universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Insights into the evolution of Gene Regulatory networks from the bioinformatic analysis of transcriptomic and genomic data“

verfasst von / submitted by

fil.mag. Rohit Dnyansagar

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 685 437
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet	Biologie
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Biol. Dr. Ulrich Technau

Preface

This thesis was prepared at the Department of Neuroscience and Developmental Biology of University of Vienna, Austria. The PhD work was funded by grants from the Austrian Science Fund (FWF) to my supervisor Prof. Ulrich Technau. Besides his direct mentorship, my work was directly supervised by Dr. Bob Zimmerman in the Technau group. I would like to thank Bob for providing every support possible to me, technical and otherwise. I would also like to thank all my colleagues at Technau lab who have provided excellent and fun working environment. I would also like to thank Dr. Grigory Genikhovich for his guidance and support. He is my go-to person for anything related to cnidaria, evolution, development, and other experimental biology questions. I would also like to thank many of my colleagues in Techanu lab, Patricio Ferrer Murguia, who is also part of brachyury project and a dear friend. My other colleagues Ekaterina Pukhlyakova, Daniela Praher and Julia Steger made working in the Technau lab, a friendly, supportive environment. Finally, I would like to thank my parents for all their support and a special mention to my wife Harshada, who has been a backbone of my support environment at home.

Table of Contents

Abstract

Diese Dissertation stellt die Arbeit dar, die an der Abteilung für Neurowissenschaften und Entwicklungsbiologie der Universität Wien, Österreich, durchgeführt wurde. Diese Dissertation besteht aus zwei verschiedenen Projekten, die sich auf die bioinformatische Analyse von zuvor erhobenen Daten konzentrieren.

Im ersten Projekt, das in Kapitel 2.2 beschrieben wird, wird untersucht die genetischen Beziehungen zwischen unserem Hauptmodellorganismus *Nematostella vectensis* und zwei eng verwandten Mitgliedern der Seeanemonenfamilie Edwardsiidae, *Edwardsiella carnea* (von der skandinavischen Küste) und *Edwardsiella lineata* (von der ostamerikanischen Küste). Übrigens hat *E. lineata* ein interessantes fakultativ parasitisches Planula-Larvenstadium, während über den Lebenszyklus von *E. carnea* nur wenig bekannt war. Morphologisch sind diese beiden Arten praktisch identisch, so dass ein genetischer Vergleich erforderlich war, um sie zu unterscheiden. Unsere vergleichenden Transkriptomanalysen von *E. lineata*, *E. carnea* und parasitären Planulae unbekannter genetischer Herkunft von der schwedischen Küste ließen uns zu dem Schluss kommen, dass *E. lineata* und *E. carnea* zwei sehr eng verwandte, aber dennoch unterschiedliche Arten sind und dass *E. carnea* auch ein Larvenstadium besitzt, das Ctenophoren parasitiert.

Das zweite Projekt, das in einem Manuskript in Kapitel 3.2 ausführlich beschrieben wird, befasst sich mit der Evolution des Genregulationsnetzwerks des Transkriptionsfaktors Brachyury. Bei Wirbeltieren wird Brachyury seit langem wegen seiner entscheidenden Rolle bei der Entwicklung einiger lebenswichtiger Körperstrukturen wie Notochord, Muskel-Skelett- und Herzgewebe untersucht (King et al., 1998; Li et al., 2021). Brachyury ist nicht nur eine wichtige mesodermale Determinante, sondern wurde vor kurzem auch als Signaturgen für eine bipotente Zellpopulation, die so genannten neuromesodermalen Vorläuferzellen (NMPs), im Schwanzknospenbereich des Wirbeltierembryos identifiziert.

Außerhalb der Bilateralen wird Brachyury auch in einigen Pilzen, einzelligen Eukaryoten und vielen basalen Metazoen exprimiert, doch seine Funktion in nicht-bilateralen Arten ist unklar. Um die Evolution der Brachyury-Funktion nachzuvollziehen, analysierten wir ChIP-seq- und RNA-seq-Daten, die in unserem Labor für *N. vectensis* und im Labor von Ina Arnone (Stazione Zoologica Neapel) für *S. purpuratus* generiert wurden, sowie zuvor veröffentlichte Datensätze von verschiedenen Chordaten und einzelligen Eukaryonten. Wir identifizierten eine uralte Wnt-Brachyury-Rückkopplungsschleife, die bei den meisten Tieren an der Achsenbildung und -musterung beteiligt war. Wir haben auch gezeigt, dass sich wichtige Zielgene, die die

mesodermale Funktion von Brachyury in Wirbeltieren vermitteln, erst an der Basis von Chordaten oder Wirbeltieren entwickelt haben, was auf eine signifikante Verschiebung der Funktion dieses konservierten Transkriptionsfaktors hinweist.

This Ph.D. thesis represents the work carried out at the Department of Neurosciences and Developmental Biology of University of Vienna, Austria. This thesis is comprised of two distinct projects focusing on the bioinformatic analyses of previously acquired data.

In the first project, described in a paper in chapter 2.2, we investigated the genetic relationships of our main model organism *Nematostella vectensis* and two closely related members of the sea anemone family Edwardsiidae, *Edwardsiella carnea* (from the Scandinavian coast) and *Edwardsiella lineata* (from the East American coast). Incidentally, *E. lineata* has an interesting facultative parasitic planula larva stage, while the life cycle of *E. carnea* was mostly limited. Morphologically these two species are virtually identical and therefore required genetic comparison to distinguish them. Our comparative transcriptomic analyses of *E. lineata*, *E. carnea*, and parasitic planulae of unknown genetic origin from the Swedish coast allowed us to conclude that *E. lineata* and *E. carnea* are two very closely related, yet distinct, species, and that *E. carnea* also possesses a larval stage, which parasitizes ctenophores.

The second project described in detail in a manuscript in Chapter 3.2 looks at the evolution of the gene regulatory network of the transcription factor Brachyury. In vertebrates, Brachyury has been studied for a long time for its crucial role in the development of some of the vital body structures like notochord, musculoskeletal, and cardiac tissues (King et al., 1998; Li et al., 2021). In addition to being a key mesodermal determinant, recently *brachyury* has been identified as a signature gene for a bipotent cell population called neuromesodermal progenitors (NMPs) in the tailbud of the vertebrate embryo.

Outside bilaterians, *brachyury* expression is shown in some fungi, single-celled eukaryotes, and many basal metazoans, however, its function in non-bilaterian species is obscure. To trace the evolution of Brachyury function, we analyzed ChIP-seq and RNA-seq data generated for *N. vectensis* in our lab and the lab of Ina Arnone (Stazione Zoologica Naples) for *S. purpuratus* as well as previously published dataset from several chordates along with single-celled eukaryotes. We identified an ancestral Wnt-Brachyury feedback loop that was involved in axis formation and patterning of most animals. We also showed that key target genes conveying the mesodermal function of Brachyury in vertebrates only evolved at the base of chordates or vertebrates, indicating a significant shift in function of this conserved transcription factor.

Chapter 1: INTRODUCTION

1.1 Why study marine species?

Our understanding of the evolutionary mechanisms is mainly based on the studies of terrestrial and freshwater organisms (Beheregaray et al., 2015). Although there is a significant increase in the studies focusing on the marine species, the studies focusing on speciation in marine environments are still limited. This might create a strong bias in our understanding of speciation processes. Land and fresh-water environments are significantly different than oceans. The primary factor is the continuous environment of the oceans with additional factors like oxygen concentration, density, and solubility of water etc. The continuous nature of the oceans significantly reduces the effectiveness of natural barriers. Furthermore, marine species more often possess eggs and/or larvae that can disperse over a long-range or sometimes can have asexual propagates capable of surviving months in water. All these factors affect an organism's dispersal potential and the gene exchanges and thus speciation. Many ecological speciation mechanisms remain to be examined in the marine environment, which is unfortunate as the marine environment more often pose a serious challenge to the theory of allopatric speciation. In the marine environment, there is no absolute barrier to the gene flow, therefore, populations, which are far apart can still be connected genetically. Additionally, the population size in marine environment tends to be large, which may slow the genetic divergence of populations. Considering all these factors one would expect speciation with allopatric speciation infrequent and slow, yet observation of marine habitat shows an abundance of the species variety (Bowen et al., 2013; Hilbish, 1996). Therefore, it is essential to study speciation in the marine environment to understand how reproductive isolation evolves in the marine speciation events.

1.2 Phylum Cnidaria

1.2.1 General Classification

Within Metazoa, most animals belong to the Bilateria, which are animals with bilateral symmetry and three germ layers (ectoderm, endoderm, mesoderm). Cnidaria forms a sister phylum to the Bilateria, which branched off about 650 million years ago (Dunn et al., 2008). The phylum Cnidaria is characterized by the phylum-specific cell type, the cnidocytes (also

known as nematocytes), which are stinging cells with a single giant secretory vesicle containing the cnidocyst. These cnidocytes are essential to capture food and for the self-defense.

There are five groups[i] within phylum Cnidaria, namely Anthozoa, Cubozoa, Hydrozoa, Scyphozoa and Staurozoa. Scyphozoa, Cubozoa, Staurozoa and Hydrozoa are united as Medusozoa due to their normal ability to form medusa, unlike the representatives of the class Anthozoa (corals, sea anemones, sea pansies, sea fans, and sea pens), which only form polyps (Kayal et al., 2013; Zapata et al., 2015).

The class Anthozoa, which contains more than 7,500 species (Daly et al., 2007), is of special interest in this study. The anthozoan body plan is essentially a tube-shaped sac with a pharynx at the oral disk. The oral disk is surrounded by the tentacles to catch prey with the help of cnidocytes. Anthozoans like sea anemones can be solitary wherein they will either attach themselves to a solid surface or burrow in the sand for support or be colonial like corals that build large colonies more commonly known as coral reefs. Commonly anthozoans are further subclassified into two classes, namely Alcyonaria (or Octocorrallia) and Zoantharia (or Hexacorallia)(Kayal et al., 2013; Zapata et al., 2015).

1.2.2 Cnidarian model organisms

The phylogenetic position of the Cnidaria in the tree of life as a sister lineage to Bilateria is crucial to study the origin of many bilaterian traits.

The simple cnidarian nervous system, mostly consisting of neural nets provides insights into the ancestral conserved genetic repertoire necessary for the development of the nervous system. Similarly, although cnidarians are often considered as radially symmetric, anthozoans have, unlike medusozoans, two axes of symmetry, first an oral-aboral axis (Fritzenwanker et al., 2007) and a second axis, termed the directive axis (Rentzsch et al., 2006; Saina et al., 2009). To understand the origin and evolution of bilaterality in animals, the role of BMP signaling has been studied in *N. vectensis*. These studies have revealed a surprising complex network of interactions of BMPs and BMP antagonists, leading to a gradient of pSMAD along the directive axis, which controls the staggered expression of several Hox genes (Genikhovich and Technau, 2017; Genikhovich et al., 2015; Technau and Genikhovich, 2018). Thus, cnidarians have provided valuable insights into the ancestral gene regulatory network required for axis formation.

Cnidarians are also highly informative regarding the origin of germ layers. They consist only of two cell layers, commonly termed ectoderm, and endoderm. Consequently, the third germ

layer, the mesoderm was lacking in Cnidaria. This has prompted research with the goal to understand the evolution of the mesoderm from a diploblastic ancestor.

To trace the origin of mesoderm, diploblastic cnidarians have been investigated for the presence of key mesodermal factors. Notably, most of the bilaterian developmental regulator genes of mesoderm have been identified in various cnidarian model organisms. These included *brachyury*, *snail*, *twist*, *mef2*, and *MyoD*-like transcription factors. The presence of these marker genes and their expression patterns led to the discussion of three scenarios that could explain the lack of mesoderm in cnidarians and the origin of mesoderm in bilateral animals. In the first scenario, these genes played no role in germ-layer specification in this scenario, and a neo-functionalization of these genes in mesoderm formation in Bilateria was considered. In the second scenario, these genes were thought to have a role in the germ-layer specification in diploblastic organisms, in the third scenario, the common ancestor of cnidaria-Bilateria was considered triploblastic and loss of middle germ layer, mesoderm in cnidarians was discussed (Martindale et al., 2004). Interestingly, in a recent study (Steinmetz et al., 2017), the authors show the presence of gene expression program of all three germ layers in Cnidarian model organism *N. vectensis* in topologically isolated regions challenges the canonical germ layer perspective of Cnidaria. These studies provided some valuable insights into the components of the ancestral gene repertoire and regulation of mesoderm and its derivatives.

Apart from their sister lineage relationship to bilaterians and their ideal position for tracing the origins of many bilateral traits, cnidarians have many fascinating intrinsic properties such as regenerative capacity, slow/lack of ageing and other niche properties like the presence of cnidae with venom and in some cases parasitism (Darling et al., 2005; Frank et al., 2001; Genikhovich and Technau, 2009; Kuhn et al., 1996; Miller and Ball, 2000; Putnam et al., 2007; Steele, 2002). Properties of regeneration in cnidarian have been known from ancient Greeks and remain a fascinating aspect of cnidarian study while other properties like ageing, and parasitism are recently noticed, but they are appealing to many in science.

In recent times the ecological impact of cnidarians such as corals, jellyfish has been another area of intense research. The effect of climate change on the corals around the world has been a matter of concern for ecologists and environmentalists alike. This also highlighted the importance of coral reefs as part of an important marine ecosystem, which provides valuable resources for marine life. A recent study pointed out the importance of jellyfish as an important source of food for many marine animals, which earlier was ignored aspect of jellyfish (Hays et al., 2018; Thiebot et al., 2016). For all these reasons the importance of studying Cnidaria as a

model system cannot be overstated and, in this regard, some of the well-known model systems are emerging from the phylum Cnidaria like *Nematostella vectensis* and *Acropora millepora* belonging to class Anthozoa while as *Clytia hemisphaerica, Hydractinia symbiolongicarpus* and *Hydra magnipapilata* belonging to class Hydrozoa, among others.

1.2.3 Nematostella vectensis as a model organism

Among the rising model organisms from the phylum Cnidaria is the anthozoan sea anemone *N. vectensis* from the family Edwardsiidae. Found in brackish waters of both Atlantic and Pacific Ocean this species was first described by T.A. Stephenson in 1935 (Stephenson, 1935). In the early 1990s, by the work of Hand and Uhlinger (Hand and Uhlinger, 1992, 1994, 1995) and with molecular biology techniques, the species was established as a model organism and later also became the first non-bilateral species to have its genome sequenced (Putnam et al., 2007). The genome sequence revealed that its surprisingly high genetic complexity, which is much closer to the vertebrates than other non-vertebrate model organisms such as *C. elegans* or *D. melanogaster*. Its remarkably simple tissue organization combined with complex gene structure made it a particularly suitable candidate for tracing many bilateral features such as central nervous system, bilateral symmetry, musculoskeletal system, etc. It can reproduce both sexually and asexually and can be maintained in the laboratory with relative ease. The animals can reach sexual maturity within 3-6 months. With many protocols established to use modern molecular biology and bioinformatics techniques *N. vectensis* continue to represent a major non-bilateral model organism.

1.2.4 Cnidarian Germ Layers

Non-bilaterian animals like sponges, cnidarians, placozoans and ctenophores form two germ layers and are therefore called diploblastic animals. In these organisms, the middle layer mesoderm is lacking, instead they form a gelatinous, more often acellular layer, called mesoglea. The mesoglea is sometimes mistakenly considered as "mesodermal layer"; however, it is an extracellular matrix with a composition similar to the basal lamina in vertebrates.
The debate about the cnidarian germ layers and its homologous structures in vertebrates has been ongoing for a long time due to evidences found in both major groups of cnidarians i.e., medusozoans as well as anthozoans. Among the hydrozoans, medusae of several species bud off from polyp or rarely from the medusa during the lateral budding process. During which the ectodermal material, called entocodon gets detached from epidermis and lies in between

epidermis and gastrodermis. This entocodon develops a cavity by developing epithelial layer from mass of cells. Entocodon is only a transient stage before the epithelium organization of the daughter medusa. Among medusozoans, some hydrozoans develop a discrete tissue called entocodon, which appears during the lateral budding process of medusae from polyps. (Korschelt and Heider, 1890). The origin of the entocodon is thought to be ectodermal, however the resulting entocodon is independent of both ectoderm and endoderm. Many studies (Fautin and Mariscal, 1991; Martindale et al., 2004; Seipel and Schmid, 2004)
have reported the presence of epitheliomuscular cells/striated muscles in entocodon and therefore correlated it to the mesoderm. The striated muscles are mesodermal derivatives in triploblastic animals; therefore, the hypothesis of entocodon-mesoderm homology was put forth.

