic girdle of *S. purpuratus* is formed from projections connected to the ambulacral plates called auricles. While these structures ultimately have very similar functional roles, they develop in different ways, suggesting a difference (or at least alteration) in evolutionary origin.

Taken together, these experiments have highlighted that despite belonging to the same class, cidaroids and euechinoids display significant differences across multiple aspects of development, from early embryonic patterning to organ system development to juvenile test formation. These differences are clearly not restricted to a single developmental system but instead demonstrate entire shifts within the general developmental program. Therefore, further cataloguing these differences in structures, and the possible underlying molecular causes, may provide novel insights concerning how and in what ways the echinoid class and its members have diverged and evolved.

Developmental knowledge gaps and justification

Despite the powerful explanatory power cidaroids hold due to their phylogenetic positioning, there has been relatively little research done using them as systems to investigation evolutionary development. While the development of specific organ systems and isolated life stages have been catalogued, only two cidaroid species have had their development tracked to metamorphosis with an emphasis on broad morphological changes. *E. tribuloides* in particular lacks standardized husbandry protocols or a well-defined developmental staging scheme from

fertilization through metamorphosis. This limits reproducibility of experimental results and the communication of said results between groups and to the general scientific community. In addition, without a baseline understanding of developmental patterns and morphology, it is difficult to draw meaningful conclusions from studies investigating developmental gene expression; in other words, any line drawn between organismal genotype and phenotype is rather tenuous.

To address these limitations and begin the establish of *E. tribuloides* as a comparative evo-devo model, this study attempts to establish a comprehensive developmental timeline by integrating the observations of development of multiple organ systems and structures through sequential developmental stages. We aim to provide more defined developmental reference points and specific husbandry protocols to facilitate future research efforts. Ultimately, our goal is to provide a baseline that can be used for cross-species comparative examinations and lay the foundation for further investigations into the evolution of echinoid morphological development.

Methods

Adult Husbandry

Adult *E. tribuloides* specimens were collected from the Florida Keys (KP Aquatics, LLC) and shipped to Auburn University (Auburn, AL). Sea urchins were housed in a recirculating aquarium with artificial seawater (ASW; Instant Ocean) and maintained at 35 ppt and 25-27 °C. Water quality was assessed weekly by measuring pH, nitrate, and ammonia levels. Animals were fed once to twice a week and maintained on a mixed diet consisting of dulce algae, scallop meat, and freeze-dried anchovies (rehydrated in ASW). To prevent the overgrowth of adult teeth and discourage chewing on aquarium equipment, a variety of mollusk shells and rocks were added to

the aquarium. We found that these animals preferentially gnawed on mussel shells. Adult coelomocytes were collected for imaging as described in Methods in Cell Biology by Smith et al. (*Echinoderms. Part A*, 2019).

Spawning, Gamete collection, and Fertilization

Spawning was induced by injecting adult animals with 1– 3 mL of a 0.5 M KCl solution. Injections were performed by inserting a syringe needle through the peristome tissue surrounding the mouth at a shallow angle (~45°) to the mouth to avoid penetrating the gut (see figure 1.1). Animals were additionally mechanically stressed by shaking. Sperm was collected dry off the animal using a pipette and stored dry in a 1.7 mL centrifuge tube at room temperature. It should be noted that sperm generally was not viable after about 12 hours in this condition. Eggs were collected by placing female sea urchins anus side down over the opening of a small beaker (typically 50 mL) containing ASW. Females were occasionally rinsed with ASW to ensure all eggs had been washed into the collection beaker.

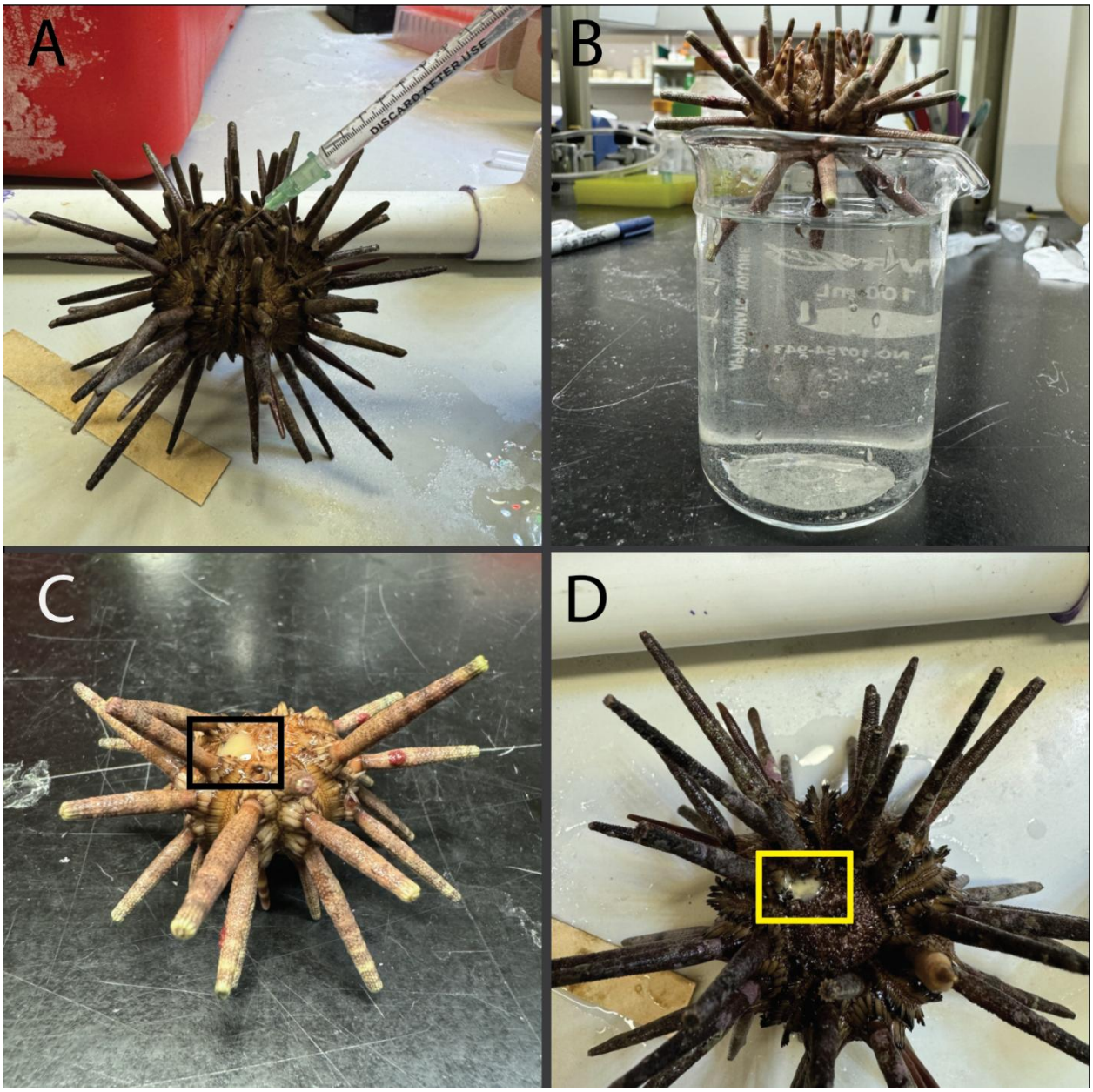

Figure 1.1: Spawning *E. tribuloides*. **(A)** A syringe of 0.5 M KCl solution was injected at a 45° angle through the peristome tissue. **(B)** Female sea urchin dropping eggs into beaker of ASW. The eggs are nearly visible and appear as debris making the water cloudy. **(C)** Female sea urchin with eggs being released from the gonopore. Eggs are clear with a slight yellow tint (black square). **(D)** Male sea urchin spawning. Sperm (yellow square) appears as an opaque white or off-white fluid being released from the gonopores.

To fertilize the eggs, sperm was activated by diluting ~1:500 in ASW (room temperature; 35 ppt).. The diluted sperm was then added to the beaker containing the eggs and gently swirled. Fertilization rates were assessed 2-3 minutes after sperm addition by visualizing a small quantity of eggs using a dissecting scope to monitor for the presence of an expanded vitelline membrane.

Larval Husbandry

After fertilization, zygotes were transferred to plastic vessels containing 3 L of ASW (35 ppt). Cultures were stirred using a rotating paddle. Paddles were constructed using an empty pipette tip tray adhered to a serological pipette using plastic cable ties. The serological pipette was then cut to an appropriate length so that the paddle would be able to spin freely within the culture bucket and affixed to a 12-volt rotating motor (26 rpm).

Embryo and larval cultures were maintained at 26 ℃ with a 12h:12h light/dark cycle (light beginning at 7am). Larvae were fed *Rhodomonas sp. ad libitum* and ~35% water changes were performed every three days using reverse filtration with a 200 μm filter. We found that this strategy was the gentlest method and resulted in the least number of damaged larvae.

Microscopy

The presence of adult structures (primordial tube feet, juvenile spines, pedicellaria) was used to assess metamorphic competency. Due to the gentle water changes, a thick layer of algal growth and microbial biofilms accumulated on the bottom of the culture vessels during the course of larval development. This provided sufficient substrate to stimulate metamorphosis. After metamorphosis, the paddle was removed and replaced with an air stone. Juvenile sea urchins grazed on the algae and no other food was provided. To perform water changes on post-metamorphosed cultures, water was simply removed/added with a beaker.

Larvae were collected using a 200 µm filter with a transfer pipette, then placed into 0.2 µm filtered ASW at 35 ppt to minimize debris when imaging. Larvae were transferred to a microscope slide in 100 – 200 µL droplets and imaged under cover slips elevated by clay feet. After imaging, larvae were returned to their culture vessels. Images were taken using the DIC channel of a Zeiss ApotTome.2 compound microscope with a Zeiss Axiocam 506 camera. Once imaging was complete, larvae could be safely returned to the culture vessel.

Samples for scanning electron microscopy (SEM) were prepared according to previous protocols by Jake Tatum (unpublished). Briefly, coelomocytes were fixed using a two-step glutaraldehyde fixation solution, dehydrated through a graded ethanol series, and transitioned through increasing concentrations of hexamethyldisilazane (HDMS). Air dried cover slips were gold sputtered. Imaging was performed using a Zeiss EVO50 scanning electron scope with magnification up to 50,000x.

Results

We found that it took *E. tribuloides* roughly 182 days (~6 months) to reach metamorphosis post fertilization. Of the six cultures made, two successfully metamorphosed. We found that *E. tribuloides* demonstrated significant seasonality in spawning behavior. Adults did not produce gametes from December to mid-May. Average culture size was roughly 45,000 eggs with high fertilization rates (~90%). It should be notated that *E. tribuloides* cultures demonstrated significant asynchronicity in development. As such, while developmental times will be included, we will emphasize developmental stage as a more apt descriptor.

Stage 1: Fertilization and Cell Cleavage

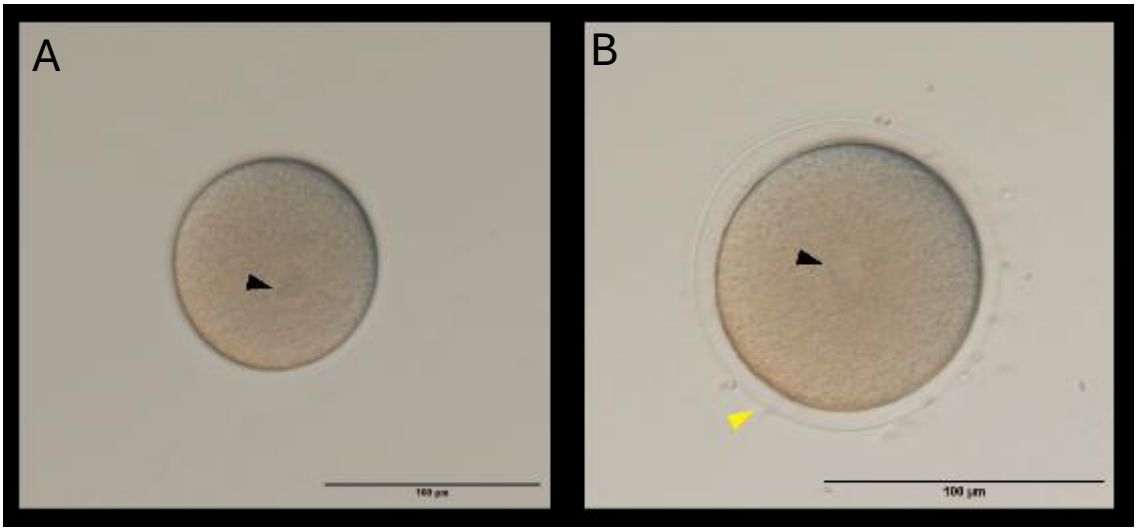

Figure 1.2: Fertilization of *E. tribuloides* egg. Scale bar = 100 µm **(A)** Unfertilized egg. Black arrow indicates cell nucleus. **(B)** Fertilized egg. Black error indicates cell nucleus, which has been centralized. Yellow arrow points towards the vitelline membrane.

Before fertilization, the nucleus of the egg can be observed to be off centered. After syngamy has occurred, the nucleus becomes centralized and the vitelline membrane expands. In figure 1.2, it can be observed that multiple sperm have been caught on the vitelline membrane post expansion, indicating that this membrane likely functions in preventing polyspermy, similar to vitelline membranes of other echinoid eggs. The first cleavage furrow forms generally by 1 hour post fertilization (hpf), with most eggs having fully cleaved by 2 hpf. *E. tribuloides* embryos undergo cell division through the expected radial cleavage pattern common to deuterostomes(Valentine, 1997). The first two cleavages occur meridionally, and by the third cleavage stage, the embryo consists of 8 cells of equal sizes, similar to euechinoids (figure 1.3). The fourth round of cell division marks where *E. tribuloides* and other cidaroids(Bennett et al., 2012a; Emlet, 1988) depart from the expected sea urchin developmental scheme. Unlike other euechinoids, which by the 16-cell stage have four clearly discernable micromere cells, *E. tribuloides* and other studied cidaroids appear to have a variable number of micromere cells that are inconsistent in size. In most cases, we were unable to differentiate between embryos that had

normal micromeres, and embryos that had undergone cell cleavage incorrectly and so were unviable. By the 32-cell stage (sometimes referred to as the morula) cells of the micromere lineage were indiscernible from the rest of the cell population. During this stage, the blastocoel begins to form and grows in diameter as subsequent cell divisions resulted in increasingly smaller cells (figure 1.3).

Stage 2: Blastula

By about 6 hpf most embryos reached the blastula stage. By this point, larvae have a fully open blastocoel. By ~9 hpf, most animals have shed the vitelline membrane and were free

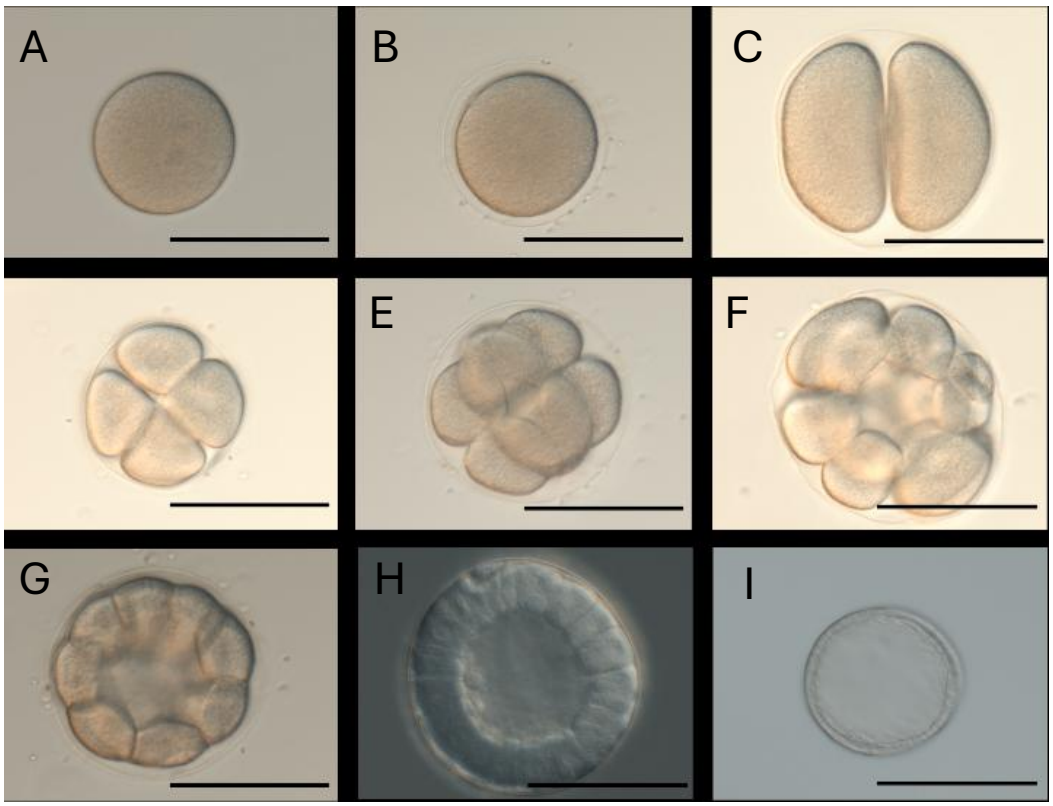

Figure 1.3: Fertilization and cleavage stages of *E. tribuloides*. Scale bar = 100 µm. **(A)** Unfertilized egg. **(B)** Fertilized egg, vitelline membrane present. **(C)** First cellular cleavage results in 2 equal sized cells. **(D)** Second cellular cleavage, resulting in 4 equal cells. **(E)** Third cellular cleavage, resulting in 8 equal sized cells **(F)** Fourth cleavage resulting in 16 cells, some of different sizes. **(G)** Fifth cleavage stage resulting in 32 cells **(H)** Later cleavage stage, unhatched blastula. **(I)** Mature hatched blastula stage, the vitelline membrane no longer present.

swimming. Embryos maneuvered through the water column using cilia. Typically around 12 hpf, blastulae had begun to reach the mesenchyme blastulae stage, in which the cells at the vegetal pole on the animal begin to thicken (figure 1.4). During this stage, body polarity becomes morphologically visible to the viewer. It should be noted that at no point did any animals demonstrate the precocious ingression of primary mesenchyme cells, which is in good agreement with past studies of early *E. tribuloides* embryology. As the blastula stage progressed, many animals were observed to shift from a spherical shape to a gentle oval shape, where the pole opposite of the vegetal pole (the animal pole) formed a mild point.

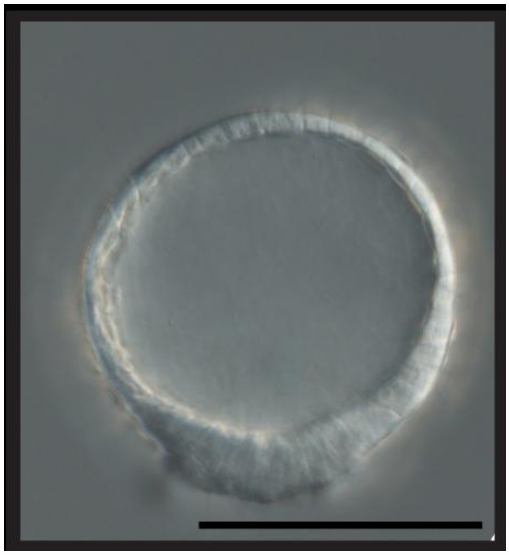

Figure 1.4: Mesenchyme blastula stage of *E. tribuloides*. Scale bar = 100 µm. Embryo orientated so the vegetal pole is pointing downwards. Advanced thickening of cells at the vegetal pole of the blastula.

Stage 3: Gastrula

Gastrulation generally began around 18hpf, with the ingression, or inward movement. of the sheet of cells at the vegetal pole (figure 1.5). This led to the formation of the blastopore, the precursor structure to the anus. As the archenteron elongated, mesodermally derived cells

delaminated off the tip and localized either to the ectoderm or to the blastocoel depending on the population (figure 1.6). As gastrulation continued, the archenteron began to curve towards the ventral plane of the animal. During this time, the axes of the animal becomes readily apparent.

Figure 1.5: Progression of gastrulation of *E. tribuloides* embryo. Scale bar = 100 µm. Embryo orientated in accordance to the body axis on the right. **(A)** Beginning of gastrulation, blastopore visible. **(B)** Elongation of the archenteron, cells are delaminating to populate the blastocoel. **(C)** Late gastrulation, pigment cells have begun forming. **(D)** Lateral view of late gastrulation.

Figure 1.6: Gastrula stage of *E. tribuloides*. Scale bar = 100 µm. Cells can be observed clearly delaminating off the tip of the archenteron.

The planes of the sea urchin embryo have been described in different ways, so we have used what appears to be the most common naming scheme to maintain consistency within the published sea urchin development literature. As aforementioned, the archenteron bends towards the ventral plane, where the mouth will eventually form. The dorsal face is directly opposite the ventral side. Gastrulation begins on the anterior end of the animal, where the anus will develop. The anterior side is opposite to the posterior. The animal swims with the posterior end forward. The left/right axis is defined as the animal's left and right side when viewed from the ventral surface.

The first pigment cells begin developing at approximately 20 – 24 hpf. Pigment cells can be identified through the presence of echinochrome A, a protein that has antibiotic properties and a conspicuous red/orange color. These cells populate the embryonic ectoderm. Pigment cells may develop within the blastocoel and travel to the ectoderm or develop directly within the ectoderm itself. *E. tribuloides* pigment cells are mostly punctate in shape, though some do adopt a slight stellate form.

Between 21 – 24 hpf, we observed some embryos begin to develop the coelomic pouches, hollow organs that are used for the eventual development of adult features. Some cells on the tip of the archenteron begin to delaminate to form one or two groups on the left and right sides of the archenteron, which eventually form hollow spaces. It appeared inconsistent if the two coelomic pouches would develop concurrently or in a staggered fashion.

Stage 4: Prism and Early 2 Arm

The prism stage is defined as the point at which two triradiate spicules form in the posterior left and right of the embryo. This is coupled with a slight left/right bulge in the ectoderm as the embryonic skin stretches to accommodate the growing spicules. The archenteron continued to grow in length until it is in the anterior one-third of the animal. The tip of the archenteron became considerably wider as the mouth developed. The body of the animal became distinctly "cupped-hand" shaped in which there was a mild depression on the middle ventral plane.

