UC Davis
UC Davis Electronic Theses and Dissertations

Title

Understanding homology across animal history: deep time evolutionary reconstruction at the cellular and genetic level

Permalink

https://escholarship.org/uc/item/2040k2hs

Author

Sierra, Noemie

Publication Date

2023

Peer reviewed|Thesis/dissertation

Understanding homology across animal history: deep time evolutionary reconstruction at the cellular and genetic level

By

NOEMIE SIERRA

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics & Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

David Gold, Chair

_____

Celina Juliano

_____

Rachael Bay

_____

Todd Oakley

Committee in Charge
2023

Abstract

Homology – or the similarity of the structure, physiology, or development between different species owing to common descent – is a fundamental concept in all biological research. That the model organisms we use are an appropriate model for human biology – and that studying the basic biology of these organisms has something to teach us about ourselves – underlies the way in which biological and medical research is conducted. But homology is multi-faceted, and at best, tricky to define. Often defined as similarity stemming from common ancestry, in truth it can be defined at multiple levels of biological hierarchy. Traits that are homologous at one level–such as anatomical–may not be homologous at another level–such as developmental. The framework of observation can therefore affect whether or not a system is defined as "homologous". The research presented in this thesis aims to investigate the effectiveness of various techniques of assessing homology of non-structural features; with a specific focus on traits shared across animals, such as the origin of novel cell types and "conserved" genetic responses. In this chapter I will lay out the specific research questions and aims of this dissertation, its significance and its limitations. I conclude by arguing for the value of using character identity over character state (following McKenna et al. 2021) in assessing homology of non-structural characters, as well as potential ways of contextualizing these features.

Acknowledgements

When I began my PhD I never could have imagined the personal and academic challenges that lie ahead, including a complete pivot in my personal life, readjusting an academic program to accommodate the onset of a global pandemic, and intense societal upheaval for the reimagining of a better future. I would like to express my deepest gratitude to the exceptional individuals who supported and guided me throughout this journey, enabling me to successfully obtain my doctorate despite the many unforeseen obstacles that arose.

First and foremost, I would like to thank my PhD supervisor, Prof. David Gold, whose constant support and guidance has been invaluable throughout the entire process. From the initial stages of refining my research proposal to the final submission of my thesis, your guidance has been instrumental in getting me to the finish line, even when I always wanted to add one more thing and read one more paper. It's through your voice that I've finally learned the balance necessary to get things done, and to recognize my own abilities and find my voice in the scientific community. I am also incredibly grateful to Professors Todd Oakley and Celina Juliano, who encouraged and believed in me during my earliest stages and without whose support I would not have begun this journey. I am thankful for the opportunity to be a part of an exceptional research institute that encourages research as interdisciplinary as mine, and the warm welcome I received at the Earth & Planetary Sciences Dept despite the fact that when I first joined the department I didn't know a metamorphic from an igneous rock.

More personally, I would also like to express gratitude to my partner Jessica, who helped see me through tremendous personal growth during this time and helped me keep my head on straight, even during the toughest parts of this experience. And finally of course, this process would not have been possible without the support and encouragement of my parents, and especially my dad who listened to me ramble incessantly for years about phylogenetic theory.

Lastly, I would like to thank my funding bodies, the NSF and the University of California - Davis, for both project funding and the fellowships that provided security and freedom for me to advance considerably in my research. As I bring my five-year-long PhD journey to a close, I can genuinely say that I am immensely proud of my accomplishments. This journey has not only shaped me as a researcher but also as an individual, and I am deeply grateful to all the individuals mentioned, as none of this would have been possible without their guidance and encouragement.

Contents

List of Figures and Tables

Chapter 1

The case for character identity in non-structural homology

INTRODUCTION

Homology is a fundamental assumption of modern biological and medical research, and as such, deserves careful nuance and consideration when making claims about the conservation of processes. "…[W]hen we argue that discoveries about a roundworm, a fruit fly, a frog, a mouse or a chimp have relevance to the human conditions – we have made a bold and direct statement about homology" (Wake, 1994). Our ability to generalize insights from model systems to both our fundamental understanding of life, and its application to human biology is founded in the assumption that the similarities between species are more than superficial – we are assuming that what we learn from one can be the basis to understand the other, and that this can enable effective human responses like therapies.

The simplest definition of homology is a shared trait due to common ancestry; yet shared traits do not equal common descent. The similarity in appearance of a character does not necessarily imply homology, as many paths can lead to a superficially similar trait. A stricter definition of homology is the possession of a trait due to its presence – in its entirety – in the last common ancestor of both descendants. So while a morphologically similar trait may be constructed of homologous parts, for the trait itself to be considered homologous it must have been present in its assembled state in the last common ancestor. If those homologous parts were assembled into a system, now found in both descendants, after the divergence of both lineages, it would be an example of homoplasy, a case of arriving at a similar solution to a common problem. Homoplasy is evident in the comparison of bird and bat wings, where similar wing structures have evolved independently in these two distantly related groups. Despite the outward resemblance, the underlying skeletal structures and genetic pathways that give rise to these wings differ significantly, indicating convergent evolution rather than shared ancestry and underscoring the principle that morphological similarity does not necessarily imply homology. The challenge lies in differentiating homology and homoplasy over deep time – even "well-conserved" traits have experienced change over time as each organism evolves, obscuring shared historical identity.

Homology is also present at multiple levels of biological organization simultaneously, and a meticulous distinction in defining the level of homology within biological systems becomes imperative to avoid misinterpretations and inaccuracies. Consider the example of genes participating in a specific biological response; while the genes themselves may be homologous due to shared ancestry, the encompassing process might not necessarily be homologous. This is evident when a single gene serves multiple functions across different temporal or spatial contexts within an organism, leading to a scenario where the genes are homologous, but the process itself is not. Conversely, a genetic response could have evolutionary roots, originating from a common ancestor, while the individual gene responsible for a particular role has undergone substitution with a functionally similar yet different gene or even a paralog, resulting in a non-homologous gene functioning within an ostensibly homologous process. This example underscores the necessity of precision when discussing homology: to avoid inaccuracies, it's imperative to specify the level being considered, as something homologous at one level might diverge significantly at another, underscoring the dynamic nature of evolutionary connections.

In structural traits, homology is established through comparison of position and often through the accompanying developmental program. A recent review by McKenna et al. (2021) reiterates an important perspective about how homology is assessed in structural features –

differentiating the state, or the actual features and structure of a trait, from its identity –

where its identity, the nomination linking it as equivalent to a similar trait in other organisms

(such as the designation autopod), is based on its "regional identity manifest" or its locational

context within a greater spatial or temporal organization plan. They argue that a structural

entity can be linked as homologous to structurally related features in other organisms based

on relative reliability of the position, reflective of conservation of patterning mechanisms. For

example, in contrast to the standard body plan of insects, dipterans (flies) possess a single pair

of wings on the standard T2 (mesothoracic) segment, and an accompanying balance organ -

know as the haltere - on T3 (the metathoracic segment). The locational link between the

second wing of the standard insect plan and the haltere is what establishes it as a homologous

structure to the second wing, changing our understanding of its evolutionary origin and

designating the haltere as a modified wing rather than a chitinous growth (such as a sensory

hair). While focusing on contextualization is nothing new (rather a tradition established in the

19th century (Panchen, 2007), its extrapolation to non-structural traits and other levels of

biological organization requires careful nuance. McKenna et al. create an effective analogy

comparing the identity of a genetic locus to allelic variation at that location: the allele

represents the current form of a gene, with its accumulated mutations making it

morphologically unique from other alleles. While the structure and expression result of two

alleles may differ (representing the current state), they can be homologized by their common

contextual location, the gene locus – such that no matter how much difference has accumulated between these different alleles, their shared evolutionary history can be identified by their "regional identity manifest", the locus, demonstrating how the principle of contextual assessment can be applied to other types of traits beyond structural.

For many complex characters, our current assessments of homology are still based on the state of the character: we assess the homology by examining whether the features of each are the same, rather than attempt to contextualize them. For example, in determining whether a genetic response to a stimulus is homologous across different organisms, we examine the differential gene expression to determine if the genes expressed are the same. However, as this is elaborated in later sections, this is an effort to assess homology based on state rather than context driven identity. Rather, like structural traits, homology of genetic responses (and other non-structural traits such as cell identity/lineage or genetic responses) should be assessed through contextualization, and the contextualization we determine is unique to each organizational level we are looking at.

Chapter 2: Homology and Cell Identity

Novel cell identities do not always exist within a fixed bodily context, such as an organ. In the absence of a specific organ or tissue that is shared through common ancestry, it could be

difficult to contextualize cell types without some form of "regional identity manifest" (McKenna et al. 2021). Neurons across metazoa offer an excellent example; while most bilaterians share a common tissue and architectural structure based around the bilateral body plan, it is hard to compare to the neurobiology of ctenophores (comb jellies), a group placed further back in the animal tree and possessing radial symmetry and a decentralized nerve net, a question pressingly relevant because of the uncertainty of the last few years as to the position of porifera (sponges) or ctenophores as the sister group to all other animals (Marlow & Arendt, 2014; Ryan, 2014; Moroz & Kohn, 2016). Morphology-based taxonomy paints a very smooth evolutionary picture linking choanoflagellates – a type of colonial single-celled organism that lies just outside the metazoan tree – to sponges, due to the significant morphological similarity between choanoflagellates and choanocytes (sponge collar cells) as well as the lack of organ and tissue differentiation in sponges. In concert with cellular similarity, the lack of tissue differentiation in sponges has at one time been interpreted as a "stepwise" progression in complexity, from a group of single cells capable of intercellular communication to a single multicellular individual lacking tissue differentiation. This makes for a very convenient story of neuron origin, suggesting neurons could have originated once along the animal tree, carried down since its origin to subsequent phyla. However, we know from other studies that a "step-wise" or gradual and linear approach to complexity, however appealing the elegance of the story, is not always the case (Oakley & Plachetzki, 2011; Picciani

6

et al. 2018). As molecular phylogenetic evidence has called into question the position of sponges at the base of the metazoan tree (Schultz et al. 2023), these alternative phylogenetic hypotheses imply very different scenarios for the evolution of neurons and raise doubts about the potential shared ancestry of neurons from ctenophores, cnidarians, and bilaterians.

With morphological comparison insufficient to distinguish between convergent evolution and shared ancestry of cell types, an alternative approach could be the use of phylogenetics and ancestral state reconstruction. Trait reconstruction incorporates broad information across a larger tree (not a one-to-one species comparison) and multiple types of state information to estimate the likelihood that a trait was present in the last common ancestor. While this reconstruction technique is still based on state, the integration of diverse types of data allows us to more robustly estimate the likelihood of observing a similar state of multiple types of associated features given its presence (or not) in the last common ancestor.

In chapter 2, I use molecular phylogenetics to trace the history of diverse forms of a unique and novel cell type in Cnidaria, the cnidocyte. These stinging cells are ubiquitously present across the clade and have been modified into more than 30 forms for a diversity of functions. The combination of cnidocyte types a species possesses, known as the cnidome, is thoroughly documented as a characteristic to define and identify species, particularly due to the relevance

of venom delivery (enacted by cnidocytes) and the danger these animals can pose to humans. These cell types are entirely unique to the phylum and are not isolated to particular organs or regions of the body, and their distribution and type can vary across life stages. The lack of associated "regional manifest" of where and at what life stage these cell types are expressed, combined with the enormous diversity of types, makes this cell an ideal candidate to explore potential convergence through phylogeny. While some types – such as the ptychocyst of Ceriantharia or rhopalonemes of Siphonophora – are clearly restricted to a particular clade, others – such as the stenotele, the classic cnidocyte with complex barbature – are broadly observed in very disparate clades, leading to the question of whether they were frequently lost or if they could have been repeatedly evolved. Using a combination of five genes and extensive morphological data on cnidomes across the cnidarian tree, I generated a time-calibrated molecular clock of cnidarians and performed ancestral state reconstruction on twelve well-documented cnidocyte types to determine 1) whether the cnidocyte at the base of the tree possessed the simplest morphology, and 2) whether the more complex cnidocyte types had likely evolved a single or multiple times. These explorations allow us to determine the role of simplification or cell type loss in the diversity of cnidomes, as well as the potential role of convergent evolution in the patterns of expression of these complex cell types. My study concluded that the first cnidarians had likely only the simplest and least specialized cnidocyte type (the isorhiza). I also found evidence to support that many of the more complex types with

elaborate morphologies, such as stenoteles and euryteles, likely evolved once, and that the diversity of their expression across the tree is frequently owing to the loss of structures rather than frequent convergent evolution. These results were robust to changes in the molecular clock model and are significant in demonstrating that while some complex characters can arise multiple times convergently (Picciani et al. 2018), loss of a complex character evolved once can also play a significant role in diversity of expression of a character in a large tree.

The various types of neurons have been historically treated as homologous, yet morphological comparisons alone often fall short in distinguishing between convergent evolution and shared ancestry. Phylogenetic reconstruction incorporates multiple data points, including genetic sequences and molecular markers, facilitating a broader assessment of traits and reducing reliance on individual morphological features that might be subject to convergent evolution. Through the application of statistical models, phylogenetic methods rigorously evaluate the likelihood of observed traits under various evolutionary scenarios, enabling discrimination between scenarios favoring homology or homoplasy. Furthermore, the use of character evolution models within phylogenetic reconstruction accommodates different ways traits can change over time, providing a robust statistical framework for assessing the origins of trait similarities. Phylogenetic reconstruction provides a more comprehensive and statistically

robust framework for addressing this challenge, offering a deeper understanding of the evolutionary relationships underlying shared traits.

Chapter 3: Homology and Genetics

Genetic response is another non-structural trait whose homology is routinely investigated to understand the conservation and commonality of biological processes in diverse organisms. Identifying homologous genetic responses can confirm the relevance of findings from model organisms and facilitate the translation of knowledge gained from model organisms to humans, as well as provide significant insight into our evolutionary history and biology of genes. The most common way to investigate the homology of genetic responses is differential gene expression, where the timing and identity of genes expressed under a common stimulus are compared to determine if the same genes are used to enact a similar response in different species. To meet our objective of studying the homology and conservation of deeply established non-structural traits, we focused on a notable challenge to studying the differential gene expression of distantly related organisms, which is data collection. To create a comparable dataset for traits one would require a comparable experimental regime, which would require not only extensive collection setup but careful planning to account for the significant organismal differences. When observing ancient traits in particular, the genetic background in distantly related species has likely changed so drastically that trying to

understand if the responses are homologous (likely to have descended, fully assembled, from a last common ancestor) can be especially difficult. For example, one of the deep-established, supposedly conserved processes that we addressed in this thesis, heat shock response, is likely to differ based on such basic traits as homeothermic capacity or even organism size, and the nature of the stimulus (temperature of heat shock as well as exposure length) would have to be accounted for. For instance, in our study of thermal shock the temperature difference required to elicit an acute shock response in the Atlantic Salmon (Salmo salar) required a 10°C difference for 24hrs, while the cnidarian Acropora millepora was exposed to only a 4℃ difference for 4hrs (Shi et al. 2019), highlighting the impact of physiological and ecological differences on thermal tolerance. Creating a comparable dataset using timing of data collection and exposure constitutes a significant challenge when comparing differential gene expression across diverse organisms.

To address this, we created a statistical method that would allow us to compare datasets from different existing experiments (publicly available on NCBI) with not only different biological regimes appropriate to the specific organism, but also different technical regimes (amount of data collected,sequencing depth, sequencing type, etc). Our approach exploits the multitude of p-values generated by a differential expression analysis for a group of genes, treating each as a data point that can contribute to a larger picture of expression significance. The differentially

expressed genes from each original experiment are clustered into conserved ortholog groups (COGs), or clusters of homologous gene families or gene groups. The associated p-values from the original experiment are treated as data values for each gene, such that the p-value for each gene within each ortholog group can then be treated as independent significance tests of the hypothesis that the broader COG is differentially expressed. We tested the method on two "well-conserved" processes, tissue regeneration and heat shock responses, and our investigation turned up two notable things. The first was that neither of these so-called "conserved" processes identified a conserved set of orthologs underlying them. This could be indicative of an issue in the nature of our statistical methodology, however the second notable observation leads us to hypothesize that there is a more fundamental concern in our approach than the statistical method, which is that differential gene expression may not represent an appropriate way of assessing homology. Two of the original acute heat shock datasets comparing very closely related species (one comparing three breeds of chickens (Lan et al. 2016) and one comparing three species of rice planthoppers (Huang et al. 2017)) failed to identify a significant set of differentially expressed genes within the original dataset. For a process that is supposed to be deeply conserved, it is concerning that differential gene expression is unable to recover common genes in very closely related organisms. It is possible that these processes are more variable than previously believed, and that attempting to glean a homologous response from these features is ineffectual.

A different interpretation of this observation however, is that differential gene response may not be an appropriate method of assessing homology. The genes currently being expressed in a response, including the timing and location in which they are expressed actually represents the state of the feature; the way a feature has evolved and been modified over time to suit contemporary needs. As such, DGE may not be an appropriate methodology for differentiating between a convergent response and true homology. Like other non-structural traits, it's critical to appropriately contextualize identity rather than state to assess homology, but unlike isolated cell types, gene expression has a form of "regional identity manifest", a structural context in which it is found that can help physically contextualize it, which is the overlying regulatory network in which it exists.

A gene regulatory network is a system of interconnected genes and their regulatory interactions that govern expression. While individual nodes represent genes or components, their true significance emerges from how they interact within the network, and analyzing network architecture allows us to uncover shared motifs, connectivity patterns, hub nodes, and modular structures that contribute to the network's behavior. Therefore the architecture of the network, like phylogenetic inference, represents the incorporation of additional layers of information increasing our capacity to contextualize the expression of individual genes and assess homology. Specifically, comparing the architecture of two networks and their similarity

will inform us about the likelihood that they were established convergently or not: the likelihood that two networks evolve into a significantly similar architecture by convergent evolution diminishes with increased network complexity. Directly comparing network architecture using well-established machine learning techniques has the potential to inform us about the likelihoods that the networks share an evolutionary history independently of changes to the nodes (genes) themselves. The identity of individual nodes in a network may change over time through paralog substitution or substitutions of structurally appropriate genes, even if the overlying network remains largely unchanged. Focusing on the expression of particular genes, particular nodes instead of the architecture of the network, can result in both over or underestimation of homology, for instance if the same gene is used to serve a similar function, but the overlying network regulating that gene or its downstream is vastly different, or if the genes are found not to be the same because an individual node has experienced a substitution a process may be deemed non-conserved, despite the governing process potentially being otherwise similar. Entire regulatory networks meant to enact a similar response may be overall more conserved than individual genes within that network. Comparing the architecture of two gene regulatory networks provides a nuanced and comprehensive insight into their similarities that surpasses the understanding gained from focusing solely on individual nodes. By examining the shared architectural features, we gain a deeper understanding of the underlying principles governing the networks' behaviors,

enabling us to identify common regulatory mechanisms, signaling pathways, and biological functions that are conserved across different organisms or contexts.

Chapter 4: Combining cell type and genetics to define homology

In chapter 2 I investigated homology of an organ-independent cell type using phylogenetic methodology, and in chapter 3 I investigated the homology of gene response (and concluded that gene regulatory networks are a far more appropriate approach than DGE). In the final chapter, I attempted to use single cell sequencing to use developmental trajectories as a way of contextualizing cell type evolution and identify a previously unidentified, putatively homologous cell type between Aurelia and Hydra – the interstitial stem cell. The polyps of Aurelia share a similar regenerative capacity as Hydra, yet Aurelia is part of a group of cnidarians with as-of-yet undescribed stem cells.

Using single cell sequencing gives us an additional layer of information beyond transcription profiles of two cell types – in organisms such as cnidarians, where cell replacement is occurring routinely at a high rate, single cell transcription plots can show the dynamic changes of expression as cells evolve from one type to another and visually demonstrate the developmental trajectory of cell lineages. My aim was to use the single cell data to identify the precursor for neurons and cnidocytes in Aurelia from plot architecture, and compare the

expression profiles between this precursor and the interstitial stem cell of Hydra. Limitations in our data collection prevented us from collecting sufficient cells to transcriptionally differentiate each of the terminal cell types, however our results suggest that Aurelia has a cnidocyte/neurosensory cell precursor comparable in its developmental trajectory to Hydra. The next steps of this work will require increased cell sampling to characterize this cell type's morphology and function, and a further description of its transcriptional identity of these cells. Temporal developmental trajectories are an effective tool for determining the homology of cell types due to their ability to reveal conserved developmental patterns, transitional states, and key regulatory mechanisms across different species. By examining how cells progress from common progenitors to mature forms, researchers can identify shared gene expression profiles, signaling pathways, and morphological changes that indicate evolutionary relatedness. These trajectories also unveil transitional cell states that might not be evident in mature cell types but provide critical intermediate links that allow us to identify unifying features between species. Additionally, the identification of master regulatory genes that govern cell fate decisions along these trajectories can showcase functional conservation despite distinct morphologies or even location. Ultimately, developmental trajectories, such as those identified in Cnidarians through single cell RNA-seq offer a more comprehensive view of cellular differentiation, enabling us to differentiate the homology of cell types and their evolutionary relationships.