In the study of striated muscles in jellyfish medusa development, which is remarkably reminiscent of vertebrate development, authors suggested that the diploblastic stage of the development in these species may be derived one and coin term for the common ancestors of triploblasts and diploblasts, the Urtriploblast (Seipel and Schmid, 2005).

However, there is lack of sufficient evidences to support this hypothesis, for instance there is hardly any similarity between developmental process of entocodon and mesoderm and therefore the entocodon-mesoderm homology hypothesis is in large part based solely on the presence of striated muscles in entocodon of few hydrozoans (Burton, 2008). Therefore, the scenario where ancestors of both cnidarians and triploblastic animals possessing striated muscles while entocodon and mesoderm evolved independently is more likely.

By contrast, earlier arguments, put forth by the Hertwig brothers (Hertwig and Hertwig, 1878), suggested the presence of a mesenchymal layer in Anthozoa reminiscent of the mesoderm of higher phyla. Yet, this mesenchymal layer is the mesoglea, which forms between the inner and the outer cell layer as an extracellular matrix and which should not be confounded with the third germ layer.

In a recent study by Steinmetz and colleagues (Steinmetz et al., 2017), lineage tracing of the blastula and gastrula cells as well as a broad in-situ hybridization screen of many endodermal and mesodermal markers including 33 transcription factors and 17 other important markers such as digestive enzymes suggested the homology of cnidarian endoderm and triploblastic mesoderm on the one hand and of cnidarian pharyngeal ectoderm and triploblastic endoderm on the other hand. With this observation the authors introduce a new dimension to the concept of germ layers where diploblastic cnidarians have already a topological separation into three germ layers (Steinmetz et al., 2017).

Therefore, the debate of germ layer homology continues and must be updated with additional evidence. Regardless, the simple diploblastic cnidarians provide an excellent opportunity for comparative genomic studies with bilaterian models to trace the origins of some of bilaterian characteristics.

Chapter 2: SPECIES DISTINCTION WITHIN THE EDWARDSIIDAE: THE CASE OF THE GENUS EDWARDSIELLA

"No one definition has yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species" -- Charles Darwin (On the Origin of Species, Chapter 2)

2.1 Introduction

2.1.1 The sea anemone family Edwardsiidae

In the first chapter we looked at the phylum Cnidaria, which consists of five classes, namely Anthozoa, Cubozoa, Hydrozoa, Scyphozoa and Staurozoa. Within the class Anthozoa, we are interested in the family of sea anemones Edwardsiidae, which is characterized by four pairs of mesenteries and named in the honor of French zoologists Henri Milne-Edwards. This family belongs to the order Actiniaria comprised by several types of sea anemones. Almost all adult anthozoans are attached or embedded to the seabed or rocks while their larvae are free swimming. The family Edwardsiidae is the most speciose among the order Actiniaria with about nine accepted genera and more than 90 species which are widely distributed in diverse habitats, ranging from warm tropical to polar environments (Izumi and Fujita, 2018). Members of this family are also found in waters with high salinity and at all depths. Apart from the diversity of body forms and distribution, members of the Edwardsiidae family also show diversity of life cycles, ranging from the
solitary independent life cycle of *Nematostella vectensis* to the more complex, parasitic life cycles of *Edwardsiella lineata*. In a fascinating recent discovery, a life cycle of one of the sea anemones, *Tempuractis rinkai gen. et sp. nov.* was found to be symbiotic with marine sponge species (Izumi et al., 2018).

2.1.2 Genetic relationship of Nematostella and Edwardsiella species

One of the emerging model organisms from the phylum Cnidaria is the starlet sea anemone *Nematostella vectensis*, which has become a major model system in EvoDevo, EcoDevo, neuroscience, regenerative biology, stress response etc. The genome of *Nematostella vectensis* has been sequenced (Putnam et al., 2007; Zimmermann et al., 2021), the next closest relative with sequenced genome was that of stony coral Acropora. Currently the closest relative to *N.*

vectensis with sequenced genome is the sea anemone *Exaiptasia pallida* which belongs to the monotypic genus *Exaiptasia* which belongs to the family Aiptasiidae.

For comparative genomic reasons, we needed to identify and get sequence information of close relatives. The genomic comparison with closest species would help us understand evolutionary question, for instance whether a gene loss or gene gain in *N. vectensis* is specific to *N. vectensis* or it is a feature of genus or family. Genetic comparison within a family will also help us define the family specific gene repertoire.

Amongst the nine genera that constitute *Edwardsiidae*, there are several, which are morphologically similar to *Nematostella*, one of them is the genus *Edwardsiella* consisting of around 14 accepted sea anemone species characterized by the lack of tenaculi, nemathybomes and nematosomes (Daly et al., 2013; Izumi et al., 2018). In *N. vectensis* tenaculi, nemathybomes are also absent but nematosomes are present. A similar genus of sea anemones, *Edwardsia* has similar morphological characteristics, however, phylogenetic analysis suggest that the genus *Edwardsia* is paraphyletic (Daly, 2005).

We obtained specimen of *Edwardsiella carnea Gosse* 1856 from the Skagerrak region close to the Swedish coast. This species has been renamed several times starting with *Edwardsia carnea*, *Fagesia carnea*, *Halcampa microps Gosse* and *Milne-Edwardsia carnea*. Its current allocation to *Edwardsiella* genus is based on the molecular phylogenetic analysis during the systematic analysis of Edwardsiidae family (Daly, 2005). While the discussion about its classification continued for a long time, its life cycle aspects remained largely unexplored. A single record suggests that it could be a parasite of one of the local comb jellies, *Bolinopsis infundibulum* (Selander et al., 2010).

Incidentally, *E. carnea* is morphologically remarkably similar to another sea anemone species, *Edwardsiella lineata*, found at other coast of Atlantic Ocean, the American East Coast. Unlike *E. carnea*, the American sea anemone *E. lineata* has been studied regarding its life cycle and its transcriptome (Reitzel et al., 2009; Stefanik et al., 2014). The fascinating biphasic life cycle of *E. lineata* involves a parasitic phase and a sessile life phase. The larva of the *E. lineata* infects the comb jelly *Mnemiopsis leidyi* and feeds off the host. However, when outside the host, the larva is capable of developing into an adult sessile polyp. Interestingly, the opportunistic adult polyp is also capable of reverting to the larval stage if it is able to infect another comb jelly. The morphological similarity between these two *Edwardsiella* species makes it difficult to distinguish between the two. Among the 60 morphological characters compared between these two species, there are only two differences found in the retractor muscle region (Daly, 2005).

Therefore, the primary question we try to address here is, whether these are two separate species or whether they are populations of the same species at separate locations? The host of *E. lineata* parasites, *M. leidyi* is native to the American East coast, however, in the 1980s it was accidentally introduced to the Black Sea with ballast water of the ships and since 2002 its periodic bloom has been observed also in the North Sea (Selander et al., 2010; Shiganova et al., 2001). Notably, we also found the invasive comb jelly *M. leidyi* at the Swedish coast infected with putative *Edwardsiella sp.* parasites. The parasite within the comb jelly is vermiform and devoid of any other morphological feature, therefore, identification based on morphology is difficult.

This raised another question, whether the invasive ctenophores have carried the parasitic larvae of the American *E. lineata* to the Swedish coast, or whether the parasitic larvae originated from the Swedish population of *E. carnea.* Therefore, the two species of *Edwardsiella* and the parasite (mentioned henceforth as E. Parasite sp.) may form a cryptic species complex. Since the morphological comparison failed to provide any characteristic to distinguish between the two *Edwardsiella* species and was also unable to identify the parasite we decided to use transcriptomic data to answer these questions.

2.2 Manuscript Edwardsiella

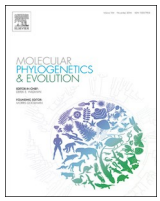
For this project many people have contributed towards its completion. Per Sundberg, Line Friis Möller and Ulrich Technau collected the Edwardsiella carnea specimens, Daniela Praher and Yehu Moran isolated RNA. Afterwards I and you and Bob Zimmermann did all the bioinformatic analyses. The first draft of the manuscript was prepared after the cross-species transcriptome comparison, where also I assembled transcriptomes of Edwardsiella parasite sp. and *Edwardsiella carnea*.

Molecular Phylogenetics and Evolution 126 (2018) 346–35 5

Contents lists ScienceDir

Molecular Phylogenetics and Evolution

journal www.elsevier.com/locate/y



Dispersal and speciation: The cross Atlantic relationship of two parasitic cnidarians

Rohit Dnyansagar[a,1], Bob Zimmermann[a,1], Yehu Moran[a,b], Daniela Praher[a], Per Sundberg[c], Lene Friis Møller[d], Ulrich Technau[a,*]

[a] Department of Molecular Evolution and Development, University of Vienna, Austria [b] Department of Ecology, Evolution and Behavior, Hebrew University of Jerusalem, Israel [c] Department of Marine Sciences, University of Gothenburg, Sweden [d] Danish Shellfish Centre, DTU Aqua, Technical University of Denmark, Denmark

ABSTRACT

How dispersal strategies impact the distribution of species and subsequent speciation events is a fundamental question in evolutionary biology. Sedentary benthic marine organisms, such as corals or sea anemones usually rely on motile larval stages for dispersal and therefore have a relatively restricted distribution along coasts. Edwardsiella lineata and Edwardsiella carnea are virtually indistinguishable edwardsiid sea anemones native to the east American and the Northern European coast, respectively. E. lineata is a facultative parasite to the ctenophore Mnemiopsis leidyi, while the life cycle of E. carnea is unknown. Recently M. leidyi was found in the Skagerrak carrying Edwardsiella sp. parasites, which raised the intriguing possibility that the invasive comb jellies acted as cargo for the facultative E. lineata parasites to establish a new population in Northern Europe. Here, we assessed the genetic differences between these two cryptic Edwardsiella species and isolated parasites from the invasive comb jelly M. leidyi in Sweden by comparing rRNA, whole transcriptomes, SNPs, ITS2 sequences and the gene complements of key developmental regulators, the Wnt gene family. We show that E. carnea and the parasite transcriptomes are more than 99% identical, hence demonstrating that E. carnea has a previously unknown parasitic life stage. ITS2 sequence analysis of E. carnea and E. lineata suggest that they may not be reproductively isolated. The transcriptomes of E. lineata and E. carnea are ~97% identical. We also estimate that the species diverged between 18.7 and 21.6 million years ago.

Introduction

How new species arise and how populations can spread over large distances and conquer new habitats is one of the fundamental questions in evolutionary biology and ecology. For marine benthic or sedentary organisms, the oceans can be unsurpassable hurdles, if the pelagic stage is either too short or not motile enough to cross the ocean. Among those animals are the Actiniaria (sea anemones), member of the class Anthozoa within the phylum Cnidaria. Cnidaria is the sister lineage to Bilateria consisting of very diverse body plans including sea anemones, corals and jellyfish. The Anthozoa, although commonly considered monophyletic, may actually represent a paraphyletic group, with the Hexacorallia (including Actiniaria) as the sister group to the Octocorallia plus Medusozoa (Kayal et al., 2013). The Anthozoa, although commonly considered monophyletic, may actually represent a paraphyletic group, with the Hexacorallia (including Actiniaria) as the sister group to the Octocorallia plus Medusozoa (Kayal et al., 2013). Cnidaria is the sister lineage to Bilateria consisting of very diverse body plans including sea anemones, corals and jellyfish. While Medusozoa generally form both polyps and medusae, Anthozoa is characterized by the absence of the medusa stage, hence lacking a long-lived pelagic stage, which would allow the distribution over large distances (Bridge et al., 1992, 1995). The only pelagic stage of the sedentary sea anemones is the planula larva, which, depending on the species, lasts for a few days to several weeks before transforming into a polyp (Nyholm, 1943). The sea anemone Nematostella vectensis recently became one of the major model organisms among cnidarians for the study of comparative genomics, developmental biology and ecology (for reviews see

* Corresponding author at: Department of Molecular Evolution and Development, Centre of Organismal Systems Biology, University of Vienna, Althanstrasse 14, 1090 Wien, Austria. E-mail address: ulrich.technau@univie.ac.at (U. Technau). [1] Equal contribution.

https://doi.org/10.1016/j.ympev.2018.04.035
Received 21 October 2017; Received in revised form 12 March 2018; Accepted 23 April 2018
Available online 25 April 2018

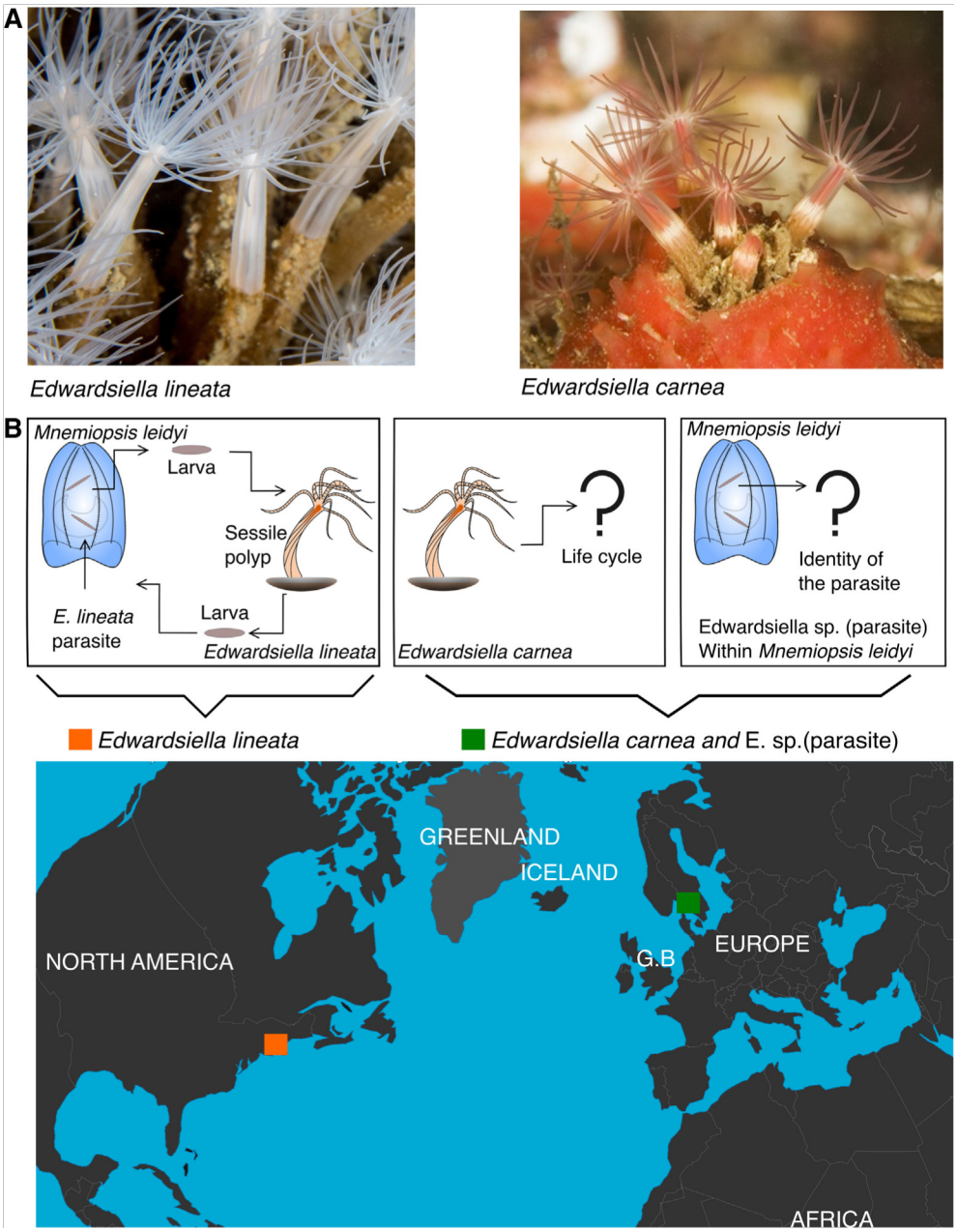


Fig. 1. A: Edwardsiella lineata (Image credit: Alex Shure) Edwardsiella carnea (Image credit: Kåre Telnes) and are two morphologically very similar sea anemones found at the opposite end of the Atlantic Ocean. E. carnea is found on Swedish West coast while
E. lineata is found on American East coast. B: E. lineata is known to have a parasitic life cycle within the ctenophore M. leidyi as vermiform larvae. Outside of the ctenophore host E. lineata is able to develop into adult sessile polyp. Although E. carnea was described in 1856, the life cycle of the E. carnea is not fully understood. We also found M. leidyi ctenophore close to the Swedish west coast infected with the Edwardsiidae parasite.