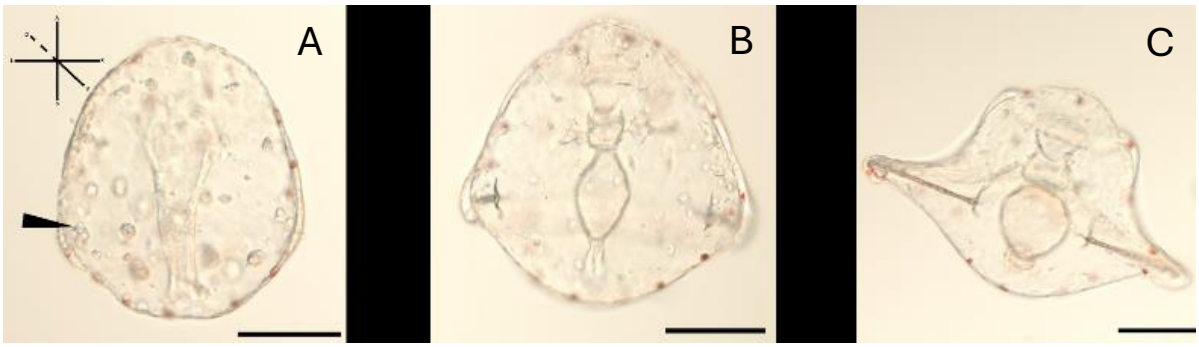

Figure 1.7: Progression of prism stage to the early two arm stage. Scale bar = 200 μm. **(A)** Black arrow indicates the first developed skeletal spicule. **(B)** Elongation of post-oral arms. **(C)** Two-arm stage larvae.

The two-arm stage was defined as the point when the animal's mouth has fully opened and the skeletal spicules have developed four clearly distinguishable elements. At this stage, the animal transitioned from an embryo to a feeding pluteus larvae. The larvae consumed *Rhodomonas* algae using cilia to guide the algal cells into the mouth. As the larvae fed, the stomachs became distinctly rotund as food entered and was digested. A thick, box liked esophagus developed directly posterior to the mouth. The digestive tract became more curved or C-shaped compared to the embryo. The tissue around the tips of the emerging postoral arms thickened and became populated by pigment cells. As the two-arm stage proceeded, the tissue of the oral hood (see figure 1.8) became increasingly more folded over the mouth, potentially being

driven by the extension of the anterolateral skeletal rods. Eventually, curve of the oral hood became less severe and more squared out as the anterolateral arm buds formed.

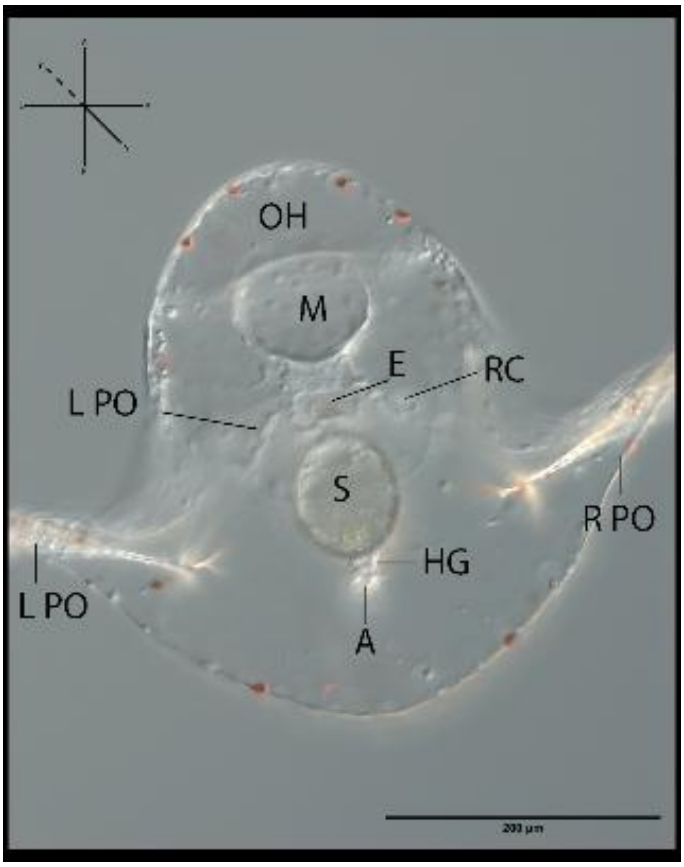

Figure 1.8: Anatomy of a two-arm stage *E. tribuloides* pluteus larva. Scale bar = 200 µm. OH (oral hood), M (mouth), E (esophagus), S (stomach), HG (hindgut), A (anus), LC (left coelomic pouch), RC(right coelomic pouch), (L PO (left postoral arm), (R PO (right postoral arm).

Stage 5: 4 arm stage

The anterolateral arms are the next set of larval arms to fully form. These arms are significantly shorter compared to the conspicuously long postoral arms, which continue to elongate during this phase. It appeared that the number of pigment cells continued to increase during stage, though currently we do not have specific numbers of this cell population. The

general body morphology of the animal continued to change, in which the body trunk (the middle of the animal) became more “pinched in” and the oral hood became increasingly squarer. The stomach became a darker colour, which would persist through the rest of larval life. It was unclear the exact causes of this colour change. Towards the end of this stage, the posteriodorsal spicules began to form, which were skeletal elements completely separate from the main skeletal structure.

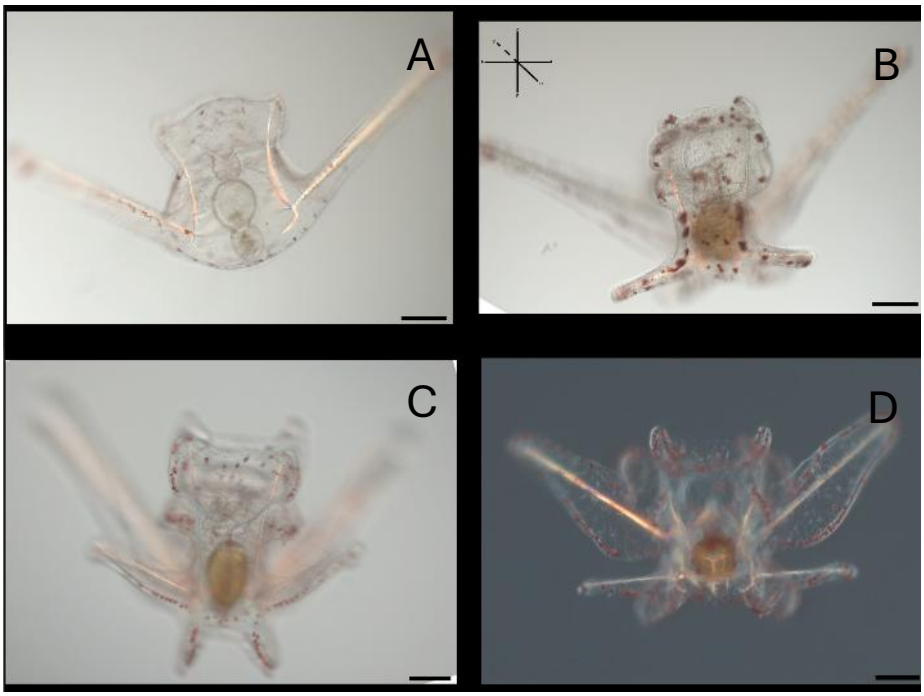

Figure 1.9: Progression of arm development. Scale bar = 200 µm. **(A)** Four arm stage, ventral view. **(B)** Six arm stage, ventral view. **(C)** Eight arm stage, dorsal view. **(D)** Eight arm stage, ventral view.

Stage 6: 6 arm stage

The six-arm stage of the *E. tribuloides* larval life cycle is major period of skeletal growth and increased skeletal morphological complexity. At this point, the sixth pair of arms, the posteriodorsal arms, have elongated and are well defined. During this phase, an additional skeletal element, the dorsal arm, has also become well developed (see figure 1.10). The dorsal arch is a Y-shaped structure that exists on the dorsal plane of the animal and extends both anteriorly and posteriorly. At the bifurcation point of the structure exists a complex web of interconnecting skeletal projections. The left and right anterioventral/posterioventral and anteriodorsal/posteriodorsal transverse rods extend to surround the stomach. It was seen that there might be some variation in the timing/symmetry of these transverse rods, as some individuals were missing a variable number of these skeletal elements. Whether this was caused by an endogenous source or by the rearing environment was not investigated. During the six-arm stage, the digestive tract also began to change. The mouth took on a distinctly square shape. The stomach elongated and became more oval in shape, as compared to the spherical shape it existed in during earlier stages.

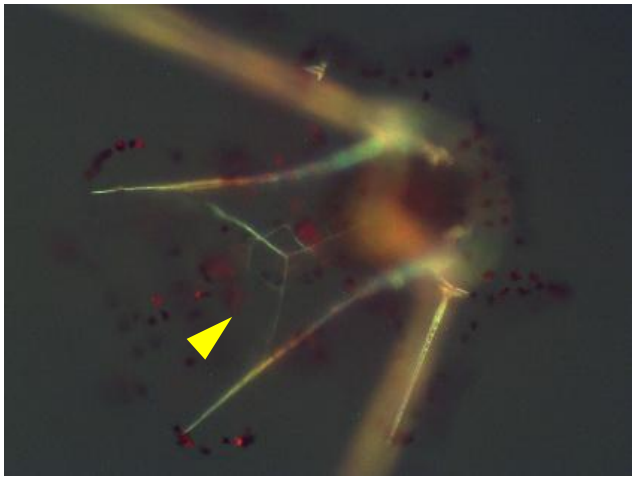

Figure 1.10: Six arm stage larva, dorsal view. Image taken using polarized light to make the larval skeleton more visible. The yellow arrow indicates the dorsal arch.

Stage 7: 8 arm stage

The eight-arm stage is marked by the growth of the final pair of arms: the pre-oral arms. These arms are supported by the pre-oral skeletal rods of the dorsal arch. The pre-oral skeletal rods curve in between the anterolateral skeletal rods in such a way that the pre-oral arms project ventrally in front of the anterolateral arms. There is an increase of epithelial tissue growth during this time as well in which 5 pairs of epithelial lobes (sometimes referred to as epilates) grow along the dorsal plane of the larval body. Most larvae during this stage exhibited distinct arm flexing motions, particularly in the post-oral arms. This supports previous observations of arm flexing in *Eucidaris tribuloides* (MacNeil et al., 2017). The exact cause of this motion was unclear, but this behavior was often seen in response to when the water holding the animal for observation was physically disturbed. The coelomic pouches were observed to migrate posteriorly.

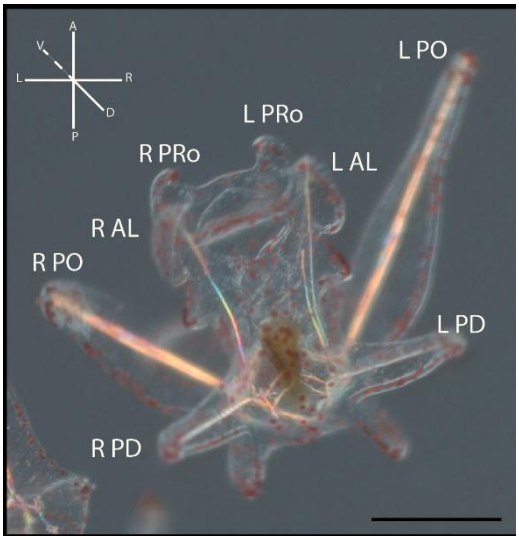

Figure 1.11: Anatomy of eight arm stage larva, dorsal view. Scale bar = 200 µm. RPO (right postoral arm), LPO (left postoral arm), RPD (right posteriodorsal arm), LPD (left posteriodorsal arm), RAL (right anterolateral arm), LAL (left anterolateral arm), RPRo (right pre-oral arm), LPRo (left pre-oral arm).

Stage 8: Pre-competency

Key features of stage nine development were a dramatic increase in the size of epithelial lobes and the post-oral and posteriodorsal arms. All ten epithelial lobes became extremely well defined and took on a "petal" like appearance. The tissues surrounding the post-oral and posteriodorsal arms became much thinner as these structures elongated. The anterolateral arms and pre-oral arms also increased in length. The tissue surrounding the gut became increasingly more pigmented or opaque, which made it difficult to observe the coelomic pouches during this stage. The animals still exhibited arm flexing behavior and were occasionally seen participating in substrate driven locomotion along the bottom of the observation glass, though we still consider them to swimming during this time.

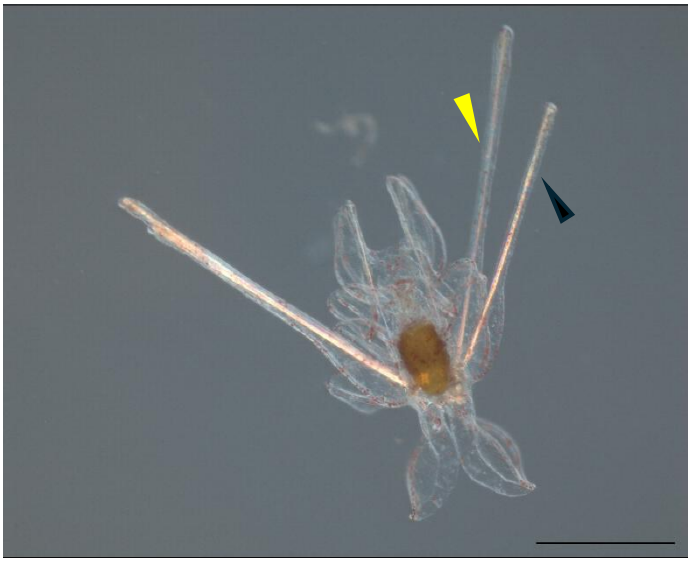

Figure 1.12: Pre-competency stage larva. Scale bar = 500 µm. Black arrow points towards posteriodorsal arm. Yellow arrow points towards post-oral arm. The well-developed epithelial lobes can be viewed. The animal has not yet begun forming juvenile structures.

Stage 9: Competency and Metamorphosis

Competency for *E. tribuloides* larvae was defined at the point in which the animal had at least one pedicellaria, one tube foot, one juvenile spine, and all ten epithelial lobes present. Given the opacity within the animal, it is difficult to describe what was occurring in the coelomic pouches. However, given that the closely related *E. thourasi* develops indirectly via the growth of juvenile structures from the coelomic pouch, it is likely that a similar process occurs in *E. tribuloides*. As metamorphosis began, the epithelial lobes were observed to recede into the center of the animal, however, no assay was performed to see if larval epithelium persisted in the post-metamorphosed juvenile. The tissues surrounding the post-oral and posteriodorsal arms receded or thinned out entirely, leaving the skeletal rods naked to the environment. The rods subsequently broke off. As metamorphosis continued, more juvenile structures became apparent. Eventually, the animal began using tube feet to locomote and from this point on the animals are considered to be benthic. Over the next several days, the juvenile spines became more robust and greater in

number and the individuals began eating using a structure very similar to an Aristotle's lantern. At this point, metamorphosis was considered to be complete, and the animal was now a fully formed juvenile *E. tribuloides*

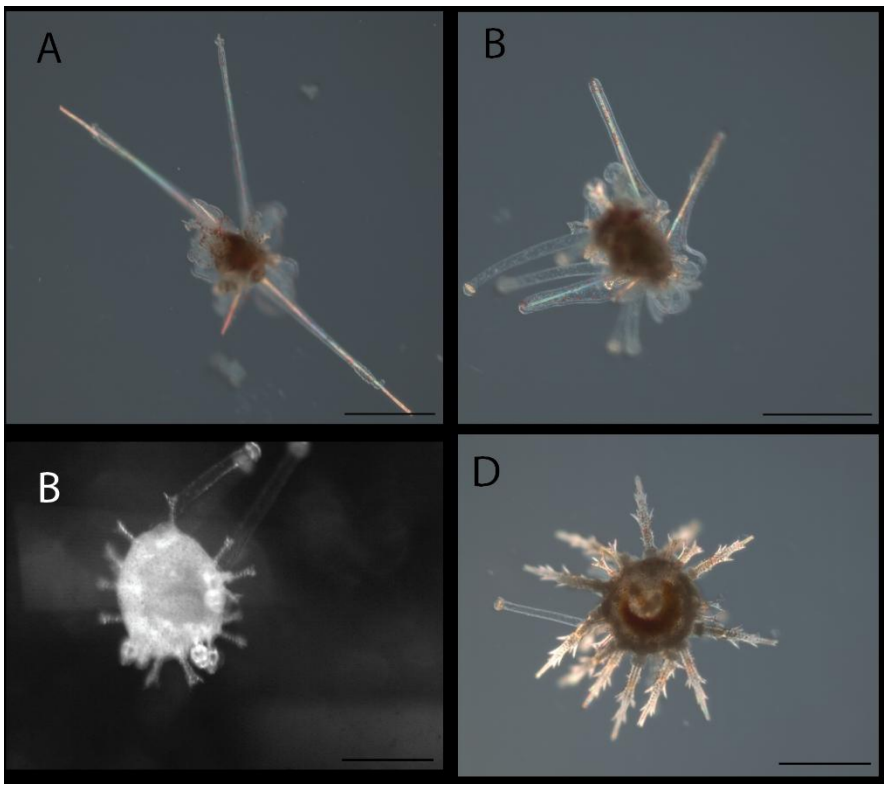

Figure 1.13: Progression of metamorphosis and juvenile *E. tribuloides*. Scale bar = 500 µm. **(A)** The beginning stage of metamorphosis. The bare skeletal arms are snapping off and juvenile structures are visible. **(B)** Multiple tube feet have erupted out of the animal. **(C)** All larval skeletal spines have snapped off. **(D)** Juvenile *E. tribuloides*, several days post metamorphosis.

Adult coelomocyte morphology

Given that both the *E. tribuloides* larval and adult skeleton/test structures have several key differences in comparison to *S. purpuratus*, we wanted to see if the adult coelomocytes also differed in morphology (figure 1.14). Based off the coelomocyte morphologies previous studies have characterized in *S. purpuratus* (Johnson, 1969; Matranga et al., 2005), we identified four

major coelomocyte types: vibratile cells, red spherule cells, colourless spherule cells, and phagocytes. The vibratile cells had a large flagellum characteristic of this cell type in other systems. Red spherule cells had two major morphologies- a punctate/spherical form and an ameboid or asymmetrical form. Colourless spherule cells were identified as separate from vibratile cells by the absence of a flagellum. These cells were also generally larger and had a conspicuous granular or "bumpy" texture. Phagocytic cells had long, outstretch filipodia, giving some a stellate appearance. Some phagocytic cells took on a petaloid appearance, similar to what is seen in *S. purpuratus.* We collected coelomocyte samples for scanning electron microscopy, though we were unable to confidently identify most cells in these images (figure 1.15).

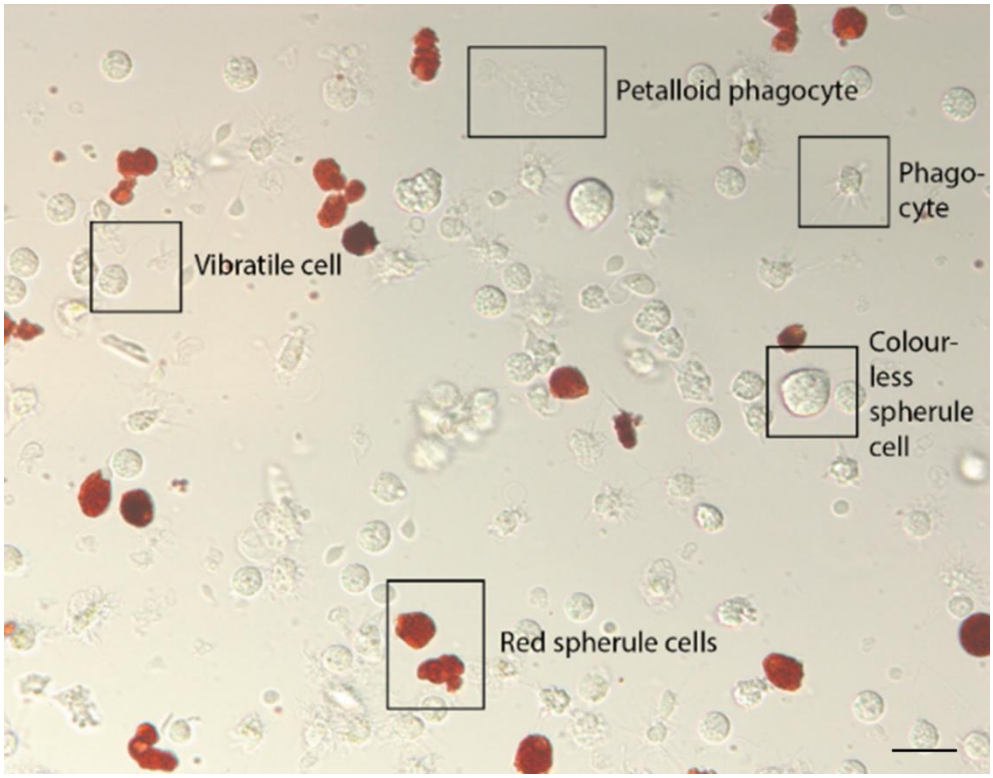

Figure 1.14: Adult *E. tribuloides* coelomocytes. Scale bar = 20 µm Cell types have been identified based off of morphological similarities to *S. purpuratus* coelomocytes. Cells that have been identified include vibratile cell, red spherule cell, phagocyte, petaloid phagocyte, and a colourless spherule cell.