CONCLUSION

The comprehensive exploration of homology across various levels of biological organization underscores its fundamental importance in modern biological and medical research. As Wake (1994) eloquently states, the assertion that findings from diverse organisms, such as roundworms, fruit flies, frogs, mice, or chimps, hold relevance for human conditions relies on the bold assumption of homology. This assumption is built upon the belief that shared traits reflect shared ancestry and that insights gained from one organism can be extended to another. However, the simplicity of the definition of homology as a shared trait due to common ancestry can be deceiving. Shared traits can result from convergent evolution, where superficially similar traits evolve independently in different lineages due to similar ecological pressures, rather than from a shared ancestor. Thus, discerning genuine homology from homoplasy requires meticulous contextualization.

In the realm of cell identity, challenges emerge when examining lineages that exist outside of an overlying organ or "regional identity manifest", as illustrated by the example of cnidocytes and their diverse forms across cnidarians. The application of phylogenetic methods and ancestral state reconstruction offers an alternative avenue to differentiate between convergence and shared ancestry. By scrutinizing larger datasets and integrating diverse data

points, these techniques provide a robust framework to assess homology that does not rely solely on morphological comparison, but rather incorporates layers of additional information to improve statistical assessment of potential homology.

Similarly, investigations into genetic responses illuminate the limitations of traditional differential gene expression analyses in discerning genuine homology. While structural homology can be established by comparing position and developmental pathways, the distinction becomes murkier when assessing non-structural traits like genetic responses. Genetic homology is traditionally explored through differential gene expression analyses, yet this approach faces challenges in interpreting conservation due to ecological, physiological, and technical variations. A critical insight arises from the realization that homology isn't solely determined by shared genetic sequences, but by the broader gene regulatory networks that orchestrate gene interactions. The intricate interplay of physiological differences, evolutionary changes, and complex regulatory networks necessitates a shift towards assessing homology through network architecture rather than isolated gene expression. This understanding propels the need to focus on network architecture when assessing homology, as shared regulatory mechanisms and signaling pathways may persist despite individual gene changes.

As demonstrated through my dissertation, defining homology is an intricate endeavor that requires careful consideration of context and level of organization. Whether in structural traits, genetic responses, or cell identities, the conventional methods of assessing homology may occasionally fall short. The integration of multiple layers of information, from phylogenetics to gene regulatory networks, and the focus on contextualizing information beyond the current morphology of a trait provides a more holistic understanding of homology. By delving beyond superficial similarity and accounting for the dynamic nature of evolutionary connections, the scientific community can refine its assertions about homology, ensuring that the foundation upon which biological research rests remains robust and nuanced.

Chapter 2

The evolution of cnidarian stinging cells supports a Precambrian radiation of animal predators

Cnidarians—the phylum including sea anemones, corals, jellyfish, and hydroids—are one of the oldest groups of predatory animals. Nearly all cnidarians are carnivores that use stinging cells called cnidocytes to ensnare and/or envenom their prey. However, there is considerable diversity in cnidocyte form and function. Tracing the evolutionary history of cnidocytes may therefore provide a proxy for early animal feeding strategies. In this study we generated a time-calibrated molecular clock of cnidarians and performed ancestral state reconstruction on twelve cnidocyte types to test the hypothesis that the original cnidocyte was involved in prey capture. We conclude that the first cnidarians had only the simplest and least specialized cnidocyte type (the isorhiza). A rapid diversification of specialized cnidocytes occurred through the Ediacaran (~654-574 mya), with major subgroups developing unique sets of cnidocytes to match their distinct feeding styles. These results are robust to changes in the molecular clock model and are consistent with growing evidence for an Ediacaran diversification of animals. Our work also provides insight into the evolution of this complex cell type, suggesting that convergence of forms is rare and typically restricted to the loss of structures.

INTRODUCTION

Cnidarians are one of the oldest living groups of carnivorous animals. With over 10,000

described species, cnidarians demonstrate a dizzying array of morphologies, ecological roles,

and life history strategies, yet they are unified by the presence of a unique cell type called the

cnidocyte (Figure 1). Cnidocytes are colloquially called "stinging cells," as they allow

cnidarians to deliver their infamous painful, venomous stings. Cnidocytes are dominated by a

single, pressurized organelle called a capsule, which houses a coiled, hollow tubule. When

triggered, the tubule can be discharged from the cell at microsecond-timescales with a force

greater than 500,000,000g (Nüchter et al., 2006). This force allows the tubule to penetrate the

tissue of other animals, and many cnidocyte tubules are barbed and/or associated with toxins.

Nearly all living cnidarians use cnidocytes for prey capture, generally feeding on zooplankton

or larger animals. The complexity of cnidocytes and their ubiquity across the Cnidaria

strongly suggests that their last common ancestor also harbored cnidocytes. Yet, molecular

clocks and fossil data (Liu et al., 2014; Iten et al., 2016) suggest that the first cnidarians evolved

in the Neoproterozoic, well before the diversification of the bilaterian animals they currently

feed on (Erwin et al., 2011; Iten et al., 2016). This begs the question; what were the earliest

cnidarians doing with stinging cells if there was no prey to sting?

21
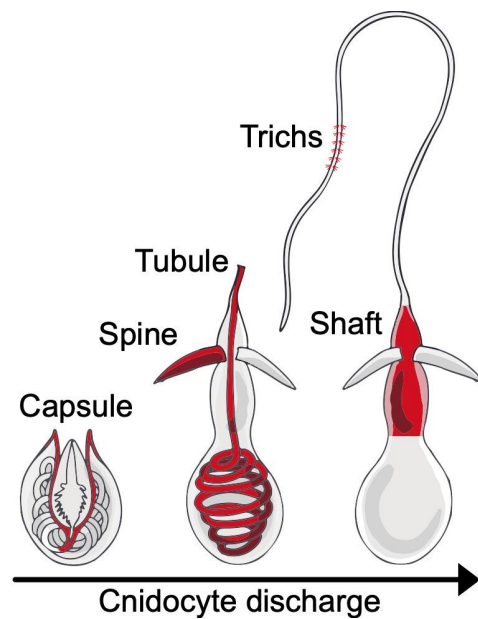
Figure 2.1: **The general structure of a cnidocyte cell, based on the stenotele form.** The images from left to right demonstrate the process of cnidocyte discharge (or firing). Components of the cell discussed in this paper are labeled and colored red.

Although cnidocytes are best known for ensnaring and/or envenomating prey (Kass-Simon and Scappaticci, 2002; Boillon et al., 2004; Damian-Serrano et al., 2021), they come in a variety of forms with distinct functions. This includes substrate attachment (Siddall et al., 1995), construction of tube-dwellings (Mariscal et al., 1977a), self-defense / intraspecific competition (Kass-Simon and Scappaticci, 2002), and even mating (Garm et al., 2015). Reflecting these different functions, cnidocytes are extremely diverse, with over 30 "types" currently described. Historically, cnidocytes have been organized based on their morphology and used as a tool for

cladistics (Östman, 2000; Boillon et al., 2004; Gershwin, 2006). Classification systems of

cnidocytes have varied across studies as new types are identified and microscopy techniques

improve, allowing for nuanced discussions of the location, size and shape of barbs (or trichs),

the position of swellings along the shaft, and the pleating within the capsule (Mariscal et al.,

1977b). However, the inconsistency of classification schemes and descriptions makes

understanding the relationships of cnidocyte types a challenge. Most cnidarian species have

more than one type of cnidocyte, and their repertoire—known as the cnidome—can change

over their lifetime, often in association with diet and ecological niche (Purcell, 1984; Carrette

et al., 2002). Still, this diversity of cnidocyte form and function provides a potentially useful

tool; by reconstructing the evolutionary history of this cell type we should be able to constrain

its original function and determine when new ecological capabilities evolved in Cnidaria.

RESULTS

The results of our analyses are summarized in Figure 2 and Tables S4-S5. For clarity, the first

time we refer to a taxonomic clade in the following text, we include the names of two species

in our tree whose last common ancestor defines the group. We also include the date of that

last common ancestor based on our molecular clock, resulting in the following format:

"taxonomic clade" ("species 1" + "species 2"; "95% confidence interval on ancestral node in

'mya', or millions of years ago"). We chose P=0.7 as a semi-arbitrary probability cutoff. This

balances our desire for confidence with the observation that uncertainty naturally increases at

deeper nodes of the tree, so demanding a higher probability cutoff results in an

improbably-high number of independent cnidocyte gains. Detailed output for each ancestral

state reconstruction can be found Supplementary Figures S1-S13.
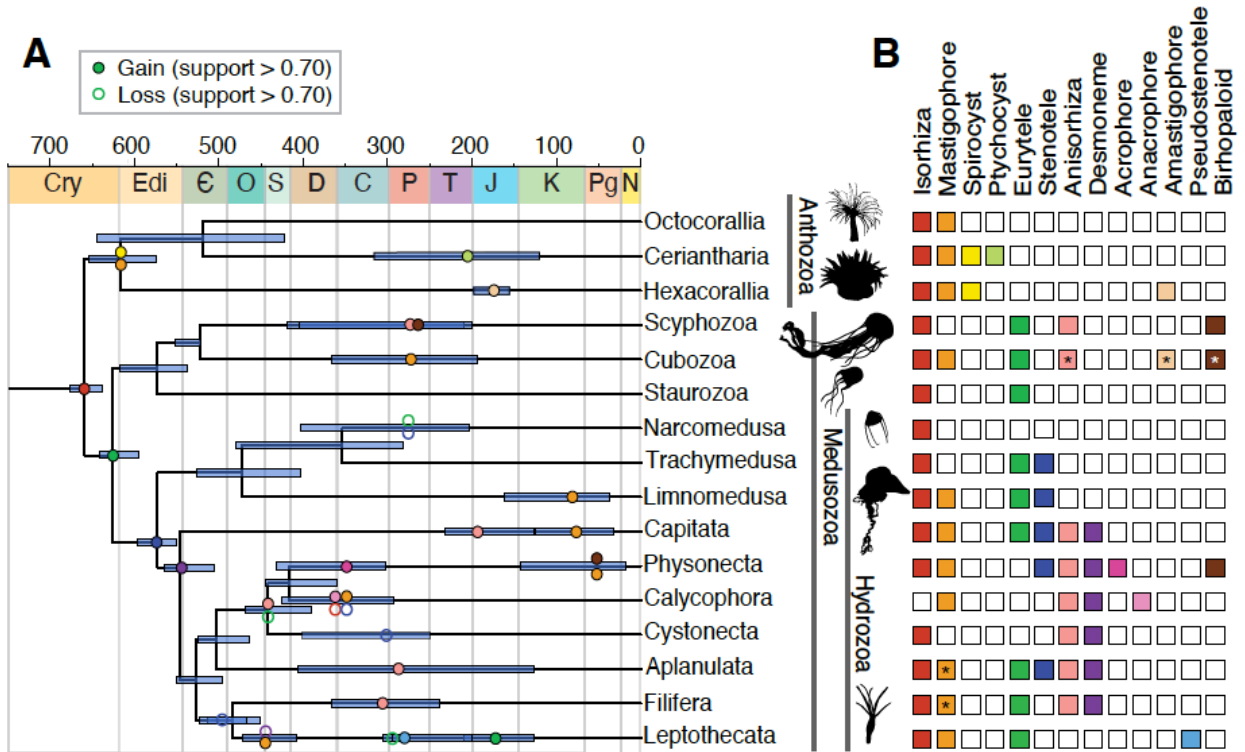
Figure 2.2: Graphic summary of cnidocyte analysis. (**A**) A molecular clock of cnidarians, with colored circles at nodes noting the timing of important gains and losses of cnidocyte types. The color of each circle signifies cnidocyte types, as detailed in section (**B**). Purple bars represent 95% confidence intervals for the divergence time of the represented node. Abbreviations: "Cry" Cryogenian; "Edi" Ediacaran; "Ͼ" Cambrian; "O" Ordovician; "S" Silurian; "D" Devonian; "C" Carboniferous; "P" Permian; "Tr" Triassic; "J" Jurassic; "K" Cretaceous; "Pe" Paleogene; "N" Neogene. (**B**) The distribution of cnidocyte types in

the major lineages of cnidarians visualized as a presence/absence matrix. Boxes that contain asterisks indicate that the cnidocyte type was too rare in our dataset to determine the timing of its origin.

Cnidarian-wide Cnidocytes

We recovered three types of cnidocytes that were present across the major cnidarian clades, but only one type—the isorhiza—could be identified in the last common ancestor with statistical confidence (P ≥ 0.7).

Isorhiza: Isorhiza are the morphologically simplest and most ubiquitous cnidocytes, characterized by an isodiametric thread and no shaft. In our analysis, isorhiza were the only cnidocyte with statistical support for being present (P=0.70) in the last common ancestor of Cnidaria (*Abietinaria filicula + Abyssoprimnoa gemina*; 676-637 mya).

Mastigophores: Mastigophores are characterized by a shaft, or a cylindrical enlargement of the thread proximal to the capsule (see Figure 1; (Carlgren, 1940; Mariscal, 1974; Tardent, 1995; Östman, 2000)). Mastigophores are ubiquitous within the sampled Anthozoa (*Abyssoprimnoa gemina + Epizoanthus illoricatus*; 654-574 mya), and broadly present in Hydrozoa (*Aegina citrea + Abietinaria filicula*; 597-552 mya), particularly Leptothecata (*Lafoea dumosa + Abietina filicula*; 471-408 mya). Mastigophores have also been described in some Cubozoa (*Alatina moseri + Chiropsalmus quadrumanus*; 467-409 mya). This cell type is absent from all Scyphozoa (*Catostylus*

mosaicus + Nausithoe atlantica; 546-429 mya) and Staurozoa (*Calvadosia campanulata + Manania uchidai*; 486-374 mya) represented in this study. Our analysis did not provide strong support for mastigophores in the last common ancestor of Cnidaria (P=0.47), suggesting they evolved independently in Anthozoa, Cubozoa and Hydrozoa. The first well-supported appearance of mastigophores is in the last common ancestor of Anthozoa (P=0.82) sometime between the end-Cryogenian and early Ediacaran. The earliest well-supported appearance of mastigophores within the Cubozoa is in the Chirodropida (*Chironex fleckeri + Chiropsalmus quadrumanus*; 365-192 mya); in Hydrozoa, the first well-supported appearance is in the calycophoran siphonophores (*Chuniphyes multidentata + Praya dubia*; 423-290 mya).

Amastigophores: The amastigophore possesses a cylindrical and pointed shaft similar to that of the mastigophore, but it lacks a tubule. While a small tubule may be visible in the undischarged cnidocyte (Cutress, 1955; Östman, 2000) upon discharge the distal tubule breaks off at or close to the shaft (Östman, 2000) so that no distal tubule (Carlgren, 1940; Mariscal, 1974; Östman, 2000) or a very short one (Fautin, 2009) is visible after discharge. These cnidocytes are present in several species of Cubozoa and Anthozoa. Our analyses strongly suggest that amastigophores evolved convergently in both classes, as the probability of their presence at their last common ancestor is P=0.00.

Anthozoan-specific Cnidocytes

Anthozoans (corals and anemones) lack the swimming medusa stage found in their sister clade, the Medusozoa (**Figure 2**). In our analysis two cnidocyte types were exclusive to anthozoans, the spirocyst and ptychocyst.

Spirocysts: Spirocysts are adhesive cnidocytes lacking spines which upon eversion create a web of fine, adhesive fibrils (Mariscal et al., 1977b). They are nearly ubiquitous among the Hexacorallia (*Acanthopathes thyoides + Epizoanthus illoricatus*; 589-513 mya) and Ceriantharia (*Ceriantheomorphe brasiliensis + Isarachnanthus nocturnus*; 314-117 mya). Our results provide strong support for spirocysts being present in the last common ancestor of Anthozoa (P=0.90), and little support for their presence in the last common ancestor of Cnidaria (P=0.01). Spirocysts have not been described in anthozoan octocorals (*Abyssoprimnoa gemina + Ellisella schmitti*; 508-208 mya), which has led some to hypothesize the cell type was a hexacoral novelty (Rifkin, 1991). Our tree—which follows the topology of Picciani *et al.* (Picciani et al., 2018)—places the Ceriantharia as sister to Octocorallia, and as a result our analysis prefers an early anthozoan origin for spirocysts. If this topology is incorrect, then spirocysts would have evolved slightly later, in the last common ancestor of Hexacorallia instead of Anthozoa.

Ptychocysts: Ptychocysts are used in construction by tube-dwelling anemones of the subclass Ceriantharia, and are characterized by a pleated, rather than helical, organization of the tubule thread. First described by Carlgren and Schmidt as "atrichs" (atrichous isorhiza) (Carlgren, 1912, 1940; Schmidt, 1974), they were eventually distinguished by Mariscal, Conklin & Bigger (Mariscal et al., 1977a) as a separate type due to the visible folds in the length of the tubule which likely help to intertwine the discharged threads forming a tightly woven mesh around the anemone. Our results support the hypothesis that ptychocysts are a novelty of the Ceriantharia (*Cerianthus membranaceus + Ceriantheomorphe brasiliensis*; 256-117 mya).

Medusozoan-specific Cnidocytes

Unlike anthozoans, many medusozoans (jellyfish, hydroids, and siphonophores) contain a swimming, sexually reproductive medusa life stage. The prevailing hypothesis is that the earliest cnidarians lacked a medusa stage (Kayal et al., 2018) although research on genomes and gene expression patterns in some medusozoans may challenge this hypothesis (Gold et al., 2019a; Leclère et al., 2019). Regardless, the medusa life stage is associated with a diversification of feeding strategies, and thus a greater range of cnidocyte types.

Euryteles: Gershwin (Gershwin, 2006) notes that the name "eurytele" has been interchangeably used to represent either cnidocytes with a single swelling along the shaft, or a

distal swelling on the shaft (this despite the Greek etymology: *eury* = widened *tele* = distal).

While it is possible that these shifting definitions indicate a conflation of multiple cnidocyte

types, the absence of images in some relevant scientific papers prevented us from

distinguishing subtypes. Under the assumption that all described euryteles are homologous,

our results strongly support the hypothesis that euryteles are a medusozoan novelty, with a

high degree of support for its presence in the last common ancestor of Medusozoa (P=0.99) and

very low support for its presence in the last common ancestor of Cnidaria (P=0.02) We

identified several potential instances of re-evolution, such as within Kirchenpauridae

(*Oswaldella stepanjantsae + Kirchenpaueria pinnata*, 212-128mya), where euryteles were most

likely absent in an ancestral clade (*Aglaophenia elongata + Abietinaria filicula*, 307-288mya;

P=0.08) but reappear with significant confidence (P=0.99).

Birhopaloids: Birhopaloids are characterized by the presence of two swellings within the shaft

(Östman, 2000; Gershwin, 2006). These cell types are present in several species of Scyphozoa,

hydrozoan siphonophores, and one cubozoan in our dataset. Our analysis rejects the

hypothesis that birhopaloids were present in the ancestor of Medusozoa (P=0.00), suggesting

this cell type evolved independently in each clade. The earliest well-supported appearance of

birhopaloids in Scyphozoa is within clade Rhizostomeae (*Stomolophus meleagris + Catostylus*

mosaicus, 418-208mya); in Hydrozoa its earliest appearance is in the genus *Apolemia* (*Apolemia lanosa + Apolemia rubriversa*, 145-18mya).

Stenoteles: Stenoteles are an iconic cnidocyte featuring prominent armature and are capable of envenomation. This cnidocyte type possesses both a long thread as well as a shaft of similar length to the capsule divided by a central constriction. It is distinguished from other cnidocytes by the sizable anchor-like spines known as stylets along the proximal end of the shaft (Tardent, 1995). Östman (Östman, 2000) considered stenoteles a Hydrozoan novelty, and our results lend strong support (P=0.91) to this hypothesis. We found evidence of multiple losses of stenoteles within the Hydrozoa, including the clade encompassing Leptothecata and Filifera (*Abietinaria filicula + Brinckmannia hexactinellidophila*; 545-481 mya), multiple families within the Limnomedusae (*Aglauropsis aeora + Olindias sambaquiensis*; 421-287 mya) and Narcomedusae (*Aeginura grimaldii + Aegina citrea*; 375-201 mya). There are isolated reports of stenoteles outside of the Hydrozoa, but these papers either lack images of this cell type (Carrette et al., 2014) or the images they provide lack the necessary stylets (Cengiz and Killi, 2021). There is little support for stenoteles in the last common ancestor of Medusozoa (P=0.13), suggesting any similarity between "stenoteles'' in hydrozoans, scyphozoans and cubozoans (assuming they exist at all in the latter two) is a case of convergent evolution.

Anisorhiza: Very similar in morphology to isorhiza, the anisorhiza has a non-isodiametric thread with a slight taper from capsule to tip. This cnidocyte is present sporadically in the Medusozoa, and our analyses suggest it may have independently evolved anywhere from 5-9 times. The earliest appearance of this cell type within the Medusozoa is in the Siphonophora (*Agalma clausi + Physalia physalis*; 469-390 mya).

Pseudostenoteles: Pseudostenoteles are differentiated from stenoteles by the lack of a prominent constriction in the shaft, which may taper more gradually toward the thread (Bouillon et al., 1986). Like stenoteles, pseudostenoteles generally possess three sizable spines, however they may display additional rows of spines along the shaft body. Our data suggests that no species possesses both stenoteles and pseudostenoteles, pseudostenoteles being found exclusively within the Leptothecata, where stenoteles have been lost. The earliest appearance within the Leptothecata is in the genus *Halecium* (*Halecium mediterraneum + Halecium beanii*, 290-199 mya).