Darling et al., 2005; Genikhovich and Technau, 2009; Layden et al., 2016; Rentzsch and Technau, 2016; Steele et al., 2011; Technau and Schwaiger, 2015) N. vectensis originates from the estuarine habitats along the east coast of America (Hand, 1990; Hand and Uhlinger, 1992; Darling et al., 2005) and it belongs to the family of Edwardsiidae. Sea anemones of this family have long slender bodies, which are buried in sediments or crevices in the rocks.

Two sea anemone species of this family, closely related to Nematostella, are Edwardsiella lineata and Edwardsiella carnea (Fig. 1). E. lineata is found at the east coast of North America. Interestingly, planula larvae of E. lineata can enter the gut of ctenophores, mostly Mnemiopsis leidyi, and transform into a worm-like parasitic stage without tentacles or mesenteries (Fig. 1B) (Crowell, 1976). Due to its facultative parasitic life cycle in M. leidyi, E. lineata has been studied in some detail (Reitzel et al., 2006; Reitzel et al., 2007; Reitzel et al., 2009) and its transcriptome is available (Stefanik et al., 2014). E. carnea, which is found along the Atlantic coast of Sweden and Norway (Gosse, 1856; Daly, 2002) is extremely similar morphologically to E. lineata. Indeed, they differ only by two out of sixty morphological characters (Daly, 2002), raising the question of their phylogenetic relationship. E. carnea was first identified as Edwardsia carnea (Gosse, 1856). Since then the species has been renamed several times as Halcampa microps (Gosse, 1856), Milneedwardsia carnea (Carlgren, 1921) and Fagesia carnea (Delphy, 1938). The name Edwardsiella carnea is currently accepted by The World Register for Marine Species, which is based on The Marine Fauna of the British Isles and North-West Europe. Although E. carnea has been identified in 1856 (Daly et al., 2002; Gosse, 1856; Daly, 2002; Selander et al., 2009), virtually no information is available about its life cycle. In contrast to E. lineata, parasitic stages were never described for E. carnea.

The American ctenophore M. leidyi (Agassiz, 1865) was for the first time recorded in Northern Europe near Kiel in 2006 (Javidpour et al., 2006) and shortly thereafter in the North Sea (Faasse and Bayha, 2006; Boersma et al., 2007) and in the Skagerrak region of the Swedish West Coast (Hansson, 2006). Since then it has been found most years in Swedish waters (Selander et al., 2009). During the peak abundance of M. leidyi in 2007 endoparasitic sea anemone larvae were observed for the first time attached inside the invasive ctenophore in Swedish waters (Selander et al., 2009) and have been frequently observed since then. In their native environment infection of the comb jellies by E. lineata is common (Crowell, 1976; Bumann and Puls, 1996), yet, the identity of the parasites found in the invasive ctenophore in Sweden remained unclear. It could be the parasitic stage of E. lineata, hosted and carried along by the invasive ctenophore, or, alternatively, represent a parasite belonging to E. carnea or another edwardsiid species. However, the comb jelly Bolinopsis infundibulum is also native to the Skagerrak area (Selander et al., 2009) and a single record in the literature also suggests the possibility that E. carnea may be a parasite of B. infundibulum (Stephenson, 1935). Identification of the parasite based on morphology is even more challenging as it is vermiform and devoid of most morphological features used to characterize sea anemones such as tentacles and mesenteries. Therefore, E. lineata, E. carnea and E. sp. (parasite) may form a species complex, consisting of closely related, morphologically almost indistinguishable species. These species complexes have the ability to showcase the successfully maintained morphology despite underlying variable genetics, behavior, or physiology. These species also reveal the limitations of morphology based traditional methods and accentuate the need of novel methods to classify organisms. Commonly used methods to distinguish species include phylogenetic comparison of rRNA sequences, nuclear and mitochondrial genes, microsatellite genotyping, comparative genomics, and Internal Transcribed Spacer 2 compensatory base change analysis. In order to resolve the Edwardsiella species complex, we generated transcriptomes of E. carnea and the

parasites isolated in Sweden and compared them to the published transcriptome dataset of E. lineata (Stefanik et al., 2014). Using a variety of methods our results show that E. carnea and E. lineata are genetically very similar, suggesting a very close species relationship. Moreover, we show that the parasite found in an American ctenophore at the shore of Sweden stems from E. carnea polyps, demonstrating that this species also has a parasitic life stage, which can enter foreign hosts.

Materials and methods

2.1. Collection of animals

E. carnea were collected by dredging at 40 m depth on the Swedish West Coast (58°°21′N, 11°06′E; 58°21′N, 11°07′E) August 2012. E. carnea were found at the bases of the soft coral Alcyonium digitatum and on rocks. The sea anemones were kept in sea water at Sven Loven Marine Center, Kristineberg, Fiskebäckskil and fed with cultured Artemia nauplii. They were starved for three days and then flash frozen in liquid nitrogen and kept in −80 °C until used. The ctenophore M. leidyi (Agassiz), infected by endoparasitic larvae, was collected in the Gullmar Fjord off the Swedish West Coast (58°21′N, 11°24′E). Larvae attached close to the pharynx or stomach of M. leidyi were gently removed from the ctenophore and preserved in RNAlater (Ambion).

2.2. Generation and quality control of transcriptome datasets

The E. lineata transcriptome and read data were obtained from EdwardsiellaBase (Stefanik et al., 2014). The read data consist of ~340 million paired end Illumina reads. Total RNA was extracted from four E. carnea adults using the VWR Omega-BioTek E.Z.N.A molluscan RNA isolation kit (catalog number R6875-00). Transcriptome sequencing of E. carnea was performed at GENEWIZ South Plainfield, NJ with Illumina Hiseq 2000, which resulted in ~180 million paired end reads of length 125. Total RNA was extracted from six E. sp. specimens with Trizol reagent according to the manufacturer's instructions. Library preparation and sequencing of the mRNA was performed at VBCF NGS unit (www.vbcf.ac.at) using an Illumina HiSeq 2500 sequencing system which resulted in ~660 million paired end reads. The quality control tool FASTQC for the high throughput sequencing data was used to assess the read quality. Due to the absence of genome data, de novo transcriptome assembly was performed with Trinity (Grabherr et al., 2011) in strand specific mode with the additional option of adapter filtering using Trimmomatic (Bolger et al., 2014). We used Transdecoder (Haas and Papanicolaou, 2012) utility to predict open reading frames (ORF) in transcripts and discarded transcripts of ORFs, which were less than 100 amino acids. To assess biological completeness of the transcriptomes, we subjected the transcriptomes to the BUSCO (Simão et al., 2015) analysis using 843 well-curated protein-coding genes. While the transcriptomes of E. sp. (parasite) and E. lineata transcriptomes contained 52% and 47% of the BUSCO marker genes, respectively, the E. carnea transcriptome is the most complete dataset among the transcriptomes under study with 92% (single copy and duplicated) marker genes (Supplementary Fig. 1). To construct the 'pseudo-genome' dataset, orthologous sequences were identified with OrthoMCL (Li et al., 2003) and used for further comparative genomic and variant analysis. Marker sequences such as 18S rRNA and Wnt genes were identified with blast homology search. Assembled transcriptomes and raw sequencing reads for E. carnea and E. sp. Parasite have been deposited to NCBI Transcriptome Shotgun Assembly (TSA) database under the accessions GGGD00000000 and GGGB00000000 respectively.

2.3. Analysis of sequence identity and genetic distance

We used the OrthoMCL pipeline (Li et al., 2003) to find a reliable set of orthologous sequences from E. lineata, E. carnea, N. vectensis and E. sp. (parasite). OrthoMCL uses homology searches with BLAST along with the Markov Cluster aLgorithm (MCL) to determine homologs. The resulting groups of homologs were further processed within house python script (https://github.com/dnyansagar/edwardsiella). In order to avoid comparisons between fragmented sequences, orthologous groups sharing less than 90% of the sequence length among all group members were filtered out. Pairwise alignments of the transcript sequences were done using lastz (Harris, 2007) and percent identity scores obtained for each alignment were used to calculate the average percentage identity between the datasets. To account for the size differences of transcripts being aligned, the '–noytrim' flag available in lastz algorithm was used, which extends alignment to the end of longer sequences. We also used the orthologous sequences obtained through OrthoMCL to calculate genetic distances using FDNADIST algorithm from the EMBOSS package (Rice et al., 2000). In order to assess the stability of the percent identity distance differences, 95% confidence intervals were calculated via bootstrapping the source alignments. Standard deviations from the mean were estimated from the distribution percent identities of 1000 bootstrap replicates of the alignments (Supplementary data 2). A calculation of the confidence intervals of the LogDet distances were estimated based on the approximate, normal theory confidence intervals as described in (Cai et al., 2015) as implemented in the heplots R package. 2.4. Protocol validation/verification

To verify our comparative genomics analysis protocol, we tested the protocol on well-studied group of apes. We obtained 14,068 orthologous groups via OrthoMCL pipeline from Pan troglodytes, Homo sapiens, Gorilla gorilla data from the Ensembl BioMart (Kinsella et al., 2011). Additionally, we created 12 pseudo transcriptomes from each of our read dataset, namely E. lineata, E. carnea, E. sp. (parasite) and then followed the same protocol of comparative genomics.

2.5. Phylogenetic analysis

For the phylogenetic comparison between the closest sea anemone species, we obtained 18S rRNA sequences of Edwardsianthus gilbertensis, Edwardsia andresi, Edwardsia japonica, Edwardsia elegans, Edwardsia sipunculoides, Edwardsia timida, Edwardsia tuberculata and Nematostella vectensis from NCBI (See Supplementary data 3 for accession). 18S rRNA sequences of these species were aligned with MAFFT (Katoh and Standley, 2013) using the E-INS-i algorithm. The resulting alignment was subjected to the maximum likelihood phylogenetic analysis using IQ-TREE (Nguyen et al., 2015) with 1000 bootstrap samples. IQ-TREE tree topology and branch lengths were inferred under the Tamura-Nei (TN) substitution model with allowance for the proportion of the invariable sites (+I), which was selected with standard model selection (not including FreeRate models). The same alignment was used for the Bayesian phylogenetic analyses. Bayesian analysis with Mrbayes (Ronquist and Huelsenbeck, 2003) was run with the default nucleotide substitution model (4 × 4). Each analysis was set to run for 2×10^6 generations and every hundredth tree was sampled. MCMC generation was terminated when the standard deviation of split frequencies fell below 0.01. The first quartile of the sampled trees was discarded to assure better sample quality. The resulting tree has an identical topology to the tree obtained with maximum likelihood therefore the posterior support values for the branches are merged.

To search for Wnt sequences in the E. lineata, E. carnea and E. sp. (parasite) transcriptomes, N. vectensis Wnt sequences (Kusserow et al., 2005) were used as baits. The tblastx algorithm of NCBI Blast (Altschul et al., 1990) was used with 1e-3 as e-value cut-off. The sequences then subjected to MAFFT alignment with using the L-INS-i algorithm. The resulting alignment was edited manually using Jalview (Waterhouse et al., 2009). The resulting

alignment was subjected to the maximum likelihood phylogenetic analysis using IQ-TREE (Nguyen et al., 2015). Best-fit model LG + I + G4 was chosen by the IQ-TREE according to BIC (Bayesian Information Criterion). The tree topology was confirmed with 500 bootstraps. The same alignment was also subjected to Bayesian phylogenetic analysis with MrBayes and the posterior probability mapped on the maximum likelihood tree.

2.6. Variant calling

For variant calling aligned reads were mapped to the orthologous gene set created earlier using bowtie2 with default settings (Langmead and Salzberg, 2012). The mapped reads were then subjected to the SAMtools mpileup program (Li et al., 2009), which provides a summary of the coverage of mapped reads on a reference dataset at a single base pair resolution. This summary was piped to VarScan (Koboldt et al., 2012) to call SNPs. A potential source of bias in the SNP analysis is the use of 'pseudo-genome' we created as a reference for the variant calling, however, we ensured with our strict filtering techniques for read data that only high quality reads will be used for variant calling and each SNP will have sufficient read support. Criteria used for filtering SNPs in order to avoid false positives are (i) read support for the SNP position should be more than 100 (ii) p-value for the SNP should be less than 0.01 (iii) Phred Quality Score for the base call should be more than 15 (call is > 90% accurate) Further, we used an in-house python script to compare the locations of SNPs in each transcript. We counted the occurrences of such events where transcripts from two species have SNP in same location.

2.7. Internal transcribed spacer 2 (ITS2)

Curated metazoan ITS2 sequences were downloaded from the ITS2 Database (http://its2.bioapps.biozentrum.uni-wuerzburg.de/) (Merget et al., 2012). Reads of E. carnea, E. lineata and E. sp. (parasite) were mapped against these sequences using bowtie2 (Langmead and Salzberg, 2012). RNA sequence for N. vectensis was obtained from RNAcentral database (Bateman et al., 2011). All the putative ITS2 sequences were subjected to the "Annotate" tool available at the ITS2 database to determine the boundaries of ITS sequences. ITS2 sequences were then subjected to RNAfold program (Lorenz et al., 2011) to predict the secondary structure of the ITS2. An option to predict best secondary structure based on minimum free energy and partition function was selected from the RNAfold program. To compare and find compensatory base change (CBC), Input data in the XFasta format (fasta sequence with secondary structure) was prepared and used in 4SALE program (Seibel et al., 2008). To evaluate the efficiency of the method we applied the method to some of the known species groups such as Apes, Rodents and Drosophila (Supplementary data 1).

2.8. Divergence time estimates

For divergence time estimation, we followed the approach used by Peterson and colleagues (Peterson et al., 2004). In this approach seven conserved nuclear genes (ATP Synthase, Eukaryotic Translation Elongation Factor 1 Alpha 1, Methionine Adenosyltransferase 1A, Triosephosphate Isomerase, Catalase, Aldolase Fructose biphosphate, Phosphofructokinase) identified from taxa were used to calculate the divergence time of Edwardsiella species, with the poriferan Oscarella carmela as an outgroup. Sequences from N. vectensis, E. carnea, E. lineata, E. sp. parasite, A. millipora and A. digitifera were found via local BLASTP searches of the 15-taxon data set downloaded from GenBank accessions AY580167-AY580307 (Altschul et al., 1990). The transcriptome of Oscarella carmela was downloaded from http://www. compagen.org/ (Ereskovsky et al., 2017). The complements of Drosophila melanogaster and Anopheles gambiae were found via online NCBI BLAST searches ("Database Resources of the National Center for Biotechnology Information," 2017). Multi-

way sequences from N. vectensis, E. carnea, E. lineata, E. sp. parasite, A. millipora and A. digitifera were found via local BLASTP searches (Altschul et al., 1990). lignments of individual genes were performed with MAFFT in E-INS-i mode (Katoh and Standley, 2013) and trimmed using trimAl in automated1 mode (Capella-Gutiérrez et al., 2009). A maximum likelihood tree was inferred using IQ-TREE using the model LG + I + G4 as selected using ModelFinder (Nguyen et al., 2015). Date estimates were determined using r8s version 1.81 using the Langley-Fitch likelihood method (Sanderson, 2003). Ranges were estimated by fixing the age of the bilaterian split between 555.0 and 641.7 Mya (dos Reis et al., 2015).

Results

3.1. 18s rRNA phylogenetic analysis shows close relationship of E. carnea, E. sp. (parasite) and E. lineata

In an effort to find species closely related to the model organism Nematostella vectensis, we collected four specimens of the edwardsiid sea anemone E. carnea on the coast of Lysekill, Sweden. Likewise, we pooled twelve specimens of parasites from the ctenophore M. leidyi collected off the coast of Sweden (Gullmar Fjord) and sequenced the transcriptomes of both. To assess the phylogenetic relationships of the two species and the parasite we carried out a phylogenetic analysis based on 18S rRNA sequences using both the maximum likelihood (IQTREE multicore version 1.5.5) (Nguyen et al., 2015) and Bayesian (MrBayes v3.2.6) (Ronquist and Huelsenbeck, 2003) approaches (Fig. 2). Edwardsiella lineata, Edwardsiella carnea and the Edwardsiella sp. parasite form a monophyletic clade with N. vectensis, within the family Edwardsiidae, supporting their close relationship. Additionally, upon examination of the alignments we found that E. carnea and E. sp. (parasite) are completely identical to each other, while E. carnea and E. lineata differ only by two base change across 1896 bases.