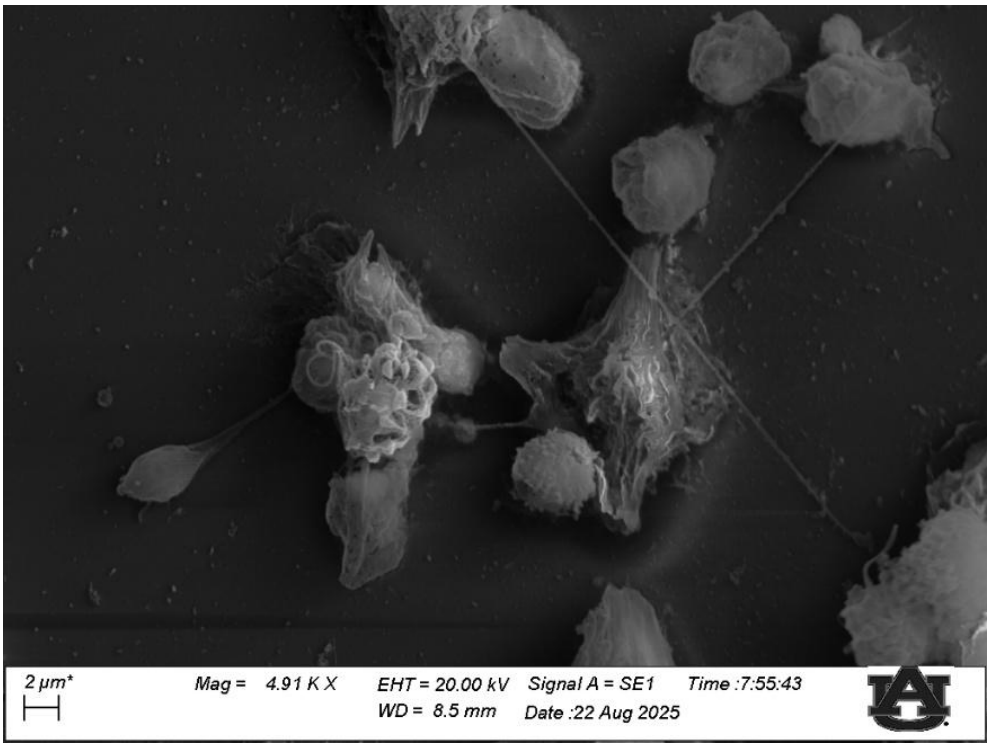

Figure 1.15: Scanning electron micrograph of mixed adult *E. tribuloides* coelomocytes. Scale bar = 2 µm.

Discussion

This study represents to the best of our knowledge the first in depth developmental staging scheme of *E. tribuloides* from fertilization through metamorphosis. Though the embryonic/early larval stages have been described, and the development of isolated organ systems has been tracked, there currently is no broad view of *E. tribuloides* development published. Current literature about *E. tribuloides* development has minimal description of husbandry protocols, making them difficult to reproduce. Here, we have been able to identify novel intermediate developmental phases and have generated a reference for developmental phases/husbandry strategies that will facilitate future investigations into the details of *E. tribuloides* growth.

Morphological comparisons among sea urchin larval and metamorphose stages

To our knowledge, only two other cidaroid species have ever had their development tracked from fertilization through metamorphosis: *Eucidaris thourasi* (Emlet, 1988) and *Cidaris blakei* (Bennett et al., 2012b). Comparisons between larval stages of these cidaroids and *E. tribuloides* highlight significant similarities and differences between this sea urchin species. First, it appears that the inconsistency in micromere cell number and size is a shared trait within Cidaroidea, having been observed in all three species. From the two-arm stage on, all three species share similar morphology, though *E. thourasi* developed significantly faster than either *E. tribuloides* or *C. blakei*. Other shared features include a lack of a hyaline layer and a lack of pre-gastrulation ingression of primary mesenchyme cells. *C. blakei* differed most of the three species in terms of early development. This animal boasted conspicuous dimpling during the blastula stage, of which the purposes of these invaginations are currently unknown. The general early pluteus morphology of this species also may be unique, in that is more "T-shaped" than what is overserved in other species.

When comparing the very late larval stages of *E. tribuloides* to other cidaroids and to euechinoids (Formery et al., 2022; George et al., 2004; Nesbit & Hamdoun, 2020), it becomes apparent that most indirectly developing sea urchins share very similar morphologies, especially during competency and immediately post-metamorphosis. As these different species grow into their adult forms, they once again become very different in terms of phenotype. Together, it appears that within sea urchins (or at least indirectly developing sea urchins), there is great diversity in developmental processes and morphologies during the embryonic and very early larval stages, less diversity during late larval and metamorphosis stages, and then again great diversity in adult morphologies. The reasons for this are currently unknown, but merits future

investigations to elucidate. One possibility is that there is great environmental pressure during the late larval/early juvenile stages that leads the animals to evolve the same general phenotype over again. It is also possible that this is due to a shared ancestry among echinoids and that the gene regulatory networks that direct these phases of development are extremely deeply conserved.

Eucidaris thourasi

Perhaps unsurprisingly, *E. thourasi* and *E. tribuloides* are incredibly similar in the ways that development proceeds. At this point in time, it appears that these species are nearly identical at every stage, from fertilization to metamorphosis. Though these species belong to the same genus, they have been geographically separated by the Isthmus of Panama for at least 3 million years. *E. thourasi* is found on the Pacific side(G., 1936) of landmass while *E. tribuloides* is found on the Atlantic side (McPHERSON, 1968). Despite this separation, these animals seem to still have very similar physiologies, body structures and temperature ranges, differentiated only by subtle variations in test morphologies during adults (G., 1936). Since the construction of the Panama Canal, a potential anthropogenic corridor has been connecting the Pacific and Atlantic oceans through Central America for the last 109 years. While Panama Canal is not a free flowing channel between these two oceans, there have been reports of marine organisms being transferred from one side through another by shipping (usually in the ballast water) (Carlton, 2011; Cohen, 2006). While it is unlikely that the benthic sea urchin adults have been moved across this canal in such a way, it is possible that the planktonic larvae of *E. tribuloides* and *E. thourasi* have been transported from one side of Panama to the other. While there are currently no reports of this occurring, we believe that the potential for biological flow should be investigated. If *E. tribuloides* and *E. thourasi* were to come into second contact through either

natural or anthropogenic means of dispersal, such an event would provide a unique opportunity to study reproductive compatibility, potential for hybridization, and the extent of developmental conservation between animal species that are closely related phylogenetically but separated by a geographical barrier.

Cidaroids and a changing ocean

It is well known that Earth's oceans have generally become more acidic over time, particularly due to anthropogenic sources of carbon dioxide emissions(Doney et al., 2009). This is cause for concern as the calcium carbonate structures of marine invertebrates are prone to dissolve in acidic environments(Morse et al., 2007), and indeed this pattern has been observed across many taxa including corals(Eyre et al., 2014), mollusks(Gazeau et al., 2013), and echinoderms(Dubois, 2014). A recent study has found that, in comparison to euechinoids, *E. tribuloides* is more resilient to ocean acidification(Dery et al., 2017). Specifically, it was demonstrated that across the range of tested pH's (8.2 – 7.2), *E. tribuloides* test plates showed very minimal signs of erosion, and that the primary spines maintained their mechanical strength. These findings are corroborated by another study that also saw a reduced dissolution response of the spines of cidaroid urchin *Phyllacanthus imperialis* to an acidic environment(Dery et al., 2014).

These observations are interesting in the context of the evolutionary history of cidaroids. Fossil records have established that these animals survived through the ocean acidification events of the Permian-Triassic period. Thus, it is possible that members that survived this period possessed adaptations that allowed them to be resistant to acidic environments. While these traits have been studied in adult cidaroids, at this point there exists no study in which the larval response to acidification has been observed. Now that developmental staging schemes for *E.*

tribuloides have been established, the physiological and biochemical responses of these organisms to ocean acidification during early and middle life stages can be determined. These future efforts could provide novel insights into how different marine invertebrates are able to successfully respond to changing ocean conditions.

Similar structures, different development

We observed several adult coelomocyte cell types that have similar morphologies to known coelomocyte populations in *S. purpuratus*. This may be significant, as it has been demonstrated that while the way mesodermally derived tissues in *E. tribuloides* have altered developmental patterns in comparison to euechinoids, some tissues may exhibit more conserved developmental patterns than others (i.e., the skeleton differs more between species than the coelomocytes). This could point towards the sea urchin immune system being under more restrictive selective pressure than the skeleton in that there is a narrower range of biologically acceptable/viable coelomocyte morphologies. The next step in this research would be to characterize the immune response of *E. tribuloides*, as though these cell types look familiar, it is possible they could demonstrate an altered functionality.

Chapter 2

Introduction

Overview of Gene Regulatory Networks

A developmental gene regulatory network (GRN) is the collection of regulatory modules, transcription factors, signaling molecules, downstream genes and *cis*-regulatory sequences which determine developmental states and drive these processes forward. By controlling gene expression patterns, GRN's are able to temporally and spatially regulate developmental processes at all levels, from cell identity to cell proliferation to tissue formation to arrangement of body parts. These *cis*-regulatory and signaling molecule interactions are often represented as a network of interactions, as a single transcription factor is capable of binding to multiple *cis*-regulatory modules, one module often requires multiple transcription factor inputs, and transcription factors and signaling molecules are able to interact with each other and modulate binding affinity(Levine & Davidson, 2005). An important aspect of these regulatory networks is that the structure is inherently hierarchical- major circuits are made of subcircuits. This is significant in that multiple layers of regulation allow for gene expression to be finetuned, and that any upper-level change has the potential to reverberate down through the network as a regulatory cascade. It can also be observed that the structure of the network, the organization of linkages, defines the function of that network or battery. Specific inputs are necessary for cell fates to be decided, and in that way GRN's provide a causal mechanism that connects a genotype to any given phenotype.

Regulatory Logic and Mesodermal Development

Gene regulatory networks are extensive webs of complex interactions. As these networks' functions are defined by their internal organization, it is necessary to understand the structural architecture of these programs to be able to accurately predict how regulatory information will be processed during development(Davidson et al., 2002; Davidson, 2006). Maternal and early zygotic developmental inputs establish the initial regulatory states of tissues(Kipryushina & Yakovlev, 2020; "Setting the Stage for Development," 2023). As development proceeds, intermediate inputs help to refine spatial domains. Later, downstream regulatory modules activate cell-differentiation genes, and thus the result is population of cells with specific functions and clear boundaries between different types of tissues. It is important to understand, however, that this hierarchy does not exist in a simple, linear fashion. The regulatory sequences of developmental genes often include binding sites that are recognized by their downstream transcription factor targets. In this way, feedback and feedforward loops(Mangan & Alon, 2003) can be established, so that the network both begins and maintains its own signal(Hart et al., 2012). Downstream targets may even interact negatively with their upstream signals so that they repress their own expression and achieve tightly controlled levels of transcription. A major consequence of this sometimes-cyclical interactive behavior is that transcription factors and *cis*-regulatory binding sequences that control expression domains are, in turn, controlled by themselves. Transcriptional regulators are in turn both the cause and effect of that respective developmental state. In other words, the network contains the instructions, encoded, to control itself. Another crucial point to understand about GRN structure is that regulatory genes encode multiple binding sites that can accommodate both multiple of the same transcription factor and multiple different transcription factors. What will occupy these DNA sequences then relies on

what transcription factors are present, and in want concentration, at a given time (Istrail & Davidson, 2005).

As *cis*-regulatory modules integrate multiple transcription factor inputs, certain types of interactions or regulatory logic materialize, which govern ultimately how, and in what ways, a gene may be expressed. There are two main types of regulatory effects: transcription factors can either directly or indirectly act as activators of genes or they can be repressors. From this fairly simple foundation, more complex patterns of relationships are observed, often referred to as network motifs(Shen-Orr et al., 2002). For the past several decades, Boolean logic has been used to model these motifs. Using Boolean logic, developmental signal transduction events can lead to cascades of regulatory events single or combination of multiple and/or/not statements (Arnosti & Ay, 2012). For example, A *and* B must bind for C to be expressed, or X must be present but *not* Y for Z to be transcribed. Currently, it is known that seven general types of network motifs contribute to sea urchin endomesoderm development(Peter & Davidson, 2009), but many of these motifs have also been discovered in other animals(Borotkanics & Lehmann, 2015; Defoort et al., 2018; Yagita & Okamura, 2000), plants(Joanito et al., 2018; Jones & Vandepoele, 2020), and prokaryotes(Alon, 2007; Eichenberger et al., 2004).

Of these seven, two types of motifs, or subcircuits, are particularly relevant to the development of mesodermally derived tissues: the double negative gate and the reciprocal repression loop (see figure 2.1). A double negative gate can be thought of as the repression of a repressor. The major function of this subcircuit is to establish defined boundaries between different tissues and segregate cell fates. Reciprocal repression loops are generally used to

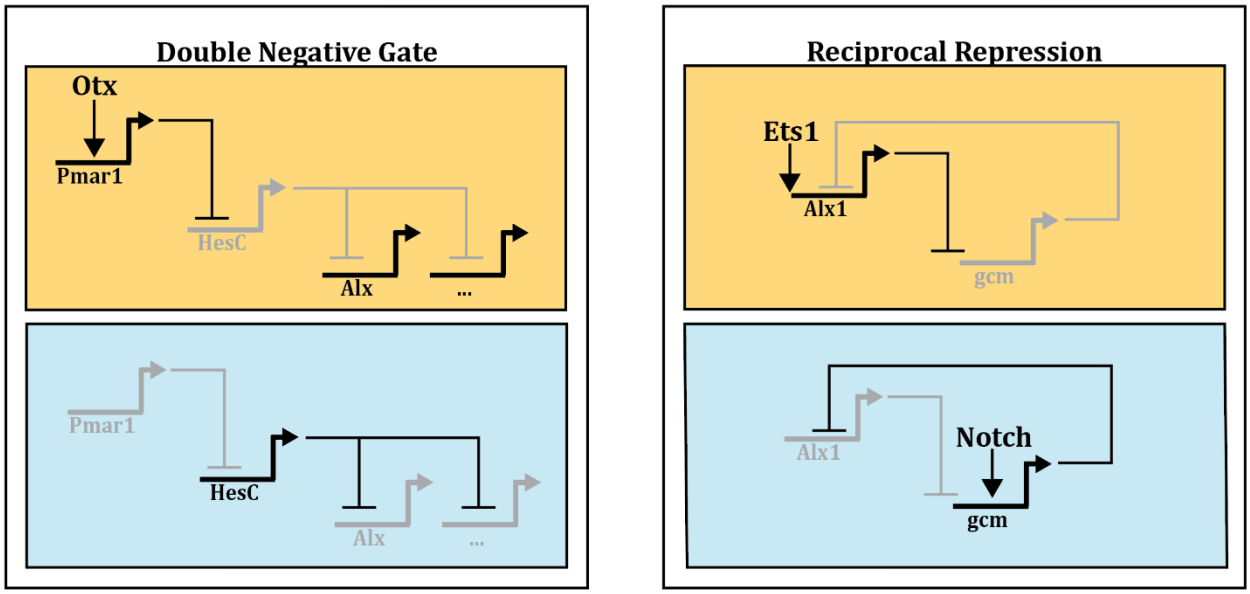

Figure 2.1: Examples of GRN motifs as they apply to canonical echinoid mesoderm development. Grey text indicates those genes are not being expressed. Ellipses indicates that more genes are involved but not shown. For the double negative gate, when Pmar1 is present, HesC is repressed. There is nothing to repress Alx and other skeletogenic genes. When HesC is present, it represses the expression of Alx and other skeletogenic genes. For reciprocal repression, when Alx1 is being expressed, it acts as a repressor of non-skeletogenic genes, such as gcm. When gcm is present, it acts as a repressor of skeletogenic genes, such as Alx. Together, these motifs act to define spatial and temporal domains during development of tissues.

reinforce and stabilize cell fates by preventing alternative gene batteries from being expressed. By using these motifs to generate sharp boundaries between different mesodermal tissue fates, organisms are able to tightly regulate the temporal and spatial development of specific tissues and ensure that the correct cells are organized into the correct developmental domain.

Development of the larval skeleton begins during the 16-cell stage of the sea urchin embryo during which time the micromeres form. Within this cell lineage, the accumulation of β – catenin within the nucleus, its interaction with TCF, and the additional input of Otx drive the expression of Pmar1. Pmar1 then acts as a transcriptional repressor, preventing HesC from being expressed in the micromeres. This in turn is what allows for major transcriptional enhancers of skeletogenic genes, especially Alx1, Ets1, and Tbr, to be expressed, setting up this cell lineage to become specified into the primary mesenchyme cells. Ultimately, these primary mesenchyme

cells will ingress into the blastocoel and differentiate into skeletal mesenchyme cells through the interactions of downstream genes.

Non-skeletal mesenchyme cells, on the other hand, arise from the veg2 mesomeres during the cleavage stage embryo. These cells lack the Pmar1 signal, and so HesC is active and represses Alx1 and other genes used for skeletal development. Delta/Notch signaling induces a subset of the veg2 mesomere cells to become non-skeletal mesenchyme cells by acting as transcriptional inputs for gcm and gata genes. Gcm is a major regulator of the non-skeletal of pigment cell fate, by acting as a main upstream enhancer for the expression of pigment cell differentiation genes, such as Pks and Sult. It should be observed that Gcm and Alx1 act as reciprocal repressors, preventing cells of mixed mesodermal lineages from developing.

GRN evolution

Gene regulatory networks provide a powerful substrate on which evolution can act upon leading to the production of modulated or entirely novel animal body plans. Evolutionary developmental research over the past several decades has produced a paradigm that postulates it is the rewiring of GRN interactions, and particularly those changes in *cis*-regulatory modules, rather than invention of new genes, that is primarily responsible for altered morphology (Davidson & Erwin, 2006, 2009; Erwin & Davidson, 2009). It can be observed that while GRNs can provide a mechanistic means for evolution, they also provide limitations. As alluded to above, there are generally only a small number of GRN subcircuits common to animals. These subcircuits also differ in their stability and probability of providing biologically meaningful phenotypic changes. The most stable GRN circuits are often referred to as “kernels” and are very unlikely to change. Other subcircuits, especially those found at the termini of networks, are more flexible and likely to change. In this way, the probability of changing differs through

developmental time. It can also be observed that alterations at different layers of regulation will have different functional consequences Changes in the types, locations, and timing of interactions act as primary ways that development can evolve.

At the genomic level, the *cis*-regulatory sequences of GRNs may be mutated in several fashions to give rise to an altered regulatory state. Point mutations and base insertions/deletions can impact transcription binding affinity, cause singular or multiple gains or losses of binding sites, and change the types of interactions between *cis*-regulatory sequences(Erwin & Davidson, 2009; Stone & Wray, 2001). Entire *cis*-regulatory modules may be inserted into new genomic regions through the actions of transposable elements or even retroviruses(Britten & Davidson, 1969; Erwin & Davidson, 2009; Sundaram et al., 2017). Replication errors, such as slippage, can cause the duplication of *cis*-regulatory modules, which may lead to sub-functionalization of the duplicated module and even the invention of novel network subcircuits (Voordeckers et al., 2015). These genomic mutations of the developmental network can have a multitude of functional consequences depending on what type of mutation occurred (Voordeckers et al., 2015) and specifically *where* this mutation occurred within the network hierarchy (Erwin & Davidson, 2009).

At the broadest level, GRN change can be thought of in terms of alterations of gene product, which have been organized into four major categories. Heterotypy is the change in the protein (in this context a transcription factor or signaling molecule) that allows for a gain and/or loss in binding activity and is typically caused by mutations in the coding sequence. Heterotopy is the change in where a regulatory gene is being expressed and may be brought on by insertion or movement of *cis*-regulatory modules, often by the action of mobile elements. Heterometry is a change in how much protein is being made, commonly caused by insertions or duplications of

multiple binding sites in the *cis*-regulatory sequence. Heterochrony is the change in *timing* of the expression of a gene or regulatory module. While all these module perturbations can have drastic consequences on the development and evolutionary trajectory of a given system, heterochronistic change is one of the most relevant in terms of sea urchin mesoderm development. About 250 million years ago, the echinoid ancestor began expressing Pmar1 in micromere cells. This allowed for the adult skeletogenic program to be expressed during the embryonic stage through the novel double negative gate described above.

Developmental Regulation in Cidaroidea

While it has been known for several decades that cidaroid urchins develop their mesodermally derived tissues (particularly the larval skeleton) differently from euechinoids, descriptions of the transcriptional regulatory interactions underlying these patterns have only just begun. In general, it appears that while some core features of general mesoderm development are conserved within Echinoidea, euechinoids demonstrate several occurrences of regulatory rewiring that has lead to a derived mesodermal state.

Of the two general mesoderm tissue types, the primary mesenchyme cells and later skeleton specific cells have the most conspicuous alterations. It has been observed that Et-erg is expressed more broadly and for a longer amount of time, relative to developmental stage, than Sp-erg, which eventually becomes PMC specific (Erkenbrack et al., 2016). Et-hex shares a similar spatial and temporal expression pattern to Et-erg, which is suggestive that these transcription factors use the same unknown activator signal. As in echinoids, Et-tgif is maternally deposited and expressed early in putative micromere cells. Together, it appears that while the use of erg/tgif/hex in mesodermal development is a deeply conserved GRN subcircuit within echinoids (and echinoderms in general), the distinct expression domains of these genes to

primary mesenchyme cells in euechinoids is novel trait specific to this group (Erkenbrack et al., 2016). One of the most severe differences in skeletogenesis between cidaroids and euechinoids is the Pmar1/HesC double negative gate. Some studies have found Pmar1 to be entirely absent from *E. tribuloides* and postulated that cidaroids lack pmar1 entirely (Erkenbrack & Davidson, 2015), though more recently, a putative Pmar1 transcript was found in the cidaroid sea urchin *Prionocidaris baculosa* (Yamazaki et al., 2020). Though the presence of a cidaroid Pmar1 may be contentious, all groups agree that the Pmar1/HesC double negative gate behaves differently in Cidaroidea. Perhaps the strongest line of evidence for this was the discovery that *E. tribuloides* expressed HesC alongside Alx1 and other skeletogenic genes, indicating that HesC does repress skeletogenesis in this system at all. It has also been observed that while Alx1 does function in regulating skeletal development in *E. tribuloides*, it is expressed much later compared to *S. purpuratus*. Neither Delta nor tbr appear to play major roles in the cidaroid larval skeleton specification, and while Ets1 is present, it was found to be broadly expressed throughout the early mesoderm.

The non-skeletal mesenchyme tissue development subcircuits appear to be more conserved, but also behaves somewhat differently between these major sea urchin lineages. The *E. tribuloides* nodal, ese, and gcm genes have similar spatial/temporal expression patterns and play similar roles to the euechinoid homologs in that they drive non-skeletal mesenchyme polarity. Downstream genes in these networks (GataE, GataC, Scl, and Prox) are generally delayed in expression and have somewhat different expression domains. Interestingly, some skeletogenic genes, including ets1 and tbrain were found to be expressed in the same regions as these non-skeletal genes, showing that the general cidaroid mesoderm tissue is less segregated than in euechinoids.

Advancing our Understanding of Cidaroid Mesodermal Development

Current echinoid GRN research is focused primarily on euechinoid systems. While this has allowed for the creation of some of the most well understood developmental networks to date, it does mean that the gap in our understanding of how these networks can change and evolve within Echinodermata has been largely ignored. This is especially clear when looking at the evolution of mesoderm development. While it is understood that there are fundamental differences in the entire mesodermal network, the ways in which specific features, like the double negative gate, have evolved is unknown. Perhaps as consequence, the genomic and genetic resources available for cidaroids is extremely minimal, especially when compared to that of *S. purpuratus*. In fact, of the 85 echinoid genomes currently on the National Center for Biotechnology Information genome database, there is only one genome for Cidaroidea (*E. tribuloides*; accession number = JZLH000000000). This genome is highly fragmented and lacks any annotation. Available transcriptomic data for this group is similarly sparce. In order for the cidaroid developmental GRN to be fully realized, a better grasp of the genetic components of these animals must be developed.