Rhopalonemes (acrophores and anacrophores): Acrophores and anacrophores are thought to be siphonophore-specific novelties that function as adhesive "clubs'' for catching prey (Boillon et al., 2004; Damian-Serrano et al., 2021). The shaft is much larger in volume than the capsule, and acrophores and anacrophores can be distinguished by the presence of a small thread at the

top of the shaft in the former. Our analysis confirms previous results from Damian-Serrano et al. (Damian-Serrano et al., 2021), with the acrophore subtype present in the Physonectae (*Marrus orthocanna* + *Agalma elegans*; 434-299 mya) and the anacrophore subtype (no apical thread) present in the Calycophorae (*Praya dubia* + *Lensia conoidea*; 411-282 mya). Both cnidocytes appear to be highly conserved, with little evidence of loss in our dataset.

Desmonemes: Desmonemes possess a prominent thread, which coils into a corkscrew formation at discharge that has been observed to ensnare (Carlgren, 1912; Purcell, 1984) and/or adhere to prey (Damian-Serrano et al., 2021). In our analysis, desmonemes are unique to the Hydrozoan subclass Hydroidolina (*Asyncoryne ryniensis* + *Abietinaria filicula*; 566-508 mya) with extremely strong support for the cell type being present in the last common ancestor (P=1). Desmonemes are broadly conserved in Hydroidolina except for a few notable losses including within Leptothecata, the genus *Eudendrium* (*Eudendrium racemosum* + *Eudendrium californicum*; 352-124 mya) as well as a subset of Capitata (*Cladocoryne floccosa* + *Asyncoryne ryniensis*; 244-162 mya). Our analysis supports one case of convergent evolution of desmonemes in the thecate hydroid *Amphinema dinema*, but we note the placement of this species in our tree (within the Leptothecata) disagrees with traditional taxonomy.

A more conservative molecular clock corroborates the Ediacaran radiation of cnidocytes

The molecular clock that our analysis is based on puts the origin of cnidarians between ~655.6 and 730 Mya. Some paleontologists might express reservations with this estimate, given the gap between it and the oldest known cnidarian fossils, as well as the paucity of clear eumetazoan fossils in the early-Ediacaran Lantian and Doushantuo biotas (Benton et al., 2015). We therefore performed a second molecular clock analysis using a more conservative model for the rate of evolution, as well as hard lower bounds for the age of cnidarians, bilaterians, and animals based on recommendations from the Fossil Calibration Database (Benton et al., 2015). The results of this clock are provided in Figure 3. Despite substantially younger fossil calibrations, the genetic data strongly supports older origins for the Anthozoa and Medusozoa, with the ancestral Cnidaria pressing against the hard lower bound. While the timing of animal radiations is notably truncated in this scenario, the timing of early cnidocyte evolution is unchanged. We still recover isorhiza, mastigophores, euryteles, and stenoteles evolving in the Ediacaran. Desmonemes may have also evolved in the late Ediacaran, although, like the former analysis, most of the 95% confidence interval rests in the Cambrian. This suggests our major finding–the Precambrian radiation of cnidocyte forms–is robust to variations in the molecular clock analysis.
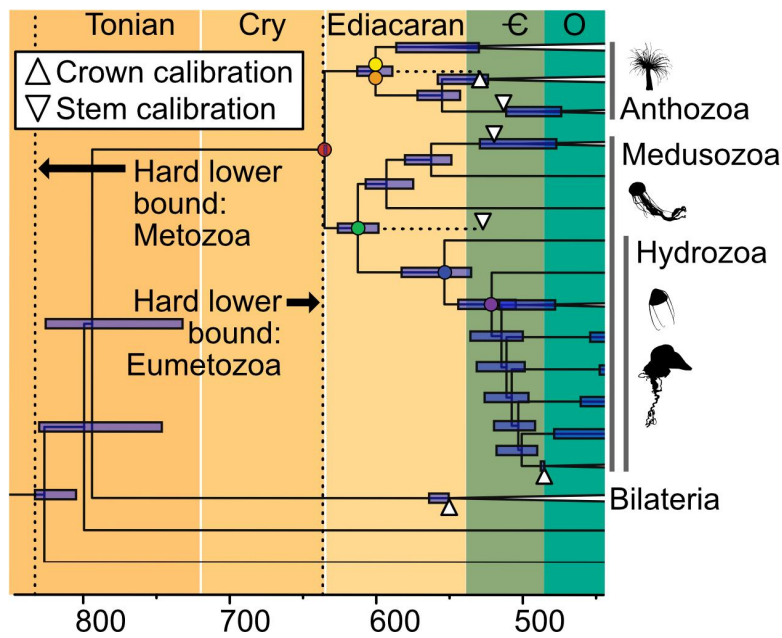
Figure 2.3: The results of a conservative molecular clock analysis, with a focus on the Neoproterozoic through Ordocivian. The placement of fossil calibrations is noted with triangles. Abbreviations: "Cry" Cryogenian; "Ⲥ" Cambrian; "O" Ordovician.

DISCUSSION

The goal of this paper is to reconstruct the evolution of stinging cells in the Cnidaria, testing the hypothesis that the first cnidocytes were used for the capture of animal prey. Looking at living cnidarians the answer seems obvious; nearly all species contain cnidocytes and nearly all of them use cnidocytes to capture animal prey. However, cnidarians are thought to have evolved before the

diversification of most major animal lineages. There is some support for this hypothesis in the fossil record; while most animal phyla make their first, unambiguous appearance in the Cambrian (~541–485 Ma), the medusozoan *Auroralumina attenboroughii* is found in the late Ediacaran (~557–562 mya) with a variety of other putative cnidarian fossils stretching back to ~580 mya (Chen et al., 2000; Liu et al., 2014; Dunn et al., 2022). Additional evidence comes from molecular clock studies, which consistently place crown group cnidarians in the Ediacaran or earlier (Erwin et al., 2011; dos Reis et al., 2015; Gold et al., 2015; Dohrmann and Wörheide, 2017). Our molecular clock similarly places the origin of crown Cnidaria in the late Cryogenian, between 676-637 mya. This temporal discrepancy between the first cnidarians and the diversification of their prey was the impetus of this study.

Using ancestral state reconstruction, our study supports the hypothesis that the earliest, Cryogenian-age cnidarians contained isorhiza—the least specialized and morphologically simplest cnidocyte. It is quite possible that the first cnidocytes were only minimally involved in prey capture; modern isorhiza serve a diverse set of non-predatory functions, including intraspecific competition (Kass-Simon and Scappaticci, 2002), defense (Bouillon et al., 1986) and locomotion (Ewer and Fox, 1947). Competition for space may have been a critical selective force for early cnidarians. The earliest known fossil communities of large organisms—which includes putative cnidarians and sea sponges—were benthic and competition for space appears to have been an important driver of their morphology and distribution (Sperling et al., 2011; Ghisalberti et al., 2014; Dunn et al., 2021).

Our results also support the rapid appearance of predatory cnidocytes through the Ediacaran, including anthozoan spirocysts and mastigophores, medusozoan euryteles, and hydrozoan stenoteles and desmonemes. The expansion of cnidocyte types is consistent with current differences in ecology and feeding behavior within the major groups. In living anthozoan hexacorals, spirocysts and mastigophores are often found together in specialized feeding tentacles, with the adhesive spirocysts ensnaring the prey (Mariscal et al., 1977b) and the barbed mastigophores delivering toxins (Kass-Simon and Scappaticci, 2002). This is distinct from sweeper tentacles found in some hexacorals, which are specialized for inter- and intra-species competition and are composed largely of isorhiza (Kass-Simon and Scappaticci, 2002). Feeding tentacles allow anthozoans to catch large animals and is one of the adaptations that permit the large size of sea anemones compared to other cnidarian polyps. The rise of large anthozoans in the Ediacaran with specialized cnidocytes used in modern taxa for larger prey is consistent with a growing body of evidence that mobile animals were diversifying at this time, at least on the seafloor (Parry et al., 2017; Bobrovskiy et al., 2018; Runnegar, 2022).

In contrast, the polyps of medusozoans are generally small and only specialize in large prey during the swimming medusa life stage. Euryteles, which likely evolved in the first medusozoans, are specialized for cracking animal exoskeletons (David et al., 2008). Hydrozoans—the most numerous and diverse clade in the Medusozoa— evolved the desmonemes and stenoteles quickly (within ~250my), which like anthozoan spirocysts and mastigophores combine an adhesive cnidocyte type with a penetrating type

(Kass-Simon and Scappaticci, 2002). In *Hydra*, desmonemes work together with stenoteles (Purcell, 1984; Kass-Simon and Scappaticci, 2002) by lassoing prey; the mechanical action of the struggling prey then triggers the discharge of venomous stenoteles (Kass-Simon and Scappaticci, 2002). Tardent (Tardent, 1995) details the penetrating function of stenoteles, which crack a hole in the exoskeletons of prey and are then ejected by the unfurling of the heavy stylets (spines) of the distal shaft. This ejection allows the remainder of the tubule to unfurl into the hole to deliver venom through the open tubule tip. While the fossil record provides little evidence for pelagic animals prior to the Cambrian (Gold, 2018), molecular work suggests that swimming euarthropod larvae were likely present in the Ediacaran (Wolfe, 2017). This coevolution of exoskeleton-cracking cnidocytes and a pelagic life stage in early Medusozoa supports the hypothesis that a radiation of swimming bilaterian prey had begun by the Ediacaran.

In addition to Earth history considerations, this study provides insight into cell type evolution in animals. Evidence from comparative genetics and developmental biology suggests that cnidogenesis is an offshoot of neurogenesis (Tardent, 1995; Babonis and Martindale, 2014; Sebé-Pedrós et al., 2018; Siebert et al., 2019; Babonis et al., 2022). In many cnidarians multiple cell lines are capable of transdifferentiating into cnidocytes, though in the hydrozoans cnidogenesis became increasingly restricted to set-aside interstitial stem cells (Denker et al., 2008; Beckmann and Özbek, 2012; Babonis and Martindale, 2014; Gold et al., 2019b). Our analyses are consistent with the current hypothesis that

all cnidocyte types can be traced to a single ancestral cell type, and most major types evolved once. Those cnidocytes that do evolve multiple times are generally defined by their lack of trait found in more complex cnidocytes (e.g., anisorhiza, amastigophores). The exception to this appears to be the mastigophore and the birhopaloids, where the presence of type-defining swellings of the shaft appear simple enough to have evolved multiple times. Elucidating how this original cell evolved to display so many nuanced and diverse types has the potential to inform us broadly how specific organelles or complex cell types might evolve and diversify over deep time.

CONCLUSIONS

By reconstructing the evolutionary history of cnidarian stinging cells, we present a hypothesis regarding the Precambrian rise of animal predation. The earliest, Cryogenian-age cnidarians were likely small, benthic animals (Kayal et al., 2018) that used a simple cnidocyte type (the isorhiza) to compete for space on the ocean floor and perhaps capture large particulate matter. As animals diversified and increased in size during the Ediacaran, cnidarians took two distinct approaches to taking advantage of this food source. Anthozoans became larger and developed the spirocysts and mastigophores, which are combined in the feeding tentacles of many sea anemones. Medusozoans developed a swimming life stage with a different set of specialized cnidocytes—the euryteles, desmonemes and stenoteles—which allow for the ensnaring of prey and cracking of exoskeletons.

This allowed for specialization on swimming arthropods and other pelagic prey. Our results support the hypothesis that animal diversification has a Precambrian origin, and suggest the cnidaria had a significant impact on the early evolution and trajectory of food webs and ecological structure.

MATERIALS AND METHODS

Tree Selection

Analyses were performed on a fossil-calibrated molecular clock based on a gene tree sourced from Picciani et al. 2018 (file:

https://github.com/npicciani/picciani_et_al_2018/blob/master/Analyses/10_Time_calibration/tree_cnid_635.tre). This tree was selected because of high species coverage (1106 cnidarians) and the availability of associated genetic data (a minimum of 2 genes for each species on the tree). Some aspects of the topology were manually corrected using Mesquite v.3.51 to reflect current knowledge about higher taxonomic relationships. Specifically, the tree was manually edited to change the relationships within medusozoa to (Hydrozoa, (Staurozoa, (Cubozoa, Scyphozoa))) based on (Zapata et al., 2015; Kayal et al., 2018) and outgroup relationships were adjusted to (Porifera,(Placozoa, (Cnidaria, (Deuterostomia, Protostomia)))) based on (Laumer et al., 2019). Because cnidomes can vary even between sister species, undescriptive species (such as species names "spp") were pruned from the tree. This species list was used to query the WoRMS

database (accessed on 10-29-2021) to update species nomenclature, resulting in 904 cnidarian species plus 4 outgroups.

Cnidome Data Collection

The cnidomes of each species were determined from literature searches, with a focus on the 904 species in the starting tree. The results of this search are available in Table S1. When conflicting cnidome data occurred in the literature, we prioritized (1) multiple separate observations over stand-alone observations, (2) cnidome surveys over individual species reports and (3) microscopy-based assignments over drawings and/or written descriptions. We ultimately recovered cnidocyte data for 477 species, and the data was converted to binary [absence=0, presence=1] format for each cnidocyte type. While as many as 30 cnidocyte types have been described (Mariscal, 1974; Tardent, 1995) we restricted our analyses to thirteen cnidocyte types that occurred in more than three species in our dataset, and whose definition has been consistent over the last century. We also excluded tumiteles, mesoteles, and trirhopaloids from our analysis, as they did not appear in a sufficient number of species on our tree for meaningful ancestral state reconstruction.

Molecular clock analysis

The molecular clock was produced using BEAST v1.10.4 (Suchard et al., 2018). Genetic data

was downloaded from the GitHub account associated with Picciani et al. (Picciani et al., 2018)

(file: "3_processed_alignments.fa"). Species names were adjusted to match changes to the tree,

and NCBI accession numbers were removed from headers. All 908 species in the pruned tree

were used for this analysis. Although only 477 species contained cnidocyte data (and therefore

could be used in the ancestral state reconstruction) we included all taxa for the molecular

clock to provide more data and ideally improve resolution. IQTree v1.6.12 (Kalyaanamoorthy et

al., 2017; Minh et al., 2020) was run on each separate gene alignment to determine the optimal

substitution model for each gene. Genes were then concatenated into a single supermatrix for

use with BEAST.

Calibrations for 7 nodes were set based on data from the fossil record (Table S2) (Johnson and

Richardson, 1968; Fedonkin and Waggoner, 1997; Ezaki, 2000; Baron-Szabo et al., 2006; Fuller

and Jenkins, 2007; Han et al., 2010, 2016; Jell et al., 2011; Dong et al., 2013). Age ranges for the

calibrated nodes were modeled under a lognormal distribution, using a lower (younger) bound

from the fossil record and an upper (older) soft bound based on a lognormal probability curve.

A starting chronogram consistent with these calibrations was generated using the starttree

script (included in the GitHub project folder). A Yule model of speciation was chosen given the incompleteness of species represented in the tree. For the Bayesian analysis, two chains were run for 10,000,000 generations, logging the results every 1,000 generations. A consensus chronogram from the two runs was produced with the TreeAnnotator application packaged with BEAST, using a 25% burnin and mean node heights. The XML file used to run BEAST is provided on GitHub.

A second, more conservative molecular clock was generated from the same dataset. In this setup, hard lower bounds were set for the age of bilateria (636.1 mya), Cnidaria (636.1 mya), and animals (833 mya) based on the Fossil Calibration Database (Benton et al., 2015). Any prior distributions for fossil calibrations that exceeded these dates were converted into uniform priors constrained between the age of the fossil and the hard lower bounds. A local clock model was used instead of a relaxed lognormal clock, which required an unlinking of the clock and site models for each gene in the dataset. The results from this second analysis produced date estimates for Hydrozoa that were inconsistent with the fossil record. We therefore re-ran the clock, adding an additional calibration for the Leptothecata using the upper-Cambrian hydroid *Palaeodiphasia* (Song et al., 2021). After the additional calibration was added, we ran the clock in triplicate, with each run including two chains and 10,000,000 generations, logging the results every 1,000 generations. The results were combined using the LogCombiner

application packaged with BEAST, with a 25% burnin for each dataset. A consensus tree was generated using the TreeAnnotator application packaged with BEAST, using mean node heights.

Ancestral State Reconstruction

The chronogram from BEAST was pruned to the 477 species with cnidocyte data using phytools v0.7.80 (Revell, 2014). Model testing was performed using the pruned chronogram and binary character data for each cnidocyte using the R package corHMM v2.7 (Beaulieu et al., 2020). For each cnidocyte type we compared four substitution models: (1) an equal-rates model (ER) with 1 hidden rate category (2) an ER model with 2 hidden rate categories (3) an all-rates-different model (ARD) with 1 hidden rate category, and (4) an ARD model with 2 hidden rate categories (for details on hidden rates models, see (Boyko and Beaulieu, 2021)). The resulting rate matrices are provided in Table S3. Model preference was assessed using AIC criteria. The rate matrix for each model (ER1cat, ER2cat, ARD1cat, ARD2cat) was created following the corHMMv2.1 vignette on CRAN (https://cran.r-project.org/web/packages/corHMM/) and provided in Table S4.

Stochastic mapping was used to simulate evolutionary histories along the time-calibrated phylogeny, using the preferred model of evolution for each analysis as predicted by corHMM.

The specific data for each model (rates, stationary frequencies, etc) for each cnidocyte was

retrieved from the corHMM output and set as the model input for stochastic mapping using

corHMMv2.7::makeSimmap(). The character data was formatted for input with the script

0_data_cleanup.Rmd. The root prior of the tree was initialized as 0 (absent) for each cnidocyte,

under the assumption that cnidocytes were not present in the last common ancestor of

animals. The analysis was run with 10,000 histories. The stochastic mapping results were then

converted to phytools format using functions available from thackl/thacklr

(https://rdrr.io/github/thackl/thacklr/src/R/phylo.R) and summarized using

phytools::describe.simmap(). Plot summaries were generated for the internal node results

(sim.obj$ace results) demonstrating the frequency at which each node was in each state

(presence or absence, within hmm1 or hmm2 in relevant 2category hmm cases) and producing

a density map demonstrating the frequency of transitions within the branch length.

Data and Resource Availability: Starting data, scripts and intermediate files used in this

analysis can be found in https://github.com/DavidGoldLab/2023_Cnidocyte_Ancestral_State.

Information gathered from cnidome literature review and related files are also provided in

Table S1.

Chapter 3

A novel approach to comparative RNA-Seq does not support a conserved set of orthologs underlying animal regeneration

Molecular studies of animal regeneration typically focus on conserved genes and signaling pathways that underlie morphogenesis. To date, a holistic analysis of gene expression across animals has not been attempted, as it presents a suite of problems related to differences in experimental design and gene homology. By combining orthology analyses with a novel statistical method for testing gene enrichment across large datasets, we can test whether some fundamental processes common across animals share transcriptional regulation. We applied this method to a meta-analysis of 6 publicly-available RNA-Seq datasets from diverse examples of animal regeneration. We recovered 160 conserved orthologous gene clusters, which are enriched in structural genes as opposed to those regulating morphogenesis. A breakdown of gene presence/absence provides limited support for the conservation of pathways typically implicated in regeneration, such as Wnt signaling and cell pluripotency pathways. Such pathways are only conserved if we permit large amounts of paralog switching through evolution. Overall, our analysis does not support the hypothesis that a shared set of ancestral genes underlie regeneration mechanisms in animals. After applying the same method to heat shock studies and getting similar results, we raise broader questions about the ability of comparative RNA-Seq to reveal evolutionarily conserved gene pathways.

INTRODUCTION

Why regeneration occurs in some animals and not others remains an enigma in biology. It is well known that certain groups can readily regenerate lost tissues and body parts (e.g. planarian worms, salamanders, cnidarians), while regeneration in others is restricted to specific organs or developmental stages (e.g. nematode worms, insects, mammals). Animals with strong regenerative capabilities are distributed across the evolutionary tree without a clear pattern (Alvarado 2000), and even closely related species can demonstrate dramatically different capacities (Bely and Sikes 2010; Zattara et al. 2019). These observations lead to two competing evolutionary scenarios: body regeneration is either an ancient, conserved animal trait that has been lost to varying degrees across multiple lineages, or it is a derived trait that multiple lineages have converged upon independently. Resolving these competing hypotheses has profound consequences for the goals of comparative regenerative biology: are we searching for unifying principles, or trying to determine how various animals deal with the universal problem of bodily damage?

While many studies focus on putative candidate genes underlying animal regeneration, a growing body of literature challenges any simplistic interpretation. Some genes and pathways

commonly reoccur in studies. Wnt signaling, for example, offers a compelling candidate for a "master regulator" of stem cell dynamics during regeneration (Clevers et al. 2014), as it has been shown to play a critical role in planarian worms (Sikes and Newmark 2013; Umesono et al. 2013), fish (Stoick-Cooper et al. 2007), amphibians (Lin and Slack 2008), and mammals (Bielefeld et al. 2013; Sanges et al. 2013; Takeo et al. 2013). Other studies suggest that key components of regeneration might be dissimilar across major groups. For example, a MARCKS-like protein that initiates limb regeneration in axolotl salamanders appears to be a vertebrate novelty (Sugiura et al. 2016). Regeneration in newts, a different set of amphibians, involves genes not found in the axolotl (Looso et al. 2013). Finally, genes such as the Oct4/POU5F1 "master regulator" of stem cell pluripotency appear absent in invertebrates (Gold et al. 2014). It is unclear whether these observations represent anomalies obfuscating a conserved set of shared genes, or if they hint at the true evolutionary convergence driving animal regeneration.