3.2. Assessment of genetic distance using comparison of 'pseudo-genomes'

Since rRNA is a strongly conserved molecule and may not reveal hidden variation underlying recent speciation events, we sought to analyze the whole transcriptome in more detail. Transcriptome data for E. lineata was recovered from published databases (Stefanik et al., 2014). Assemblies of E. lineata, E. carnea and E. sp. (parasite) comprise 117,890, 296,463 and 186,572 transcripts respectively (Supplementary

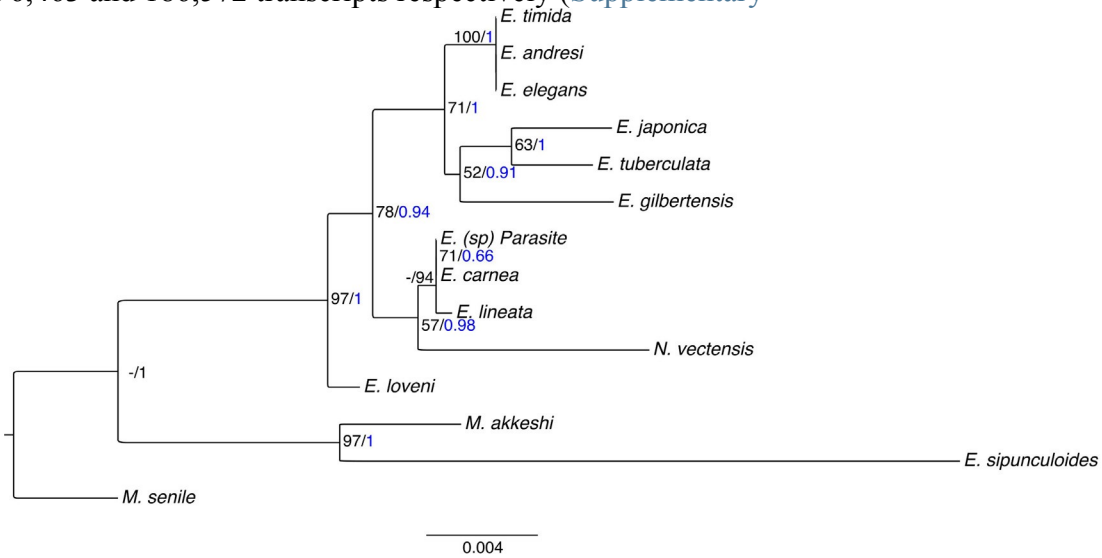


Fig. 2. Phylogenetic relationship among Edwardsiidae species based on 18S rRNA sequences. The phylogenetic tree was constructed with the maximum likelihood method implemented in IQ-TREE using Metridium senile as outgroup. The first value at the nodes indicates the Bootstrap support from the maximum likelihood method while the second value (shown in blue) indicates posterior probability support from the Bayesian inference method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 2). We checked for the presence of BUSCO metazoan marker genes and found that the transcriptomes of E. sp. (parasite) and E. lineata transcriptomes contained between 81% (E. lineata) and 97% (E. carnea) of the gene set in partial or full-length form (Supplementary Fig. 1). This shows that all three transcriptomes are close to saturation.
In order to evaluate nucleotide-level sequence identity, de novo transcriptomes generally are not sufficient because of the potential misassembly of transcripts. Therefore, we constructed a 'pseudo-genome' from well-assembled transcripts, which were highly represented in the transcriptomes. We included N. vectensis as an outgroup reference to aid in comparative analysis of data. We selected 7021 orthologous sequence groups using the OrthoMCL pipeline (Li et al., 2003) and inhouse Python scripts and created separate datasets for each species to pairwise compare the percentage identity and genetic distance with LogDet distance measure. Amongst several genetic distance measures available we selected LogDet distance due to its robustness to the composition biases (Massingham and Goldman, 2007) that may occur in transcriptome data (Zheng et al., 2011). The percentage identity measure calculated with lastz (Fig. 3) indicates that the transcriptomes of E. carnea and E. lineata are ~97% identical, while E. carnea and E. sp. (parasite) share more than 99% nucleotide identity. E. carnea and E. sp. (parasite) are also equidistant from E. lineata (~97%) and N. vectensis (~73%). Both E. lineata and E. carnea are ~73% identical to N. vectensis. The LogDet distance calculated with FDNADIST program from the EMBOSS (Rice et al., 2000) package is also shown in the Fig. 3. The FDNADIST program reads in DNA sequences and outputs a distance matrix. The LogDet distance calculated here between E. carnea and E. sp. (parasite) is 0.008, thus two orders of magnitude smaller than the distance between E. carnea and E. lineata (0.252). N. vectensis is the farthest from other species according to the measures in the study, although it is closer to E. lineata (0.793) than to E. carnea (1.126) or the E. sp. (parasite) (1.12612). This distinction in the relationship is not evident in the percentage sequence identity measure and thus provides a more effective measure to assess genetic differences among species. Our validation by dividing our datasets in twelve subsets shows very small standard deviation indicating the robustness of our comparative transcriptomic analysis protocol (Supplementary Fig. 3). To further validate our analysis, we also compared humans with chimpanzee and gorilla. The comparative genomic analysis between human and chimpanzee shows 99.03% identity and 0.013 LogDet distance. Human and gorilla

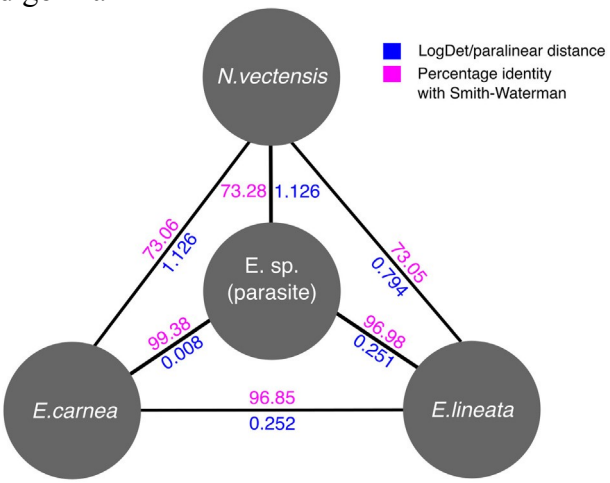


Fig. 3. A: Comparative transcriptomic analysis of E. lineata, E. carnea, N. vectensis and E. sp. (parasite). The LogDet distance (blue) measure of evolutionary divergence is calculated using the FDNADIST program from the EMBOSS package. The alignment of orthologous sequences used to create 'pseudogenome' is used as input to FDNADIST. Percentage similarity (pink) calculated using Lastz program with same orthologous sequences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

show 98.45% identity and 0.022 genetic distance, while chimpanzee and gorilla show 98.32% identity and 0.024 genetic distance (Supplementary Fig. 4). Additionally, our transcriptomic estimation of chimpanzee-human substitution rates closely reflects those observed genome-wide (The Chimpanzee Sequencing and Analysis Consortium, 2005). These results demonstrate the robustness of our comparative analysis protocol and show that E. carnea and E. lineata are slightly more divergent than humans and chimpanzee, suggesting that they are two distinct species.

3.3. Variant analysis reveals a common origin of E. carnea and E. sp. (parasite) populations

The analyses above suggest a very close relationship of E. carnea and the parasite. However, this does not preclude two very closely related species. Single nucleotide polymorphism (SNP) analysis has been widely applied to answer population genetics questions in other model and non-model organisms (Morin et al., 2004; Stetz et al., 2016; Vendrami et al., 2016). Since demographic and geographic changes leave their signatures in the genome of a species, studies on genetic diversity with SNP genotyping have been extremely effective to trace the evolutionary history of different populations (Tishkoff et al., 2009; Campbell and Tishkoff, 2010; Choudhury et al., 2014). SNPs can serve as excellent markers to distinguish populations or closely related species. We reasoned that if our E. carnea and E. sp. (parasite) samples are representing the same species, they should share many more SNPs than the closely related species E. lineata or E. carnea. To investigate whether the similarities observed in the analyses above are reflected in the genetic variance of individuals, we called and compared the SNP locations between the transcriptomes of the species under investigation thereby looking for SNPs, which are unique or shared with at least two of the three samples. The constructed 'pseudo-genome' was used as a reference genome to call SNPs using a cascade of tools, i.e. bowtie2 (Langmead and Salzberg, 2012), SAMtools (Li et al., 2009), VarScan (Koboldt et al., 2012) and finally in-house python scripts, to filter the variants. After the filtering, we collected 14,363 SNPs (1.52/kb) in E. carnea, 21,251 SNPs (2.14/kb) in E. lineata and 7925 SNPs (1.25/kb) in E. sp. (parasite). No discernible correlation between called SNPs and read coverage was found, ruling out the possibility that differences in library complexity or sequencing error gave rise to differences in the number of SNPs called (Supplementary Fig. 5).
We found that there are 2061 common SNP sites between E. carnea and E. sp. (parasite), compared to only 167 shared between E. lineata and E. carnea and 107 between E. lineata and E. sp. (parasite) (Fig. 4). Thus, our analysis uncovers a considerable number of common SNP sites between the E. carnea and the E. sp. (parasite), suggesting a common origin of the populations sampled. Only a moderate number of common SNP sites were found common between E. carnea and E. lineata and between E. lineata and E. sp. (parasite), indicating a common evolutionary history shared by these two species. The common SNP sites between the species also support our earlier results, which indicate that E. carnea and E. sp. (parasite) are the same species.

3.4. The two Edwardsiella populations diverged 18.7–21.6 Mya

Next, we determined the divergence time of the two Edwardsiella populations by a molecular clock approach. We constructed a phylogenetic tree using seven nuclear genes (Supplementary Fig. 6) and node constraints calibration points as used in Peterson et al to estimate the divergence times (Peterson et al., 2004). Based on an estimated split of

bilaterians between 555 and 642 Mya (dos Reis et al., 2015), the split between Edwardsiella species and N. vectensis dates back to 184.1–212.7 Mya, while the calculated divergence time of the E. carnea and E. lineata lies between 18.7 and 21.6 Mya, ruling out a human impact in the distribution of the two populations.

3.5. Are E. lineata and E. carnea reproductively isolated?

Biological species are commonly defined by their ability to hybridize and produce viable and fertile offspring (Queiroz, 2005). However, given the difficulty to obtain live specimens of E. carnea and the lack of a spawning induction protocol for E. carnea and E. lineata, such a test between E. carnea and E. lineata is not possible. However, reproductive isolation also correlates with the number of compensatory base changes (CBC) of the rRNA, in particular with the one of the internal transcribed spacer 2 (ITS2). Indeed, the presence of CBC in helix II or helix III has been correlated with the reproductive isolation between the populations and therefore CBC in the ITS2 sequence in specific locations has been proposed as a molecular barcode for species identification (Coleman 2003). An early study to test this correlation

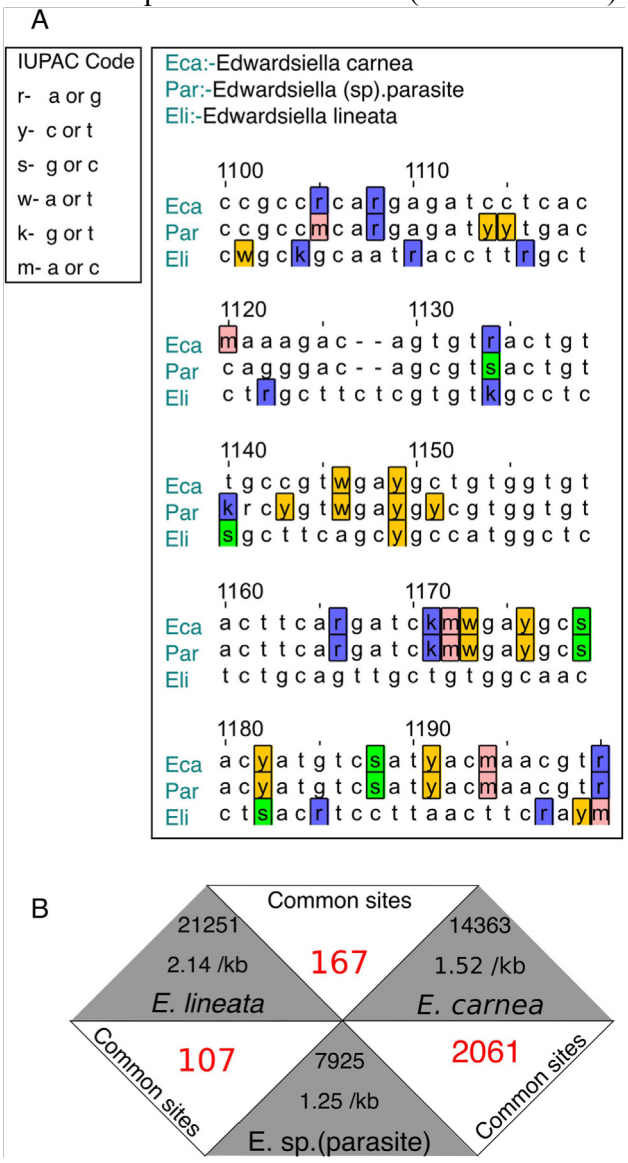


Fig. 4. Analysis of SNPs. A: An illustration of SNP locus comparison wherein multiple sequence alignment consisting of E. carnea, E. lineata and E. sp. (parasite) sequences with the SNP positions highlighted in IUPAC nucleotide code. SNPs were called with combination of tools such as Bowtie, SamTools, VarScan. The SNPs shown in the figure are called by the VarScan based on the read support and after applying support filters, eliminating varying nucleotides with read support less than 100 or the quality scores less than 15. Note that the highlighted nucleotides are examples of the underlying polymorphism at this

position. B: Summary of the SNP locus comparison in 5830 orthologous sequences along with SNP per kb. There are 2061 common SNP sites between E. carnea and E. sp.(parasite) while as 167 common SNP sites between E. lineata and E. carnea. We found only 107 common sites between E. sp.(parasite) and E. lineata.

investigated 1300 closely related species and found that the presence of CBC distinguished ~93% of pairs in separate species, however, the absence of CBC could only group ~77% of pairs merged to single species (Müller et al., 2007). Another study found a clear correlation between having two or more CBCs and reproductive isolation (Pawłowska et al., 2013). Using the ITS2 sequences from the ITS2 database as reference, we identified ITS2 sequences from N. vectensis, E. lineata and E. carnea. We found CBCs between N. vectensis and E. carnea and between N. vectensis and E. lineata (Supplementary data 1), suggesting reproductive isolation. Notably, we did not find any CBC between E. carnea and E. lineata, raising the possibility that these two species may still have the potential to hybridize if they were in the same environment.

3.6. Expression of Wnt genes and Wnt phylogeny

Lastly, we wished to gain insights into the genetic regulation of development by analyzing the Wnt genes. Wnt genes encode signaling molecules which are important developmental regulators involved in early axis formation, in stem cell biology and regulation of cellular differentiation processes (McMahon and Moon, 1989; Smith and Harland, 1991; Sokol et al., 1991; Christian et al., 1991; Steinbeisser et al., 1993; Wylie et al., 1996). The WNT pathway is highly conserved in all animals, but not found outside of the animal kingdom (Kusserow et al., 2005; Adamska et al., 2007). A total of 13 distinct subfamilies of WNT ligands have been characterized. While humans possess 12 of the 13 subfamilies, Drosophila melanogaster and Caenorhabditis elegans have only 6 and 3 Wnt genes, respectively (Prud'homme et al., 2002). Interestingly, the edwardsiid Nematostella vectensis expresses also 12 of the 13 known subfamilies (Kusserow et al., 2005). Hence, this sea anemone has maintained virtually all of the ancestral full complement of Wnt genes, which was substantially reduced in flies and nematodes. While Wnt proteins form highly conserved intra-molecular cysteine bridges, large regions of the amino acid sequence of Wnt proteins are fairly divergent, which are suitable to detect recent speciation events. Therefore, we identified the Wnt protein-coding transcripts from the transcriptomes of E. lineata, E. carnea and E. sp. (parasite) and constructed a phylogenetic tree (Fig. 5). As in N. vectensis, no Wnt9 could be detected in the investigated edwardsiids. Remarkably, although only one developmental stage was sampled in the case of E. carnea and E. sp. (parasite), almost all Wnt ligand subfamilies known from N. vectensis were also expressed in E. carnea and E. lineata. Only Wnt2 was missing from E. carnea, while Wnt7, Wnt10 and Wnt11 were missing from E. sp. (parasite). Generally, the Wnt protein tree corroborates the phylogenetic relationships of the Edwardsiella species as the 18S rRNA phylogeny. Wherever a Wnt subfamily was present in all three species/ specimens, E. carnea and the E. sp. (parasite) were almost identical. However, E. carnea/E. sp. (parasite) were very close to E. lineata, with N. vectensis as the closest homologs.