Thus, our goal with this study is two-fold. We aim to increase to available genetic data for cidaroids, using *E. tribuloides* as the group representative, through the assembly of a developmental transcriptome. Developmental transcriptomes are a necessary resource for GRN structure discovery, as they provide a catalogue of what genes are being expression when for a given developmental stage. By annotating this assembly, we aim to provide a dataset that can facilitate the reconstruction of the ancestral echinoid GRN and the identification of cidaroid regulatory modules. Sequencing transcripts from multiple early *E. tribuloides* developmental stages can allow for cross-species gene expression comparisons to be made, which will be

essential for identifying the ways in which the echinoid developmental GRN has changed (heterochrony, heterometry, and so on). *E. tribuloides* is uniquely positioned in that it belongs to a basal echinoid lineage, has clearly marked differences in developmental patterns and resulting morphology, but it not so distantly related to euechinoids that it would be extremely difficult to identify homologous regulatory modules. This study marks a critical step forward in understanding how conserved regulatory subcircuits can be altered through differences in timing of expression and regulatory interactions.

Methods

RNA Collection

We developed two cultures (culture A and culture B) from mating pairs female 1 and male 1/female 2 and male 2. These adult urchins were unrelated to one another. Larval RNA was collected from both cultures such that each sample had a biological replicate. Adult RNA was collected from gut and coelomocyte tissues of the adult female of culture A and adult male of culture B.

Larval RNA collection

We collected larval RNA from six major developmental stages: cleavage stage (16/32 cell stage), blastula stage, mesenchyme blastula stage, early gastrula stage, prism stage, and the two arm stage. For each collection point, roughly 5,000 embryos or larvae were collected using a 40 µm mesh cell strainer. The animals were then centrifuged, the supernatant was aspirated, then the pellet was re-suspended in 600 µL of TRIzol. RNA could be extracted immediately or the sample could be stored in TRIzol at -80°C until needed. For RNA extraction, embryos were allowed to sit in TRIzol for 5 minutes at room temperature. 120 µL of chloroform was added, then the

solution was vigorously shaken for 15 seconds or until the solution turned light pink. This was allowed to sit at room temperature for 5 minutes. Next, the solution was centrifuged at 12,000 rpm for 15 minutes at 4°C. After centrifugation, three layers were visible: a clear layer on top, a white layer of lipids, proteins and other debris, and a pink layer of excess TRIzol at the bottom. The clear layer was transferred, taking care to not disturb the underlying white layer, into a clean 1.7 mL tube. 300 µL of 200 proof ethanol was added to the tube, the solution was mixed, then transferred into a Zymo-Spin™ IC column, along with 300 µL of Zymo RNA wash buffer. Sample was centrifuged for 1 minute at 12,000 rpm and the flowthrough was discarded. An additional 400 µL was Zymo RNA prep buffer was added to the column and the previous step was repeated. Then Zymo RNA wash buffer was brought to 4°C and 700 µL was added to the column. The sample was centrifuged for 2 minutes at 12,000 rpm. 400 µL of wash buffer were added to the column and the centrifugation step was repeated. The column was then transferred to an RNase free tube and the RNA was eluted using 12 µL of RNase free water.

Adult RNA Collection

Gut tissue and coelomocytes were collected from the adult female of culture A and the adult male of culture B. To extract cut tissue, the sea urchin was cut using dissection scissors starting from the peristome. The animal was cut along an ambulacral plate, where we then cut straight across the periproct and down the opposite ambulacral plate, effectively splitting the animal into two equal halves. The intestines were then collected using forceps and placed into a 15 mL conical tube. We centrifuged the tissue, removed excess ASW, and resuspended the issue in 2 mL of TriZol. The tissue was physically disrupted using a homogenizer, which had been treated with 70% ethanol and rinsed with milli-Q filtered deionized water. The tissues were continuously for several seconds and allowed to sit for 5 minutes. The solution was then briefly

centrifuged to remove unnecessary solid tissues from the supernatant, as we found allowing it to stay severely clogged the Zymo column in later steps. We then split the 2mL supernatant between two 1.7 mL centrifuge tubes and added 200 µL of chloroform to each tube. We shook the tubes by hand for 15 seconds or until the mixture turned light pink and allowed the solutions to sit at room temperature for 5 minutes. The solution was then centrifuged at 12,000 rpm for 15 minutes at 4°C and the clear layer was pipetted into a clean 1.5 mL tube. We found that the solutions still contained excess debris, so the solutions were briefly centrifuged a second time at 10,000 rpm. The resulting supernatant was collected, and RNA was extracted using the same methods as described above. Coelomocytes were collected using the methods described in Methods in Cell Biology by Smith et al. (*Echinoderms. Part A*, 2019), with one modification. Namely, we extracted as much coelomic fluid from the adults as possible, as animal viability was not a concern. The coelomocytes were briefly centrifuged at 10,000 rpm, supernatant removed and resuspended in 600 uL of TriZol. RNA extracted proceeded as described above.

DNA contamination removal

For all samples, DNA was removed using the routine DNase treatment procedure described in the *TURBO DNA-free Kit*™ manual by Thermo Fisher Scientific.

Pre-assembly processing

A summary of our workflow can be viewed in figure 2.2 below. RNA samples were submitted to Novogene Corporation Incorporated (Sacramento, CA, USA) for library preparation and sequencing. After cDNA synthesis and subsequent library preparation, samples were sequenced using the Illumina platform and PE150 technology, producing paired end reads that were 150 base pairs in length. We performed an initial quality control step on the sequenced

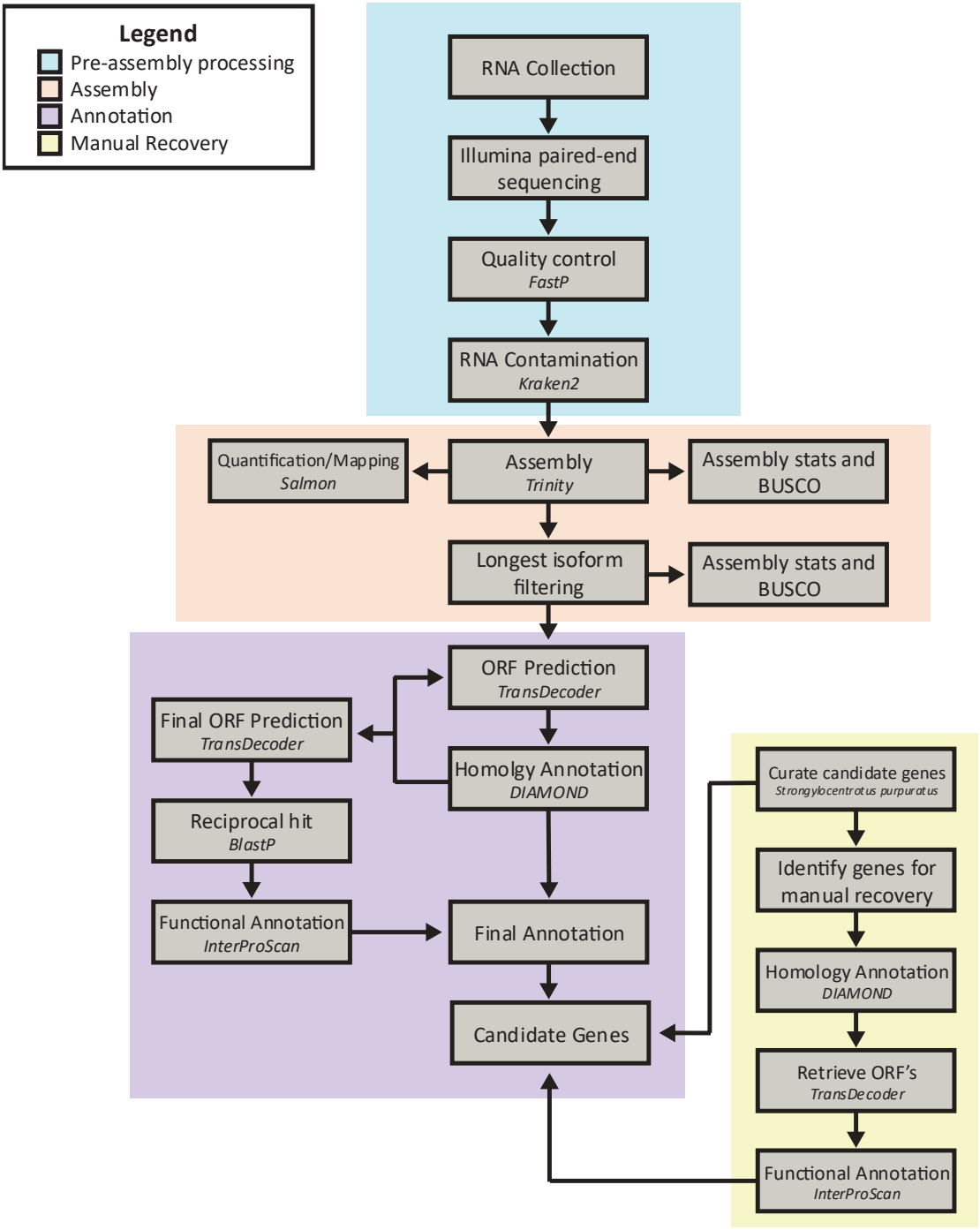

Figure 2.2: Summarized flowchart of computational transcript assembly and annotation of *E. tribuloides*

FASTQ files using fastp(Chen et al., 2018). Fastp was used to trim bases that had a quality score of less than Q20, remove reads where 40% or more of bases had a quality score of less than Q20, remove reads that had more than five ambiguous bases, remove reads that were less than 50 base

pairs long, remove artificial poly-X tails, and remove adapter sequences from reads. Using Kraken2 (Wood et al., 2019) and the standard Kraken2 database, we identified contaminating transcripts that closely matched known bacterial, fungal, and mammalian sources. We found a negligible amount of contamination across all samples (<1 – 2%).

Assembly

While a scaffold level genome for *E. triubuloides* is available through the National Center for Biotechnology Information (accession number JZLH000000000), we found it too fragmented (637,071 scaffolds; N50 = 39.2 kb) and computationally expensive to use as a reference genome for RNA-seq alignment. Therefore, we assembled our transcriptome using the *de novo* RNA-seq assembler Trinity v2.14.0 (Grabherr et al., 2011), using nearly all default parameters. To decrease assembly time and resource demand, we used the built-in *in silico* normalization function Trinity has using the flag (--normalize_by_read_set). Highly expressed genes can result in transcripts of unworkable read depth, which demands unnecessarily expanded computing power and time to assemble (Raghavan et al., 2022). Therefore, normalization is an essential step during the assembly process. We collected general assembly metrics using the TrinityStats.pl and contig_ExN50_statistic.pl, which are perl scripts included in the Trinity suite. An N50 score can be a problematic method to gauge an assembled transcriptome's quality, as these assemblies by their nature have many relatively short sequences, and this score can be influenced by expression level, the number of isoforms present, and the length of isoforms. It has been suggested that the ExN50 (weighted expressed N50) score is a more appropriate metric, as it represents the mean isoform N50 statistic per top X% expressed transcripts(Geniza & Jaiswal, 2017). To gauge the completeness of our assembly, a BUSCO (Simão et al., 2015) score was calculated using the metazoa_odb10 database.

We quantified transcript expression using Salmon(Patro et al., 2017) using all default parameters. Salmon is a quasi-alignment tool that relies on *k-mer* similarity to map reads to an assembly and quantify them. Salmon was chosen over strictly base-to-base alignment tools for its faster runtime and comparable accuracy (Jin et al., 2017).

A feature of Trinity is that it identifies a number of isoforms per gene cluster. In order to decrease computational resource demand and facilitate data analysis, we chose to represent each gene cluster by the longest isoform present. We did not filter by the most highly expressed isoform, because most highly expressed isoform may change over developmental time or tissue type. To collapse gene clusters, we used the get_longest_isoform_seq_per_trinity_gene.pl script available in the Trinity suite. This longest isoform or unigene (unique gene) dataset was assessed for quality in the same manner as the complete Trinity assembly.

Transcriptome Annotation

The unigene dataset was used as the input for Transdecoder (Haas, BJ. https://github.com/TransDecoder/TransDecoder) in order to identify predicted open reading frames (ORFs) and amino acid sequences for each trinity gene. The longest ORF's were predicted using TransDecoder.LongOrfs and the subsequent protein sequences were generated using TransDecoder.Predict using default settings. Using the output peptide file, homology annotations were produced by using the DIAMOND (Buchfink et al., 2015) alignment tool. For this, we specified the –blastp flag for protein/protein alignment function against the SwissProt protein database, filtering for one hit per sequence with a minimum e-value of 1e-5. These hits were then used to fine tune our predicted protein sequences by feeding them back into TransDecoder.Predict using the (--retain_blastp_hits) function.

Next, we performed a reverse homology annotation using a reciprocal best hit (RBH) approach between the finalized, DIAMOND supported unigene dataset and the *Strongylocentrotus purpuratus* genome in order to identify high confidence orthologs in our assembly. To do this, we generated DIAMOND custom protein databases using our predicted protein file and the *S. purpuratus* protein fasta (v5). Then, we used the –blastp function to align the datasets to each other to find the RBH's, or sequences that were each other's top alignments for each dataset.

InterProScan (Buchfink et al., 2015) is an umbrella tool that was used in combination with our *E. tribuloides* unigene peptide file to produce functional annotations for our assembly. Using this tool, we utilized multiple protein signature databases, including Pfam, SMART, SUPERFAMILY, PROSITE, and TIGRFAMs, to identify conserved protein domains, functional motifs, and gene ontology (GO) terms.

A list of 377 transcription factors, signaling molecules, and receptors that have been implemented in *S. purpuratus* developmental regulatory processes was curated. The protein ID's were compared to our RBH list to begin identifying candidate proteins in *E. tribuloides*. Of these 377 proteins, 70 were missing from the RBH's. To manually recover candidate sequences for these proteins, we generated a small custom DIAMOND database from the *S. purpuratus* proteins and aligned our unigene fasta file against this database and filtered by the single best hit per sequence using bitscore and e-value. A subsequent reciprocal blastp step demonstrated that all of our recovered *E. tribuloides* sequences also aligned to the same *S. purpuratus* proteins. These sequences were annotated using InterProScan as above.

WGCNA

In order to gain an overview of the regulatory processes and coordinated regulatory gene expression events that occur during *E. tribuloides* development, we performed a Weighted Gene Co-expression Network Analysis (WGCNA) in R using the WGCNA v1.47(Langfelder & Horvath, 2008) using the larval sample expression counts generated by Salmon. As the adult tissues were not part of the devised developmental scheme and could introduce artificial variation, they were removed prior to count normalization using DESeq2 and variance stabilization. To decrease computational resource demand and decrease noise, this developmental dataset was further filtered to maintain only the Trinity reads that we in our final annotation matrix and that had expression values above 0 transcripts per million (TPM). After normalization, reads that had a variation score of less than 0.10 were removed. After a quality check using goodSampleGenes, for which all of the remaining reads and samples passed, this matrix was used to perform sample clustering to identify any outlying samples that may obfuscate results. We found that all of our samples clustered so that similar developmental stages were together and that there were no outlier samples.

An important step to performing a WGCNA is choosing an appropriate value for the soft-thresholding power. This power (β) is used to transform the pairwise gene correlation matrix generated by the WGCNA package into a gene co-expression network by raising the correlation value of gene expression profiles by power (β). Ultimately, the goal of this step is to emphasize strong correlation values between genes while reducing the influence of weakly correlated genes. As part of this step, we found that maintaining the top 18,000 most variable genes provided the best balance of reducing computational time, retention of candidate regulator genes (280 genes kept), and decreasing noise. To choose our soft-thresholding power, we used this filtered dataset analyze the scale-topology fit (R^2) and mean connectivity values. We chose a power of 12, as this

maximized the R^2 correlation values (figure 2.3) for our data while maintaining sufficiently high mean network connectivity (figure 2.4).

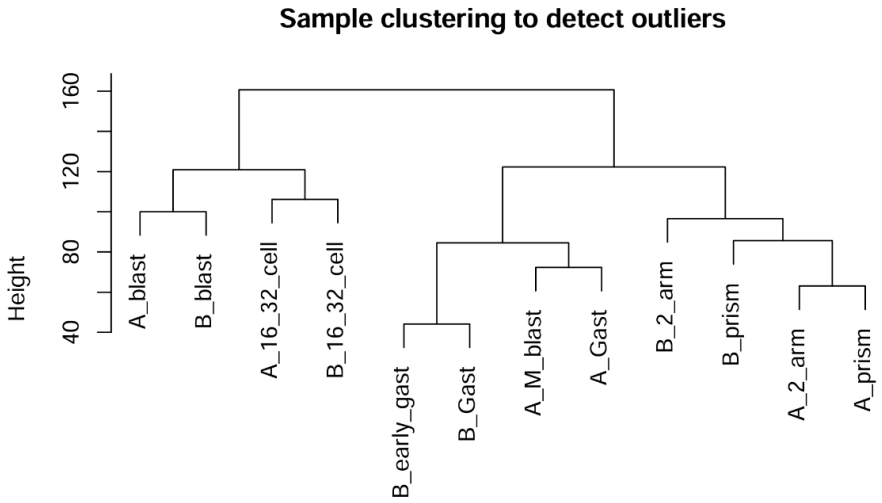

Figure 2.3: Hierarchical clustering of larval RNA samples by relative similarity of transcript expression values.

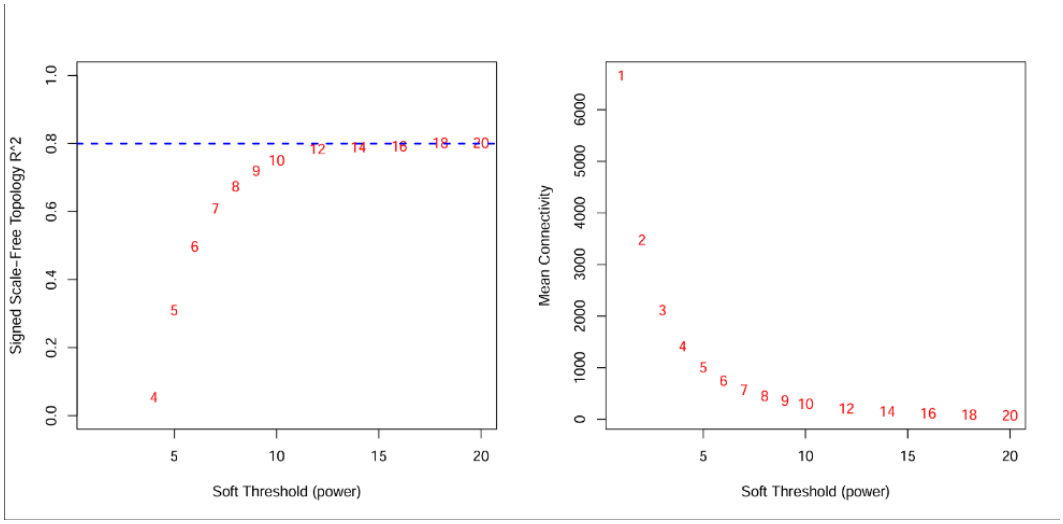

Figure 2.4: Soft-thresholding power analysis to identify appropriate power value among candidate values to maintain high R^2 (cut off R^2=0.80, indicated by blue dotted line) and mean connectivity value.

Results

Raw Assembly

Sequenced RNA isolated from two biological replicates of six developmental stages and the gut and coelomocyte tissue from dam 1 and sire 2 (16 samples in total) were used to perform a *de novo* transcriptome assembly of *E. tribuloides* (summary of statistics in table 2.1). Of the 1.2 billion raw reads, 1.185 (98.6%) were kept after trimming and quality filtering by fastp, indicating that most reads were high quality. In total, 2.3 million transcripts produced across approximately 1.35 Gb of assembled bases. In total, this assembly had about 1.27 million contigs. Approximately 30 – 60 million paired reads were produced per sample. The mean transcript length was 572.99 base pairs (bp) and the median transcript length was 367 bp. This assembly had a nearly complete BUSCO score (99.2% complete) with a high percentage of duplicated orthologues (92.6%). We observed a normal distribution of transcript lengths (figure 2.5). The assembly also had very high read-mapping rates, with an average of 91% of mapped reads across all 16 samples (figure 2.6). 35.86% of all transcripts were between 0 and 299 bp in length and about 55.20% of all transcripts were between 0 and 399 bp in length. The complete assembly had a contig N50 score

	Trinity.Assembly	Unigene.Assembly
Number of transcripts	2366929	1276924
Number of genes	1276924	1276924
Total assembled bases	1356221671	608026060
Mean transcript length	572.99	476.16
Median transcript length	367	322
Longest transcript	37004	37004
Contig N50	714	529
Transcripts >500 bp	78048	313471
Transcripts > 1000 bp	275907	92722
BUSCO Complete (%)	99.2	98.3
BUSCO Single-copy (%)	6.6	94
BUSCO Duplicated (%)	92.6	4.3
BUSCO Fragmented (%)	0.6	1.3
BUSCO Missing (%)	0.2	0.4

Table 2.1: Summary of assembly statistics for the raw transcriptome assembly and the unigene filtered assembly.

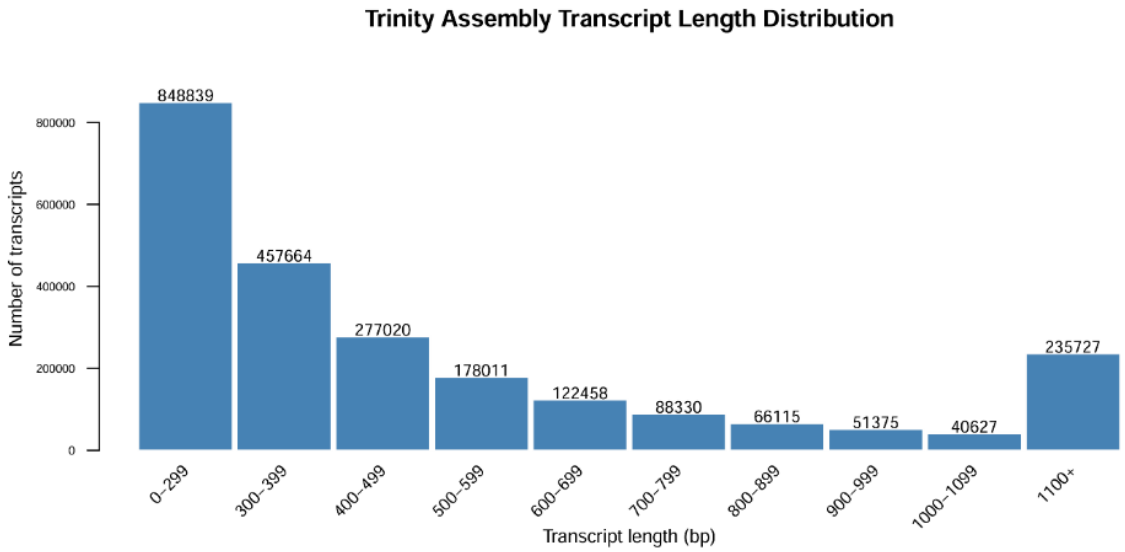

Figure 2.5: Distribution of transcripts per base pair length category of raw transcript assembly. Transcript length values generated by Trinity.