Whether the molecular mechanisms of regeneration are conserved across animals rests, in part, on what counts as a "conserved" (i.e. homologous) gene. Homologous genes can be subdivided into orthologs—genes related by vertical descent from a common ancestor—and paralogs—genes that arise by duplication events. Orthologs or paralogs may perform similar functions, but in evolutionary biology, common ancestry is what defines conservation. Paralogs by definition cannot be traced back to a single gene in a last common ancestor; this means the

utilization of paralogs by different species during regeneration does not necessarily support the hypothesis of a conserved ancestral function, as it may reflect evolutionary convergence achieved *after* gene duplication. Further complicating this matter, the ortholog/paralog distinction is contingent on the organisms being studied. As more distantly related species are analyzed, families of paralogous genes are often collapsed into a single orthologous clade (see Figure S1 for an illustration of this phenomenon). Tests of molecular conservation therefore require careful consideration of the evolutionary history of genes.

The problem is compounded when using RNA-Seq technology to identify "conserved" genes between distantly related taxa undergoing similar biological processes. The first issue is a biological one: genes rarely share one-to-one homology across distantly related species. An ancestral gene might, over the course of evolution, undergo multiple, lineage-specific rounds of duplication. The second issue is technical: RNA-seq studies have varying temporal resolutions, time-scales, and depths of sequencing. When looking for significant differences in gene expression, these two issues result in a heterogeneous list of statistical tests that are problematic to compare between studies. As an example, imagine a conserved orthologous gene group, where a sea sponge has one gene sampled at three time points, while an axolotl has 5 genes sampled at seven time points. If all time points are compared to each other, this would result in 3 statistical tests for the sea sponge compared to 105 tests for the axolotl.

To address this discrepancy we used a Lancaster P-value aggregation method, which provides a systematic way of collapsing multiple statistical tests for significant differential expression from multiple homologous genes into one value (Yi et al. 2018). This allows us to cluster genes into putative conserved ortholog groups (COG), and then see which COGs are statistically enriched during the regenerative process for each species. The method takes the p-values generated by a differential expression analysis for a group of genes, and essentially treats each as an independent significance test of the hypothesis that the broader COG is differentially expressed. Intuitively it may be the case that no single p-value from a set of independent tests registers as significant, however many borderline-significant values can be aggregated to determine significance. These aggregation methods take advantage of the fact that many independent p-values generated by the null hypothesis should follow a uniform distribution on the interval (0,1). Consequently, we can test the *uniformity* of the set of p-values to determine their likelihood of being generated from the null hypothesis. In other words, the tests of a non-significant COG should create a random distribution of p-values, while a COG with one or more significant components will statistically deviate from this distribution. Mathematically, the appropriate test statistic for uniformity can be computed from the sum of inverse cumulative distribution function with p-values and raw read counts as inputs (See Yi et al. 2018 for details, and Additional File 1, part 0.5 for Python code). The result of this process is a table with entries corresponding to taxon-COG pairs, and an associated aggregated p-value for each COG. Figure

S2 illustrates our approach to applying the Lancaster method to aggregate p-values across orthologous genes within each RNA-Seq experiment (Lancaster 1961; Yi et al. 2018). This approach allows us to make statistically honest comparisons of differential gene expression between diverse studies, and elucidates what conserved genes are shared across animals during regeneration.

In this study, we compared publicly available RNA-Seq datasets spanning wildly different organisms and structures undergoing regeneration (Figure 1) to determine if an underlying core set of genes could be elucidated. The datasets analyzed include tissue regeneration in sea sponges (Kenny et al. 2017), oral/aboral body regeneration in sea anemones (Schaffer et al. 2016), head/tail regeneration in planarian worms (Kao et al. 2013), regeneration of "Cuvierian tubules" in the respiratory system of sea cucumbers (Sun et al. 2013), hair cell regeneration in zebrafish (Jiang et al. 2014), and limb regeneration in axolotl salamanders (Wu et al. 2013: 201). These datasets are highly divergent in their sampling regimes but cover the relevant early window between wound healing and blastema formation/ cell proliferation (Figure 1). Despite the limitations inherent in comparative RNA-Seq (considered in detail in the Discussion), this study provides a first-order analysis to clarify what is conserved in animal regeneration at a transcriptional level.
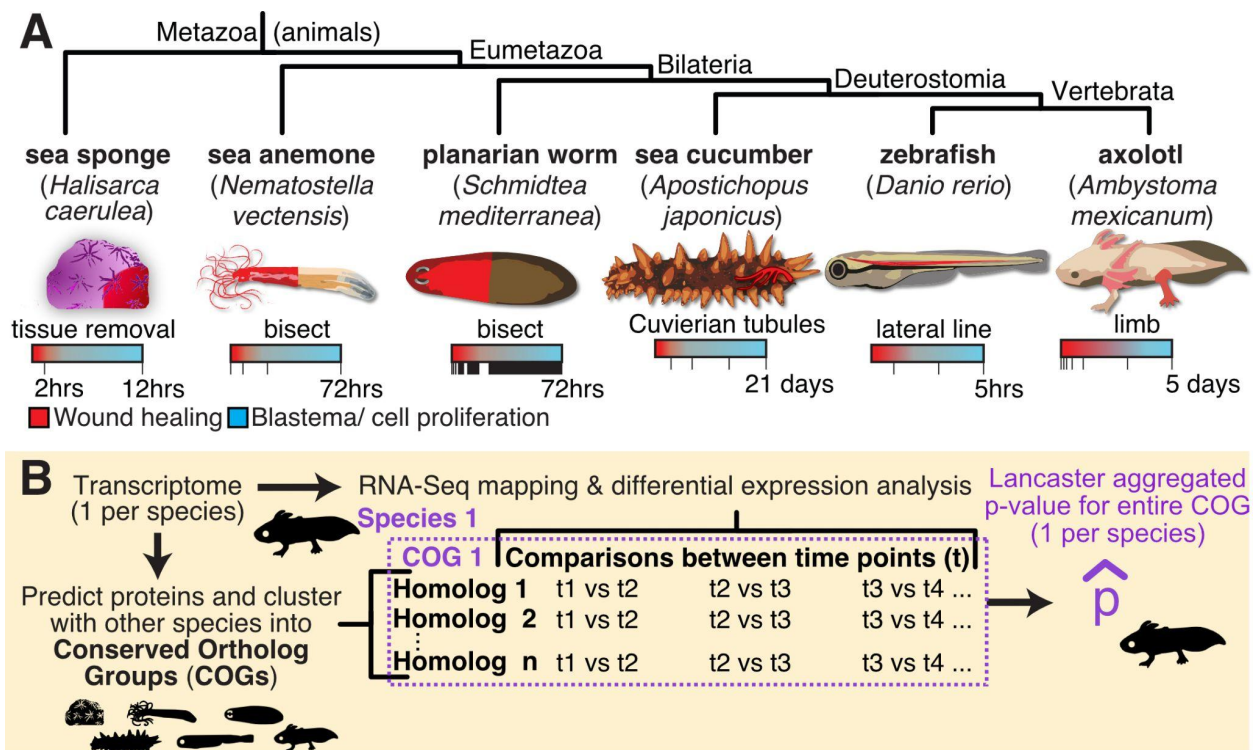
Figure 3.1: Cases of animal regeneration included in this study. (A) The 6 animals analyzed in this paper, organized by their evolutionary relationships. The region of each organism undergoing regeneration is highlighted in red and is described underneath the image of each animal. The RNA-Seq sampling regime from each study is visualized with a bar; each time point that was sampled is represented by a notch in that bar. Despite the different absolute time ranges, the studies are comparable in that the time points span the early key stages of regeneration: starting with wound healing (red) and transitioning into blastema formation / cell proliferation (blue). (B) A simplified overview of the methodology used to define differentially expressed conserved orthologous groups (deCOGs). A more detailed version is provided in Figure S2.

RESULTS

P-value aggregation method does not reveal conserved ortholog groups involved in regeneration across animals tested

The first step was to organize all proteins from our 6 datasets into clusters of putative orthologs. We used OrthoFinder (Emms and Kelly 2015) to assign orthology, as this program combines amino acid sequence similarity and phylogenetic relationships to reconstruct the evolutionary history of gene families. OrthoFinder assigned 266,324 proteins across our 6 datasets into 16,116 conserved orthologous groups or "COGs" (see Table S1 for detailed OrthoFinder results). These COGs were typically large, with a mean of 16.5 genes per COG. This reflected the large number of gene models in certain datasets (particularly the axolotl and zebrafish) as well as the wide evolutionary vantage taken in this study. Because we assigned orthology at the pan-animal scale, many paralogs in vertebrates or eumetazoans collapsed into a single COG in this study. As discussed later, this phenomenon is particularly important when interpreting our results. After genes were assigned to COGs, we used the Lancaster method to aggregate all p-values per dataset per COG into one $\hat{p}$-value (Yi et al. 2018). If that $\hat{p}$-value met a false-discovery adjusted threshold of 0.05, we considered the COG differentially expressed for that particular dataset.

To test how robust the assignment of differentially-expressed COGs (deCOGs) was to differences between datasets, we examined how adding and removing datasets impacted the final number of deCOGs (illustrated in Figure 2). Removing any particular dataset from the study increased the number of deCOGs shared across the remaining 5 datasets by an additional 26 to 196. We did not find any correlation between the quality of the RNA-Seq study and the number of additional deCOGs recovered when a dataset was removed. For example, removing the sea anemone from the analysis provided the greatest increase in deCOGs, even though this dataset included four RNA-Seq time points with biological replicates, as well as a well-annotated genome to work from. Conversely, the sea sponge had the poorest sampling regime, yet its removal resulted in one of the smallest gains (49 deCOGs). While some deCOGs could be lost due to incomplete sampling of gene expression during regeneration, our analyses do not suggest an obvious bias caused by the quality of the datasets under consideration.
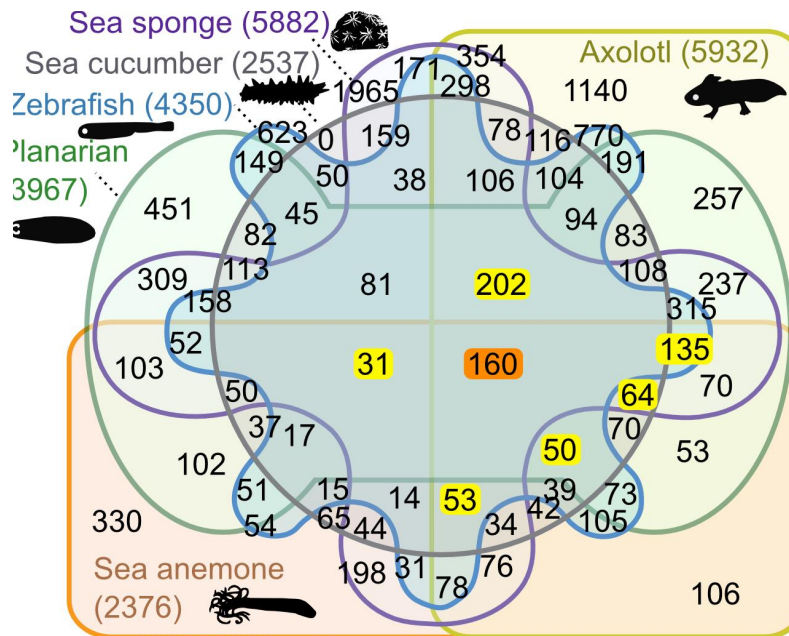
Figure 3.2: An Edwards-Venn diagram demonstrating the number of overlapping differentially expressed conserved orthologous groups (deCOGs) across all 6 datasets. The number of deCOGs common across all 6 cases (160) is highlighted in orange. Additional deCOGs that are recovered when individual case studies are removed are highlighted in yellow.

Following this check on the data, we proceeded with a holistic assay of COGs, and discovered

that the 6 datasets exhibit dramatically distinct gene repertoires. We used presence/absence data

to construct a correlation matrix that illustrates the total number of COGs shared across

datasets (Figure 3A) and a second matrix restricted to COGs that are differentially expressed in

one or more datasets (Figure 3B). Both matrices organize the taxa on evolutionary relationships

—albeit imperfectly—with vertebrates forming one major clade and the invertebrates forming a

second. If genes expressed during regeneration represented an evolutionarily conserved

network, we would anticipate the deCOG correlation matrix in Figure 3B to show greater

similarity than the full COG matrix in Figure 3A. Instead, there appears to be even less

correlation between datasets in Figure 3B compared to 3A. This suggests that genes expressed

during regeneration are no more similar across datasets than the gene repertoires as a whole.
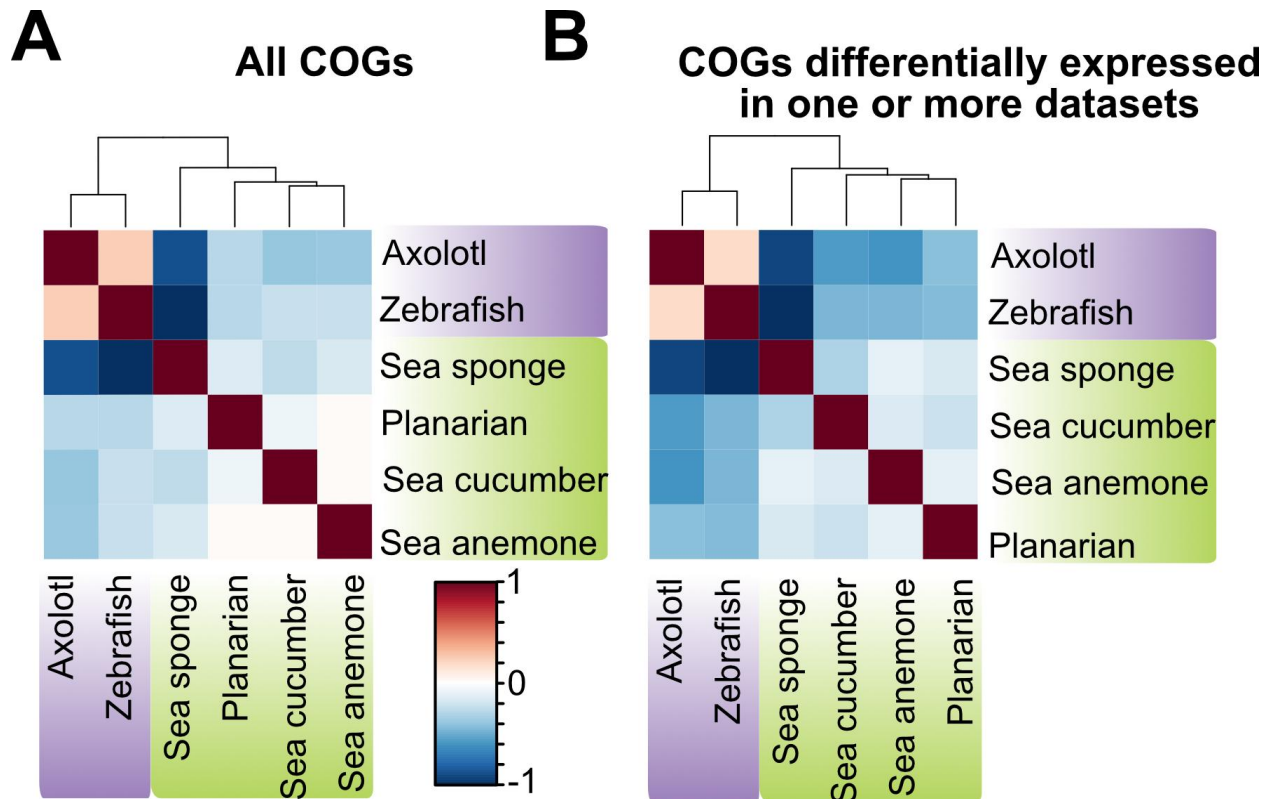
Figure 3.3: **Correlation matrices based on the presence/absence of COGs across taxa**. (A) Matrix derived from all COGs as assigned by OrthoFinder. (B) The same analysis, but restricted to differentially expressed COGs (deCOGs).

One of the patterns seen in Figure 3 is that the vertebrates (the axolotl and zebrafish) appear more similar to each other than any other combination of taxa. This raises the possibility that regeneration in vertebrates is driven by vertebrate-specific genes. To test this hypothesis, we assigned each deCOG a phyletic origin, illustrated in Figure 4. At all nodes of the phylogeny, the majority of deCOGs can be found across diverse eukaryotes. In other words, regeneration in most animal groups does not appear to require much input from novel, animal-specific genes. While this observation holds true in the vertebrates, ~9% of all deCOGs unique to this clade do appear to be vertebrate-specific novelties. This suggests that while the genetic control of animal regeneration is largely driven by the co-option of ancient genes, regeneration in vertebrates also requires input from genes unique to the group.
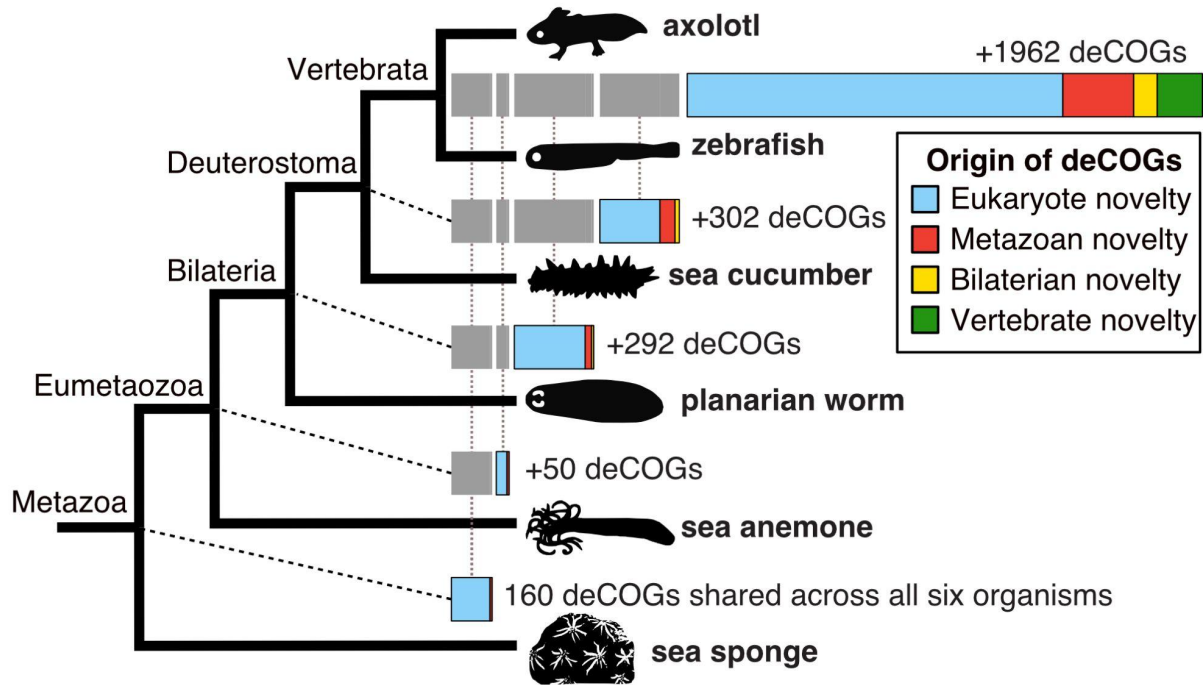
Figure 3.4: **Evolutionary** (**phyletic**) **origin of deCOGs**. The total number of deCOGs recovered at each node of the evolutionary tree is indicated by a bar chart to the right. Novel deCOGs at each node are broken down by their phyletic origin; for example, deCOGs that are a "bilaterian novelty" contain genes that have no significant sequence similarity to genes outside of the Bilateria.