Discussion

In this study, we aimed to gain genetic and phylogenetic distinction between two edwardsiid sea anemones, Edwardsiella carnea and Edwardsiella lineata, which occur at great distance from each other, along the west coast of Sweden and the east coast of North America, respectively. E. lineata is known to have a facultative parasitic stage between the planula larva and the polyp stage, which inhabits the gut of the American ctenophore Mnemiopsis leidyi (Crowell, 1976). By comparison, the life cycle of E. carnea has been unclear. Since Mnemiopsis leidyi carrying parasitic edwardsiids has been detected regularly during the last

ten years in the North Sea and the Baltic Sea, this raised the question, whether this American invasive species might have drifted with the Gulf Stream to Northern Europe and acted as carriers for the parasitic cnidarians. This raised also the possibility that E. carnea forms a cryptic species of E. lineata, as a population that was established by the delivered parasites. It was therefore imperative to determine whether E. carnea and E. lineata are two distinct species and whether the parasite found in the ctenophore M. leydi at the Swedish coast belongs to one of them.

4.1. E. lineata and E. carnea are closely related, yet distinct species

When E. lineata and E. carnea were compared morphologically the demarcation between the species or distinctive characteristics for the species appeared weakly defined (Daly, 2002). In order to determine the phylogenetic relationship of E. lineata and E. carnea, we first carried out a phylogenetic analysis using 18S rRNA sequences. This showed that both E. lineata and E. carnea are very closely related species and together they form the closest sister species to Nematostella vectensis, a widely-used model organism for comparative genomics. This result was corroborated by a global comparison of the transcriptomes assembled as 'pseudo-genomes'. Results of percentage similarity and LogDet distance measures make it evident that, although E. lineata and E. carnea are very close they are distinct on transcriptome level and share about 97% nucleotide identity.

It is clear that there is no perfect way to define a species. One way is to test reproductive isolation by crossing representatives of both populations, according to the biological species concept, although a number of counter examples are known. However, testing even enforced hybridization was not possible because obtaining the specimens of E. carnea is very difficult and moreover, there is no reliable spawning protocol in the lab for E. carnea and E. lineata. Cytochrome c oxidase subunit I (COI), which was used earlier as a phylogenetic marker, shows very low sequence divergence in cnidarian species (Hebert et al., 2003). Therefore, we used the number of compensatory base changes (CBC) of the ITS2 of the rRNA as a proxy for reproductive isolation, although it is based on correlation alone (Coleman, 2007) and there are other limitations of this method such as its dependence on the secondary structure prediction algorithms (Caisová et al., 2011). Some studies suggest that 2 or more CBCs in the ITS2 sequence correlate with reproductive isolation of distinct species. Yet, while the CBC method can distinguish most animal species in over 90% of the cases, in Cnidaria only 77% of the pair-wise comparisons CBCs correlate with distinct species (Yao et al., 2010). As the calculated divergence time two Edwardsiella populations is roughly 20 Mio years we conclude that the two populations are indeed two distinct, yet very closely related species that are geographically isolated. Nevertheless, we do not find any CBC between E. lineata and E. carnea, raising the possibility that they could still interbreed. We conclude that these two populations are so close that the CBC method does not recognize them confidently as two separate species.

4.2. Edwardsiella carnea also has a parasitic stage, which can infect foreign ctenophores

When we included the parasite in these analyses, we unambiguously found that the parasite within the comb jelly is genetically almost identical to E. carnea. In the global transcriptome analysis the parasite is 99.38% identical to E. carnea; in the phylogenetic tree of the 18S rRNA and the Wnt genes, E. carnea is always identified as the closest relative, in many cases indistinguishable from it. This is further supported by analysis of the SNPs, where about 20 times more are shared between E. carnea and the parasite. Although it was not feasible to collect multiple specimens from different locations, it is likely that the minor differences

between the polyps of E. carnea and the parasite reflect the intra-population variation. Therefore, we can rule out our initial hypothesis that the parasite stems from E. lineata. Thus, the parasite isolated from M. leidyi in Sweden is a parasitic stage of E. carnea. This observation is remarkable as the parasite is native to the North Sea while its host is predominantly found on the North American east coast, though periodically has been found in the North Sea. Our findings suggest that E. carnea is a non-selective parasite to ctenophore species as we found the parasite within M. leidyi, which is probably not its common host. Indeed, earlier evidence suggested it parasitized B. infundibulum (Stephenson, 1935; Selander et al., 2009). Interestingly, Edwardsiella species might have a wider host range than initially appreciated, as at least one so far unconfirmed study reported the occurrence of a parasitic stage of an edwardsiid-like organism in a scyphozoan jellyfish of the genus Aurelia in Croatia (Chiaverano et al., 2015).

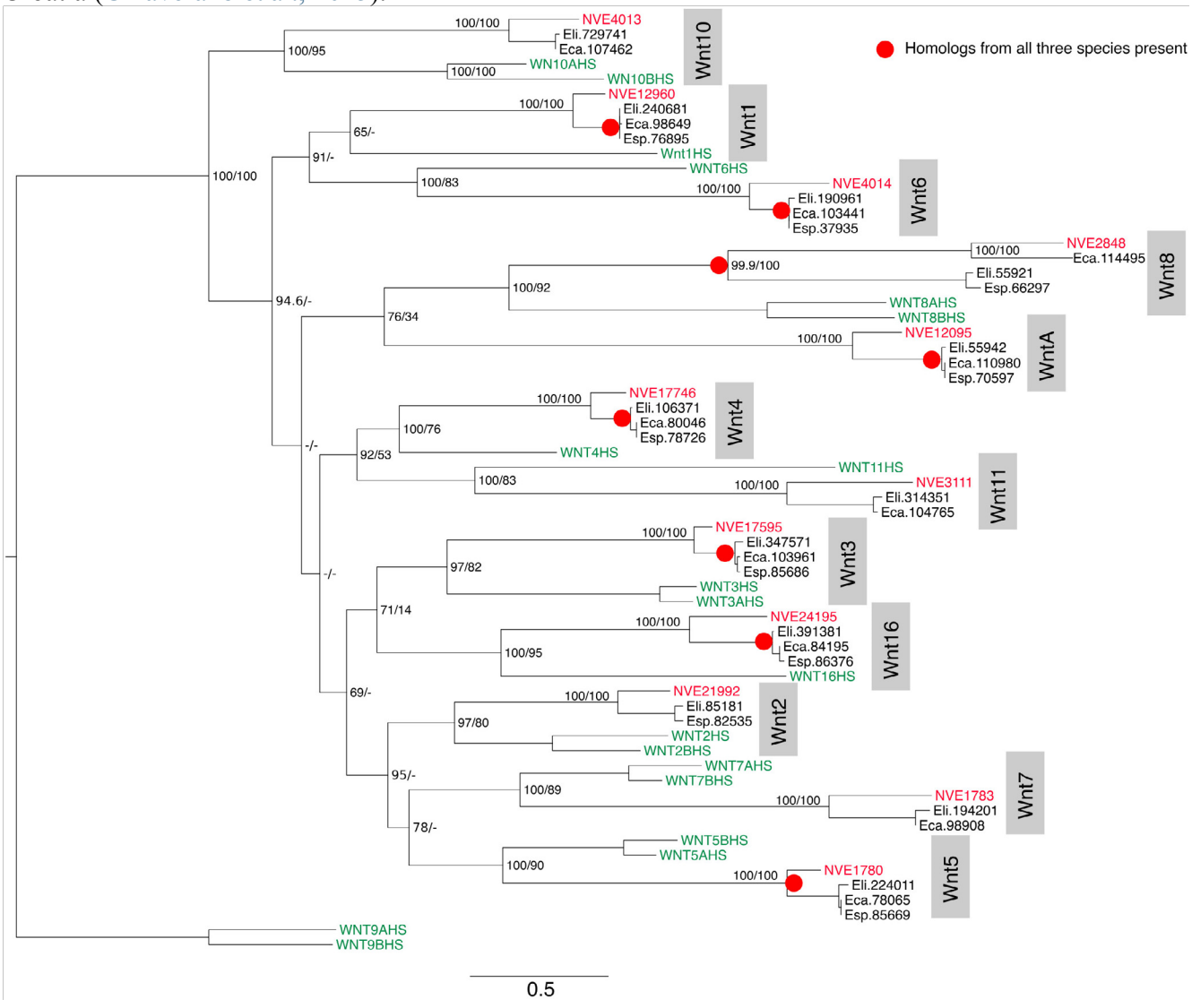


Fig. 5. Phylogenetic analysis of Wnt family genes of Homo sapiens (green), Nematostella vectensis (purple), Edwardsiella lineata, Edwardsiella carnea, Edwardsiella sp. parasite. The tree topology shown here is supported by both maximum likelihood and Bayesian inference. The ML method implemented in IQ-tree (Nguyen et al., 2015) was used. The algorithm selects the best substitution model for the alignment, which in this case was LG + I +G4. The first values shown at the nodes are percentage of 500 bootstrap runs supporting the node. The second values are the posterior probability values from the Bayesian inference. Bootstrap values below 50% and percentage posterior probability values less than 70 are not considered, indicated with "–". (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Biogeography and speciation of a facultative parasitic cnidarian

Accepting that E. lineata and E. carnea form two very closely related, yet distinct species, it remains striking that such closely related species populate habitats that are almost 6000 km apart. This raises the question, how the two populations were established to give rise to two species.

There are studies estimating the range of drifting cnidarian larvae: In a recent study of coral reefs it was estimated that the dispersal range of coral larvae ranges from ~10 km to ~50 km (Markey et al., 2016). The sea anemone Isarachnanthus nocturnus may have a range of 2000–4000 km for the dispersal of the free swimming larvae, although the range is limited along the coast and not crossing an ocean. I. nocturnus also has longer larval stage of 63–118 days (Stampar et al., 2015). Since the non-parasitic form of the edwardsiid larva is nonfeeding, crossing the Atlantic seems impossible for planula larvae. In the case of host-mediated transfer via M. leidyi the speed would be faster than the one of the larva, considering the larger size of the comb jelly. However, if we consider factors like the ocean currents such as Gulf Stream the speed would increase substantially. The Gulf Stream with an average speed of 6.4 km/hr would make the host-mediated transport seem plausible under optimal conditions, although probably extremely rare.

However, the Gulf Stream only established after the rise of isthmus of Panama in the Cenozoic era, which caused the separation of the two oceans. While some recent studies proposed that the rise isthmus of Panama might have occurred between 7 and 23 Mya (Brady, 2017; Bacon et al., 2015; Montes et al., 2015), a more recent rise of the isthmus between 2 and 5 Mya appears to be the currently accepted view in the field (O'Dea et al., 2016). If the latter is the case, then the divergence time predates this event and we can rule out the scenario that the European population was established through the drift of the host by the Gulf Stream. On the other hand, the supercontinent Pangaea has already split long before the divergence of the two Edwardsiella populations giving rise to the American, the Eurasian and the African continent (Dietz and Holden, 1970a, 1970b). It is conceivable that in the northern territories North America and Europe remained much closer together than today, even at times of divergence of the two Edwardsiella species. In this scenario, it is possible that a common ancestor population that was contiguously distributed along the northern coasts of America and Europe became split by further separation of the continents. Further sampling of these species in Canada, Greenland and Northern Scandinavia would be required to evaluate the precise phylogeographic distribution of these two species.

Conclusions

We here show that the Edwardsiella carnea found at North Sea close to Sweden and Norway also has a facultative parasitic stage and is a non-selective parasite to ctenophore hosts. Our work unravels the phylogenetic relationship of two edwardsiid species, which are closest to the non-parasitic model cnidarian Nematostella vectensis, and the possible impact of their parasitic life cycle on the speciation events and the resulting biogeography of the species. Our analysis suggests that Edwardsiella carnea and Edwardsiella lineata are two distinct species with the possibility of crossbreeding. Moreover, we prove that the parasite found in Mnemiopsis leidyi is Edwardsiella carnea, for which no parasitic stage has been described to date.

Acknowledgements

This work was supported by a grant of the Austrian Science Fund FWF to U.T. (P24858).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.04.035.

References

Adamska, M., Degnan, S.M., Green, K.M., Adamski, M., Craigie, A., Larroux, C., Degnan, B.M., 2007. Wnt and TGF-β Expression in the sponge Amphimedon Queenslandica and the origin of metazoan embryonic patterning. Plos One 2, e1031. http://dx.doi. org/10.1371/journal.pone.0001031.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. http://dx.doi.org/10.1016/S0022-2836(05) 80360-2.

Bacon, C.D., Silvestro, D., Jaramillo, C., Smith, B.T., Chakrabarty, P., Antonelli, A., 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proc. Natl. Acad. Sci. U.S.A. 112, 6110–6115. http://dx.doi.org/10.1073/ pnas.1423853112.

Bateman, A., Agrawal, S., Birney, E., Bruford, E.A., Bujnicki, J.M., Cochrane, G., Cole, J.R., Dinger, M.E., Enright, A.J., Gardner, P.P., Gautheret, D., Griffiths-Jones, S., Harrow, J., Herrero, J., Holmes, I.H., Huang, H.-D., Kelly, K.A., Kersey, P., Kozomara, A., Lowe, T.M., Marz, M., Moxon, S., Pruitt, K.D., Samuelsson, T., Stadler, P.F., Vilella, A.J., Vogel, J.-H., Williams, K.P., Wright, M.W., Zwieb, C., 2011. RNAcentral: a vision for an international database of RNA sequences. RNA 17, 1941–1946. http:// dx.doi.org/10.1261/rna.2750811.

Boersma, M., Malzahn, A.M., Greve, W., Javidpour, J., 2007. The first occurrence of the ctenophore Mnemiopsis leidyi in the North Sea. Helgol. Mar. Res. 61, 153. http://dx. doi.org/10.1007/s10152-006-0055-2.

Bolger, et al., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics.

Brady, S.G., 2017. Army ant invasions reveal phylogeographic processes across the Isthmus of Panama. Mol. Ecol. 26, 703–705. http://dx.doi.org/10.1111/mec.13981.

Bridge, D., Cunningham, C.W., DeSalle, R., Buss, L.W., 1995. Class-level relationships in the phylum Cnidaria: molecular and morphological evidence. Mol. Biol. Evol. 12, 679–689. http://dx.doi.org/10.1093/oxfordjournals.molbev.a040246.

Bridge, D., Cunningham, C.W., Schierwater, B., DeSalle, R., Buss, L.W., 1992. Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. PNAS 89, 8750–8753. http://dx.doi.org/10.1073/pnas.89.18.8750.

Bumann, D., Puls, G., 1996. Infestation with larvae of the sea anemone Edwardsia lineata affects nutrition and growth of the ctenophore Mnemiopsis leidyi. Parasitology 113, 123–128. http://dx.doi.org/10.1017/S0031182000066361.

Cai, T.T., Liang, T., Zhou, H.H., 2015. Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. J. Multivariate Anal. 137, 161–172. http://dx.doi.org/10.1016/j.jmva. 2015.02.003.

Caisová, L., Marin, B., Melkonian, M., 2011. A close-up view on ITS2 evolution and speciation – a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). BMC Evol. Biol. 11, 262. http://dx.doi.org/10.1186/1471-2148-11-262.

Campbell, M.C., Tishkoff, S.A., 2010. The evolution of human genetic and phenotypic variation in Africa. Curr. Biol. 20, R166–R173. http://dx.doi.org/10.1016/j.cub. 2009.11.050.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. http://dx.doi.org/10.1093/bioinformatics/btp348.

Chiaverano, L.M., Graham, W.M., Costello, J.H., 2015. Parasites alter behavior, reproductive output, and growth patterns of Aurelia medusae in a marine lake. Mar. Ecol. Prog. Ser. 540, 87–98. http://dx.doi.org/10.3354/meps11513.