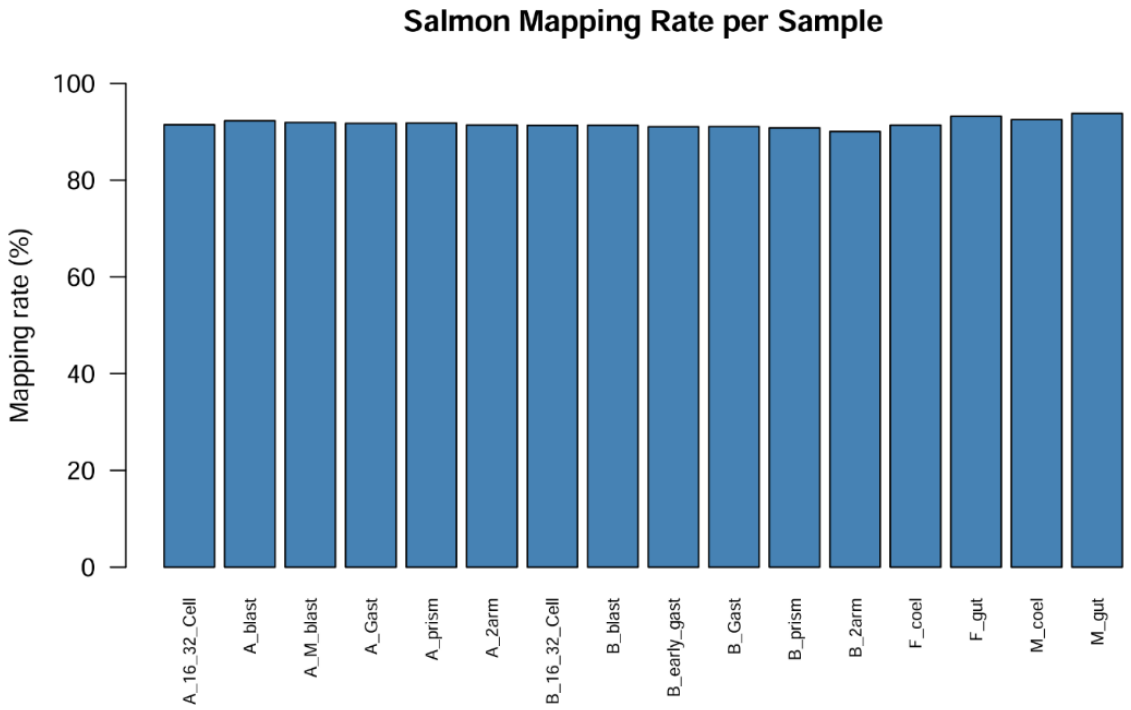

Figure 2.6: Salmon generated mapping rates for all 16 *E. tribuloides* larval and adult RNA samples.

of 714. To further examine assembly quality, we plotted the ExN50 values, which are generated by Trinity, for increasingly larger fractions of the total normalized expression of the transcriptome. The ExN50 plot reached its highest values at lower expression percentiles (figure 2.7). This indicates that the top 30-40% of expressed transcripts are found on contigs 715-800+ bp long. The top 50%-80% of expressed transcripts are found on contigs between 700 to 715 bp long. Overall, the plotted ExN50 values indicate that the majority of highly expressed transcripts were assembled onto contigs of relatively consistent length, indicating a generally consistent transcriptome assembly.

After filtering the assembly for unigenes, 1.27 million transcripts were kept, with a mean transcript length of 476.16 bp, and median transcript length of 322, and an N50 value of 529. The BUSCO completeness score for this unigene assembly decreased slightly (98.3%), which was

expected and showed that this filtered dataset was still very nearly complete. Importantly, the percentage of single copy BUSCO genes increased dramatically to 94%, and the percentage of duplicated BUSCO genes decreased to only 0.4%.

Homology Annotation

TransDecoder.LongOrf was used to predict 151,915 open reading frames from the 1.27 million transcripts found in the unigene assembly. These open reading frames were aligned to

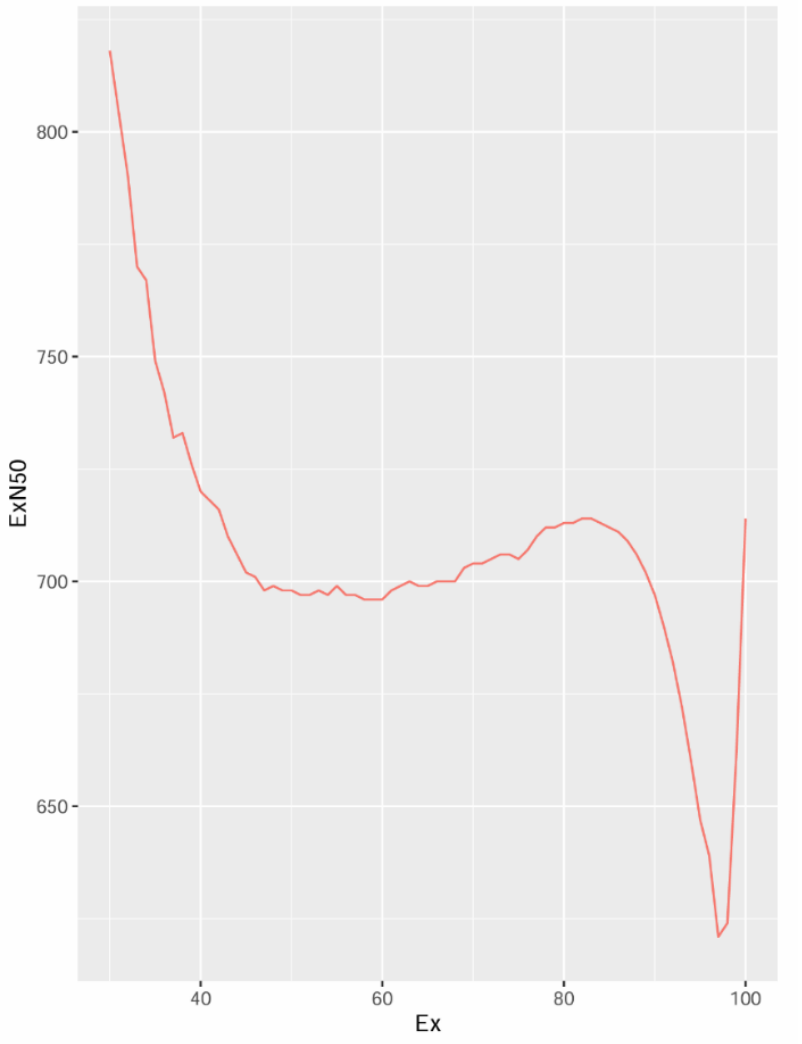

Figure 2.7: Plotted ExN50 values for raw transcript assembly. Top 50% - 80% expressed transcripts are between 700 – 715 bp long.

known protein sequences from the SwissProt database using the blastp function of DIAMOND. TransDecoder.Predict was then ran on these open reading frames, and using homology support, we generated a final protein prediction catalogue of 28,331 sequences. Given that other sea urchins generally have between 25,000 – 30,000 protein coding genes, we found our final prediction set to be a reasonable length. BUSCO returned a completeness score of 95% on this set, which indicates that the majority of what we removed were spurious and redundant sequences and that the vast majority of biologically relevant coding sequences were maintained. 28,331 homology supported sequences were analyzed by InterProScan for identification of protein domains, functional motifs, and GO terms. 27,613 (97.5%) were able to generate at least one functional annotation. To further investigate protein homology and conservation, a reciprocal best hit analysis was conducted using a custom protein database generated from the 5.0 *S. purpuratus* genome assembly. We found 9,908 (32.1%) of our sequences to be shared between *E. tribuloides* and *S. purpuratus*, suggesting an orthologous relationship between these genes.

As transcription factors, signaling molecules, and receptors are the key functional components of the echinoid regulatory network, we specifically sought out to curate a list of proteins known to have some kind of developmental function. In total, 374 proteins that have been identified in *S. purpuratus* were used to search for candidate regulatory proteins in *E. tribuloides*. Using the homology annotations, functional annotations, and reciprocal best hit results, 304 candidates were identified. For the remaining 70 proteins, a manual recovery had to be implemented in which we aligned our predicted open reading frames against the 70 *S. purpuratus* protein sequences. With this homology support, we then ran TransDecoder.Predict to get the predicted protein sequences. SwissProt is a vertebrate-dominated protein database, so it is possible that these missing proteins lack vertebrate homologues or that the vertebrate

homologues have diverged enough that the alignment did not meet out minimum threshold. Regardless, from this manual recovery process, 67 sequences were successfully annotated. Three genes (Sp-paxA, Mta, and SP-NFX) had no alignment support whatsoever. While this is insufficient evidence to declaratively say that *E. tribuloides* lacks these genes, it is a possibility that should be explored in greater effort in the future.

Candidate of pmar1

Interestingly, we were able to identify a candidate gene of Pmar1 during the manual recovery process. When performing a reciprocal blastp function, our candidate gene, TRINITY_DN120398_c0_g1 (233 amino acid residues), aligns to the *S. purpuratus* pmar1 gene NP_999673 (Oliveri et al., 2002) with a alignment score of 80.1, an e-value of 4e-18, and was 44.36% identical. When doing a protein alignment against NCBI's Echinodermata database, the top alignment was the pmar1 related protein BBK09440 (Yamazaki et al., 2020) found in the cidaroid urchin *Prionocidaris baculosa*. This had an alignment score of 391, and e-value of 3e-139, and was 85.41% identical. Our candidate pmar1 sequence was also predicted to have key functional domains associated with the canonical pmar1 gene, including homeobox domain signatures, and was assigned GO terms associated with transcriptional regulation by RNA polymerase II.

A maximum likelihood tree was produced by a MEGA(Kumar et al., 2024) sequence alignment and tree building software using our candidate pmar1 predicted amino acid sequence, known pmar1 sequences from other echinoderms, and other paired homeobox box class transcription factors. The bootstrap consensus tree was made using 500 replicates using the

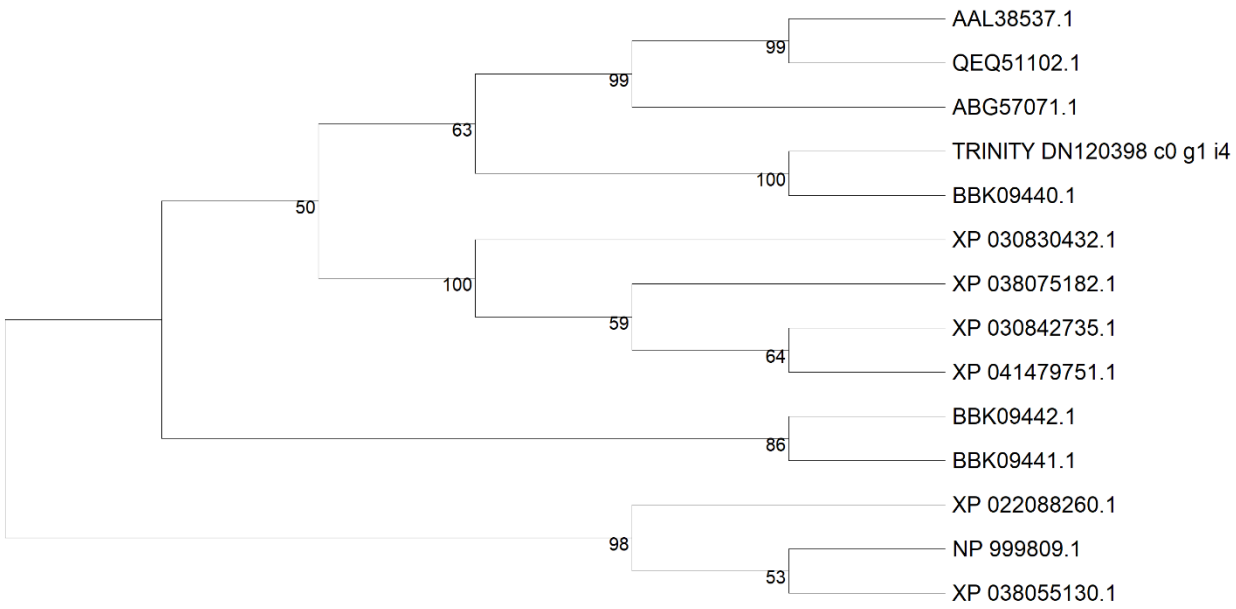

Figure 2.8: Maximum likelihood tree of echinoid Pmar1 and paired homeobox facotrs using WAG model. Alx1 sequences (NP 999809.1 and XP 038055130.1) are outliers. Tree was generated using 500 bootstraps. A support value of 63 was assigned for cidaroid pmar1 being the sister group to echinoid pmar1. A support value of 100 was assigned to the node in which Et-pmar1 and *P. baculosa* pmar1 are sister groups.

Whelan and Goldman (WAG) model. The support value, the percentage of trees in which those taxa are clustered together, is found beneath each node. Sequence positions in which there was less than 95% coverage were deleted from this analysis.

Our ML tree indicates that all tree replicates were in agreement that our candidate pmar1 sequence in *E. tribuloides* clusters with *P. baculosa* pmar1 (support value = 100). The majority of replicates were in agreement that this cidaroid gene pmar1 lineage is sister to the known euechinoid pmar1 sequences (support value = 63). Half of tree replicates placed this entire pmar1 group as the sister taxon to the Phb sequence cluster.

The expression of TRINITY_DN120398_c0_g1 was averaged between our stage replicates and plotted as transcripts per million (TPM) (see figure 2.9). While the overall expression pattern matches that of *S. purpuratus* pmar1 in that the highest peak is during the

cleavage stage of the early embryo and expression thereafter dramatically decreases, there are some notable differences. There appears to be a slight increase in expression during the mesenchyme blastula/early gastrula stage of the embryo, whereas in *S. purpuratus*, expression plateaus between the blastula and mesenchyme blastula stages before decreasing to 0 TPM for the remainder of the animal's development. While TRINITY_DN120398_c0_g1 is extremely low after the mesenchyme blastula stage in *E. tribuloides*, it does not stop being entirely expressed and is even present in low quantities during after the larval arms have formed.

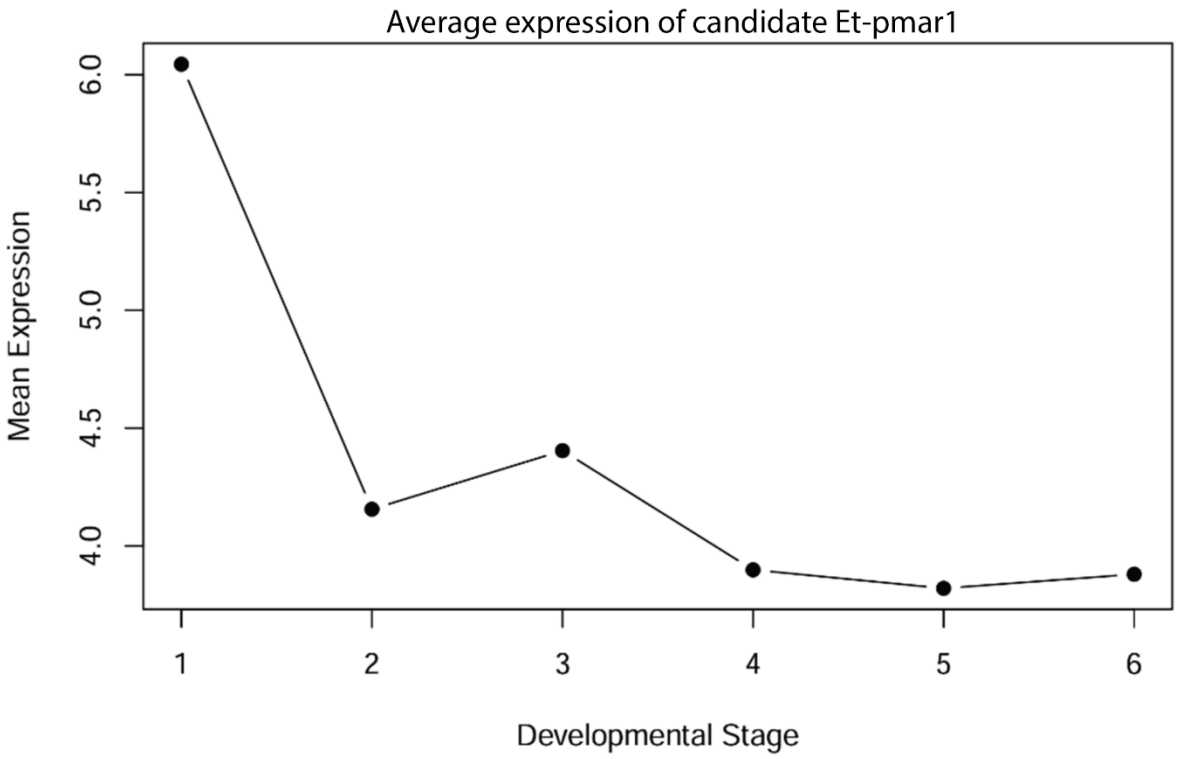

Figure 2.9: Averaged expression of potential Et pmar1 transcript TRINITY_DN120398_c0_g1 averaged across development stage. Mean expression is in transcripts per million (TPM). Stages 1 through six correspond with developmental stages 16/32 cleavage stage through two-arm stage.

Functional Annotation

Using InterProScan, we classified three ontologies of GO terms: biological process (BP), cellular component (CC), and molecular function (MF). Under biological processes, 564 genes were assigned GO terms associated with DNA transcription regulation. Under cell components, 2,480 genes were identified as being involved or physically locating to the cell nucleus. Of the molecular function GO terms, 397 genes were identified as having DNA transcription factor binding activity. Our GO terms demonstrate that our annotated unigene assembly contains many sequences that function in developmental regulation through transcriptional control.

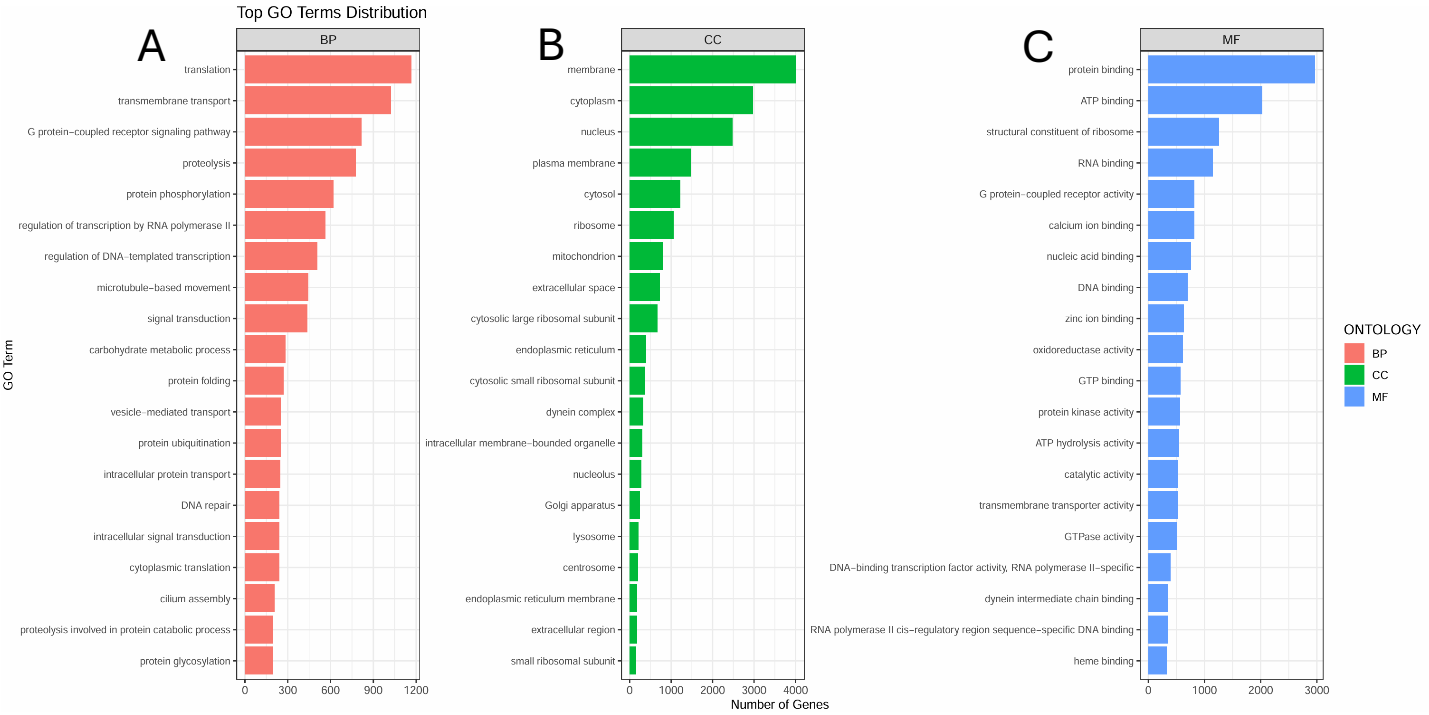

Figure 2.10: Distribution of Gene Ontology terms for annotated *E. tribuloides* predicted protein sequences from unigene filtered transcript assembly. **(A)** Distribution of biological process terms (red). **(B)** Distribution of cellular component terms (green). **(C)** Distribution of molecular function terms (blue).

Expression variation driven by developmental progression

Transcript reads were quantified using Salmon and imported in R using tximport. And normalized using DESeq2(Love et al., 2014). A principals component analysis was performed using our entire dataset for all 16 samples (figure 2.11). The first principal component captured 56% of the variance among samples and the second principal component captured 21% of variance. The samples demonstrate tight clustering between replicates. Adult tissues group together distinctly from developmental stage samples, suggesting large global differences in gene expression between these two groups.