Using our methodology, we recovered 160 deCOGs present in all 6 datasets. While 160 deCOGs might appear noteworthy, our approach purposefully takes a generous view of what counts as "conserved." We have, for example, ignored differences in expression direction or timing, meaning a COG is considered "conserved" if the same gene is upregulated during wound healing in one dataset and downregulated in blastema formation in another. It is unlikely that such a gene actually has a conserved biological function. Moreover, the inclusion of distantly

related animals in this analysis means that many large gene families have been reduced to a single COG. A good example of this latter issue comes from the Wnt family of genes, which are recovered as a single deCOG in our analysis. The gene tree produced by OrthoFinder is reprinted in Figure 5. Our analysis suggests that sponge Wnt genes cannot be assigned to known subfamilies, resulting in all Wnts collapsing into one COG (see (Borisenko et al. 2016) for similar results). Ignoring the sponge, only one of the Wnt subfamilies (Wnt8/9) is present in all organisms in our analysis, and no Wnt subfamily demonstrates differential expression across all taxa. So while Wnt genes are differentially expressed in every example of regeneration, each organism uses a different combination of paralogs. This result could be interpreted as evidence that diverse Wnt paralogs can be removed and integrated into a conserved regeneration gene network (i.e. paralog switching), or alternatively, that different organisms have independently integrated Wnt signaling into regeneration. Either way, this case study illustrates that a deCOG is not synonymous with a conserved gene and offers no support that an ancestral Wnt protein has a conserved function in regeneration across animal evolution.
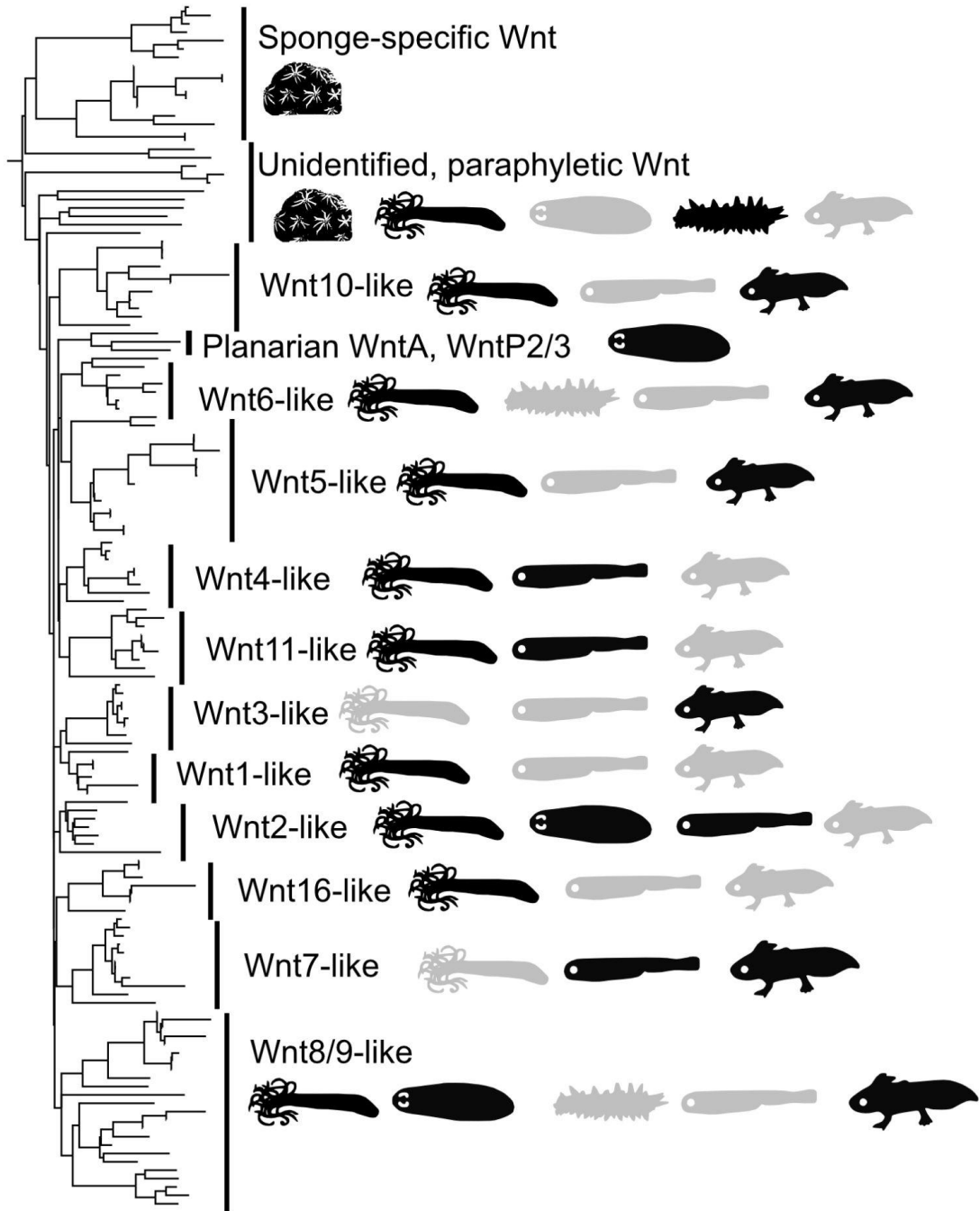
Figure 3.5: The presence of Wnt genes in the 6 RNA-Seq datasets analyzed (produced by OrthoFinder). Wnt genes were recovered as a single deCOG in our analysis, which can be subdivided into a minimum of 13 previously described subfamilies. The presence/absence of these subfamilies in each taxon is demonstrated by silhouettes. Gray silhouettes show the subfamily is present in the organism's transcriptome; black silhouettes show that the subfamily is present and differentially expressed in the relevant RNA-Seq study. Note that no subfamily is present and differentially expressed across all taxa.

To explore the possible significance of the 160 deCOGs recovered across all taxa, we used two highly-cited web resources, STRING (Szklarczyk et al. 2014) and DAVID (Dennis et al. 2003), to perform functional enrichment analyses. We focused on the zebrafish for these analyses, as it represents the best-studied model organism in our dataset. The 160 deCOGs include 2,182 zebrafish transcripts, 554 of which could be considered differentially expressed (using the generous cutoff of an unadjusted p-value < 0.01). We compared this list of genes against the zebrafish genome to look for enriched biological pathways using the comprehensive and highly-cited Kyoto Encyclopedia of Genes and Genomes (KEGG) database (see Additional File 1, part 4 for full results). According to STRING and DAVID analyses, the 554 differentially expressed zebrafish genes are enriched in basic cell processes, including melanogenesis, regulation of the actin cytoskeleton, phagosomes, and focal adhesion. Regarding KEGG pathways, Notch signaling is recovered in both analyses, while Wnt, FoxO, and mTOR pathways are enriched in the STRING analysis. However, these enriched pathways are suspect, as they are primarily driven by multiple homologs from the same COG. For example, Wnt and Frizzled homologs represent 9 out of 11 genes driving "Wnt enrichment" in STRING and 9 of the 15

genes driving "mTOR enrichment". Similarly, "Notch enrichment" is driven by the presence of 8

differentially expressed genes, 7 of which are Delta/Jagged homologs. If these pathways were

truly playing an important role in regeneration, we would anticipate more genes from these

pathways being differentially expressed. Re-running the analysis on larger lists of deCOGs

following the removal of individual datasets (see Figure 2) did not have a major impact on the

pathways recovered. However, when we restricted our analysis to deCOGs shared between the

vertebrates, we found a dramatic increase in the number of Wnt pathway genes represented (58

genes). Furthermore, FoxO (65 genes) and p53 signaling (32 genes) were also recovered as

significantly overrepresented pathways. All of these pathways have been implicated in

vertebrate regeneration (Di Giovanni et al. 2006: 53; Tothova and Gilliland 2007; Yun et al. 2013;

Martins et al. 2016). These results further support the hypothesis that a conserved regeneration

network might exist across vertebrates, even though there is little evidence for conservation

across the animals as a whole.

Given the longstanding interest in stem cell dynamics as a critical regulator in animal

regeneration, we decided to conclude our study by exploring the representation of these

pathways in our data. Figure 6 presents a simplified version of the KEGG stem cell pluripotency

network (KEGG 04550), colored to indicate the number of datasets with one or more

differentially expressed genes in the relevant COG. Few molecular signaling components were

differentially expressed across all 6 datasets, and most downstream signaling targets were

expressed in fewer than four datasets. Additionally, the ultimate target of these pathways—the

core transcriptional network driving mammalian stem cell pluripotency (Li and Belmonte

2017)—were largely absent, with two of the genes missing from all datasets (Oct4/POU5F1 and

Nanog). At first glance, some interesting signaling and receptor proteins appeared to be

conserved across all 6 datasets. However, detailed analysis of the relevant COGs revealed that

every example involved well-described paralogs being collapsed into a single pan-metazoan

COG, as described previously for Wnt. Examples include "Activin" and "BMP4" being part of a

single deCOG that also contains BMP2/4/5/6/8/15/16, and the "SOX2" deCOG, which also

contains SOX1/3/9/14 (see Table S2, Figure S5, and Additional File 1, part 7 for details). We

therefore find limited support for conserved genes in the cell pluripotency network and find

"conservation" in a few pathways only in the context of equating paralogs across various

evolutionary lineages.
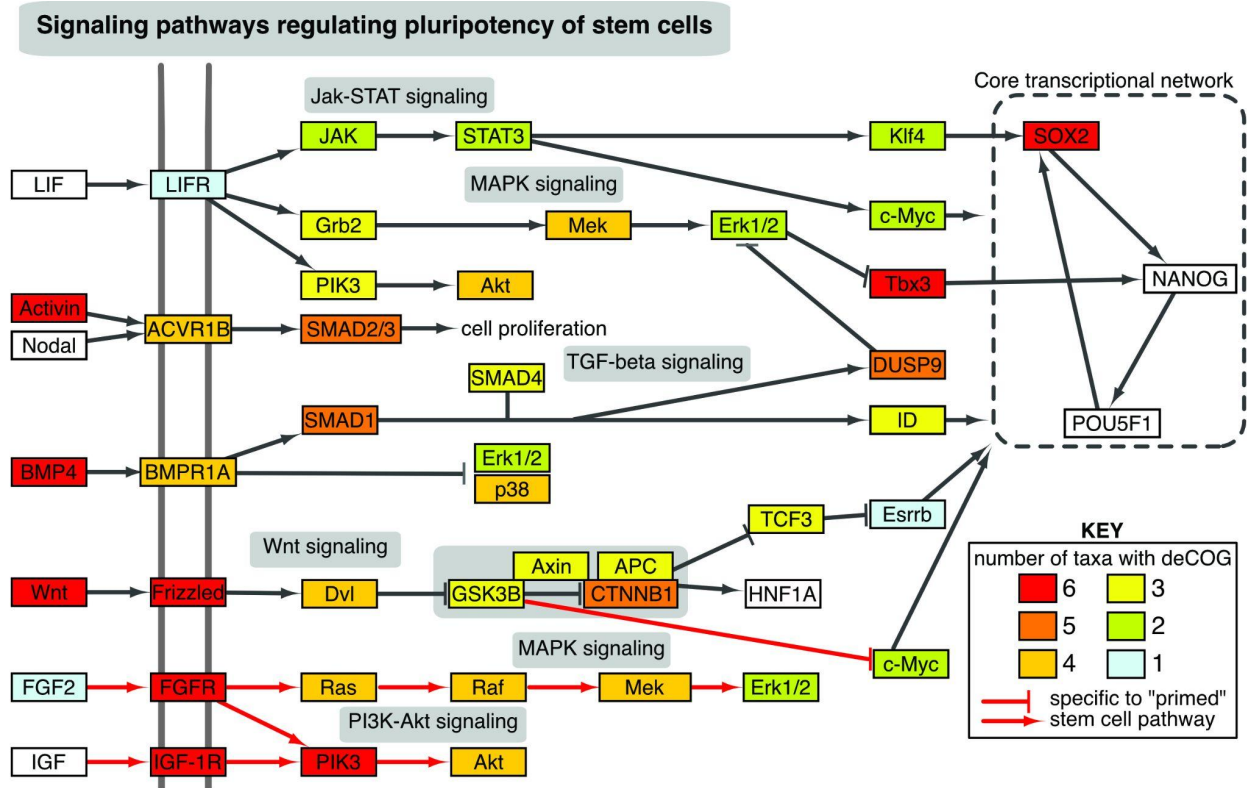
Figure 3.6: The presence of deCOGs within the stem cell pluripotency network. The network has been reproduced and simplified from KEGG pathway 04550. The color of each box indicates the number of datasets with one or more differentially expressed genes within the relevant COG. Red arrows indicate pathways that are specific to "primed" stem cells (e.g. human embryonic stem cells, human induced pluripotent stem cells, mouse epiblast-derived stem cells), gray arrows indicate pathways also found in "naïve" stem cells (e.g. mouse embryonic stem cells, mouse induced pluripotent stem cells).

A heat shock "positive control" suggests the problem of identifying conserved orthologs from comparative RNA-seq is not restricted to regeneration

It has been several years since this study was originally posted on a preprint server. One of the

reasons for the delay was an early reviewer's recommendation that we look for a "positive

control," demonstrating how the Lancaster method described here can recover conserved gene

sets from diverse RNA-Seq datasets. After testing many datasets, we were unable to find a

compelling control. An instructive example is our study of heat-stress, which we anticipated

would reveal COGs enriched in heat-shock responses. In this project, we analyzed six datasets

covering the relevant window of acute stress response to short-term heat shock in a diverse set

of organisms. Our data included expression profiles of liver response of the Atlantic salmon (Shi

et al. 2019), haemocyte transcriptive response in Pacific oysters (Yang et al. 2017), whole

organism response in the Sahara silver ant (Willot et al. 2018), whole adult somatic tissues of a

demosponge (Guzman and Conaco 2016), and a comparison of the liver transcriptome response

of three breeds of commercial chickens (Lan et al. 2016). Similar to our original analysis, we were

unable to recover a core set of genes governing the heat shock response. We could not find any

COGs shared across all datasets, and therefore focused our analysis on the comparison that

produced the most results–the oyster and sponge (Supplementary File, section 7). Enrichment

analysis of the 105 COGs found little evidence for functional conservation. DAVID Functional

Annotation recovered evidence for an enrichment of the "stress response" biological process.

This was driven by the differential expression of six heat shock transcripts, four of which are

annotated as "heat shock protein 68". Otherwise all enrichment terms were related to basic cell

processes, muscle/actin activity, and melanogenesis (Supplementary File, section 7.3). If we used

transcript IDs from the chicken–the best studied of the five species involved–we also recovered

enrichment of MAPK signaling, which was driven by 18 genes, 11 of which were calcium

voltage-gated channel auxiliary subunits, and 5 of which were RAS guanyl releasing proteins.

A dearth of conserved heat shock genes is not unique to our study; similar results have been

found in more traditional RNA-Seq analyses, even in closely related organisms. For example,

one study looking at heat stress transcriptomics in three genera of planthopper insects found

only 7 differentially expressed genes in common, out of a total of 331, when the animals were

exposed to high temperatures (Huang et al. 2017). In the chicken study cited earlier (Lan et al.

2016), only 9 out of 753 differentially expressed genes were conserved between all three lines (in

our analysis, we combined deCOGs from all three breeds into our "chicken" dataset). Similar

problems in identifying conserved genes have been discussed in other biological processes, such

as animal biomineralization (Gold and Vermeij 2023). Given these results, we are inclined to

argue that the inability to identify conserved orthologs is a general problem in comparative

RNA-Seq studies, compounded with increasing phylogenetic distance.

DISCUSSION

In this study, we have found little evidence for a shared "core" network of orthologous genes

across 6 RNA-Seq datasets related to regeneration. Our forgiving analysis design combined with

the fact that each dataset includes hundreds to thousands of differentially expressed genes

65

makes it remarkable that so few deCOGs were recovered. The fact that we found similar results in an analysis related to heat shock response suggests this pattern may apply broadly in comparative RNA-Seq.

There are several ways to interpret our results. One possibility is that a shared genetic network underlies these processes, but we failed to recover it because of insufficient RNA sampling. However there are several arguments against this interpretation. Firstly, while it is true that the datasets included in this study had markedly different sampling regimes, they were chosen to capture overlapping critical time frames in the regeneration processes. Secondly, removing any single taxon had minimal impact on the ortholog group content or the recovered list of differentially expressed genes (Figure 2). Finally, the observation that phylogenetic relatedness is more predictive of gene content than the RNA sampling regime (Figure 3) suggests that sampling variation is insufficient to explain the differences in gene expression. So although we cannot reject the hypothesis that deeper RNA sampling could increase the number of shared genes, we feel confident that our results reflect a real signal in the data.

A second possibility is that regenerative abilities across the animals is a function of convergent evolution. The conserved biological processes identified across our datasets–without common transcripts driving them–reflect the common challenges multicellular organisms must address

in exposure to bodily injury. The lack of conservation in Wnt downstream pathway targets supports this hypothesis; presence of Wnt signaling genes across our datasets (and across studies of regeneration more broadly) could reflect the fact that there are a limited number of cell signaling pathways that animals use to pattern tissues.

A final interpretation is that an originally conserved process has been obscured over the course of evolution through developmental system drift (True and Haag 2001). In this scenario, such drift can happen when non-homologous but functionally similar genes are recruited to perform equivalent functions (Koonin 2005a). Non-orthologous gene displacement could explain, for example, why different animals appear to use different Wnt paralogs in regeneration. When paralogs are first generated by gene duplication events, they are likely to be functionally redundant at first. This can lead to a variety of complex evolutionary dynamics, including the rapid evolution of one of the two gene copies (neofunctionalization), the substitution of one paralog with another in different lineages (paralog switching), conservation of both copies (redundancy-based dosage regulation), or differences in situational deployment (subfunctionalization) (Koonin 2005b; Veitia 2005). Cases such as these could open the possibility for paralog substitutions through pseudo-redundancy, where even distantly diverged paralogs may retain the ability to perform each other's functions if substituted within the relevant functional gene network. Our work adds to a growing body of evidence that

non-orthologous gene displacement is commonplace in deep-time evolution (Tarashansky et al. 2021), and challenges the "ortholog conjecture" that assumes orthologs are better predictors of shared function than paralogs (Nehrt et al. 2011; Stamboulian et al. 2020).

The question, then, is how we distinguish paralog switching from evolutionary convergence? In other words, if one organism uses Wnt3 to regenerate lost tissue, and another uses Wnt4, are we gaining insight into an ancient function of Wnt genes, or revealing how Wnts can be co-opted into the process of tissue repair? We conclude that distinguishing between these competing hypotheses will require the laborous reconstruction of gene regulatory networks. Similarities in regulatory binding sites and structure of network interactions, independent of paralog choice, would provide support for a common and conserved architecture in the regenerative process. An example of this comes from the careful dissection of the EGR-driven regeneration pathway in the acoel *Hofstenia* (Gehrke et al. 2019), identifying specific downstream pathways as well regeneration-responsive chromatin regions governing the deployment of the pathway. These binding motifs and the regulatory network architecture can be specifically compared to synonymous EGR-driven regeneration networks in other organisms that are capable of similar feats of regeneration, such as sea stars and planarians.

Our results add to a growing body of literature suggesting that the molecular components of regeneration are dissimilar across major animal clades. While it is possible that conserved gene regulatory networks exist, our results suggest that extensive paralog switching must have taken place, and that mere comparisons of gene presence/absence from RNA-Seq experiments will prove insufficient to reveal such networks. We note that the non-homology of animal regeneration at the transcriptional level does not negate the value of comparative studies across diverse taxa. Perhaps animal regeneration is homologous at another level of biological hierarchy (e.g. cell type regulation, tissue coordination), and the molecular logic coordinating this process evolved in an *ad hoc* manner across tissues and organisms. Evidence of this may come from a recent comparison of regeneration across a sea star, planarian, and hydra, in which the authors found ample evidence of conserved gene ontologies without deeply exploring the relationships of the underlying transcripts (Cary et al. 2019). In this scenario, how conserved processes could be regulated by different molecular machinery would be the great challenge going forward. Alternatively, our results could signify true evolutionary convergence, in which case dozens—perhaps hundreds—of animal lineages have independently evolved solutions to bodily damage with varying degrees of success. Such a scenario puts a greater emphasis on natural selection actively driving regenerative capabilities, as opposed to such abilities being lost to genetic drift or countervailing selective forces. Given the apparent advantages of regeneration, how and why natural selection drives this trait in specific lineages would be the great challenge

going forward. Detailed studies across diverse animals is the only way to distinguish between these competing paradigms and determine what the great questions are in regeneration biology.

MATERIALS AND METHODS

Transcriptome Collection. <u>Regeneration dataset:</u> For the axolotl (*Ambystoma mexicanum*), a transcriptome was downloaded from the Broad Institute's Axolotl Transcriptome Project (https://portals.broadinstitute.org/axolotlomics/; File: "Axolotl.Trinity.CellReports2017.fasta.gz"). For the planarian (*Schmidtea mediterranea*), a transcriptome was obtained from SmedGD (http://smedgd.stowers.org/; File: "SmedSxl Genome Annotations version 4.0 Predicted Nucleotide FASTA"). For the sea anemone (*Nematostella vectensis*) a transcriptome was downloaded from NCBI (BioProjects: PRJNA19965, PRJNA12581; File: "GCF_000209225.1_ASM20922v1_rna.fna"). For the sea cucumber (*Apostichopus japonicus*), reference isotigs were downloaded from the relevant paper (NCBI accession: GSE44995; File: "GSE44995_Reference_assembled_isotig_seq.fna.gz") (Sun et al. 2013). For the sea sponge (*Halisarca caerulea*) the transcriptome was downloaded from the Figshare link provided in the original paper (File: "Halisarca_REF_trinity.fasta.zip"). For the zebrafish (*Danio rerio*), all predicted cDNAs were downloaded from ENSEMBLE release-89 (file: "GRCz10.cdna.all.fa"). <u>Heat shock response dataset:</u> The experimental transcriptome datasets for the oyster (PRJNA232944), sponge (PRJNA274004), **silver ant** (PRJNA419094), chicken strains (PRJEB13064)

and salmon ([PRJNA427772](#)) were downloaded from the NCBI SRA Database. The SRA ID list

can be found in the accessions list for each species in the associated github. Total

transcriptomes for mapping were retrieved in FASTA format from the NCBI "Genome" page for

each of the five organisms ("transcript" downloaded in FASTA format). The genes from these

transcriptomes were converted into proteins using Transdecoder (v5.0.2)[30], and are provided in

Additional File 2.