Choudhury, A., Hazelhurst, S., Meintjes, A., Achinike-Oduaran, O., Aron, S., Gamieldien, J., Jalali Sefid Dashti, M., Mulder, N., Tiffin, N., Ramsay, M., 2014. Populationspecific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. BMC Genomics 15, 437. http://dx.doi.org/10. 1186/1471-2164-15-437.

Christian, J.L., McMahon, J.A., McMahon, A.P., Moon, R.T., 1991. Xwnt-8, a Xenopus Wnt-1/int-1-related gene responsive to mesoderm-inducing growth factors, may play a role in ventral mesodermal patterning during embryogenesis. Development 111, 1045–1055.

Coleman, A.W., 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucl. Acids Res. 35, 3322–3329. http://dx.doi.org/10.1093/nar/gkm233.

Crowell, S., 1976. An Edwardsiid Larva Parasitic in Mnemiopsis. In: Coelenterate Ecology and Behavior. Springer, Boston, MA, pp. 247–250. https://doi.org/10.1007/978-14757-9724-4_26.

Daly, M., 2002. A systematic revision of Edwardsiidae (Cnidaria, Anthozoa). Invertebr. Biol. 121. http://dx.doi.org/10.1111/j.1744-7410.2002.tb00061.x.

Daly, M., Lipscomb, D.L., Allard, M.W., 2002. A simple test: evaluating explanations for the relative simplicity of the Edwardsiidae (cnidaria: Anthozoa). Evolution 56, 502–510. http://dx.doi.org/10.1111/j.0014-3820.2002.tb01361.x.

Darling, J.A., Reitzel, A.R., Burton, P.M., Mazza, M.E., Ryan, J.F., Sullivan, J.C., Finnerty, J.R., 2005. Rising starlet: the starlet sea anemone, Nematostella vectensis. Bioessays 27, 211–221. http://dx.doi.org/10.1002/bies.20181.

Database Resources of the National Center for Biotechnology Information, 2017. Nucleic Acids Res 45, D12–D17. https://doi.org/10.1093/nar/gkw1071.

Dietz, R.S., Holden, J.C., 1970a. The breakup of pangaea. Sci. Am. 223, 30–41.

Dietz, R.S., Holden, J.C., 1970b. Reconstruction of Pangaea: breakup and dispersion of continents, Permian to Present. J. Geophys. Res. 75, 4939–4956. http://dx.doi.org/10.1029/JB075i026p04939.

dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C.J., Yang, Z., 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr. Biol. 25, 2939–2950. http://dx.doi.org/10.1016/j.cub. 2015.09.066.

Ereskovsky, A.V., Richter, D.J., Lavrov, D.V., Schippers, K.J., Nichols, S.A., 2017. Transcriptome sequencing and delimitation of sympatric Oscarella species (O. carmela and O. pearsei sp. nov) from California, USA. PLOS ONE 12, e0183002. http:// dx.doi.org/10.1371/journal.pone.0183002.

Faasse, M.A., Bayha, K.M., 2006. The ctenophore Mnemiopsis leidyi A. Agassiz 1865 in coastal waters of the Netherlands: an unrecognized invasion? Aquat. Invasions. http://dx.doi.org/10.3391/ai.2006.1.4.10.

Genikhovich, G., Technau, U., 2009. The starlet sea anemone Nematostella vectensis: an anthozoan model organism for studies in comparative genomics and functional evolutionary developmental biology. Cold Spring Harb Protoc 2009, pdb.emo129. https://doi.org/10.1101/pdb.emo129.

Gosse, P.H., 1856. A manual of marine zoology for the British Isles. J. Van Voorst.

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

Haas, B., Papanicolaou, A., 2012. TransDecoder (Find Coding Regions within Transcripts).

Hand, C., 1990. Controlled Culture of a New Marine Model, the Sea Anemone Nematostella Vectensis. Biennial Report of Completed Projects 70.

Hand, C., Uhlinger, K.R., 1992. The culture, sexual and asexual reproduction, and growth of the sea anemone nematostella vectensis. Biol. Bull. 182, 169–176. http://dx.doi.org/10.2307/1542110.

Hansson, Hans G., 2006. Ctenophores of the Baltic and adjacent Seas – the invader Mnemiopsis is here! – Mnemiopsis. HGH.pdf. Aquatic Invasions. vol. 1, pp. 295–298.

Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University.

Hebert, P.D.N., Ratnasingham, S., de Waard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. London B: Biol. Sci. 270, S96–S99. http://dx.doi.org/10.1098/rsbl.2003.0025.

Javidpour, J., Sommer, U., Shiganova, T., 2006. First record of Mnemiopsis leidyi A. Agassiz 1865 in the Baltic Sea. Aquatic Invasions 1, pp. 299–302. https://doi.org/10. 3391/ai.2006.1.4.17.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. http://dx. doi.org/10.1093/molbev/mst010.

Kayal, E., Roure, B., Philippe, H., Collins, A.G., Lavrov, D.V., 2013. Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evol. Biol. 13, 5. http://dx.doi. org/10.1186/1471-2148-13-5.
Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., Flicek, P., 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011. http://dx.doi.org/10.1093/database/bar030.
Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576. http://dx.doi.org/10.1101/gr.129684.111.
Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H.A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M.Q., Holstein, T.W., 2005. Unexpected complexity of the Wnt gene family in a sea anemone. Nature 433, 156–160. http:// dx.doi.org/10.1038/nature03158.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Meth. 9, 357–359. http://dx.doi.org/10.1038/nmeth.1923.
Layden, M.J., Johnston, H., Amiel, A.R., Havrilak, J., Steinworth, B., Chock, T., Röttinger, E., Martindale, M.Q., 2016a. MAPK signaling is necessary for neurogenesis in Nematostella vectensis. BMC Biol. 14, 61. http://dx.doi.org/10.1186/s12915-0160282-1.
Layden, M.J., Rentzsch, F., Röttinger, E., 2016b. The rise of the starlet sea anemone Nematostella vectensis as a model system to investigate development and regeneration. WIREs Dev. Biol. 5, 408–428. http://dx.doi.org/10.1002/wdev.222.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189. http://dx.doi.org/10.1101/gr.1224503.
Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. Algorithms for Molecular Biology 6, 26. https://doi.org/10.1186/1748-7188-6-26.
Markey, K.L., Abdo, D.A., Evans, S.N., Bosserelle, C., 2016. Keeping it local: dispersal limitations of coral larvae to the high latitude coral reefs of the Houtman Abrolhos Islands. PLOS ONE 11, e0147628. http://dx.doi.org/10.1371/journal.pone.0147628.
Massingham, T., Goldman, N., 2007. Statistics of the log-det estimator. Mol. Biol. Evol. 24, 2277–2285. http://dx.doi.org/10.1093/molbev/msm160.
McMahon, A.P., Moon, R.T., 1989. Ectopic expression of the proto-oncogene int-1 in Xenopus embryos leads to duplication of the embryonic axis. Cell 58, 1075–1084. http://dx.doi.org/10.1016/0092-8674(89)90506-0.
Merget, B., Koetschan, C., Hackl, T., Förster, F., Dandekar, T., Müller, T., Schultz, J., Wolf, M., 2012. The ITS2 database. J. Vis. Exp. http://dx.doi.org/10.3791/3806.
Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J.C., Valencia, V., Ayala, C., PérezAngel, L.C., Rodriguez-Parra, L.A., Ramirez, V., Niño, H., 2015. Middle Miocene closure of the Central American Seaway. Science 348, 226–229. http://dx.doi.org/ 10.1126/science.aaa2815.
Morin, P.A., Luikart, G., Wayne, R.K., group the S. workshop, 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19, 208–216. http://dx.doi.org/10.1016/j.tree.2004.01.009.
Müller, T., Philippi, N., Dandekar, T., Schultz, J., Wolf, M., 2007. Distinguishing species. RNA 13, 1469–1472. http://dx.doi.org/10.1261/rna.617107.
Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. http://dx.doi.org/10.1093/molbev/msu300.
Nyholm, K.G., 1943. Zur Entwicklung und Entwicklungsbiologie der Ceriantharien und Aktinien. Zool. Bidr Upps 22, 87–248.
O'Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., de Queiroz, A., Farris, D.W., Norris, R.D., Stallard, R.F., Woodburne, M.O., Aguilera, O., Aubry, M.-P., Berggren, W.A., Budd, A.F., Cozzuol, M.A., Coppard, S.E., Duque-Caro, H., Finnegan, S., Gasparini, G.M., Grossman, E.L., Johnson, K.G., Keigwin, L.D., Knowlton, N., Leigh, E.G., Leonard-Pingel, J.S., Marko, P.B., Pyenson, N.D., Rachello-Dolmen, P.G., Soibelzon, E., Soibelzon, L., Todd, J.A., Vermeij, G.J., Jackson, J.B.C., 2016. Formation of the isthmus of panama. Sci. Adv. 2, e1600883. http://dx.doi.org/10.1126/sciadv.1600883.
Pawłowska, J., Walther, G., Wilk, M., de Hoog, S., Wrzosek, M., 2013. The use of compensatory base change analysis of ITS2 as a tool in the phylogeny of Mucorales, illustrated by the Mucor circinelloides complex. Org. Divers. Evol. 13, 497–502. http://dx.doi.org/10.1007/s13127-013-0139-1.
Peterson, K.J., Lyons, J.B., Nowak, K.S., Takacs, C.M., Wargo, M.J., McPeek, M.A., 2004. Estimating metazoan divergence times with a molecular clock. PNAS 101, 6536–6541. http://dx.doi.org/10.1073/pnas.0401670101.
Prud'homme, B., Lartillot, N., Balavoine, G., Adoutte, A., Vervoort, M., 2002. Phylogenetic analysis of the Wnt gene family. Curr. Biol. 12, 1395–1400. http://dx.doi.org/10.1016/S0960-9822(02)01068-0.
de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. PNAS 102, 6600–6607. http://dx.doi.org/10.1073/pnas.0502030102.
Reitzel, A.M., Daly, M., Sullivan, J.C., Finnerty, J.R., 2009. Comparative anatomy and histology of developmental and parasitic stages in the life cycle of the lined sea anemone Edwardsiella lineata. J. Parasitol. 95, 100–112. http://dx.doi.org/10.1645/ GE-1623.1.
Reitzel, A.M., Sullivan, J.C., Brown, B.K., Chin, D.W., Cira, E.K., Edquist, S.K., Genco, B.M., Joseph, O.C., Kaufman, C.A., Kovitvongsa, K., Muñoz, M.M., Negri, T.L., Taffel, J.R., Zuehlke, R.T., Finnerty, J.R., 2007. Ecological and developmental dynamics of a host-parasite system involving a sea anemone and two ctenophores. J. Parasitol. 93, 1392–1402. http://dx.doi.org/10.1645/GE-1250.1.
Reitzel, A.M., Sullivan, J.C., Finnerty, J.R., 2006. Qualitative shift to indirect development in the parasitic sea anemone Edwardsiella lineata. Oxford J. 46, 827–837.
Rentzsch, F., Technau, U., 2016. Genomics and development of Nematostella vectensis and other anthozoans. Curr. Opin. Genet. Dev. 39, 63–70. http://dx.doi.org/10.1016/j.gde.2016.05.024.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16, 276–277. http://dx.doi.org/10.1016/S01689525(00)02024-2.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Oxford J. Bioinf. 19, 1572–1574.
Sanderson, M.J., 2003. r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics (Oxford, England) 19, pp. 301–302. https://doi.org/10.1093/bioinformatics/19.2.301.
Seibel, P.N., Müller, T., Dandekar, T., Wolf, M., 2008. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. BMC Res. Notes 1, 1–7. http://dx.doi.org/10.1186/1756-0500-1-91.
Selander, E., Møller, L.F., Sundberg, P., Tiselius, P., 2009. Parasitic anemone infects the invasive ctenophore Mnemiopsis leidyi in the North East Atlantic. Biol Invasions 12, 1003–1009. http://dx.doi.org/10.1007/s10530-009-9552-y.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics btv351. https://doi.org/10.1093/bioinformatics/btv351.
Smith, W.C., Harland, R.M., 1991. Injected Xwnt-8 RNA acts early in Xenopus embryos to promote formation of a vegetal dorsalizing center. Cell 67, 753–765. http://dx.doi. org/10.1016/0092-8674(91)90070-F.
Sokol, S., Christian, J.L., Moon, R.T., Melton, D.A., 1991. Injected Wnt RNA induces a complete body axis in Xenopus embryos. Cell 67, 741–752. http://dx.doi.org/10.1016/0092-8674(91)90069-B.
Stampar, S.N., Morandini, A.C., Branco, L.C., da Silveira, F.L., Migotto, A.E., 2015. Drifting in the oceans: Isarachnanthus nocturnus (Cnidaria, Ceriantharia, Arachnactidae), an anthozoan with an extended planktonic stage. Mar. Biol. 162, 2161–2169. http://dx.doi.org/10.1007/s00227-015-2747-0.
Steele, R.E., David, C.N., Technau, U., 2011. A genomic view of 500 million years of cnidarian evolution. Trends Genet. 27, 7–13. http://dx.doi.org/10.1016/j.tig.2010.10.002.

Stefanik, D.J., Lubinski, T.J., Granger, B.R., Byrd, A.L., Reitzel, A.M., DeFilippo, L., Lorenc, A., Finnerty, J.R., 2014. Production of a reference transcriptome and transcriptomic database (EdwardsiellaBase) for the lined sea anemone, Edwardsiella lineata, a parasitic cnidarian. BMC Genomics 15, 71. http://dx.doi.org/10.1186/1471-2164-15-71.

Steinbeisser, H., Robertis, E.M.D., Ku, M., Kessler, D.S., Melton, D.A., 1993. Xenopus axis formation: induction of goosecoid by injected Xwnt-8 and activin mRNAs. Development 118, 499–507.

Stephenson, T., 1935. The British Sea Anemones. Ray Society Monograph No. 121. London, vol. 2, p. 426.

Stetz, J.B., smith, S., Sawaya, M.A., Ramsey, A.B., Amish, S.J., Schwartz, M.K., Luikart, G., 2016. Discovery of 20,000 RAD–SNPs and development of a 52-SNP array for monitoring river otters. Conserv. Genet. Resour. 8, 299–302. http://dx.doi.org/10.1007/s12686-016-0558-3.

Technau, U., Schwaiger, M., 2015. Recent advances in genomics and transcriptomics of cnidarians. Mar. Genom., Mar. Genom. Evol. Dev. 24, 131–138. http://dx.doi.org/10.1016/j.margen.2015.09.007.

The Chimpanzee Sequencing and Analysis Consortium, Waterson, R.H., Lander, E.S., Wilson, R.K., 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437, 69–87. http://dx.doi.org/10.1038/nature04072.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.-M., Doumbo, O., Ibrahim, M., Juma, A.T., Kotze, M.J., Lema, G., Moore, J.H., Mortensen, H., Nyambo, T.B., Omar, S.A., Powell, K., Pretorius, G.S., Smith, M.W., Thera, M.A., Wambebe, C., Weber, J.L., Williams, S.M., 2009. The genetic structure and history of Africans and African Americans. Science 324, 1035–1044. http://dx.doi.org/10.1126/science.1172257.

Vendrami, D.L.J., Shah, A., Telesca, L., Hoffman, J.I., 2016. Mining the transcriptomes of four commercially important shellfish species for single nucleotide polymorphisms within biomineralization genes. Mar. Genom. 27, 17–23. http://dx.doi.org/10.1016/ j.margen.2015.12.009.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. http://dx.doi.org/10.1093/bioinformatics/btp033.

Wylie, C., Kofron, M., Payne, C., Anderson, R., Hosobuchi, M., Joseph, E., Heasman, J., 1996. Maternal beta-catenin establishes a 'dorsal signal' in early Xenopus embryos. Development 122, 2987–2996.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., Chen, S., 2010. Use of ITS2 Region as the Universal DNA barcode for plants and animals. PLOS ONE 5, e13102. http://dx.doi.org/10.1371/journal.pone.0013102.

Zheng, W., Chung, L.M., Zhao, H., 2011. Bias detection and correction in RNA-Sequencing data. BMC Bioinf. 12, 290. http://dx.doi.org/10.1186/1471-2105-12-290.

2.3 Discussion Chapter 2

2.3.1 Species and Speciation

To understand the biological diversity around us, biologists need to quantify this diversity. One way to quantify the diversity is by counting species. This task is very arduous, as there are many definitions of species and different species-delimiting criteria.