To investigate how samples correlated in terms of expression variance, we performed a sample-to-sample Pearson correlation using variance stabilized expression data for all development samples (adult tissues were excluded) (figure 2.12). Overall, this heatmap revealed high similarity between samples, with correlation coefficients ranging from 0.75-1.00. Biological replicates are highly correlated with each other (0.94-0.97). Developmentally sequential stages shared higher coefficients overall (0.90-0.94) with each other than with distant stages (0.83-0.75). These data suggest that while these samples share core transcriptomic expression patterns, there is a gradual decline in correlation coefficient values as development progresses.

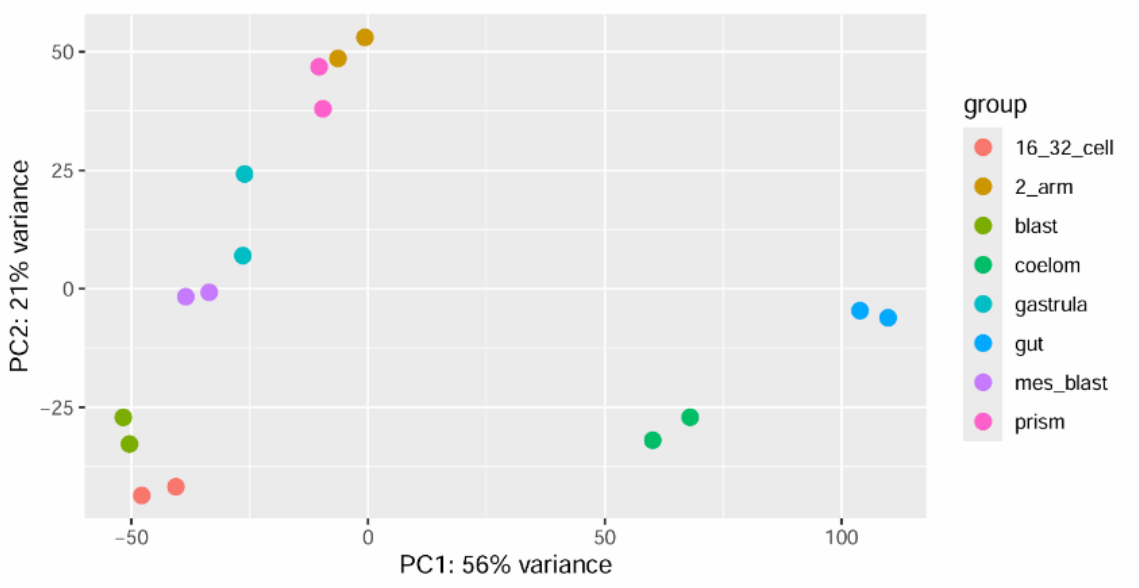

Figure 2.11: Principal component analysis (PCA) of normalized gene expression data across developmental stages and adult tissues. Samples cluster by stage/tissue type, explaining 56% (PC1) and 21% (PC2) of the total variation observed.

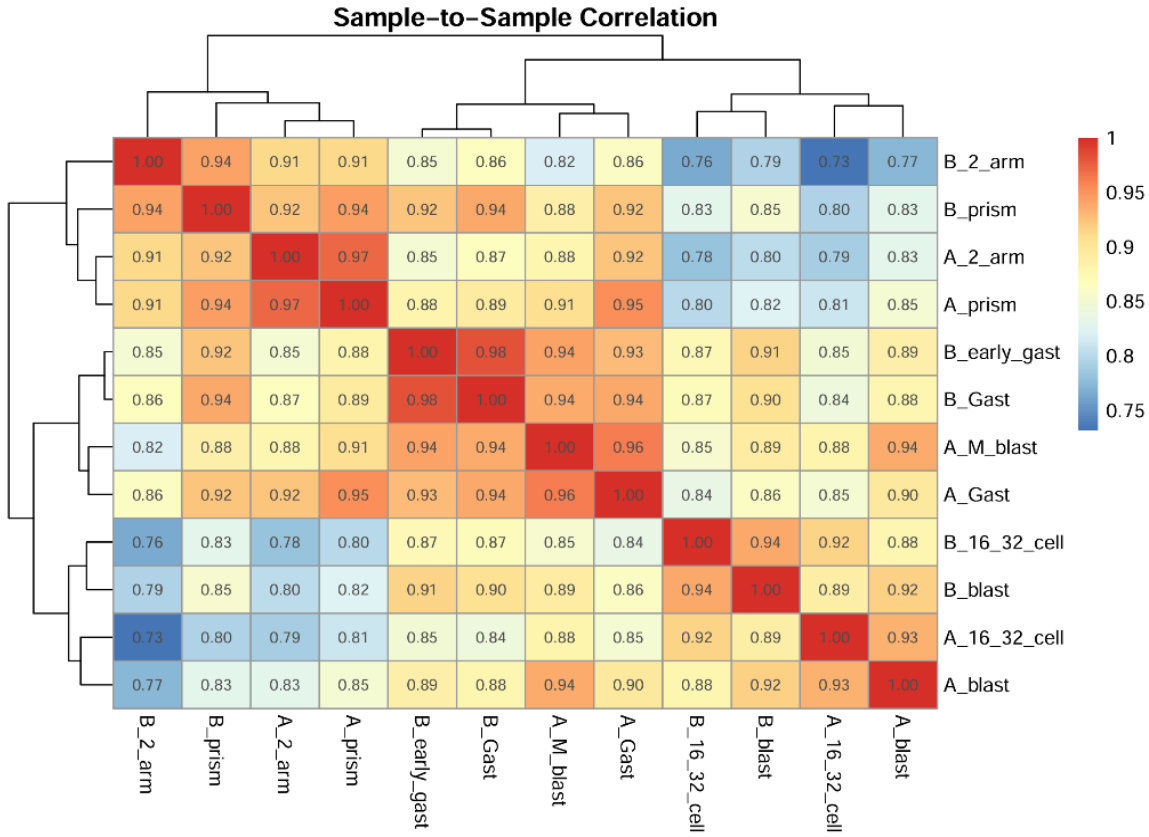

Figure 2.12: Sample to sample Pearson correlation. Biological replicates have high correlation values (0.94-0.97). Sequential developmental stages have higher correlation values (0.90-0.94) than distant stages (0.83-0.75)

Taken together, it is likely that the changes in expression patterns between embryo samples are being driven by stage specific developmental changes.

Heterochronic Shifts in the Skeleton Development Program

Reads were generated using Salmon on the entire transcript dataset. These counts were imported into R for normalization using DESeq2 and variance stabilization using the vst command. This data was filtered to maintain only genes with homology and functional support. We then removed transcripts that had 0 counts across all samples and removed the bottom 20% least variant transcripts. After these filtering steps, 326 of our candidate regulatory genes were maintained. Next, a heatmap was generated focusing on the candidate genes involved in the development of primary mesenchyme cells and the larval skeleton (figure 2.13). It was observed that there is a nearly whole temporal shift of the expression of the skeleton program. This observation may be what is causing *E. tribuloides* to develop its skeleton much later compared to euechinoids. Most of these genes became upregulated later in the larva's development (gastrula or prism stage) compared to *S. purpuratus*, in which these genes generally reach peak expression in during the blastula or mesenchyme blastula stage. This is consistent with our observations of when the *E. tribuloides* larval skeleton begins to grow, which was during late gastrula. By graphing the average expression of all candidate skeletogenic genes, it can be observed that there is a subtle, but defined peak of expression values (TPM) during the prism and 2 arm stage, which is much later than the peak of these genes in *S. purpuratus* (figure 2.14). A summary of the candidate genes and their plotted expression values can be found in table 2.2.

Non-Skeletal Mesenchyme Program

To gain a broader picture of the development of the *E. tribuloides* mesodermal tissues, we next investigated the non-skeletal mesenchyme (NSM) program (figure 2.16), particularly those genes

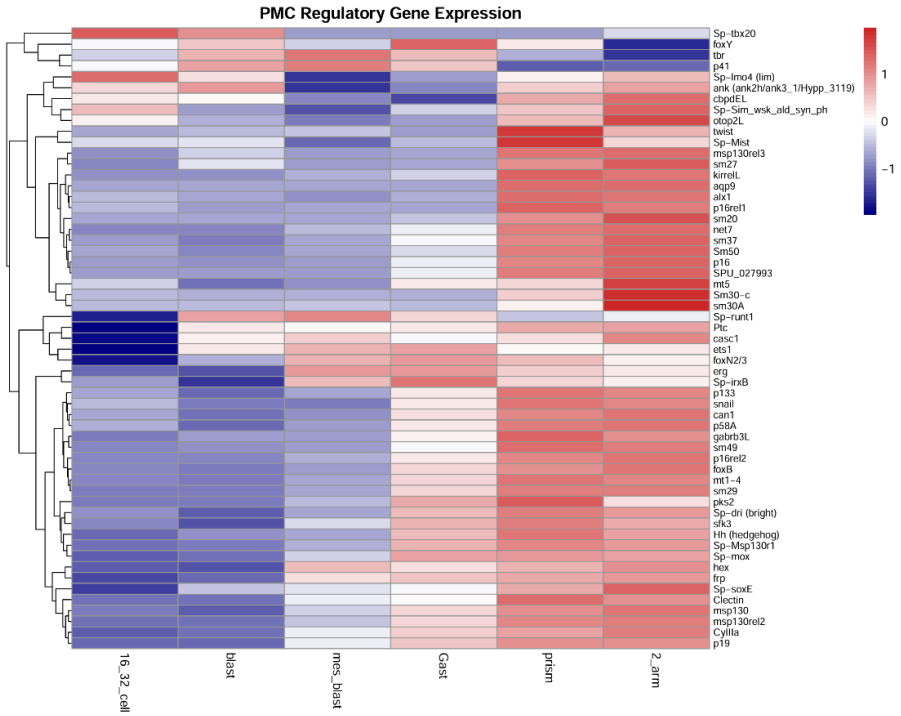

Figure 2.13: Heatmap of averaged, normalized expression levels for candidate genes involved in skeleton development of *E. tribuloides*. Red indicates higher expression and blue indicates lower expression. Rows represent genes and columns represent developmental stage.

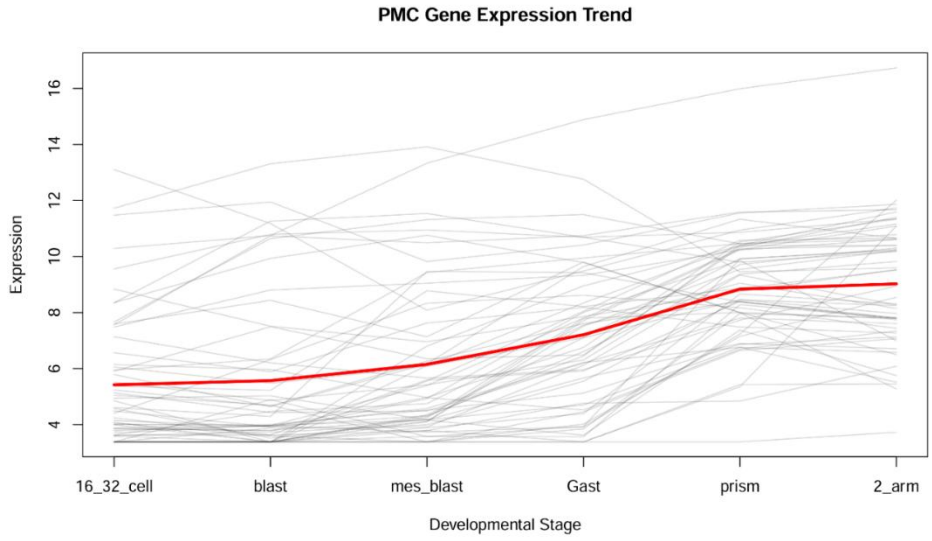

Figure 2.14 Graphed average expression trends for candidate skeletogenic genes across developmental stage of *E. tribuloides*. Red line represents the averaged expression value of all genes at that stage. Expression in transcripts per million (TPM)

gene_name	trinity_id	16_32_cell	blast	mes_blast	Gast	prism	2_arm
CyIIIa	TRINITY_DN94_c0_g1	10.291549	10.749678	13.334099	14.888022	15.991230	16.734928
alx1	TRINITY_DN61585_c0_g1	4.058568	3.596904	3.385305	3.846770	7.964334	7.804523
ets1	TRINITY_DN3054_c0_g1	7.664416	10.643650	11.319200	11.499619	10.456357	10.608936
tbr	TRINITY_DN14796_c1_g1	8.355849	9.935390	10.760906	9.788657	8.008400	6.498286
foxB	TRINITY_DN13260_c0_g1	3.585801	3.385305	4.109497	7.516784	9.541221	10.214813
hex	TRINITY_DN11422_c0_g1	5.377653	5.225123	8.778449	8.214253	8.866023	9.518176
Sm50	TRINITY_DN8666_c1_g2	5.510753	4.860984	5.423691	6.308034	9.717156	10.311327
erg	TRINITY_DN1438_c1_g1	4.612610	4.294256	9.472169	9.425495	8.378705	7.793827
twist	TRINITY_DN97967_c0_g1	3.783685	3.973840	4.166251	3.623219	7.378850	5.754123
snail	TRINITY_DN16040_c0_g1	5.137072	4.435771	4.488995	6.195510	7.805415	7.606010
foxN2/3	TRINITY_DN11915_c0_g1	4.410418	6.322221	8.337295	8.615360	8.174653	7.437744
Sp-Mist	TRINITY_DN88041_c0_g1	4.945687	5.027369	4.175165	4.701561	6.763357	5.516570
Sp-Sim_wsk_ald_syn_ph	TRINITY_DN31403_c0_g1	8.840160	7.514062	6.957597	7.864264	8.749419	9.534728
Sp-soxE	TRINITY_DN13618_c0_g1	7.486877	8.805967	9.050158	9.330404	10.244532	11.087456
Sp-tbx20	TRINITY_DN132164_c0_g1	5.051119	4.700263	3.385305	3.385305	3.385305	3.728765
Sp-dri (bright)	TRINITY_DN4262_c0_g1	5.789860	4.640812	6.170343	9.644508	11.333459	10.651459
Sp-lmo4 (lim)	TRINITY_DN1555_c2_g1	13.104839	11.179603	8.083558	9.484765	10.943115	11.735603
Sp-runt1	TRINITY_DN9728_c0_g1	8.350130	11.264262	11.537157	10.689296	9.796479	10.179828
Sp-Msp130r1	TRINITY_DN5671_c3_g1	3.880083	3.968189	4.580108	6.220865	6.758280	6.579514
Sm30-c	TRINITY_DN3381_c2_g1	4.012032	3.958833	3.737240	3.923429	7.236453	12.011655
foxY	TRINITY_DN13281_c0_g1	7.588453	8.438637	7.137351	9.805704	7.995028	5.276330
Ptc	TRINITY_DN3723_c6_g2	7.563901	10.738216	10.492164	10.748368	11.587339	11.672890
Hh (hedgehog)	TRINITY_DN38226_c0_g1	3.628669	4.476802	4.748814	7.555392	9.049832	8.153050
Sp-irxB	TRINITY_DN3721_c2_g1	6.567322	5.887764	7.627552	8.190487	7.484300	7.272808
Sp-mox	TRINITY_DN28193_c0_g1	3.706804	3.967528	5.397256	7.804633	8.005382	7.792912
ank (ank2h/ank3_1/Hypp_3119)	TRINITY_DN6_c1_g1	11.477804	11.949706	9.823783	10.411103	11.556849	11.862588
aqp9	TRINITY_DN3823_c0_g2	3.385305	3.385305	3.385305	3.385305	5.430122	5.454201
can1	TRINITY_DN5361_c2_g1	4.579715	3.518566	4.116584	7.711033	10.235687	10.616484
casc1	TRINITY_DN2722_c0_g1	9.560170	10.792830	10.951289	10.688993	10.864855	11.345646
cbpdEL	TRINITY_DN39114_c0_g1	6.145848	5.965998	4.964496	4.523928	6.696077	7.346399
Clectin	TRINITY_DN29045_c0_g1	3.620839	3.635633	5.595547	5.917522	8.399640	7.817011
frp	TRINITY_DN1142_c1_g1	5.946998	6.348631	9.431870	9.948529	10.464871	10.787208
gabrb3L	TRINITY_DN5713_c3_g1	3.385305	3.874839	3.916278	6.120879	9.339533	8.278607
kirrelL	TRINITY_DN39044_c1_g1	3.385305	3.385305	3.894578	4.392316	8.666283	8.308529
msp130	TRINITY_DN19404_c0_g1	4.079831	3.385305	6.116046	8.343956	10.383768	11.159410
msp130rel2	TRINITY_DN7069_c0_g1	3.385305	3.385305	5.561209	8.045932	10.583842	10.659559
msp130rel3	TRINITY_DN6131_c0_g1	3.385305	4.893228	3.575090	3.935688	9.913047	10.257894
mt1-4	TRINITY_DN1260_c4_g1	3.841443	3.799805	4.168458	5.668568	6.895250	6.712022
mt5	TRINITY_DN3485_c15_g1	4.241516	3.626942	3.789060	4.766402	4.842252	6.083950
net7	TRINITY_DN8537_c1_g1	4.001478	3.999766	4.518373	5.125887	6.869663	7.128711
otop2L	TRINITY_DN14633_c0_g1	7.135076	6.207836	5.660863	5.962320	7.730045	8.944944
p133	TRINITY_DN92988_c0_g2	6.055565	5.475861	6.085492	7.100915	8.456145	8.251666
p16	TRINITY_DN7681_c0_g1	4.147392	3.932212	4.320686	6.450559	10.405284	11.593477
p16rel1	TRINITY_DN19141_c0_g1	3.996040	3.385305	3.586466	3.569298	8.283535	7.803379
p16rel2	TRINITY_DN1158_c1_g1	3.385305	3.385305	4.295786	6.609830	9.370101	9.826185
p19	TRINITY_DN19059_c0_g1	3.385305	3.385305	6.803012	8.952238	10.386960	10.399688
p41	TRINITY_DN4917_c0_g1	11.726450	13.313097	13.917492	12.763180	9.454730	9.597547
p58A	TRINITY_DN33324_c0_g1	4.851340	3.385305	4.343953	7.105357	9.925315	10.219844
pks2	TRINITY_DN72695_c0_g1	3.868700	3.779822	4.928206	8.008474	9.867326	7.000021
sfk3	TRINITY_DN23935_c0_g1	5.239179	4.667368	6.268206	7.607885	8.397643	7.738081
sm20	TRINITY_DN18935_c0_g2	3.385305	3.385305	3.385305	4.015204	7.152135	8.548200
sm27	TRINITY_DN22302_c0_g1	5.909168	7.495207	6.352552	6.487176	10.331301	11.398864
sm29	TRINITY_DN1291_c0_g1	3.385305	3.385305	4.244647	7.784009	10.275531	10.393111
sm30A	TRINITY_DN14553_c0_g1	3.385305	3.385305	3.791948	3.385305	5.354046	11.086828
sm37	TRINITY_DN5918_c0_g1	4.489542	3.920489	4.974561	6.756873	10.525720	11.336563
sm49	TRINITY_DN6583_c0_g1	3.614223	3.729230	3.993574	5.550232	8.456351	7.966756
SPU_027993	TRINITY_DN121972_c0_g1	3.385305	3.385305	3.385305	4.456095	6.649738	7.070462

Table 2.2: Mean expression values in transcripts per million of skeletal development genes. Expression values have been averaged for A and B replicates. Trinity ID represents that longest isoform of that gene cluster.

involved the development of pigment cells, which are part of the larval immune system. Similar to how the skeletal regulatory genes were treated, we averaged the normalized/stabilized expression values for A and B embryonic samples. Unlike the skeletogenic battery, the genes involved in NSM do not show strong whole module shifts in expression patterns. Of these candidate genes, Pks-1, the sult genes and irf4 had the most divergent expressions when compared to these gene orthologs in *S. purpuratus*. We found that these genes had delayed upregulation, which is in good agreement with previous studies. It was also observed that Pks-1 was expressed at a relatively higher level between the gastrula and two-arm stage, where in *S. purpuratus,* this signal decreases over time. Gcm and GataC show very similar expression patterns compared to *S. purpuratus*.

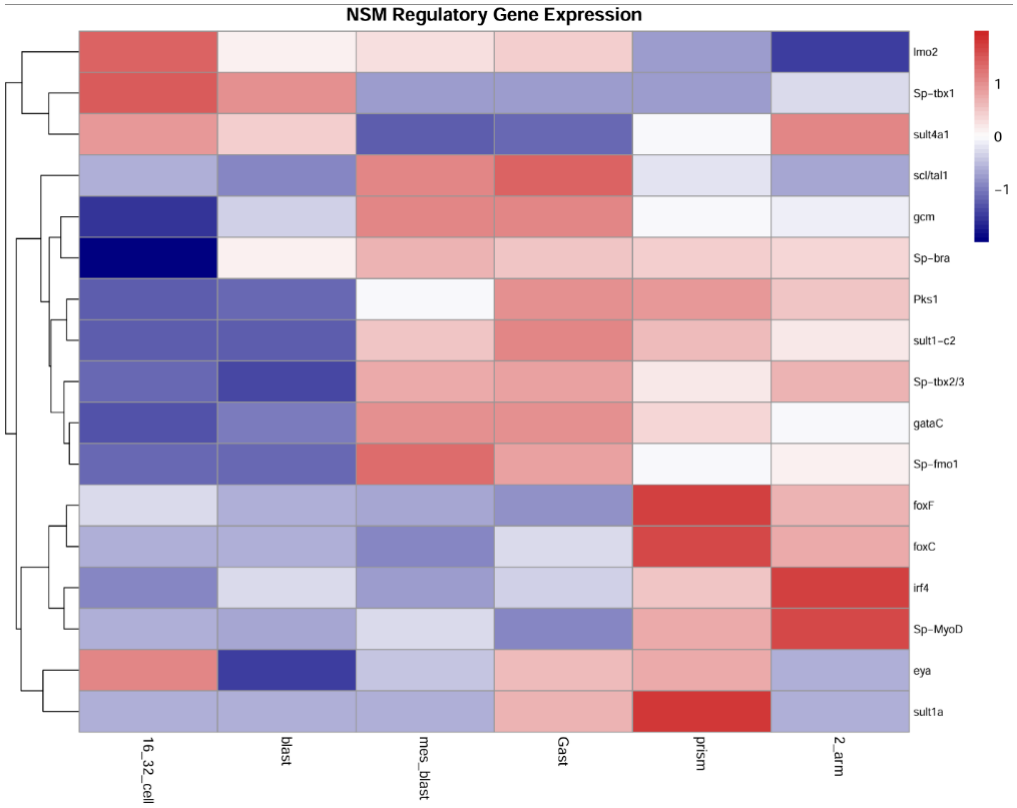

Figure 2.15: Sample heatmap of candidate genes involved in non-skeletal development.