Read Collection and Mapping. RNA-Seq reads were downloaded from the NCBI Short Read

Archive (SRA) using the "fastq-dump" program in the SRA Toolkit

(https://www.ncbi.nlm.nih.gov/sra). Table 2 provides a list of SRA IDs. The RNA-Seq reads were

aligned to the transcriptomes using HISAT-2 (Kim et al. 2015) for the regeneration dataset and

BOWTIE2 (Langmead and Salzberg 2012) for the heat shock dataset. Both were quantified using

RSEM v1.3.0 (Li and Dewey 2011). The results from RSEM mapping are provided in Table S3,

and the commands used to execute RSEM are reproduced in Additional File 1, part 0.1.

Ortholog identification. The proteins from our species datasets were grouped into orthologous

"gene sets" using the clustering algorithm OrthoFinder (Emms and Kelly 2015). The results of

orthofinder analysis are provided in Table S1. All orthogroups are provided in Additional File 1,

part 1. The resulting raw count matrices from RSEM were analyzed using EdgeR (Robinson et al.

2010). We chose EdgeR because of its ability to accept a user-defined square root-dispersion value for studies that lack biological replication. The axolotl, cucumber, and sponge datasets lack biological replicates, making it impossible to estimate gene variance within samples. To deal with this shortcoming, we used EdgeR to see how various values for the biological coefficient of variation (BCV) impacted the number of differentially expressed genes. According to the EdgeR manual, typical values for BCV range from 0.4 for human data to 0.1 for genetically identical model organisms. We therefore tested a variety of BCV values within this space; the results are shown in Figure S3. Multidimensional scaling plots of BCV distances for samples with biological replicates are shown in Figure S4. We chose the lowest value for the square root-dispersion (0.1), in part because this allowed for the largest number of differentially expressed genes, and also because the spread of differentially expressed genes at various fold-change cutoffs behave most similarly to datasets with biological replicates at this value (Figure S3). EdgeR was used to perform comparisons between adjacent time points. If a "wild-type" sample was included in the study, it was treated as equivalent to "time 0." An example of the R code used to execute EdgeR is reproduced in the Additional File 1, parts 0.2-0.3. The resulting p-values and log count-per-million values were used for downstream aggregation of p-values and are also provided as Additional File 3.

p-value Aggregation. Aggregation of the p-values produced by EdgeR was based on methods described in Yi *et al.* (2018). The method essentially treated each p-value generated from adjacent time points for a given gene as an independent significance test of the hypothesis that the broader COG was differentially expressed. Intuitively it may be the case that no single p-value from a set of independent tests registers as significant, however many borderline-significant values can be aggregated to determine significance. The aggregation methods from Yi *et al.* takes advantage of the fact that many independent p-values generated by the null hypothesis should follow a uniform distribution on the interval (0,1). Consequently, we can test the *uniformity* of the set of p-values to determine their likelihood of being generated from the null hypothesis. If the probability that the p-values as a set came from a uniform distribution is small, then we can reject the null hypothesis as having generated them. In our case, the null hypothesis corresponds to the ortholog group not being differentially expressed during regeneration for a given taxon. Mathematically, the appropriate test statistic for uniformity can be computed from the sum of inverse cumulative distribution function with p-values and raw read counts as inputs (See Yi et al. for details, and Additional File 1, part 0.5 for Python code). The result of this process is a table with entries corresponding to taxon-ortholog group pairs, and an associated aggregated p-value.

False Discovery Rate Correction. Because each taxon has hundreds to thousands of distinct COGs, individual significance testing will result in many false positives. To ameliorate this, we perform the Benjamini-Hochberg procedure to adjust p-values for false discovery rate. The p-values were adjusted based on the total number of COGs such that no more than a constant fraction were likely to be false discoveries. These adjusted p-values were used for significance testing, and result in a list of ortholog groups corresponding to genes that are likely to be differentially expressed during regeneration.

Intersection Analysis. The final step was to derive a list of deCOGs shared across datasets. We originally attempted to do this by significance testing, but found that numerical issues stemming from small p-values biased our tests such that a single p-value very close to 0 would yield a positive result, even if only one taxon showed strong results for that ortholog group. To avoid this problem, we instead used intersection analysis, looking at the presence/absence of deCOGs across datasets. This intersection method is less statistically rigorous, but has the advantage of being robust to bias from small p-values.

Correlation Plots and Venn Diagram. Overlap of COGs across taxa was visualized using correlation matrices and an Edwards Venn Diagram. A binary presence/absence table for each COG was modified from the output of OrthoFinder (provided in Additional File 1, part 2.1). A

second table focused on presence/absence of deCOGs (Additional File 1, part 2.2). These tables were used to generate the correlation plots in Figure 3 of the main text with the corrplot R library. Commands for generating the plots are provided in Additional File 1, part 2.3. The table of deCOGs was used to create an Edwards Venn Diagram using InteractiVenn (Heberle et al. 2015).

Phylogenetic Assignment of Gene Families. Ideally, the evolutionary origin of each deCOG would be determined using a phylogenetically-informed clustering analysis such as OrthoFinder. Unfortunately taking such an approach at a eukaryote-wide scale is, for the time being, computationally prohibitive. Instead we performed a series of BLAST queries and used sequence similarity of protein sequences to assign a phyletic origin for each COG.

Firstly, Uniprot Swissprot datasets were downloaded from www.Uniprot.com using the following queries: (1) Eukaryote (non-animal) dataset: "*NOT taxonomy:"Metazoa [33208]" AND reviewed:yes*" (2) Early animal dataset: "*taxonomy:"Metazoa [33208]" NOT taxonomy:"Bilateria [33213]" AND reviewed:yes*" (3) Bilaterian invertebrate dataset: "*taxonomy:"Bilateria [33213]" NOT taxonomy:"Vertebrata [7742]" AND reviewed:yes*"

Each of these datasets was turned into a BLAST databased using the *makeblastdb* command. Our query COGs were the 2,770 deCOGs present in both the zebrafish and axolotl (see Figure 4 of the main text), which also encompassed all deCOGs at broader evolutionary scales (i.e. the deCOGs shared by all vertebrates necessarily includes all deCOGs shared by deuterostomes, and so on). All protein sequences from these 2,770 deCOGs were collected and formatted into a query fasta file.

With the production of our query and database files, we proceeded with an iterative process of BLAST analyses. All proteins from the 2,770 deCOGs were queried against the "Eukaryote" database using BLASTp (command: *blastp -query Query_Proteins.fasta -db Eukaryote_Dataset -outfmt 6 -evalue 10e-5 -max_target_seqs 1 -num_threads 4 -out Results.txt*). If one or query had a hit, the entire deCOG was considered a "eukaryote novelty". Proteins in the deCOGs that did not match anything in the "Eukaryote" database were used as the query sequences for the next BLASTp analysis against the "Early animal" database. In addition, any deCOG that had no match in the "Eukaryote" database and included at least one sponge protein was automatically designated as an "animal novelty," regardless of whether or not it had a BLAST hit in the "Early animal" database. This process was repeated until all deCOGs were assigned a phyletic origin. A summary of these results is provided in Additional File 1, part 6.

Enrichment analysis of deCOGs. Our comparison between all 6 taxa resulted in 160 deCOGs. We also examined the impact of individual taxa on the deCOG list by re-running the analysis with one organism excluded. Zebrafish (*Danio*) gene IDs from the resulting deCOGs were collected from each analysis, and are provided in Additional File 1, part 3**.** We restricted enrichment analysis to zebrafish genes that had at least one uncorrected (raw) p-value less than 0.01 from the original EdgeR analysis (Additional File 1, part 0.2-0.3).

DAVID enrichment analysis was performed on the server (https://david.ncifcrf.gov). Zebrafish gene IDs were submitted using the "ENSEMBL_TRANSCRIPT_ID" identifier and a "Background" list type. STRING enrichment analysis requires a list of protein IDs, so the zebrafish transcripts were converted into protein identifiers using UniProt's "Retrieve/ID mapping" function (https://www.uniprot.org/uploadlists/). The resulting IDs are provided in Additional File 1, part 3. These IDs were submitted to the STRING server for enrichment analysis (https://string-db.org). For both analyses, we restricted our study to conserved KEGG pathways. The full results of these analyses are provided in Additional File 1, part 4.

Analysis of gene trees. In this paper, we examined the coverage of deCOGs in the KEGG stem cell pluripotency network (Figure 6). For genes present in all 6 datasets, we went back to the Orthofinder data to determine how gene families were organized into COGs, and which genes

within those COGs were differentially expressed. Species-tree corrected gene trees were collected from the Orthofinder output. These trees were manually annotated to include gene names (based on zebrafish IDs) and whether or not genes were differentially expressed (smallest uncorrected p-value < 0.01 from EdgeR output). Figure S5 shows the gene tree for *activin* and *bmp4* constructed using this method. The tree in Figure S4 and all additional, annotated trees are provided in newick format in Additional File 1, part 7.

Data and Resource Availability: All data and code used in this study are available on GitHub at

https://github.com/DavidGoldLab/2023_Comp_Regen.

Chapter 4

Single Cell RNA-Seq Identifies a Putative Interstitial Stem Cell in the Moon Jellyfish *Aurelia*

The decline of stem or progenitor cell populations leads to a loss in tissue replacement capability, the inability to maintain tissue integrity and the onset of aging phenotypes[1-3]. An important target of aging research is therefore to understand the mechanisms of maintenance of progenitor cell population integrity[4]. The moon jellyfish (Aurelia aurita) is an ideal model to explore the relationship between progenitor cell integrity maintenance and immortality as its complex life cycle includes a polyp stage which can maintain tissues indefinitely, and a medusa stage with a limited regenerative capacity that decreases over its 1-2 year lifespan[5]. These two life stages allow us to compare how aging and non-aging manifest under differential utilization of the same genome. Since Aurelia appears to lack canonical stem cells[5-9], it has been hypothesized that epitheliomuscular cells act as a progenitor cell[10] through transdifferentiation–or the conversion of one somatic cell type to another– replenishing Aurelia tissue both in injury response and for routine cell maintenance. This has yet to be conclusively demonstrated through traditional approaches[5-8], however single cell RNA-sequencing techniques will provide the opportunity to capture and visualize the unique transcriptional signature of low abundance cells or those with high transcriptomic (and potentially functional) diversity despite a lack of morphological diversity.

INTRODUCTION

Tissue maintenance is essential for the survival and function of multicellular organisms, however regenerative capacity and the ability to replace lost parts can differ remarkably across the tree of life (Alvarado 2000; Bely and Sikes 2010; Cary et al. 2019). The ability of animals to maintain and regenerate tissue varies widely depending on the species, tissue type, and age. While some can regenerate lost body parts or even the entire organism, most are limited to the replacement of a select few types of cells and tissues. Tissue replacement is most commonly driven by stem cells, cell populations which retain the ability to differentiate into necessary tissues when signaled (Alvarado and Yamanaka 2014; Zakrzewski et al. 2019). Many organisms retain only multi- or unipotent cells–those with limited differentiation potential–for the purpose of routine cell replacement, limiting their regenerative capacity following injury; however some groups can regenerate body parts or even their entire body, including cnidarians (jellyfish, corals and their relatives), planarians (flatworms), and echinoderms (sea urchins, sea stars, sea cucumbers and their relatives). Organisms capable of whole body regeneration appear to employ diverse strategies at the cellular level, often combining the use of multipotent stem cells with flexible cell programming known as transdifferentiation (Alvarado and Yamanaka 2014). This diversity poses a challenge in investigating the evolution of injurious tissue replacement and the homology of stem cells.

Cnidarians (corals, jellyfish, anemones and their relatives) display a wide range of regeneration

strategies for both maintaining and replacing tissue populations, including stem cell

differentiation and transdifferentiation of differentiated cells. For example Hydra – the

preeminent cnidarian aging and regeneration model of the class Hydrozoa – maintain multiple

stem cell populations that continually produce new tissue during homeostatic maintenance that

allow whole body regeneration (Bosch et al. 2010). Transdifferentiation also plays a relevant role

in Hydra cell maintenance; as new tissue is created near the center of the organism, older tissue

is pushed towards the morphologically distinct extremities, and locational signals cause cells to

adopt relevant roles and morphologies during this migration process (Siebert et al. 2008,

Koizumi et al. 1988, Primack et al. 2023). Cnidarians from other classes (Scyphozoa, Cubozoa)

lack morphologically identifiable stem cells, and this has been thought to be the ancestral

condition (Gold and Jacobs 2013). Even other members of the class Hydrozoa such as Turritopsis

– the so-called "immortal jellyfish" – rely entirely on transdifferentiation of committed cells

(Piraino et al. 1996). However, recent studies on the Anthozoan model Nematostella reveal a

population of quiescent cells similar to bilaterian adult stem cells maintained in the

mesenteries, which re-enter mitosis following amputation and are responsible for generating

both germline and neurons (Röttinger 2021; Miramon-Puertolas and Steinmetz 2023). The

patterns of stem cell presence across Cnidaria has led to varying hypotheses about how

stem-ness has evolved in this phylum (Gold and Jacobs 2013), and makes it difficult to homologize stem cell precursors between the major cnidarian classes.

The moon jellyfish Aurelia is a relevant model system for disentangling the evolutionary history of cnidarian stem cells. A member of the class Scyphozoa, it is part of a group that lacks visually identifiable interstitial stem cells. Yet in its morphology and regeneration capacity, the Aurelia polyp stage has many similarities to Hydra. Both are of comparable size and demonstrate continual cell turnover, which allows for constant asexual reproduction and the capacity to regenerate following injury (Gold and Jacobs 2013). But unlike Hydra, there is no morphological evidence for an undifferentiated stem cell line.

The goal of this study was to use single-cell RNA sequencing (scRNA-Seq) to identify which Aurelia cell serves as precursor to cnidocytes and neurons, comparable to the Hydra interstitial stem cell (ISC). scRNA-seq is a method for measuring the expression profile of individual cells, allowing researchers to isolate the transcriptomes of cells that make up a sample (Kolodziejczyk et al. 2015). Siebert et al. (2019) generated a cell atlas of Hydra using scRNA-Seq, and were able to identify the three known stem cell populations in their data set. One of the Hydra stem cell types is the multipotent interstitial stem cells (ISC), which is the progenitor of stinging cells (cnidocytes) neurons, gland cells, and the germ line. We focused on identifying putative ISCs in

Aurelia because it is one of the best characterized stem cell types in Medusozoa, and because the development of cnidocytes and neurons of Aurelia have been previously studied, (Nakanishi et al. 2009; Gold, Lau, et al. 2019) making identification of the cell types easier. Our hypothesis is that cnidocytes and neurons in Aurelia share a cell precursor, and that the gene expression profile of this precursor cell will have a similar gene profile to the ISCs of Hydra.

METHODS

Whole body polyp dissociation

Prior to dissociation, we prepared several solutions to incubate the animals in. The first solution was a low-calcium, magnesium free artificial seawater (Low-No-ASW). To make this, we combined 460 mM NaCl, 10 mM KCl, 1mM CaCl2, and 10mM HEPES. The seawater was raised to 1 liter using DI water and adjusted to a pH of 7.6 with NaOH. The second solution was calcium and magnesium-free artificial seawater (No-No-ASW). To make this, we combined 495 mM NaCL, 9.7 mM KCl, 27.6 mM NaHCO3, 5mM EDTA, and 50 mM Tris-HCl. The solution was raised to 1 liter using DI water, and adjusted to a pH of 8 with NaOH. Both solutions were filtered through a 0.22 um bottletop filter and autoclaved before use.

We started with four medium sized polyps for cell dissociation. These animals were clones of each other, taken from the same line that the Aurelia coerulea ("sp.1") genome was sequenced from (Gold, Katsuki, et al. 2019). The animals were washed three times in 50 mL of calcium and magnesium-free artificial seawater (No-No-ASW). The animals were left in a fourth wash of No CA-No MG ASW for five minutes. The animals were then transferred to 200uL of low-calcium no-magnesium artificial seawater (Low-No-ASW), which included 2mg of Protease from Streptomyces Type XIV (Sigma P5147-1G) for a 1:100 ratio. The animals were incubated in the solution for an additional five minutes. Cell dissociation was encouraged by pipetting the sample up and down twenty times with a disposable pipette under a dissecting microscope. The tip was cut off of the pipette to increase the diameter of the tip orifice and minimize damage to cells. The samples rested for three minutes, and then the pipetting process was repeated. After another three minutes, we pipetted the simple vigorously with a P200 pipette tip, and 100μL of Low-No-ASW was added (for 300μL total volume). After an additional 3 minute incubation time, the sample was centrifuged for 6 minutes at 200g at room temperature. The resulting cell pellet was washed with Low-No-ASW and resuspended in 200μL Low-No-ASW. The centrifuging, resuspension process was repeated one additional time, and then the sample was filtered twice through a 40 μm mesh to isolate single cells.

Hemocytometer validation and cell counting

Cell counts were estimated using a hemocytometer (Chemglass Life Sciences, CLS-4207). 100µl of the cell suspension was placed in a separate Eppendof tube. 10µl of trypan blue was added to the sample and mixed. 10ul of the sample was added to the groove of the hemocytometer , in between the coverslip. The number of live cells (those not stained blue) were counted in each grid at the four corners of the hemocytometer, and an average cell count was calculated. This number was multiplied by 104 and 10 (to account for the 1:10 dilution) to calculate the total number of cells per mL.

Chromium sequencing and protocols

The sample was sequenced on the Chromium 10X v3.0.2 platform using Single Cell 3' v3 chemistry at the UC Davis Genome Center. The core provided a preliminary analysis with read quality and cell count information, detailed in Table 1. The genome reference was constructed with Cell Ranger v3.0.2 mkref using the Aurelia Genome v1.2_11-27-18. The reads were then mapped to the reference using CellRanger count to estimate read expression frequency and construct the count matrix. The files and code used to construct the count matrix, deployed on an high performance computing cluster running Ubuntu 18.04.6 LTS (GNU/Linux

5.4.0-146-generic x86_64), is provided on GitHub at

https://github.com/DavidGoldLab/2023_scRNA-Seq.

Matrix analysis and cell profile comparison

The count matrix was analyzed using Seurat v3.1.2 under R version 3.5.1 (Satija et al. 2015).

Mitochondrial genes were filtered by removing the contig 3751, and an elbow plot was used to

restrict the analysis to the first 6 most informative principal components. These were used to

create a t-distributed Stochastic Neighbor Embedding (t-SNE) plot, showing cells clustered by

similarity of expression. The FindMarkers() function of Seurat was used to isolate the genes

driving cluster identity using Wilcoxon Rank Sum Test to quantify differential expression across

clusters. The contents of these marker lists were then compared to cluster markers defining cell

types of Hydra (Siebert et al. 2019) to assess potential similarity of cluster identity .

GO Term enrichment analysis

We looked for enriched gene ontology (GO) terms using the Trinotate package v4.0.0 (Bryant et

al. 2017). In brief, this package uses the genome and gtf file to produce transcript and gene

models. These models are compared against a curated Uniprot-Swissprot database for similarity

searches using Diamond v2.1.6.160 (Buchfink et al. 2021) and conserved protein domains are

identified using the PFamA database and HMMER v3.3.2 (Finn et al. 2011). This information is used to assign putative GO terms to each Aurelia gene. We then used GOSeq v3.17 (Young et al. 2010) to look at the gene ontology terms identified in the cluster markers of each cell type, and compared them to a background dataset consisting of all cluster markers from our analysis. This produced a set of enriched and depleted GO terms for each cell type.

Profile comparison via BLAST

To identify whether the putative stem-cell-like cluster in the Aurelia dataset shared transcriptional similarity to Hydra stem cells or stem cell progenitors, we performed reciprocal blast between Hydra and Aurelia proteomes seeded from the Aurelia cluster markers to identify best reciprocal best hits for the Aurelia cluster markers. Cluster markers identifiers for all nine Aurelia clusters were extracted from the Aurelia proteome (Gold, Katsuki, et al. 2019) and blasted against the Hydra aepLRv2 transcriptome (https://research.nhgri.nih.gov/hydra/download/transcripts/aepLRv2_transcript.fa.gz) using BLASTp (Camacho et al. 2009). The Hydra transcriptome was converted into protein models using Transdecoder v5.5.0 (Haas et al. 2013). Reciprocal BLAST was performed with the output of this blast search against the Aurelia proteome. Cluster of origin was compared between the two to determine how many Aurelia markers (for each cluster) have a reciprocal best hit with a

marker of Hydra stem cells and/or ISC progenitors, to identify whether the putative

stem-cell-like cluster in Aurelia was more likely to share Hydra ISC transcripts than other

clusters.

RESULTS

Initial collection generated linked supercluster with terminally differentiated cell types

This was the first time we attempted scRNA-seq on Aurelia, and as a result the quantity and

quality of reads was not ideal. The standard protocol for the Chromium system includes adding

freshwater to bring a PBS solution to hypotonicity, so that the cells lyse as they enter the system,

however an excess of freshwater appeared to have lysed the cells prior to sequencing, and we

experienced a considerably lower cell recovery rate than expected. Still, we recovered data from

2,239 cells, with a mean of 6,668 reads per cell. By comparison, the Hydra atlas contains 24,985

cells (Siebert et al. 2019).

Sequencing	
Number of Reads	14,930,229
Valid Barcodes	97.50%
Sequencing Saturation	38.60%

Q30 Bases in Barcode	94.70%
Q30 Bases in RNA Read	55.80%
Q30 Bases in UMI	94.70%
Mapping	
Reads Mapped to Genome	73.70%
Reads Mapped Confidently to Genome	65.60%
Reads Mapped Confidently to Transcriptome	32.30%
Cell estimations	
Estimated Number of Cells	2,239
Fraction Reads in Cells	72.80%
Mean Reads per Cell	6,668
Median Genes per Cell	337
Total Genes Detected	16,539
Median UMI Counts per Cell	656

Table 4.1: Sequencing and mapping statistics from the Chromium output.