Kevin De Queiroz lists 24 species definitions (De Queiroz, 2007) while as John S. Wilkins listed 26 species definitions (Wilkins, 2011). These species concepts are not mutually exclusive and at times use different evidence for species delimitation. From the many listed species concepts the most widely accepted concept of species at least for animals is Ernst Mayr's definition of the biological species, which defines species as population that interbreed or can potentially interbreed in nature (de Queiroz, 2005). The essence of the biological species concept is "reproductive isolation" under natural conditions. This reproductive isolation can be pre-zygotic or post-zygotic and can be tested accordingly in the controlled environment. The reproductive isolation by various means leads to the process of speciation (Figure 1A,1B).

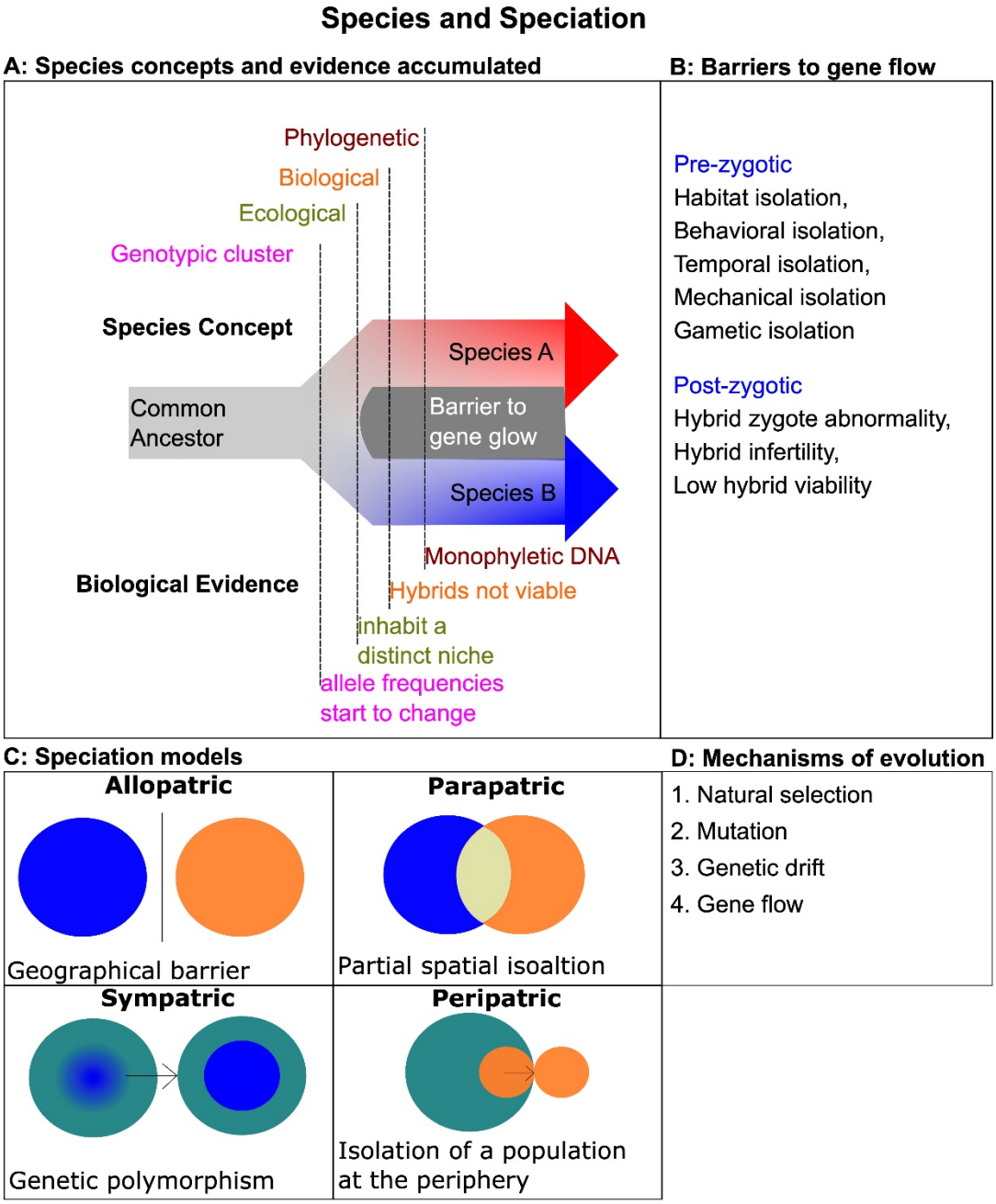


Figure1: Species and speciation
A: evolution and distinction of species. Various species concept and the biological evidences B: Lists of pre-zygotic and post-zygotic barriers of gene flow that leads to speciation events. C: Geographic speciation models for barriers D: Mechanisms that can cause a population to have a change in allele frequency in next generations.

Speciation, the origin of new species, is one of the central topics in evolutionary biology, and thus earns a significant importance. For a long time, the formation of the species is one of the central, yet very elusive subjects for the evolutionary biologists. In the 'Origin of the species',

Charles Darwin provided overwhelming support for his theory of natural selection and his theory of natural selection inherently implies the modes of generation of novel species. According to the natural selection theory, speciation can occur by two mechanisms in general, primarily through ecological evolution and secondly through adaptive divergence. In ecological evolution, there is the emergence of reproductive isolation. The premise here is that the populations that are geographically separated will differ in their environment, causing divergence from other populations and eventually become reproductively isolated distinct species (Orr and Smith, 1998). This speciation type is called 'allopatric speciation' (Figure 1C) and plays a vital role in generating biodiversity. However, geographic separation is not necessarily required for speciation, speciation can also happen in populations sharing the same geographic locations and overlapping populations notwithstanding considerable gene flow between the populations, this type of speciation is termed as 'sympatric speciation' (Dieckmann and Doebeli, 1999)(Figure 1C). Many biologists see this scenario as extremely difficult (Berlocher and Feder, 2002), however, there are many compelling examples, which suggest that this type of scenario is plausible. Examples such as insects adapting to a different type of plants/fruits (Berlocher and Feder, 2002) or hummingbirds adapting to draw nectar from particular flower for niche specialization to avoid competition (McGuire et al., 2014) or adaptation of cichlid fishes to the environments of the African Lakes (Ronco et al., 2021; Trewavas, 1947). Another geographic scenario for the speciation is 'parapatric speciation,' where two populations of a species are in non-overlapping ranges, however at the boundary between the ranges, gene flow is still possible. There are several examples, such as those of walking-stick insects (Nosil et al., 2002; Riesch et al., 2017) and salamanders (Wake and Yanev, 1986). Another mechanism of speciation is peripatric speciation, where the population at the periphery evolve into separate species. Over the course of time, there are additional mechanisms of speciation have been identified along with what Darwin had suggested, such as mutation, genetic drift, and gene flow.

2.3.2 Understanding speciation among Edwardsiella species-pair

To investigate the genetic relationship of *E.* carnea and *E. lineata*, and to elucidate whether both have a facultative parasitic life cycle we used comparative transcriptomic measures such as percentage similarity and LogDet distance, we showed that the *E. lineata* and *E. carnea* are about 97% identical on transcriptome level. In general, there are no set limits as to how much the two populations should differ genetically, for them to be considered separate species. One

way we have addressed this issue is by doing a similar analysis on the known distinct species such as primates (Human, Gorilla, Chimpanzee). Following the same protocol, we used for *Edwardsiella* species, we found that the genetic similarity between *E. lineata* and *E. carnea* is less than Human and Chimpanzee (or between Human and Gorilla), therefore the two *Edwardsiella* populations are considered as distinct species.

Using the same comparative transcriptomic analyses, we also find that the parasite found in invasive ctenophores (*M. leydii*) near the Swedish coast is more than 99% identical to *E. carnea,* thus suggesting that the parasite is indeed *E. carnea,* thereby also indicating that *E. carnea* also has a facultative parasitic life cycle and is an indiscriminate parasite to various ctenophore species. We supported this observation with SNP comparison between the datasets where we found 20 times more SNPs shared by *E. parasite sp*. and *E. carnea* than the ones shared by *E. lineata* and *E. carnea*.

Since we could not test the reproductive isolation by direct cross-fertilization as basis of biological species distinction, we employed the Internal transcribed spacer 2 (ITS2) Compensatory Base Change (CBC) test for reproductive isolation (Wolf et al., 2013). In this test the conservation of secondary structure of ITS2 in certain areas (helix II, helix III) is crucial. Any change (CBC) in these regions is correlated in 70% of the cases with reproductive isolation (Figure 2A-2B). ITS2 CBC analysis of Edwardsiella species did not find CBC in helix II or helix III between *E. lineata* and *E. carnea* suggesting a potential incomplete/lack of reproductive isolation between *E. carnea* and *E. lineata* (Figure2C-2D). However, this is based on correlative studies and the exact mechanism of action ITS2 and its effect on reproductive isolation is not yet established.

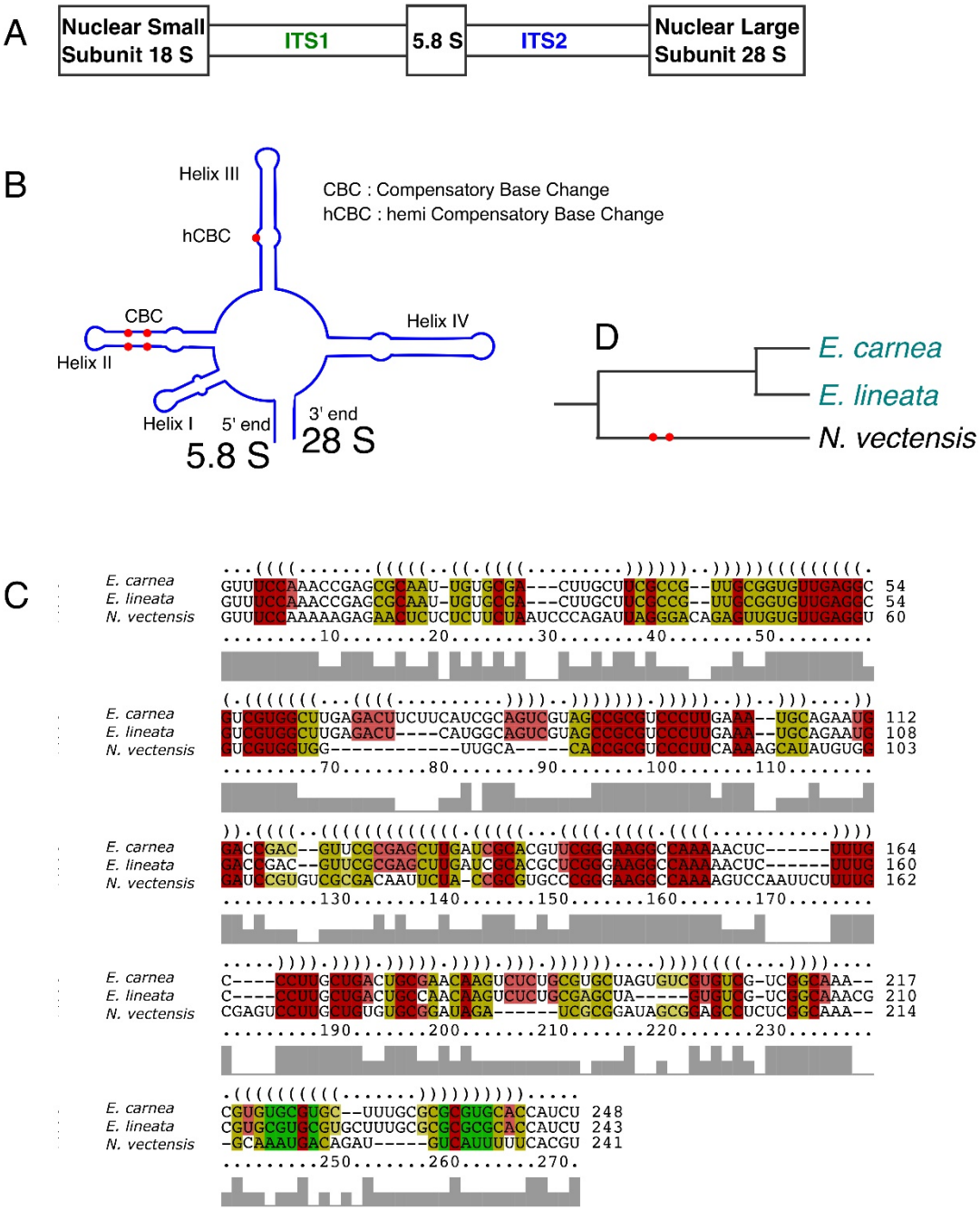


Figure2: Compensatory base change in internal transcribed spacer 2 correlates to biological concept

A: ITS2 is a subunit between 5.8S subunit and 28S subunit of rRNA. B: CBC Compensatory base change occurs when there is a mutation in the Helix II or Helix III of the ITS2 secondary structure.
C: *E lineata*, *E. carnea* and *N. vectensis* sequence alignment and secondary structure in Vienna format. D: ITS2 CBC analysis shows two CBCs between *N. vectensis* and *E. carnea* and three CBCs between *N. vectensis* and *E. lineata*.

2.3.3 Possible Speciation Event

To understand how these two populations separated and evolved into distinct species, we must consider multiple facets. Primarily, if we assumed that a population of *E. lineata* crossed the Atlantic Ocean and evolved into distinct species, we need to consider several scenarios. The only free-living mobile stage in the animal life cycle is the larva. Many marine animals including Cnidaria have free-living larvae that are known to disperse over a long range. There are even instances of larvae dispersed over 2000-4000 km, although still restricted to the coastal region (Stampar et al., 2015). The swimming speed of the *Edwardsiella* larva has not been estimated, however, we can equate it to the swimming speed of the coral larvae which has been well studied. With the modest estimated speed of 1 - 5 mm/s in all directions for the swimming speed of the larva (Stake and Sammarco, 2003), it seems highly unlikely that the free-swimming larva can cross the Atlantic Ocean. The possibility of free-swimming larva crossing the Atlantic Ocean is made even more unlikely considering the fact that the non-parasitic larva is non-feeding. Even when the aspect of ocean currents such as the Gulf Stream is considered as carrier of the free-swimming larva, the larva still needs to swim for a few hundred kilometers from the coast just to reach the Gulf Stream, which also seems exceedingly difficult and unlikely.

The next stage in the animal life cycle capable of dispersal is the parasitic stage and should be considered more likely than the larval stage. As the *Edwardsiella* parasite of the *Mnemiopsis* species is vermiform, larva can travel along with the pelagic host, known to travel large distances. *Mnemiopsis* species are very versatile and have tolerance for salinity (3.4-70 ppt), temperature (-0.7 to 35 °C), and can inhabit both coastal and estuarian waters (Powell et al., 2010). They have been found in oceanic waters and one study even recorded them in Ocean currents such as Gulf Stream (Harbison et al., 1978; Powell et al., 2010). Thus, *Mnemiopsis* could be an ideal carrier for the *Edwardsiella* larvae from the American East Coast to European coasts. However, there are records of *Mnemiopsis* species accidentally being introduced in the

1980s via the ballast water of ships to the Black Sea and then subsequently it invaded North-East Atlantic, Sea of Marmara, Caspian Sea, Aral Sea, the North, the Baltic Seas, and most recently the Mediterranean Sea (Ghabooli et al., 2013). There are no earlier records of *Mnemiopsis* species in European waters.

Importantly, calibrating the molecular clock using fossil data we estimate the separation time between the *Edwardsiella* species to be between 18.7–21.6 Mya. Therefore, any dispersal mechanism that considers this period is the more likely explanation of dispersal and speciation of these species. In that regard, we need to consider two geographical events that are very crucial. Incidentally, this period is in between two major geological events and depending on the actual divergence time either of these geographical events seems the most plausible cause for the separation of the *Edwardsiella* populations.

The first geological event is the rise of the isthmus of Panama, which led to some significant geological changes followed by some significant climatic changes (Bacon et al., 2015; Jaramillo et al., 2017; O'Dea et al., 2016). The Gulf Stream, which carries warm waters from Caribbean islands and Gulf of Mexico north across Atlantic Ocean and toward Europe has a significant role for the moderate climate in Europe. The rise of the Isthmus of Panama redirected some of the ocean currents, the ocean currents in the Pacific Ocean were directed southwards and ocean currents of Atlantic Ocean were directed northwards, forming the Gulf Stream (Figure 3A). There has been a debate about the timeline for the rise of the Isthmus of Panama (Bacon et al., 2015; Jaramillo et al., 2017; O'Dea et al., 2016). The most agreed period of the rise of about 5-3 Mya has been challenged by new evidence suggesting an earlier, more intermittent rise of the isthmus of Panama between 20-6 Mya (Bacon et al., 2015). The lower range of divergence time we estimated overlaps with the newly proposed timeline of the rise of the Isthmus of Panama. However, the new timeline of the rise of Isthmus of Panama is still debated.