WGCNA

A weighted gene co-expression network analysis (WGCNA) was performed using Salmon generated, gene level counts for the 27,613 predicted sequences that had both homology and functional support. The correlation network was designed using the top 18,000 most variable genes and a power of 12. Of this filtered dataset, 280 of our candidate regulatory sequences were kept. This resulted in 28 modules in total, 19 of which had at least one candidate gene. These 19 modules were coded as black (8 candidate genes), blue (87 candidate genes), brown (58 candidate genes), cyan (5 candidate genes), dark turquoise (4 candidate genes), green (6 candidate genes), grey60 (3 candidate genes), light cyan (4 candidate genes), light green (10 candidate genes), midnight blue (2 candidate genes), orange (1 candidate gene), pink (1 candidate genes), purple (17 candidate genes), red (37 candidate genes), royal blue (1 candidate gene), salmon (2 candidate genes), tan (8 candidate genes), turquoise (25 candidate genes), and yellow (5 candidate genes). The hierarchical clustering of these correlation modules produced a gene dendrogram (figures 2.16 – 2.18), labeled by module assignment colour. It should be observed that very few genes were placed into the unassigned or "grey" module, indicating that most of the genes in our input dataset have strong, variable patterns of expression over developmental time.

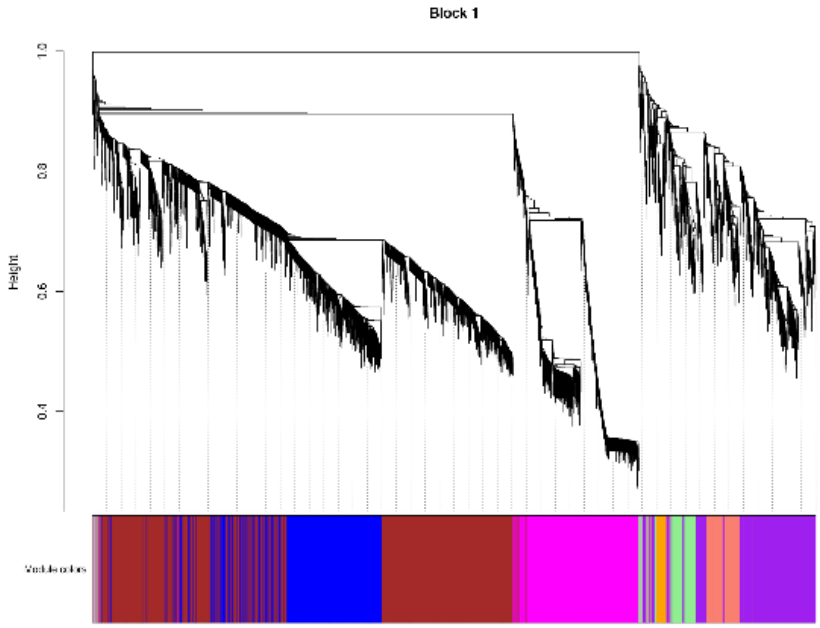

Figure 2.16: WGCNA dendrogram 1. Major modules are red, navy and magenta.

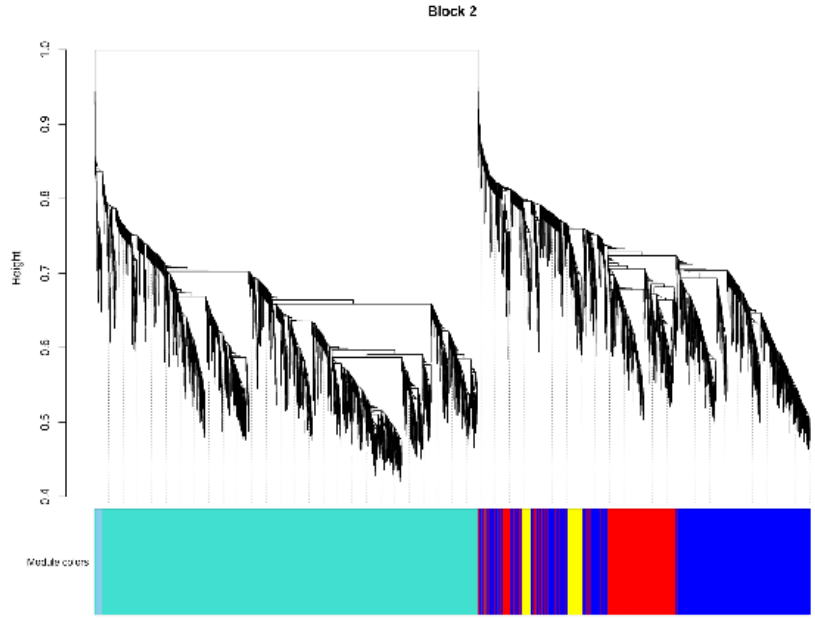

Figure 2.17: WGCNA dendrogram 2. Most major modules are cyan and royal blue

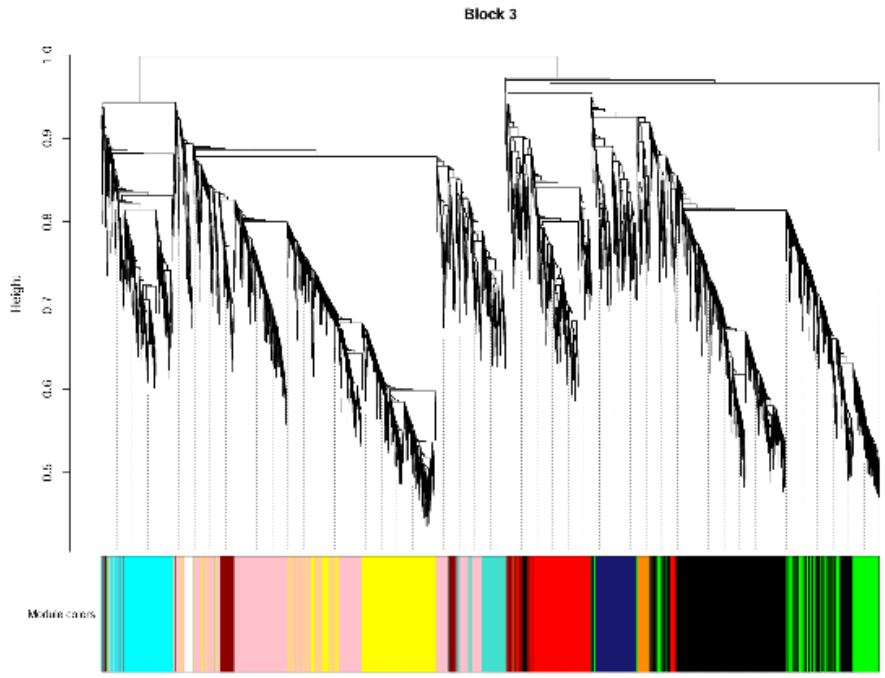

Figure 2.18: WGCNA dendrogram 3. Major modules are salmon and black.

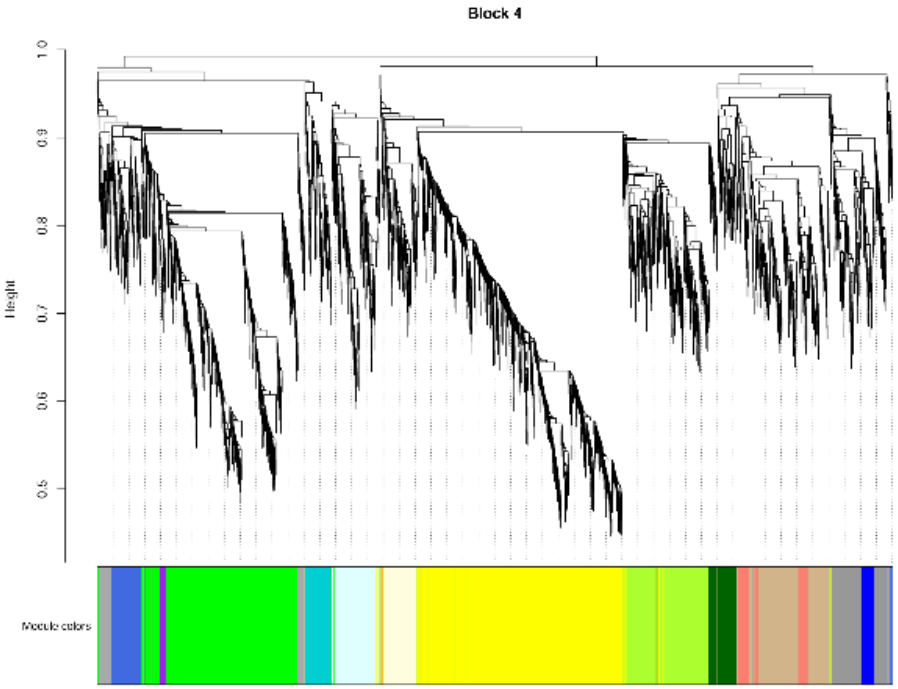

Figure 2.19: WGCNA dendrogram 4. Major modules are yellow and green.

We found that the majority of the potential mesoderm developmental genes were clustered into module blue, module brown, and module red (see figure 2.19). Interestingly, while genes involved in primary mesenchyme cell development and secondary mesenchyme cell development grouped into module blue and module red respectively, there were some genes that were mixed in between the two. This may indicate that the general mesoderm network has broadly similar expression patterns and that the developmental domains of this tissue are less well defined as they are in euechinoids.

The correlation patterns of module blue (correlation value = 0.47; 0.48) and module brown (correlation value = 0.47; 0.48) support our observations that the genes involved in skeletogenesis have a coordinated movement in the same direction (upregulation) in the latter half of the larvae's development and further substantiates that this entire subcircuit has undergone a heterochronistic shift in the echinoid evolution. Module red demonstrates that some skeletal mesenchyme and non-skeletal mesenchyme genes become mildly correlated in the same direction (upregulation) between the mesenchyme blastula to prism stages. Of these modules, only module brown is significantly associated with a particular stage sample (stage 2 arm; p-value = 0.005). None of our mesenchyme gene containing modules are significantly associated with gastrulation or prism stage samples. This is not entirely surprising, as each of these modules contain thousands of other genes that have functions outside of developmental regulation whose expression patterns may be obscuring any high association between our mesodermal genes and a particular stage.

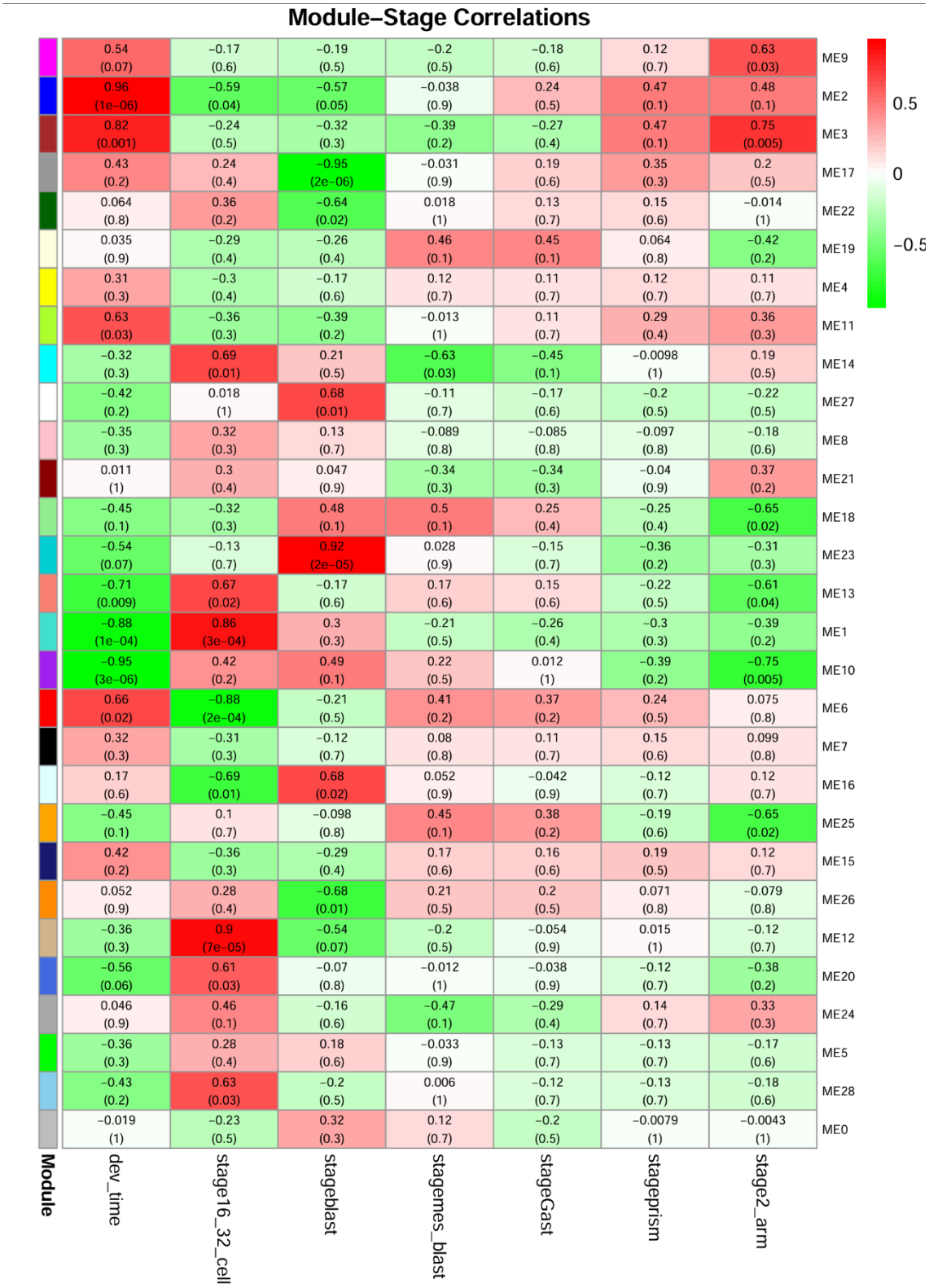

Figure 2.20: WGCNA module stage correlations. Top value per box is the correlation of modules with each other at that stage. Bottom value per box is the correlation of that module with that stage. Dev_time indicates the overall trend for that module.

Discussion

Developmental Transcriptome Assembly

In this study, the first *E. tribuloides* developmental transcriptome was generated by *de novo* assembly using samples representing six major stages during the early life of the cidaroid larvae: 16/32 cleavage stage, blastula stage, mesenchyme and early gastrula stage, gastrula stage, prism stage, and finally the two arm stage. In addition, RNA was collected and sequenced using the coelomocytes and gut tissue of the parents of the cultures. This transcriptome assembly proved to be both extremely complete (BUSCO completeness score of 95% for final protein coding sequence set), exhibits high quality bases sequences (generally only 1% of bases trimmed for quality control) and reads mapped back to the assembly at an extremely high rate (average mapping rate = 91%). Amino acid sequences were predicted for each of the longest isoform per gene cluster, of which 27,613 had produced both homology and functional annotations.

This annotated transcriptome is the first of its kind for *E. tribuloides*. One of the most essential tools for untangling the GRN of a given species is genomic resources. Previously, only a first version sequenced genome and RNA seq data for the gastrula tissue and adult spine tissue were available. As of this writing, the genome consists of 637,071 scaffolds and over a million contigs. Over the course of this research, this genome proved to be extremely difficult to work with, primarily due to the highly fragmented state it exists in. Annotation predicting software, such as AUGUSTUS, struggled to predict sequences that were split between multiple contigs and required enormous computational time and resources. The existing RNA sequencing data is limited in scope and misses many stage specific gene expressions. We believe that our annotated, assembled transcriptome will be a useful dataset moving forward and will greatly facilitate any

future endeavors in elucidating the structure of the *E. tribuloides* developmental GRN in several ways.

Developmental transcriptomes are essential to the understanding of GRNs in that they provide a catalogue of genes, both with and without regulatory functions, specifically being expressed at a given life stage. This transcriptome currently provides the highest resolution developmental gene expression data for cidaroids. While previous work done by Erkenbrack, Wray identified candidate regulatory genes in *E. tribuloides* and other cidaroid urchins, their studies were, by necessity, limited in scope. We now have an atlas of 377 potential regulatory genes which, once experimentally validated, can be used to investigate the broader network structure in this system.

Current echinoid GRN research is focused primarily in euechinoid sea urchins, such as *S. purpuratus* and *L. variagatus.* While these systems have undoubtedly been instrumental to the field, this bias means that the modern understanding of how regulation has evolved within sea urchin is incomplete. As an early diverging, monophyletic sister group, Cidaroidea is uniquely positioned to provide insights into how GRNs may evolve over time and the functional consequences of those accumulated alterations. This cidaroid transcriptome thus serves to fill in a significant phylogenetic gap in echinoderm genetics and allows for direct developmental comparisons to be made between these echinoid lineages, which could provide novel or interesting insights into the mechanisms by which GRNs evolve.

Cidaroid Pmar1

The prevailing hypothesis concerning the evolution of skeletogenesis in Echinoidea was driven by the emergence of the euechinoid Pmar1 repressor and the Pmar1/double negative gate. This conclusion was drawn from the inability to find a Pmar1 homologue in the previously

existing *E. tribuloides* transcript data or genome, experiments that demonstrated *E. tribuloides* embryos did not respond to microinjected *S. purpuratus* Pmar1 BACs, and by the observation that HesC appears to be expressed alongside genes involved in primary mesenchyme development (specifically delta and ets1). This was recently challenged by the discovery of a cidaroid Pmar1 homologue in *P. baculosa*. These experiments demonstrated that in this cidaroid system, Pmar1 acts as a repressor (though the target is unknown), HesC has some ability to repress the expression of alx1, and that cidaroid Pmar1 is able to repress *S. purpuratus* HesC when exogenously expressed in *S. purpuratus* embryos. Thus, it is suggested that HesC underwent some form of alteration that 1) placed it under the control of Pmar1 and 2) caused it to become a major repressor of the skeletogenesis gene battery. While this cidaroid Pmar1 was identified in the current *E. tribuloides* genome through BLASTP alignments, it was unable to be found in the current *E. tribuloides* transcript data.

In this study, we have found a strong candidate for Pb-Pmar1 in our developmental transcriptome. These findings are supported by nearly identical BLASTP alignments and phylogenetic clustering of the predicted protein sequences. Like with our other candidate gene sequences, this will need future experimental validation to confirm the exact action of this transcription factor. Future functional assays using *E. tribuloides* as a comparative cidaroid system to *P. bacuolosa* will be imperative in elucidating whether the existence of Pmar1 is a conserved ancestral condition or a lineage trait specific to the *Prionocidaris* genus. If it is found that *E. tribuloides* truly does lack Pmar1, this could point towards a trait reversal/deletion within the *Eucidaris* genus, or even an independent evolution event within *Prionocidaris*.

Heterochrony within the mesoderm regulatory modules

The work presented here not only supports most previous reports of shifts in timing of skeletal mesenchyme genes but has also expanded the extent to which this pattern may be observed. Previous work demonstrated that Alx1, a major driver of skeletal differentiation in euechinoids, begins its expression later and the expression peak itself is during the gastrula stage in *E. tribuloides*. This is in good agreement with our findings of our identified Alx1 transcript. In addition to the early drivers of skeletogenesis (Alx1, Ets1, Tbr), we investigated the expression patterns of nearly the entire downstream battery. It seems like almost the entire skeletal module of *E. tribuloides* has shifted its expression peak to much later in development (gastrula and prism stages) compared to euechinoids. The general morphology of the larval skeleton, particularly early during environment, is different compared to euechinoid larval skeletons, and this observation may be explained in part by these differences in development. Interestingly, the juvenile and adult tests of *E. tribuloides* have several marked differences when compared to *S. purpuratus*, particularly in regard to organization of the ambulacral/interambulacral plates and to the structures that support the Aristotle's lantern(Gao et al., 2015). It is possible that the alterations in micromere and primary mesenchyme cells specification and differentiation events during embryogenesis have had cascading effects on skeletal development that persist into later life stages on *E. tribuloides*.

These results, in conjunction with previous research, supports the hypothesis that heterochronic shifts in the expression of developmental GRNs can lead to changes in phenotype over evolutionary time(Raff & Wray, 1989). The evolutionary rewiring of the skeletal regulatory module has thus likely played a major role in the divergence and diversification of class Echinoidea, and future work investigating the temporal and spatial shifts in expression and

changes in GRN linkages of cidaroids will further illuminate the evolutionary consequences of these network alterations.

In contrast to this module wide shift of the expression of skeletal genes, genes associated with the non-skeletal mesenchyme (and especially the larval pigment cells), appear to have more minor expression alterations that create a less distinct pattern. Overall, the expression patterns for these genes support what has been found in previous studies, in that it seems like most major changes in the timing of non-skeletal mesenchyme gene expression are occurring in those genes at the terminus of the developmental module. Larval and adult coelomocytes of *E. tribuloides* seem to have similar morphologies to (presumably) homologous cells found in euechinoids, though it remains to be seen whether or not cidaroid coelomocytes have similar functions.

Regardless, the evidence presented here seems to support the hypothesis that alterations in developmental modules have different impacts on morphology depending on where in the module hierarchy the difference(s) occurred(*The Regulatory Genome*, 2006). It appears that skeletal program rewiring has occurred very early in echinoid development, and so severe differences are observed in the skeleton at nearly every stage of early development and even into the adult stages. Changes in the non-skeletal program appear to be mostly contained to the lowest hierarchical level and so the differences in final morphology are less significant. Ultimately, however, further investigation of *E. tribuloides* developmental gene linkages and functional morphology needs to be completed before declarative statements on the matter can be made.

References

Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450–461. https://doi.org/10.1038/nrg2102

Andrews, E. A. (1892). Experimental Embryology (Continued). *The American Naturalist*, *26*(307), 580–592. https://doi.org/10.1086/275556

Arnosti, D. N., & Ay, A. (2012). Boolean modeling of gene regulatory networks: Driesch redux. *Proceedings of the National Academy of Sciences*, *109*(45), 18239–18240. https://doi.org/10.1073/pnas.1215732109

Baer, K. E. von. (1999). *Über Entwickelungsgeschichte der Thiere: Beobachtung und Reflexion; 3 Bände in 1 Band* (O. Breidbach, Ed.; Nachdr. der Ausg. Königsberg, Bornträger, 1828-1888). Olms-Weidmann.