Despite the limitations of our dataset, we found compelling evidence for a cell precursor for cnidocytes and neurons. The results of our single-cell seq analysis are visualized in Figure 1 using a tSNE plot, which divides the cells into eight clusters. Clusters 0,1,2,3, and 7 are linked in a super-cluster, and likely represent the cellular transition from a precursor cell into cnidocytes and neurosensory cells. Cluster 1, which represents one of the termini of the super-cluster, likely represents cnidocytes. The most significant genetic marker in cluster 1 is a venom protein, which is a hallmark of stinging cells (Turk and Kem 2009). Other lines of evidence include GO

enrichment of peptidase activity, and the presence of multiple homologs of ZNF845 (Seg93.3; Seg2951.3) as cluster markers. ZNF845-like proteins have been described as important regulators of Nematostella cnidogenesis (Steger et al. 2022). Complicating our interpretation, Cluster 1 is also enriched in the transcription factor ELAV, which is an important pan-animal marker of neurons and is notably absent in the cnidocytes of Nematostella (CITATION). It is possible that the transcription factor works differently in Aurelia, but our preferred hypothesis is that a neural cell type is mixed with cnidocytes in cluster 1, and that we lack the depth of data to resolve the multiple types. Cluster 3, another terminal on the main super-cluster, includes neuron-specific markers including a GABA transporter (Seg4683.1) and neuropeptide ff receptor (Seg3014.2) homologs. It also contains transcription factors known to play a role in cnidarian neurogenesis, such as lhx/LIM3 (Seg4290.1) and Smaug2 (Seg1080.9) homologs (Sebé-Pedrós et al. 2018; Li et al. 2023). Cluster 7 also contains Smaug2 and is enriched in calcium-binding genes and ion channels (Seg4024.1; Seg571.22). Cluster 7 may either represent a second type of neuron, as multiple neuron types have been described in Aurelia (Yuan et al. 2008; Nakanishi et al. 2009), or else a sensory cell. Based on this data, we hypothesize that Aurelia development includes a cnidocyte / neurosensory cell trajectory that is comparable to Hydra, connected by a putative ISC-like stem cell precursor.
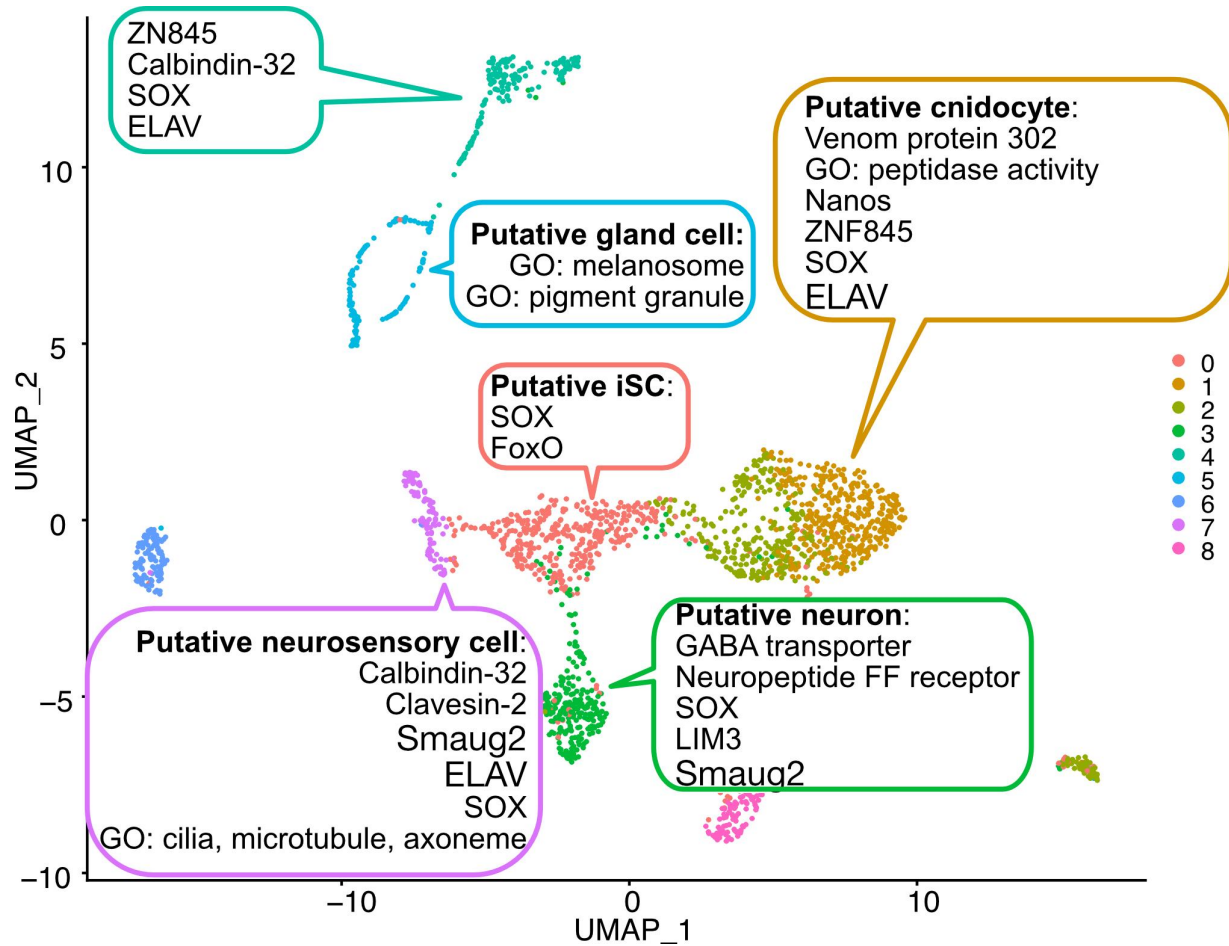
Figure 4.1: tSNE plot of single cell RNA-Seq data generated with Seurat.

While not directly related to our hypothesis, there are several additional observations from the

single-cell seq data that are worth noting. For example, Cluster 5 is likely a gland cell; this is

based on the enrichment of pigment and melanosome transcripts, which are a morphologically

characteristic of Aurelia gland cells (Yuan et al. 2008). In other cnidarians, gland cells are derived

from the same stem cell precursor as neurosensory cells. It's neighbor, Cluster 4, is enriched in

many neurosensory markers seen in super-cluster 0,1,2,3,7; this includes homologs to SOX (Seg523.9; the same transcript enriched in those clusters), ELAV (Seg3239.2), and ZN845 (Seg1955.5) transcription factors. It is plausible that Cluster 4/5 would connect to the main supercluster with better single cell sampling. Another notable observation is the presence of a nanos homolog as a marker for Cluster 1, the putative cnidocyte. In Hydra nanos is associated with multipotent stem cells and germ cells (Mochizuki et al. 2000), while in Nematostella, nanos drives the development of neurons (Steger et al. 2022). As noted earlier, it is quite possible that a neural cell line is mixed with cnidocytes in cluster 1, so it is too soon to generate a hypothesis about the significance of nanos in this cluster. Still, these observations are notable as potential areas of future research.

BLAST analysis does not reveal enrichment of Hydra stem cell markers in supercluster

To test whether cells from Cluster 0 were statistically enriched in Hydra ISC markers, we performed a reciprocal BLAST search to identify homologs between Hydra and Aurelia. Our results, summarized in Table 2, do not suggest that Cluster 0 has a higher percentage of Hydra stem cell/ISC transcripts than any other Aurelia cell type. While 20% of Cluster 0 reciprocal blast hits were markers of stem cell/ISC identity in Hydra (Table 2), the average across all Aurelia clusters was 20.71%. We also note that stem cell specific genes piwi and vasa, which have been

previously identified as useful marker genes for Hydra multipotent interstitial stem cells in

addition to nanos (Juliano et al. 2014) are absent from all Aurelia cluster markers, including

Cluster 0. This would be surprising if true and indicates a need for further data collection.

However, given its position within the tSNE plot connecting the putative cnidocytes, neurons,

and glandular and secretory cells, and its expression of FoxO, we conclude that this cluster could

play and ISC-like role functioning as a precursor to cnidocytes and neurosensory cells, and

further data collection will be necessary to investigate the transcriptomic profile of these cells.

	Number of cluster markers	Hydra blast matches	Reciprocal best hits	N sc/isc hits in recip match	Freq sc/isc hits per reciprocal match	Freq sc/isc hits per Hydra match	Freq sc/isc hits per Aurelia marker
cluster0	345	253	175	35	20.00%	13.83%	10.14%
cluster1	456	332	243	35	14.40%	10.54%	7.68%
cluster2	297	227	169	33	19.53%	14.54%	11.11%
cluster3	279	196	143	28	19.58%	14.29%	10.04%
cluster4	490	379	292	63	21.58%	16.62%	12.86%

cluster5	889	671	543	107	19.71%	15.95%	12.04%
cluster6	518	394	302	66	21.85%	16.75%	12.74%
cluster7	532	417	328	84	25.61%	20.14%	15.79%
cluster8	470	349	278	67	24.10%	19.20%	14.26%
average					**20.71%**	**15.76%**	**11.85%**

Table 4.2: Reciprocal blast comparison. Cluster 0 did not demonstrate an enrichment of reciprocal best hits corresponding to Hydra single cell or interstitial stem cell progenitor markers. Labels: n.cluster.markers = number of markers present within each Aurelia tSNE cluster; hydra.blast.matches = number of aurelia markers that had BLAST hits within the Hydra aepLv2 transcriptome; recip.best.hits = number of markers from each Aurelia cluster where the Hydra blast result matched to the original Aurelia representing a reciprocal best match between the Aurelia and Hydra protein sequence; n_isc/sc.hits = number of reciprocal best matches which were markers of single cell or interstitial stem cell progenitor identity in Hydra;

CONCLUSION

Taken together, our results suggest that Aurelia likely has a cnidocyte/neurosensory cell

precursor comparable in its developmental trajectory to Hydra. The next steps will require

characterizing this cell type's morphology and function, and a further description of its

transcriptional identity of these cells. Additional single cell data will help to improve the resolution of cell type trajectories, and may even help to reveal cryptic subtypes. In situ hybridization of FoxO and other markers from cluster 0 will help us to understand more about the spatial location and morphology of these putative progenitor cells. This should resolve whether the progenitor cell in Aurelia represents an undescribed, set-aside stem cell like the Hydra ISC, or a transdifferentiated epitheliomuscular cell as previously hypothesized (Gold and Jacobs 2013; Gold, Lau, et al. 2019)

Despite remaining uncertainty about the morphology of the Aurelia ISC-like cell, one notable observation from our analysis is a lack of evidence that the cnidocyte/neurosensory cell is enriched in ISC-like transcripts. This may be due to low scRNA-Seq sampling, but it may also support work done in this dissertation which suggests comparative transcriptomics is of limited value for identifying homologous traits across distantly related organisms (Sierra et al. 2021). Still, understanding how these cell types have diverged from a developmental and transcriptional perspective will be illuminating. The flexibility of cell identity is crucial for the diversity of regenerative strategies observed in the Cnidaria. From an evolutionary perspective, this flexibility may be linked to the evolution of novel cell types, enabling the development of new features and functions within an organism, and investigation into this plasticity will provide valuable insight into the evolution of animal complexity.

References

Introduction

Huang HJ, Xue J, Zhuo JC, Cheng RL, Xu HJ, & Zhang CX. 2017. Comparative analysis of the transcriptional responses to low and high temperatures in three rice planthopper species. *Molecular Ecology*, *26*(10), 2726-2737.

Lan X, Hsieh JC, Schmidt CJ, Zhu Q, & Lamont SJ. 2016. Liver transcriptome response to hyperthermic stress in three distinct chicken lines. *BMC genomics*, *17*, 1-11.

Marlow H, & Arendt D. 2014. Evolution: ctenophore genomes and the origin of neurons. *Current Biology*, *24*(16), R757-R761.

McKenna KZ, Wagner GP, & Cooper KL. 2021. A developmental perspective of homology and evolutionary novelty. In *Current topics in developmental biology* (Vol. 141, pp. 1-38). Academic Press.

Moroz LL, & Kohn AB. 2016. Independent origins of neurons and synapses: insights from ctenophores. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1685), 20150041.

Oakley TH & Plachetzki DC. 2011. Key Transitions During Animal Phototransduction Evolution: Co-duplication as a Null Hypothesis for the Evolutionary Origins of Novel Traits (pp. 217-237). CRC Press.

Panchen AL. 2007. Homology—history of a concept. In *Novartis Foundation Symposium 222-Homology: Homology: Novartis Foundation Symposium 222* (pp. 5-23). Chichester, UK: John Wiley & Sons, Ltd.

Picciani N, Kerlin JR, Sierra N, Swafford AJ, Ramirez MD, Roberts NG, Cannon JT, Daly M, Oakley TH. 2018. Prolific origination of eyes in Cnidaria with co-option of non-visual opsins. *Current Biology*, *28*(15), 2413-2419.

Ryan JF. 2014. Did the ctenophore nervous system evolve independently?. *Zoology*, *117*(4), 225-226.

Shi KP, Dong SL, Zhou YG, Li Y, Gao QF, & Sun DJ. 2019. RNA-seq reveals temporal differences in the transcriptome response to acute heat stress in the Atlantic salmon (Salmo salar). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *30*, 169-178.

Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, & Rokhsar DS. 2023. Ancient gene linkages support ctenophores as sister to other animals. *Nature*, *618*(7963), 110–117. https://doi.org/10.1038/s41586-023-05936-6

Wake DB. 1994. Comparative Terminology: Homology. The Hierarchical Basis of Comparative Biology. Brian K. Hall, Ed. Academic Press, San Diego, CA, 1994. xvi, 483. *Science*, *265*(5169), 268-269.

Chapter 2

Alvarado AS. 2000. Regeneration in the metazoans: why does it happen? *Bioessays* 22:578590.

Bely AE, Sikes JM. 2010. Latent regeneration abilities persist following recent evolutionary loss in asexual annelids. *Proceedings of the National Academy of Sciences* 107:1464–1469.

Bielefeld KA, Amini-Nik S, Alman BA. 2013. Cutaneous wound healing: recruiting

developmental pathways for regeneration. *Cellular and Molecular Life Sciences* 70:2059–2081.

Borisenko I, Adamski M, Ereskovsky A, Adamska M. 2016. Surprisingly rich repertoire of Wnt genes in the demosponge *Halisarca dujardini*. *BMC evolutionary biology* 16:123.

Cary GA, Wolff A, Zueva O, Pattinato J, Hinman VF. 2019. Analysis of sea star larval regeneration reveals conserved processes of whole-body regeneration across the metazoa. *BMC biology* 17:1–19.

Clevers H, Loh KM, Nusse R. 2014. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* 346:1248012.

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome biology* 4:R60.

Di Giovanni S, Knights CD, Rao M, Yakovlev A, Beers J, Catania J, Avantaggiati ML, Faden AI. 2006. The tumor suppressor protein p53 is required for neurite outgrowth and axon regeneration. *The EMBO journal* 25:4084–4096.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* 16:157.

Gehrke AR, Neverett E, Luo Y-J, Brandt A, Ricci L, Hulett RE, Gompers A, Ruby JG, Rokhsar DS, Reddien PW. 2019. Acoel genome reveals the regulatory landscape of whole-body regeneration. *Science* 363:eaau6173.

Gold DA, Gates RD, Jacobs DK. 2014. The early expansion and evolutionary dynamics of POU class genes. *Molecular biology and evolution* 31:3136–3147.

Gold DA, Vermeij GJ. 2023. Deep resilience: An evolutionary perspective on calcification in an age of ocean acidification. *Frontiers in Physiology* 14:133.

Guzman C, Conaco C. 2016. Gene expression dynamics accompanying the sponge thermal stress response. *PloS one* 11:e0165368.

Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC bioinformatics* 16:169.

Huang H-J, Xue J, Zhuo J-C, Cheng R-L, Xu H-J, Zhang C-X. 2017. Comparative analysis of the transcriptional responses to low and high temperatures in three rice planthopper species. *Molecular Ecology* 26:2726–2737.

Jiang L, Romero-Carvajal A, Haug JS, Seidel CW, Piotrowski T. 2014. Gene-expression analysis of hair cell regeneration in the zebrafish lateral line. *Proceedings of the National Academy of Sciences* 111:E1383–E1392.

Kao D, Felix D, Aboobaker A. 2013. The planarian regeneration transcriptome reveals a shared but temporally shifted regulatory program between opposing head and tail scenarios. *BMC genomics* 14:797.

Kenny NJ, de Goeij JM, de Bakker DM, Whalen CG, Berezikov E, Riesgo A. 2017. Towards the identification of ancestrally shared regenerative mechanisms across the Metazoa: A Transcriptomic case study in the Demosponge Halisarca caerulea. *Marine genomics.*

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* 12:357.

Koonin EV. 2005a. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* 39:309–338.

Koonin EV. 2005b. Paralogs and mutational robustness linked through transcriptional

reprogramming. *Bioessays* 27:865–868.

Lan X, Hsieh JC, Schmidt CJ, Zhu Q, Lamont SJ. 2016. Liver transcriptome response to hyperthermic stress in three distinct chicken lines. *BMC genomics* 17:1–11.

Lancaster HO. 1961. The combination of probabilities: an application of orthonormal functions. *Australian & New Zealand Journal of Statistics* 3:20–33.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* 9:357–359.

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

Li M, Belmonte JCI. 2017. Ground rules of the pluripotency gene regulatory network. *Nature Reviews Genetics* 18:180–191.

Lin G, Slack JM. 2008. Requirement for Wnt and FGF signaling in Xenopus tadpole tail regeneration. *Developmental biology* 316:323–335.

Looso M, Preussner J, Sousounis K, Bruckskotten M, Michel CS, Lignelli E, Reinhardt R, Höffner S, Krüger M, Tsonis PA. 2013. A de novo assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. *Genome biology* 14:R16.

Martins R, Lithgow GJ, Link W. 2016. Long live FOXO: unraveling the role of FOXO proteins in aging and longevity. *Aging cell* 15:196–207.

Nehrt NL, Clark WT, Radivojac P, Hahn MW. 2011. Testing the ortholog conjecture with comparative functional genomic data from mammals. *PLoS computational biology* 7:e1002073.

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

Sanges D, Romo N, Simonte G, Di Vicino U, Tahoces AD, Fernández E, Cosma MP. 2013. Wnt/β-catenin signaling triggers neuron reprogramming and regeneration in the mouse retina. *Cell reports* 4:271–286.

Schaffer AA, Bazarsky M, Levy K, Chalifa-Caspi V, Gat U. 2016. A transcriptional time-course analysis of oral vs. aboral whole-body regeneration in the Sea anemone Nematostella vectensis. *BMC genomics* 17:718.

Shi K-P, Dong S-L, Zhou Y-G, Li Y, Gao Q-F, Sun D-J. 2019. RNA-seq reveals temporal differences in the transcriptome response to acute heat stress in the Atlantic salmon (Salmo salar). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 30:169–178.

Sikes JM, Newmark PA. 2013. Restoration of anterior regeneration in a planarian with limited regenerative ability. *Nature* 500:77–80.

Stamboulian M, Guerrero RF, Hahn MW, Radivojac P. 2020. The ortholog conjecture revisited: the value of orthologs and paralogs in function prediction. *Bioinformatics* 36:i219–i226.

Stoick-Cooper CL, Weidinger G, Riehle KJ, Hubbert C, Major MB, Fausto N, Moon RT. 2007. Distinct Wnt signaling pathways have opposing roles in appendage regeneration. *Development* 134:479–489.

Sugiura T, Wang H, Barsacchi R, Simon A, Tanaka EM. 2016. MARCKS-like protein is an initiating molecule in axolotl appendage regeneration. *Nature* 531:237.

Sun L, Yang H, Chen M, Ma D, Lin C. 2013. RNA-Seq reveals dynamic changes of gene expression in key stages of intestine regeneration in the sea cucumber Apostichopus

japonicas. *PloS one* 8:e69441.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP. 2014. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research* 43:D447–D452.

Takeo M, Chou WC, Sun Q, Lee W, Rabbani P, Loomis C, Taketo MM, Ito M. 2013. Wnt activation in nail epithelium couples nail growth to digit regeneration. *Nature* 499:228.

Tarashansky AJ, Musser JM, Khariton M, Li P, Arendt D, Quake SR, Wang B. 2021. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *Elife* 10:e66747.

Tothova Z, Gilliland DG. 2007. FoxO transcription factors and stem cell homeostasis: insights from the hematopoietic system. *Cell stem cell* 1:140–152.

True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evolution & development* 3:109–119.

Umesono Y, Tasaki J, Nishimura Y, Hrouda M, Kawaguchi E, Yazawa S, Nishimura O, Hosoda K, Inoue T, Agata K. 2013. The molecular logic for planarian regeneration along the anterior-posterior axis. *Nature* 500:73–76.

Veitia RA. 2005. Paralogs in polyploids: one for all and all for one? *The Plant Cell* 17:4–11.

Willot Q, Mardulyn P, Defrance M, Gueydan C, Aron S. 2018. Molecular chaperoning helps safeguarding mitochondrial integrity and motor functions in the Sahara silver ant Cataglyphis bombycina. *Scientific reports* 8:1–8.

Wu C-H, Tsai M-H, Ho C-C, Chen C-Y, Lee H-S. 2013. De novo transcriptome sequencing of axolotl blastema for identification of differentially expressed genes during limb regeneration. *BMC genomics* 14:434.

Yang C, Gao Q, Liu C, Wang L, Zhou Z, Gong C, Zhang A, Zhang H, Qiu L, Song L. 2017. The transcriptional response of the Pacific oyster Crassostrea gigas against acute heat stress. *Fish & Shellfish Immunology* 68:132–143.

Yi L, Pimentel H, Bray NL, Pachter L. 2018. Gene-level differential analysis at transcript-level resolution. *Genome biology* 19:53.

Yun MH, Gates PB, Brockes JP. 2013. Regulation of p53 is critical for vertebrate limb regeneration. *Proceedings of the National Academy of Sciences* 110:17392–17397.

Zattara EE, Fernández-Álvarez FA, Hiebert TC, Bely AE, Norenburg JL. 2019. A phylum-wide survey reveals multiple independent gains of head regeneration in Nemertea. *Proceedings of the Royal Society B* 286:20182524.

Chapter 3

Babonis, L.S., Enjolras, C., Ryan, J.F., and Martindale, M.Q. 2022. A novel regulatory gene promotes novel cell fate by suppressing ancestral fate in the sea anemone Nematostella vectensis. *Proceedings of the National Academy of Sciences* 119: e2113701119.

Babonis, L.S., and Martindale, M.Q. 2014. Old cell, new trick? Cnidocytes as a model for the evolution of novelty (The Society for Integrative and Comparative Biology).

Baron-Szabo, R.C., Schafhauser, A., Götz, S., and Stinnesbeck, W. 2006. Scleractinian corals from the Cardenas Formation (Maastrichtian), San Luis Potosí, Mexico. *Journal of Paleontology* 80: 1033–1046.

Beaulieu, J., O'Meara, B., Oliver, J., and Boyko, J. 2020. CorHMM: hidden Markov models of character evolution. *R Package Version* 2:.

Beckmann, A., and Özbek, S. 2012. The nematocyst: a molecular map of the cnidarian stinging organelle. *International Journal of Developmental Biology* 56: 577–582.

Benton, M.J., Donoghue, P.C., Asher, R.J., Friedman, M., Near, T.J., and Vinther, J. 2015. Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica* 18: 1–106.

Bobrovskiy, I., Hope, J.M., Ivantsov, A., Nettersheim, B.J., Hallmann, C., and Brocks, J.J. 2018. Ancient steroids establish the Ediacaran fossil Dickinsonia as one of the earliest animals. *Science* 361: 1246–1249.

Boillon, J., Medel, M.D., Pagès, F., Gili, J.-M., Boero, F., and Gravili, C. 2004. Fauna of the Mediterranean hydrozoa. *Scientia Marina* 68: 5–438.

Bouillon, J., Boero, F., and Gravier-Bonnet, N. 1986. Pseudostenotele, a new type of nematocyst, and its phylogenetic meaning within the Haleciidae (Cnidaria, Hydrozoa). *Indo-Malayan Zoology* 3: 63–69.

Boyko, J.D., and Beaulieu, J.M. 2021. Generalized hidden Markov models for phylogenetic comparative datasets. *Methods in Ecology and Evolution* 12: 468–478.

Carlgren, O.H. 1912. Ceriantharia (B. Luno).

Carlgren, O.H. 1940. A Contribution to the Knowledge of the Structure and Distribution of the Cnidae in the Anthozoa,. (CWK Gleerup).

Carrette, T., Alderslade, P., and Seymour, J. 2002. Nematocyst ratio and prey in two Australian cubomedusans, Chironex fleckeri and Chiropsalmus sp. *Toxicon* 40: 1547–1551.

Carrette, T., Straehler-Pohl, I., and Seymour, J. 2014. Early life history of Alatina cf. moseri populations from Australia and Hawaii with implications for taxonomy (Cubozoa: Carybdeida, Alatinidae). *PloS One* 9: e84377.

Cengiz, S., and Killi, N. 2021. Nematocysts types and morphological features of some scyphozoa species in the Southwest Turkey.

Chen, J.-Y., Oliveri, P., Li, C.-W., Zhou, G.-Q., Gao, F., Hagadorn, J.W., Peterson, K.J., and Davidson, E.H. 2000. Precambrian animal diversity: putative phosphatized embryos from the Doushantuo Formation of China. *Proceedings of the National Academy of Sciences* 97: 4457–4462.

Cutress, C.E. 1955. An interpretation of the structure and distribution of cnidae in Anthozoa. *Systematic Zoology* 4: 120–137.

Damian-Serrano, A., Haddock, S.H., and Dunn, C.W. 2021. The evolution of siphonophore tentilla for specialized prey capture in the open ocean. *Proceedings of the National Academy of Sciences* 118: e2005063118.

David, C.N., Özbek, S., Adamczyk, P., Meier, S., Pauly, B., Chapman, J., Hwang, J.S., Gojobori, T., and Holstein, T.W. 2008. Evolution of complex structures: minicollagens shape the cnidarian nematocyst. *Trends in Genetics* 24: 431–438.

Denker, E., Manuel, M., Leclère, L., Le Guyader, H., and Rabet, N. 2008. Ordered progression of nematogenesis from stem cells through differentiation stages in the tentacle bulb of Clytia hemisphaerica (Hydrozoa, Cnidaria). *Developmental Biology* 315: 99–113.

Dohrmann, M., and Wörheide, G. 2017. Dating early animal evolution using phylogenomic data. *Scientific Reports* 7: 3599.

Dong, X.-P., Cunningham, J.A., Bengtson, S., Thomas, C.-W., Liu, J., Stampanoni, M., and

Donoghue, P.C. 2013. Embryos, polyps and medusae of the Early Cambrian scyphozoan Olivooides. *Proceedings of the Royal Society B: Biological Sciences* 280: 20130071.

Dunn, F.S., Kenchington, C.G., Parry, L.A., Clark, J.W., Kendall, R.S., and Wilby, P.R. 2022. A crown-group cnidarian from the Ediacaran of Charnwood Forest, UK. *Nature Ecology & Evolution* 6: 1095–1104.

Dunn, F.S., Liu, A.G., Grazhdankin, D.V., Vixseboxse, P., Flannery-Sutherland, J., Green, E., Harris, S., Wilby, P.R., and Donoghue, P.C. 2021. The developmental biology of Charnia and the eumetazoan affinity of the Ediacaran rangeomorphs. *Science Advances* 7: eabe0291.

Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., and Peterson, K.J. 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* 334: 1091–1097.

Ewer, R.F., and Fox, H.M. 1947. On the functions and mode of action of the nematocysts of Hydra. In Proceedings of the Zoological Society of London, (Wiley Online Library), pp. 365–376.

Ezaki, Y. 2000. Palaeoecological and phylogenetic implications of a new scleractiniamorph genus from Permian sponge reefs, South China. *Palaeontology* 43: 199–217.

Fautin, D.G. 2009. Structural diversity, systematics, and evolution of cnidae. *Toxicon* 54: 1054–1064.

Fedonkin, M.A., and Waggoner, B.M. 1997. The Late Precambrian fossil Kimberella is a mollusc-like bilaterian organism. *Nature* 388: 868–871.

Fuller, M., and Jenkins, R. 2007. Reef corals from the lower Cambrian of the Flinders Ranges, South Australia. *Palaeontology* 50: 961–980.

Garm, A., Lebouvier, M., and Tolunay, D. 2015. Mating in the box jellyfish C opula sivickisi—Novel function of cnidocytes. *Journal of Morphology* 276: 1055–1064.

Gershwin, L.-A. 2006. Nematocysts of the Cubozoa. *Zootaxa* 1232: 30.

Ghisalberti, M., Gold, D.A., Laflamme, M., Clapham, M.E., Narbonne, G.M., Summons, R.E., Johnston, D.T., and Jacobs, D.K. 2014. Canopy flow analysis reveals the advantage of size in the oldest communities of multicellular eukaryotes. *Current Biology* 24: 305–309.

Gold, D.A. 2018. Life in changing fluids: A critical appraisal of swimming animals before the Cambrian. *Integrative and Comparative Biology*.

Gold, D.A., Katsuki, T., Li, Y., Yan, X., Regulski, M., Ibberson, D., Holstein, T., Steele, R.E., Jacobs, D.K., and Greenspan, R.J. 2019a. The genome of the jellyfish Aurelia and the evolution of animal complexity. *Nature Ecology & Evolution* 3: 96–104.

Gold, D.A., Lau, C.L.F., Fuong, H., Kao, G., Hartenstein, V., and Jacobs, D.K. 2019b. Mechanisms of cnidocyte development in the moon jellyfish Aurelia. *Evolution & Development* 21: 72–81.

Gold, D.A., Runnegar, B., Gehling, J.G., and Jacobs, D.K. 2015. Ancestral state reconstruction of ontogeny supports a bilaterian affinity for Dickinsonia. *Evolution & Development* 17: 315–324.

Han, J., Hu, S., Cartwright, P., Zhao, F., Ou, Q., Kubota, S., Wang, X., and Yang, X. 2016. The earliest pelagic jellyfish with rhopalia from Cambrian Chengjiang Lagerstätte. *Palaeogeography, Palaeoclimatology, Palaeoecology* 449: 166–173.

Han, J., Kubota, S., Uchida, H., Stanley Jr, G.D., Yao, X., Shu, D., Li, Y., and Yasui, K. 2010. Tiny sea anemone from the Lower Cambrian of China. *PLoS One* 5: e13276.

Iten, H.V., Leme, J.M., Pacheco, M.L., Simões, M.G., Fairchild, T.R., Rodrigues, F., Galante, D., Boggiani, P.C., and Marques, A.C. 2016. Origin and early diversification of phylum Cnidaria: key macrofossils from the Ediacaran System of North and South America. In The Cnidaria, Past, Present and Future, (Springer), pp. 31–40.

Jell, J.S., Cook, A.G., and Jell, P.A. 2011. Australian Cretaceous Cnidaria and Porifera. *Alcheringa* 35: 241–284.

Johnson, R.G., and Richardson, E.S. 1968. Pennsylvanian invertebrates of the Mazon Creek area, Illinois: The Essex fauna and medusae.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A., and Jermiin, L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.

Kass-Simon, G., and Scappaticci, A.A. 2002. The behavioral and developmental physiology of nematocysts. *Canadian Journal of Zoology* 80: 1772–1794.

Kayal, E., Bentlage, B., Sabrina Pankey, M., Ohdera, A.H., Medina, M., Plachetzki, D.C., Collins, A.G., and Ryan, J.F. 2018. Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evolutionary Biology* 18: 1–18.

Laumer, C.E., Fernández, R., Lemer, S., Combosch, D., Kocot, K.M., Riesgo, A., Andrade, S.C., Sterrer, W., Sørensen, M.V., and Giribet, G. 2019. Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B* 286: 20190831.

Leclère, L., Horin, C., Chevalier, S., Lapébie, P., Dru, P., Péron, S., Jager, M., Condamine, T., Pottin, K., and Romano, S. 2019. The genome of the jellyfish Clytia hemisphaerica and the evolution of the cnidarian life-cycle. *Nature Ecology & Evolution* 3: 801–810.

Liu, A.G., Matthews, J.J., Menon, L.R., McIlroy, D., and Brasier, M.D. 2014. Haootia quadriformis n. gen., n. sp., interpreted as a muscular cnidarian impression from the Late Ediacaran period (approx. 560 Ma). In Proc. R. Soc. B, (The Royal Society), p. 20141202.

Mariscal, R.N. 1974. Scanning electron microscopy of the sensory surface of the tentacles of sea anemones and corals. *Zeitschrift Für Zellforschung Und Mikroskopische Anatomie* 147: 149–156.

Mariscal, R.N., Conklin, E.J., and Bigger, C.H. 1977a. The ptychocyst, a major new category of cnida used in tube construction by a cerianthid anemone. *The Biological Bulletin* 152: 392–405.

Mariscal, R.N., McLean, R.B., and Hand, C. 1977b. The form and function of cnidarian spirocysts. *Cell and Tissue Research* 178: 427–433.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., and Lanfear, R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.

Nüchter, T., Benoit, M., Engel, U., Özbek, S., and Holstein, T.W. 2006. Nanosecond-scale kinetics of nematocyst discharge. *Current Biology* 16: R316–R318.

Östman, C. 2000. A guideline to nematocyst nomenclature and classification, and some notes on the systematic value of nematocysts. *Scientia Marina* 64: 31–46.

Parry, L.A., Boggiani, P.C., Condon, D.J., Garwood, R.J., Leme, J. de M., McIlroy, D., Brasier, M.D., Trindade, R., Campanha, G.A., and Pacheco, M.L. 2017. Ichnological evidence for meiofaunal bilaterians from the terminal Ediacaran and earliest Cambrian of Brazil. *Nature Ecology & Evolution* 1: 1455.

Picciani, N., Kerlin, J.R., Sierra, N., Swafford, A.J., Ramirez, M.D., Roberts, N.G., Cannon, J.T.,

Daly, M., and Oakley, T.H. 2018. Prolific origination of eyes in Cnidaria with co-option of non-visual opsins. *Current Biology* 28: 2413–2419.

Purcell, J.E. 1984. The functions of nematocysts in prey capture by epipelagic siphonophores (Coelenterata, Hydrozoa). *The Biological Bulletin* 166: 310–327.

dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C., and Yang, Z. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Current Biology* 25: 2939–2950.

Revell, L.J. 2014. Package "phytools": Phylogenetic tools for comparative biology (and other things) (WWW document] URL http://cran. r-project. org/package= phytools [accessed 15 ...).

Rifkin, J.F. 1991. A study of the spirocytes from the Ceriantharia and Actiniaria (Cnidaria: Anthozoa). *Cell and Tissue Research* 266: 365–373.

Runnegar, B. 2022. Following the logic behind biological interpretations of the Ediacaran biotas. *Geological Magazine* 159: 1093–1117.

Schmidt, H. 1974. On evolution in the Anthozoa. In Proceedings of the Second International Coral Reef Symposium, 1974, (Great Barrier Reef Committee), pp. 533–560.

Sebé-Pedrós, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhé, M.-P., Renno, J., Loe-Mie, Y., Lifshitz, A., Mukamel, Z., and Schmutz, S. 2018. Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* 173: 1520–1534.

Siddall, M.E., Martin, D.S., Bridge, D., Desser, S.S., and Cone, D.K. 1995. The demise of a phylum of protists: phylogeny of Myxozoa and other parasitic Cnidaria. *The Journal of Parasitology* 961–967.

Siebert, S., Farrell, J.A., Cazet, J.F., Abeykoon, Y., Primack, A.S., Schnitzler, C.E., and Juliano, C.E. 2019. Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science* 365: eaav9314.

Song, X., Ruthensteiner, B., Lyu, M., Liu, X., Wang, J., and Han, J. 2021. Advanced Cambrian hydroid fossils (Cnidaria: Hydrozoa) extend the medusozoan evolutionary history. *Proceedings of the Royal Society B* 288: 20202939.

Sperling, E.A., Peterson, K.J., and Laflamme, M. 2011. Rangeomorphs, Thectardis (Porifera?) and dissolved organic carbon in the Ediacaran oceans. *Geobiology* 9: 24–33.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4: vey016.

Tardent, P. 1995. The cnidarian cnidocyte, a hightech cellular weaponry. *BioEssays* 17: 351–362.

Wolfe, J.M. 2017. Metamorphosis is ancestral for crown euarthropods, and evolved in the Cambrian or earlier. *Integrative and Comparative Biology* 57: 499–509.

Zapata, F., Goetz, F.E., Smith, S.A., Howison, M., Siebert, S., Church, S.H., Sanders, S.M., Ames, C.L., McFadden, C.S., and France, S.C. 2015. Phylogenomic analyses support traditional relationships within Cnidaria. *PLoS One* 10: e0139068.

Chapter 4

Alvarado AS. 2000. Regeneration in the metazoans: why does it happen? *Bioessays* 22:578590.

Alvarado AS, Yamanaka S. 2014. Rethinking differentiation: stem cells, regeneration, and plasticity. *Cell* 157:110–119.

Bely AE, Sikes JM. 2010. Latent regeneration abilities persist following recent evolutionary loss in asexual annelids. *Proceedings of the National Academy of Sciences* 107:1464–1469.

Bosch TC, Anton-Erxleben F, Hemmrich G, Khalturin K. 2010. The Hydra polyp: nothing but an active stem cell community. *Development, growth & differentiation* 52:15–25.

Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, Lee TJ, Leigh ND, Kuo T-H, Davis FG. 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell reports* 18:762–776.

Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature methods* 18:366–368.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC bioinformatics* 10:421.

Cary GA, Wolff A, Zueva O, Pattinato J, Hinman VF. 2019. Analysis of sea star larval regeneration reveals conserved processes of whole-body regeneration across the metazoa. *BMC biology* 17:1–19.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* 39:W29–W37.

Gold DA, Jacobs DK. 2013. Stem cell dynamics in Cnidaria: are there unifying principles? *Development genes and evolution* 223:53–66.

Gold DA, Katsuki T, Li Y, Yan X, Regulski M, Ibberson D, Holstein T, Steele RE, Jacobs DK, Greenspan RJ. 2019. The genome of the jellyfish Aurelia and the evolution of animal

complexity. *Nature ecology & evolution* 3:96–104.

Gold DA, Lau CLF, Fuong H, Kao G, Hartenstein V, Jacobs DK. 2019. Mechanisms of cnidocyte development in the moon jellyfish Aurelia. *Evolution & development* 21:72–81.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8:1494–1512.

Juliano CE, Reich A, Liu N, Götzfried J, Zhong M, Uman S, Reenan RA, Wessel GM, Steele RE, Lin H. 2014. PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. *Proceedings of the National Academy of Sciences* 111:337–342.

Koizumi O, Heimfeld S, Bode HR. Plasticity in the nervous system of adult hydra. II. Conversion of ganglion cells of the body column into epidermal sensory cells of the hypostome. Dev Biol. 1988 Oct;129(2):358-71.

Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. 2015. The technology and biology of single-cell RNA sequencing. *Molecular cell* 58:610–620.

Li Y, Peng S, Liu Y, He K, Sun K, Yu Z, Ma Y, Wang F, Xu P, Sun T. 2023. Single-cell transcriptomic analyses reveal the cellular and genetic basis of aquatic locomotion in scyphozoan jellyfish. *bioRxiv*:2023–02.

Miramon-Puertolas P, Steinmetz PR. 2023. An adult stem-like cell population generates germline and neurons in the sea anemone Nematostella vectensis. *bioRxiv*:2023–01.

Mochizuki K, Sano H, Kobayashi S, Nishimiya-Fujisawa C, Fujisawa T. 2000. Expression and evolutionary conservation of nanos-related genes in Hydra. *Development genes and evolution* 210:591–602.

Nakanishi N, Hartenstein V, Jacobs DK. 2009. Development of the rhopalial nervous system in Aurelia sp. 1 (Cnidaria, Scyphozoa). *Development genes and evolution* 219:301–317.

Piraino S, Boero F, Aeschbach B, Schmid V. 1996. Reversing the life cycle: medusae transforming into polyps and cell transdifferentiation in Turritopsis nutricula (Cnidaria, Hydrozoa). *The Biological Bulletin* 190:302–312.

Primack AS, Cazet JF, Morris Little H, Mühlbauer S, Cox BD, David CN, Farrell JA, Juliano CE. 2023. Differentiation trajectories of the *Hydra* nervous system reveal transcriptional regulators of neuronal fate. *bioRxiv*:2023.03.15.531610

Röttinger E. 2021. Nematostella vectensis, an emerging model for deciphering the molecular and cellular mechanisms underlying whole-body regeneration. *Cells* 10:2692.

Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* 33:495–502.

Sebé-Pedrós A, Saudemont B, Chomsky E, Plessier F, Mailhé M-P, Renno J, Loe-Mie Y, Lifshitz A, Mukamel Z, Schmutz S. 2018. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-Seq. *Cell* 173:1520–1534.

Siebert S, Anton-Erxleben F, Bosch TCG. 2008. Cell type complexity in the basal metazoan Hydra is maintained by both stem cell based mechanisms and transdifferentiation. *Developmental Biology* 313:13-24.

Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. 2019. Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science* 365:eaav9314.

Sierra N, Olsman N, Yi L, Pachter L, Goentoro L, Gold DA. 2021. A novel approach to comparative RNA-Seq does not support a conserved set of genes underlying animal

regeneration. *bioRxiv*:2021–03.

Steger J, Cole AG, Denner A, Lebedeva T, Genikhovich G, Ries A, Reischl R, Taudes E, Lassnig M, Technau U. 2022. Single-cell transcriptomics identifies conserved regulators of neuroglandular lineages. *Cell Reports* 40:111370.

Turk T, Kem WR. 2009. The phylum Cnidaria and investigations of its toxins and venoms until 1990. *Toxicon* 54:1031–1037.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology* 11:R14.

Yuan D, Nakanishi N, Jacobs DK, Hartenstein V. 2008. Embryonic development and metamorphosis of the scyphozoan Aurelia. *Development genes and evolution* 218:525–539.

Zakrzewski W, Dobrzyński M, Szymonowicz M, Rybak Z. 2019. Stem cells: past, present, and future. *Stem cell research & therapy* 10:1–22.