Bennett, K. C., Young, C. M., & Emlet, R. B. (2012a). Larval Development and Metamorphosis of the Deep-Sea Cidaroid Urchin *Cidaris blakei*. *The Biological Bulletin*, *222*(2), 105–117. https://doi.org/10.1086/BBLv222n2p105

Bennett, K. C., Young, C. M., & Emlet, R. B. (2012b). Larval Development and Metamorphosis of the Deep-Sea Cidaroid Urchin *Cidaris blakei*. *The Biological Bulletin*, *222*(2), 105–117. https://doi.org/10.1086/BBLv222n2p105

Bishop, C. D., MacNeil, K. E. A., Patel, D., Taylor, V. J., & Burke, R. D. (2013). Neural development in Eucidaris tribuloides and the evolutionary history of the echinoid larval nervous system. *Developmental Biology*, *377*(1), 236–244. https://doi.org/10.1016/j.ydbio.2013.03.006

Borotkanics, R., & Lehmann, H. (2015). Network motifs that recur across species, including gene regulatory and protein–protein interaction networks. *Archives of Toxicology*, *89*(4), 489–499. https://doi.org/10.1007/s00204-014-1274-y

Britten, R. J., & Davidson, E. H. (1969). Gene Regulation for Higher Cells: A Theory: New facts regarding the organization of the genome provide clues to the nature of gene regulation. *Science*, *165*(3891), 349–357. https://doi.org/10.1126/science.165.3891.349

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Calow, P. (1983). *Evolutionary Principles*. Springer US. https://doi.org/10.1007/978-1-4684-1518-6

Carlton, J. T. (2011). The Global Dispersal of Marine and Estuarine Crustaceans. In B. S. Galil, P. F. Clark, & J. T. Carlton (Eds.), *In the Wrong Place—Alien Marine Crustaceans: Distribution, Biology and Impacts* (pp. 3–23). Springer Netherlands. https://doi.org/10.1007/978-94-007-0591-3_1

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cohen, A. N. (2006). Chapter III Species Introductions and the Panama Canal. In S. Gollasch, B. S. Galil, & A. N. Cohen (Eds.), *Bridging Divides* (pp. 127–206). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5047-3_5

Davidson, E. H. (2006). *The regulatory genome: Gene regulatory networks in development and evolution*. Academic.

Davidson, E. H., & Erwin, D. H. (2006). Gene Regulatory Networks and the Evolution of Animal Body Plans. *Science*, *311*(5762), 796–800. https://doi.org/10.1126/science.1113832

Davidson, E. H., & Erwin, D. H. (2009). An Integrated View of Precambrian Eumetazoan Evolution. *Cold Spring Harbor Symposia on Quantitative Biology*, *74*(0), 65–80. https://doi.org/10.1101/sqb.2009.74.042

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. J., Schilstra, M. J., Clarke, P. J. C., … Bolouri, H. (2002). A Genomic Regulatory Network for Development. *Science*, *295*(5560), 1669–1678. https://doi.org/10.1126/science.1069883

Defoort, J., Van de Peer, Y., & Vermeirssen, V. (2018). Function, dynamics and evolution of network motif modules in integrated gene regulatory networks of worm and plant. *Nucleic Acids Research*, *46*(13), 6480–6503. https://doi.org/10.1093/nar/gky468

Dery, A., Collard, M., & Dubois, P. (2017). Ocean Acidification Reduces Spine Mechanical Strength in Euechinoid but Not in Cidaroid Sea Urchins. *Environmental Science & Technology*, *51*(7), 3640–3648. https://doi.org/10.1021/acs.est.6b05138

Dery, A., Guibourt, V., Catarino, A. I., Compère, P., & Dubois, P. (2014). Properties, morphogenesis, and effect of acidification on spines of the cidaroid sea urchin *Phyllacanthus imperialis*. *Invertebrate Biology*, *133*(2), 188–199. https://doi.org/10.1111/ivb.12054

Doney, S. C., Fabry, V. J., Feely, R. A., & Kleypas, J. A. (2009). Ocean Acidification: The Other CO_2 Problem. *Annual Review of Marine Science*, *1*(1), 169–192. https://doi.org/10.1146/annurev.marine.010908.163834

Dubois, P. (2014). The Skeleton of Postmetamorphic Echinoderms in a Changing World. *The Biological Bulletin*, *226*(3), 223–236. https://doi.org/10.1086/BBLv226n3p223

Echinoderms. Part A (First edition). (2019). Academic Press.

Eichenberger, P., Fujita, M., Jensen, S. T., Conlon, E. M., Rudner, D. Z., Wang, S. T., Ferguson, C., Haga, K., Sato, T., Liu, J. S., & Losick, R. (2004). The Program of Gene Transcription for a Single Differentiating Cell Type during Sporulation in Bacillus subtilis. *PLoS Biology*, *2*(10), e328. https://doi.org/10.1371/journal.pbio.0020328

Emlet, R. B. (1988). Larval Form and Metamorphosis of a "Primitive" Sea Urchin, *Eucidaris thouarsi* (Echinodermata: Echinoidea: Cidaroida), with Implications for Developmental and Phylogenetic Studies. *The Biological Bulletin*, *174*(1), 4–19. https://doi.org/10.2307/1541754

Erkenbrack, E. M., Ako-Asare, K., Miller, E., Tekelenburg, S., Thompson, J. R., & Romano, L. (2016). Ancestral state reconstruction by comparative analysis of a GRN kernel operating in echinoderms. *Development Genes and Evolution*, *226*(1), 37–45. https://doi.org/10.1007/s00427-015-0527-y

Erkenbrack, E. M., & Davidson, E. H. (2015). Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. *Proceedings of the National Academy of Sciences*, *112*(30). https://doi.org/10.1073/pnas.1509845112

Erwin, D. H., & Davidson, E. H. (2009). The evolution of hierarchical gene regulatory networks. *Nature Reviews Genetics*, *10*(2), 141–148. https://doi.org/10.1038/nrg2499

Eyre, B. D., Andersson, A. J., & Cyronak, T. (2014). Benthic coral reef calcium carbonate dissolution in an acidifying ocean. *Nature Climate Change*, *4*(11), 969–976. https://doi.org/10.1038/nclimate2380

Formery, L., Wakefield, A., Gesson, M., Toisoul, L., Lhomond, G., Gilletta, L., Lasbleiz, R., Schubert, M., & Croce, J. C. (2022). Developmental atlas of the indirect-developing sea

urchin Paracentrotus lividus: From fertilization to juvenile stages. *Frontiers in Cell and Developmental Biology*, *10*, 966408. https://doi.org/10.3389/fcell.2022.966408

G., J. S. (1936). A Monograph of the Echinoidea. *Nature*, *138*(3487), 344–344. https://doi.org/10.1038/138344a0

Gao, F., Thompson, J. R., Petsios, E., Erkenbrack, E., Moats, R. A., Bottjer, D. J., & Davidson, E. H. (2015). Juvenile skeletogenesis in anciently diverged sea urchin clades. *Developmental Biology*, *400*(1), 148–158. https://doi.org/10.1016/j.ydbio.2015.01.017

Gazeau, F., Parker, L. M., Comeau, S., Gattuso, J.-P., O'Connor, W. A., Martin, S., Pörtner, H.-O., & Ross, P. M. (2013). Impacts of ocean acidification on marine shelled molluscs. *Marine Biology*, *160*(8), 2207–2245. https://doi.org/10.1007/s00227-013-2219-3

Geniza, M., & Jaiswal, P. (2017). Tools for building de novo transcriptome assembly. *Current Plant Biology*, *11–12*, 41–45. https://doi.org/10.1016/j.cpb.2017.12.004

George, S. B., Lawrence, J. M., & Lawrence, A. L. (2004). Complete larval development of the sea urchin Lytechinus variegatus fed an artificial feed. *Aquaculture*, *242*(1–4), 217–228. https://doi.org/10.1016/j.aquaculture.2004.06.024

Goodman, C. S., & Coughlin, B. C. (2000). The evolution of evo-devo biology. *Proceedings of the National Academy of Sciences*, *97*(9), 4424–4425. https://doi.org/10.1073/pnas.97.9.4424

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Grossmann, J. N., & Nebelsick, J. H. (2013). Comparative morphological and structural analysis of selected cidaroid and camarodont sea urchin spines. *Zoomorphology*, *132*(3), 301–315. https://doi.org/10.1007/s00435-013-0192-5

Haeckel, E. & Royal College of Physicians of London. (1866). *Generelle Morphologie der Organismen: Allgemeine Grundzuge der organischen Formen-Wissenschaft, mechanisch begrundet durch die von Charles Darwin reformirte Descendenz-Theorie /*. Georg Reimer,. https://doi.org/10.5962/bhl.title.101747

Hall, B. K. (2012). Evolutionary Developmental Biology (Evo-Devo): Past, Present, and Future. *Evolution: Education and Outreach*, *5*(2), 184–193. https://doi.org/10.1007/s12052-012-0418-x

Hart, Y., Antebi, Y. E., Mayo, A. E., Friedman, N., & Alon, U. (2012). Design principles of cell circuits with paradoxical components. *Proceedings of the National Academy of Sciences*, *109*(21), 8346–8351. https://doi.org/10.1073/pnas.1117475109

Hopwood, N. (2007). A history of normal plates, tables and stages in vertebrate embryology. *The International Journal of Developmental Biology*, *51*(1), 1–26. https://doi.org/10.1387/ijdb.062189nh

Horstadius, S. (1939). THE MECHANICS OF SEA URCHIN DEVELOPMENT, STUDIED BY OPERATIVE METHODS. *Biological Reviews*, *14*(2), 132–179. https://doi.org/10.1111/j.1469-185X.1939.tb00929.x

Hörstadius, S. (1973). *Experimental embryology of echinoderms*. Clarendon Pr.

Istrail, S., & Davidson, E. H. (2005). Logic functions of the genomic cis-regulatory code. *Proceedings of the National Academy of Sciences*, *102*(14), 4954–4959. https://doi.org/10.1073/pnas.0409624102

Jin, H., Wan, Y.-W., & Liu, Z. (2017). Comprehensive evaluation of RNA-seq quantification methods for linearity. *BMC Bioinformatics*, *18*(S4), 117. https://doi.org/10.1186/s12859-017-1526-y

Joanito, I., Chu, J.-W., Wu, S.-H., & Hsu, C.-P. (2018). An incoherent feed-forward loop switches the Arabidopsis clock rapidly between two hysteretic states. *Scientific Reports*, *8*(1), 13944. https://doi.org/10.1038/s41598-018-32030-z

Johnson, P. T. (1969). The coelomic elements of sea urchins (Strongylocentrotus) I. The normal coelomocytes; Their morphology and dynamics in hanging drops. *Journal of Invertebrate Pathology*, *13*(1), 25–41. https://doi.org/10.1016/0022-2011(69)90236-5

Jones, D. M., & Vandepoele, K. (2020). Identification and evolution of gene regulatory networks: Insights from comparative studies in plants. *Current Opinion in Plant Biology*, *54*, 42–48. https://doi.org/10.1016/j.pbi.2019.12.008

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics*, *203*(3), 253–310. https://doi.org/10.1002/aja.1002030302

Kipryushina, Y. O., & Yakovlev, K. V. (2020). Maternal control of early patterning in sea urchin embryos. *Differentiation*, *113*, 28–37. https://doi.org/10.1016/j.diff.2020.04.001

Kowalevsky, A. (1871). Weitere Studien über die Entwicklung der einfachen Ascidien. *Archiv für Mikroskopische Anatomie*, *7*(1), 101–130. https://doi.org/10.1007/BF02956047

Kroh, A., & Smith, A. B. (2010). The phylogeny and classification of post-Palaeozoic echinoids. *Journal of Systematic Palaeontology*, *8*(2), 147–212. https://doi.org/10.1080/14772011003603556

Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., & Tamura, K. (2024). MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. *Molecular Biology and Evolution*, *41*(12), msae263. https://doi.org/10.1093/molbev/msae263

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559. https://doi.org/10.1186/1471-2105-9-559

Lee, Y.-H. (2003). Molecular Phylogenies and Divergence Times of Sea Urchin Species of Strongylocentrotidae, Echinoida. *Molecular Biology and Evolution*, *20*(8), 1211–1221. https://doi.org/10.1093/molbev/msg125

Lessios, H. A. (1991). Presence and absence of monthly reproductive rhythms among eight Caribbean echinoids off the coast of Panama. *Journal of Experimental Marine Biology and Ecology*, *153*(1), 27–47. https://doi.org/10.1016/S0022-0981(05)80004-8

Levine, M., & Davidson, E. H. (2005). Gene regulatory networks for development. *Proceedings of the National Academy of Sciences*, *102*(14), 4936–4942. https://doi.org/10.1073/pnas.0408031102

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

MacNeil, K. E. A., Scaros, A. T., Croll, R. P., & Bishop, C. D. (2017). Differences in Larval Arm Movements Correlate with the Complexity of Musculature in Two Phylogenetically Distant Echinoids, *Eucidaris tribuloides* (Cidaroidea) and *Lytechinus variegatus* (Euechinoidea). *The Biological Bulletin*, *233*(2), 111–122. https://doi.org/10.1086/694892

Mangan, S., & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, *100*(21), 11980–11985. https://doi.org/10.1073/pnas.2133841100

Martin, S., Richier, S., Pedrotti, M.-L., Dupont, S., Castejon, C., Gerakis, Y., Kerros, M.-E., Oberhänsli, F., Teyssié, J.-L., Jeffree, R., & Gattuso, J.-P. (2011). Early development and molecular plasticity in the Mediterranean sea urchin *Paracentrotus lividus* exposed to CO2-driven acidification. *Journal of Experimental Biology*, *214*(8), 1357–1368. https://doi.org/10.1242/jeb.051169

Matranga, V., Pinsino, A., Celi, M., Natoli, A., Bonaventura, R., Schröder, H. C., & Müller, W. E. G. (2005). Monitoring Chemical and Physical Stress Using Sea Urchin Immune Cells. In V. Matranga (Ed.), *Echinodermata* (Vol. 39, pp. 85–110). Springer-Verlag. https://doi.org/10.1007/3-540-27683-1_5

McClay, D. R. (2011). Evolutionary crossroads in developmental biology: Sea urchins. *Development*, *138*(13), 2639–2648. https://doi.org/10.1242/dev.048967

McClintock, J. B., & Watts, S. A. (1990). The effects of photoperiod on gametogenesis in the tropical sea urchin Eucidaris tribuloides (Lamarck) (Echinodermata: Echinoidea). *Journal of Experimental Marine Biology and Ecology*, *139*(3), 175–184. https://doi.org/10.1016/0022-0981(90)90145-3

McEdward, L. R., & Herrera, J. C. (1999). Body form and skeletal morphometrics during larval development of the sea urchin Lytechinus variegatus Lamarck. *Journal of Experimental Marine Biology and Ecology*, *232*(2), 151–176. https://doi.org/10.1016/S0022-0981(98)00106-3

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., & Gehring, W. J. (1984). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. *Cell*, *37*(2), 403–408. https://doi.org/10.1016/0092-8674(84)90370-2

McPHERSON, B. F. (1968). FEEDING AND OXYGEN UPTAKE OF THE TROPICAL SEA URCHIN EUCIDARIS TRIBULOIDES (LAMARCK) ,. *The Biological Bulletin*, *135*(2), 308–321. https://doi.org/10.2307/1539784

Morse, J. W., Arvidson, R. S., & Lüttge, A. (2007). Calcium Carbonate Formation and Dissolution. *Chemical Reviews*, *107*(2), 342–381. https://doi.org/10.1021/cr050358j

Moulin, L., Catarino, A. I., Claessens, T., & Dubois, P. (2011). Effects of seawater acidification on early development of the intertidal sea urchin Paracentrotus lividus (Lamarck 1816). *Marine Pollution Bulletin*, *62*(1), 48–54. https://doi.org/10.1016/j.marpolbul.2010.09.012

Nesbit, K. T., & Hamdoun, A. (2020). Embryo, larval, and juvenile staging of *LYTECHINUS PICTUS* from fertilization through sexual maturation. *Developmental Dynamics*, *249*(11), 1334–1346. https://doi.org/10.1002/dvdy.223

Oliveri, P., Carrick, D. M., & Davidson, E. H. (2002). A Regulatory Gene Network That Directs Micromere Specification in the Sea Urchin Embryo. *Developmental Biology*, *246*(1), 209–228. https://doi.org/10.1006/dbio.2002.0627

Oliveri, P., Tu, Q., & Davidson, E. H. (2008). Global regulatory logic for specification of an embryonic cell lineage. *Proceedings of the National Academy of Sciences*, *105*(16), 5955–5962. https://doi.org/10.1073/pnas.0711220105

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Peter, I. S., & Davidson, E. H. (2009). Modularity and design principles in the sea urchin embryo gene regulatory network. *FEBS Letters*, *583*(24), 3948–3958. https://doi.org/10.1016/j.febslet.2009.11.060

Raff, R. A., & Love, A. C. (2004). Kowalevsky, comparative evolutionary embryology, and the intellectual lineage of evo-devo. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *302B*(1), 19–34. https://doi.org/10.1002/jez.b.20004

Raff, R. A., & Wray, G. A. (1989). Heterochrony: Developmental mechanisms and evolutionary results. *Journal of Evolutionary Biology*, *2*(6), 409–434. https://doi.org/10.1046/j.1420-9101.1989.2060409.x

Rafiq, K., Cheers, M. S., & Ettensohn, C. A. (2012). The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development*, *139*(3), 579–590. https://doi.org/10.1242/dev.073049

Raghavan, V., Kraft, L., Mesny, F., & Rigerte, L. (2022). A simple guide to *de novo* transcriptome assembly and annotation. *Briefings in Bioinformatics*, *23*(2), bbab563. https://doi.org/10.1093/bib/bbab563

Rijsewijk, F., Schuermann, M., Wagenaar, E., Parren, P., Weigel, D., & Nusse, R. (1987). The Drosophila homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. *Cell*, *50*(4), 649–657. https://doi.org/10.1016/0092-8674(87)90038-9

Saudemont, A., Haillot, E., Mekpoh, F., Bessodes, N., Quirin, M., Lapraz, F., Duboc, V., Röttinger, E., Range, R., Oisel, A., Besnardeau, L., Wincker, P., & Lepage, T. (2010). Ancestral Regulatory Circuits Governing Ectoderm Patterning Downstream of Nodal and

BMP2/4 Revealed by Gene Regulatory Network Analysis in an Echinoderm. *PLoS Genetics*, *6*(12), e1001259. https://doi.org/10.1371/journal.pgen.1001259

Schroeder, T. E. (1981). DEVELOPMENT OF A “PRIMITIVE” SEA URCHIN (*EUCIDARIS TRIBULOIDES:*) IRREGULARITIES IN THE HYALINE LAYER, MICROMERES, AND PRIMARY MESENCHYME. *The Biological Bulletin*, *161*(1), 141–151. https://doi.org/10.2307/1541114

Scott, M. P., & Weiner, A. J. (1984). Structural relationships among genes that control development: Sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(13), 4115–4119. https://doi.org/10.1073/pnas.81.13.4115

Setting the stage for development: The maternal-to-zygotic transition in Drosophila. (2023). *Genetics*. https://doi.org/10.1093/genetics/iyad142

Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics*, *31*(1), 64–68. https://doi.org/10.1038/ng881

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Stone, J. R., & Wray, G. A. (2001). Rapid Evolution of cis-Regulatory Sequences via Local Point Mutations. *Molecular Biology and Evolution*, *18*(9), 1764–1770. https://doi.org/10.1093/oxfordjournals.molbev.a003964

Sundaram, V., Choudhary, M. N. K., Pehrsson, E., Xing, X., Fiore, C., Pandey, M., Maricque, B., Udawatta, M., Ngo, D., Chen, Y., Paguntalan, A., Ray, T., Hughes, A., Cohen, B. A., & Wang, T. (2017). Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nature Communications*, *8*(1), 14550. https://doi.org/10.1038/ncomms14550

The Regulatory Genome. (2006). Elsevier. https://doi.org/10.1016/B978-0-12-088563-3.X5018-4

Thompson, J. R., Petsios, E., Davidson, E. H., Erkenbrack, E. M., Gao, F., & Bottjer, D. J. (2015). Reorganization of sea urchin gene regulatory networks at least 268 million years ago as revealed by oldest fossil cidaroid echinoid. *Scientific Reports*, *5*(1), 15541. https://doi.org/10.1038/srep15541

Thompson, J. V. (1829). *Zoological researches and illustrations: Or, Natural history of nondescript or imperfectly known animals, in a series of memoirs / by John V. Thompson*. [s.n.]. https://doi.org/10.5962/bhl.title.14446

Valentine, J. W. (1997). Cleavage patterns and the topology of the metazoan tree of life. *Proceedings of the National Academy of Sciences*, *94*(15), 8001–8005. https://doi.org/10.1073/pnas.94.15.8001

van ’t Veer, L. J., van Kessel, A. G., van Heerikhuizen, H., van Ooyen, A., & Nusse, R. (1984). Molecular cloning and chromosomal assignment of the human homolog of int-1, a mouse gene implicated in mammary tumorigenesis. *Molecular and Cellular Biology*, *4*(11), 2532–2534. https://doi.org/10.1128/mcb.4.11.2532-2534.1984

Voordeckers, K., Pougach, K., & Verstrepen, K. J. (2015). How do regulatory networks evolve and expand throughout evolution? *Current Opinion in Biotechnology*, *34*, 180–188. https://doi.org/10.1016/j.copbio.2015.02.001

Werneburg, I., & Sánchez-Villagra, M. R. (2009). Timing of organogenesis support basal position of turtles in the amniote tree of life. *BMC Evolutionary Biology*, *9*(1), 82. https://doi.org/10.1186/1471-2148-9-82

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Yagita, K., & Okamura, H. (2000). Forskolin induces circadian gene expression of *rPer1* , *rPer2* and *dbp* in mammalian rat-1 fibroblasts. *FEBS Letters*, *465*(1), 79–82. https://doi.org/10.1016/S0014-5793(99)01724-X

Yamazaki, A., Morino, Y., Urata, M., Yamaguchi, M., Minokawa, T., Furukawa, R., Kondo, M., & Wada, H. (2020). *Pmar1* / *phb* homeobox genes and the evolution of the double-negative gate for endomesoderm specification in echinoderms. *Development*, dev.182139. https://doi.org/10.1242/dev.182139