Another geographical event to be considered as a probable cause of separation of two populations of *Edwardsiella* is the split of supercontinent Pangea that gave rise to the current continents (Figure 3B). A study based on the fossils of many marine invertebrates (including Anthozoa) suggest that the separation and merging of supercontinent Pangea has contributed significantly towards marine animal diversity (Zaffos et al., 2017). The process of continental drift continued for a long time and continues today; however, we are interested in the cretaceous

period of around 65 Mya, when the North American continent was still connected with the Eurasian continent. If the parent population of *Edwardsiella* was present at coast of combined North American-Eurasian landmass, it is plausible that the population became slowly separated along with the split of the continents. Although the period of separation of the continents is beyond our estimated time of divergence of the two *Edwardsiella* species, we consider this scenario as a plausible cause for the separation of the two Edwardsiella populations.

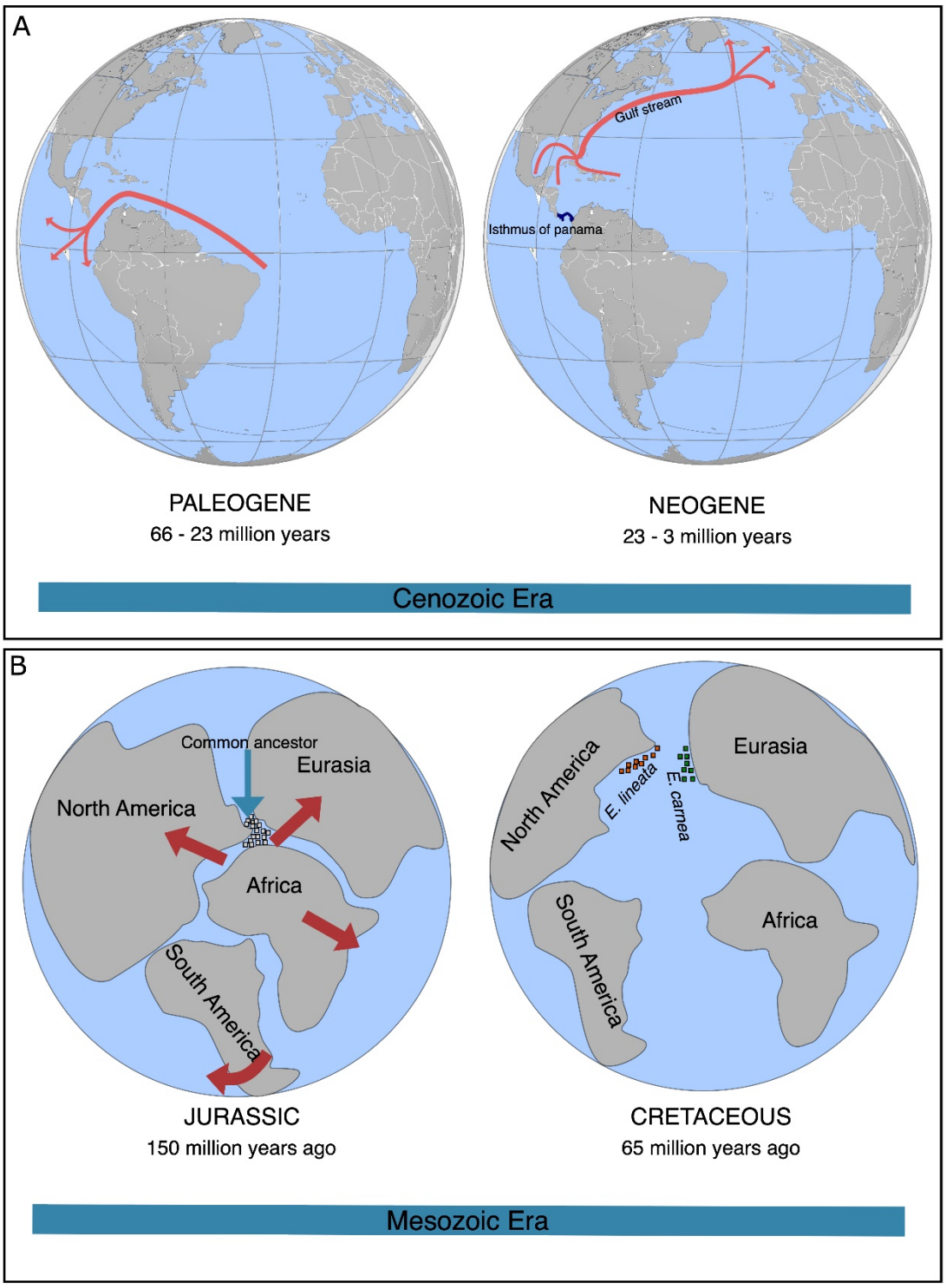


Figure3: Possible dispersion models for Edwardsiella species.
A: Closure of Isthmus of Panama led to diversion of ocean stream northwards creating existing Gulf stream
B: Separation of supercontinent Pangea

Chapter 3: THE EVOLUTION OF BRACHYURY FUNCTION

3.1 Introduction

3.1.1 Germ Layers

The germ layer theory put forward by Karl Ernst von Baer (1792–1876) suggested that all vertebrate eggs form four layers (middle two layers were later considered as single one i.e., mesoderm) during embryo development and each of these layers gives rise to distinct tissues in adult organism. (Brauckmann, 2012)

The germ layer theory provided basis for the comparative study of the developmental process across species.

In the following century many embryologists such as Ernst Haeckel, Ray Lankester have contributed towards a coherent explanation for early stages of animal development.

Ray Lankester distinguished between phyla that contain three germ layers (triploblasts) and the phyla that contain only two (diploblasts) and he also coined the terms ectoderm, endoderm, and mesoderm for the three germ layers, terms which are widely used today (Lankester, 1877).

Throughout the nineteenth century, the germ layer theory remained controversial due to the inability to assign orthologous organs arising from a germ layer across species. With the advent of technology of marking the cells and tracing it throughout the development, it became possible to study orthologous features across species.

The germ layers (primary layer of cells) arise during gastrulation, i.e., the movement of cells of a blastula into the interior of the embryo to generate inner cell layer(s). There are four primary modes of gastrulation

a) Invagination: Invagination is a process of forming a tube-like structure in an embryo by local in-pushing during gastrulation (Davies, 2013). During invagination, the folding of exterior sheet of cells towards interior occurs, akin to pushing of a finger to a balloon. The two main forms of invagination are axial and orthogonal. Axial invagination occurs at a single point while as orthogonal invagination occurs along a line. Therefore, axial invagination produces a tube while orthogonal invagination produces a trough.

b) Ingression: Ingression is a process where cells leave an epithelial cell layer. This produces animal mesenchyme cells. To achieve changes in the location or relative position of cells, cells change their program of motility and /or their adhesive relationship to the neighboring cells.

c) Epiboly: Epiboly is a process of thinning and spreading of sheet of cells. This process contributes to the expansion of the embryo surface. The process of epiboly can consist of single

layer of cell or multiple layers of cells. Both cases are followed by changes in cell shape. In case of epiboly with multiple cell-layers the process occurs via radial intercalation where two cell layers mix cell-on-cell to form a single layer (Panousopoulou et al., 2016).

d) Delamination: Delamination is a process where a sheet of cells split into separate layers

Different animals use different mechanisms to form germ layers, and most animals use more than one mechanism for gastrulation. In bilaterians, the process of gastrulation results in the formation of the three germ layers: the ectoderm, which later forms epidermis and central nervous system, followed by the middle layer, mesoderm, which leads to the formation of the muscles, bones, blood and the inner layer, endoderm, which leads to the formation of lining of the gut and associated surrounding organs (Table 1).

Table 1: Germ layer derivatives		
Endoderm	**Mesoderm**	**Ectoderm**
Gastrointestinal tract from Pharynx to upper rectum	Notochord	Ganglia of cranial nerves And spinal dorsal root
Liver	Muscular system	Major part of autonomic Nervous system
Pancreas	Skeletal system Except in head	Schwann and glial cells
Respiratory epithelium	Urinary system	Melanocytes
Middle ear epithelium	Reproductive system	Adrenal medulla
Urinary bladder Epithelium	Excretory system	Bone and connective tissue Of head
Thyroid	Dermis (Except in head and neck)	Dermis in head and neck
Parathyroid	Circulatory and lymphatic System	Nervous system
		Sensory epithelium of Eye, ear, and nose
		Oral epithelium
		Epidermis of skin and its derivatives (Including sweat glands and hair follicle)
		Mammary gland
		Epithelial lining of mouth and anus
		Tooth enamel
		Epithelium of pineal and pituitary gland

Table 1: summary of resulting germ layers and their derivatives in bilaterians.

In this study we focus on a transcription factor brachyury which is shown in vertebrates to be marker gene for the middle germ layer, mesoderm.

3.1.2 Brachyury in Vertebrates

In vertebrates, the mesoderm gives rise to many crucial body parts and tissues such as notochord, musculoskeletal, renal, reproductive, blood and cardiovascular systems (Table1). One of the defining features of chordates is the notochord. Apart from providing structural support to the developing embryo, the notochord also acts as a signaling center for patterning adjacent tissue including neural tube and somites (Stemple, 2005).

The transcription factor Brachyury is one of the crucial factors that also regulates the notochord formation in vertebrates. Brachyury (Greek Brachy=short, ury=tail) was first identified in 1927 by Nadine Dobrovolska'ia-Zavadska'ia while studying the effects of radiations on mice. The phenotype of a short, slightly kinked tail was found in heterozygous animals while homozygous mutants die *in utero*, lacking the whole posterior mesoderm (Beddington et al., 1992; Chesley and Dunn, 1936; Gluecksohn-Schoenheimer, 1944; Gruneberg, 1958). The gene was isolated by positional cloning in the 1990s and binding motif for the dimeric Brachyury protein identified (Herrmann et al., 1990; Kispert and Herrmann, 1993) by in-vitro experiments. Brachyury is also the founding member of the T-Box family of transcription factors. The expansion of the T-Box family coincides with increased complexity in multicellular animals (Sebé-Pedrós and Ruiz-Trillo, 2017).

Further studies discovered its crucial role as pan-mesodermal determinant and crucial factor for subsequent mesodermal development. It was shown to be involved in the development of posterior mesoderm, cardiac mesoderm, and notochord (Beddington et al., 1992; Kitajima et al., 2000; Martin and Kimelman, 2010; Smith et al., 1991; Stemple, 2005).

The crucial role of Brachyury in mesoderm development is conserved in all vertebrates(Beddington et al., 1992; Schulte-Merker et al., 1994).

To get insights into the underlying genetic mechanisms of its role in mesoderm formation, several labs in recent years attempted to reveal the target genes of Brachyury by genome-wide ChIP-seq or ChIP-on-chip experiments in the mouse (Koch et al., 2017; Lolas et al., 2014), *Xenopus* (Gentsch et al., 2013), zebrafish (Martin and Kimelman, 2010). The function of Brachyury in notochord development is conserved in the ascidian *Ciona intestinalis* (Takahashi

et al., 1999). Other cofactors such as *Mnx* and *Foxa* has also been shown to contribute significantly in the notochord development along with brachyury (Reeves et al., 2021).

3.1.3 Brachyury in invertebrate bilaterians

Outside of chordates, the function and target genes of Braychyury has not been very well studied. In *D. melanogaster* the *brachyury* ortholog *brachyenteron* has a vital function in ectodermal hindgut formation along with specification of caudal visceral mesoderm (Kusch and Reuter, 1999; Singer et al., 1996). In annelids brachyury seems to have role in formation of the gut where it is shown to express at ventral portion of foregut and hindgut at later stage of development. At an earlier stage it is shown to express around the blastopore (Arendt et al., 2001). In sea urchins Brachyury has a role of in the formation of the endodermal gut (Gross and McClay, 2001). Therefore, even within bilaterians Brachyury seems to have distinct functions in different phyla and may play divergent roles in the development of either ectoderm, mesoderm, or endoderm.

3.1.4 Brachyury in Diploblasts

Unlike triploblastic organisms mentioned above, the members of the early branching metazoan phyla (Porifera, Ctenophora, Placozoa, Cnidaria) are diploblasts. The existence of the gastrulation process in sponges is still a matter of debate (Adamska et al., 2007; Leys and Degnan, 2002; Nakanishi et al., 2014), and little is known about embryonic development of placozoans, however, *brachyury* is expressed at one end of the early embryo of the calcareous sponge *Sycon* (Leininger et al., 2014) as well as in few isolated cells in *Trichoplax* (Martinelli and Spring, 2003). In contrast to sponges and placozoans, ctenophores and cnidarians definitely gastrulate. In ctenophores like *Mnemiopsis leidyi* the expression of *brachyury* is ectodermal, around the gastrulation site (Yamada et al., 2010). In Cnidaria, the bilaterian sister group, the situation with *brachyury* is more complex. A single *brachyury* gene is expressed in the circumblastoporal ectoderm in the sea anemone *Nematostella vectensis* and the stony coral *Acropora millepora* (Scholz and Technau, 2003; Yasuoka et al., 2016). In contrast, in hydroids, two or three *brachyury* paralogs exist. Some of them are endodermally expressed, others are expressed in the ectoderm, however, independent of the germ layer, oral expression appears to be ancestral (Bagaeva et al., 2019; Bielen et al., 2007; Lapébie et al., 2014; Technau and Bode, 1999). Of note, as explained above, the homology of cnidarian ectoderm and endoderm to bilaterian ectoderm and endoderm has recently been questioned (Steinmetz et al., 2017).

According to this new view, oral ectoderm would correspond to endodermal tissue in Bilateria, while the inner (often termed endodermal) layer would correspond to the mesoderm of Bilateria. Thus, Brachyury would have an ancestral endodermal expression in diploblastic and invertebrate animals.

The assumption that *brachyury* is a metazoan-specific gene has been overturned by the discovery of *brachyury* in a protist *Capsaspora owczarzaki*. An orthologous sequence of *brachyury* was also found in the fungus *Spizellomyces punctatus* (Sebe-Pedros et al., 2013). The expansion of T-Box family coincides with the evolution of multicellularity and many members of this family have acquired diverse roles in animal development. In various organisms, these roles include functions like limb development, cardiac development, muscle development, morphogenic movements, and posterior identity, etc.

Taken together, in animals, Brachyury appears to have a wide range of functions and in different phyla can be found to be expressed either in the endoderm, mesoderm, or ectoderm, yet always in conjunction with the process of germ layer formation during gastrulation. Therefore, it is essential to understand ancestral function of Brachyury and evolution of Brachyury function in animal lineage.

3.1.5 Role of Brachyury in Neuromesodermal Progenitors (NMPs)

In vertebrates, Brachyury has a particularly vital role in mesoderm formation downstream of early FGF signaling, as loss of FGF signaling also causes loss of *brachyury* expression leading to defects in mesoderm formation (Kiecker et al., 2015; Schulte-Merker and Smith, 1995) . In vertebrates, most of the mesoderm is specified during the process of gastrulation, however a group of cells called Neuromesodermal Progenitor cells (NMPs) arise at caudal lateral epiblast, which continues to direct the cells to either neural or mesodermal fate depending on the inputs from Wnt and FGF signaling. The NMP cells from the tailbud contribute to both neural and mesodermal cell population involved in the elongation of the body axis posteriorly (Gentsch et al., 2013; Goto et al., 2017; Lolas et al., 2014) . While the molecular mechanism of gastrulation is well studied in vertebrate model systems (Kiecker et al., 2015), there is a renewed interest in the formation and fate of NMPs. There is significant research and debate about the NMPs, starting with the definition and characteristics of NMPs. There are several studies where the NMPs are characterized by the expression of transcription factors *brachyury*, *sox2* and *cdx1, cdx2, cdx4* (Gouti et al., 2014; Koch et al., 2017; Metzis et al., 2018).

Two potential fates follow the NMPs depending on the expression levels of two antagonistic marker transcription factors. High *sox2* expression and lower levels of Bra/T drive cells towards neural fate (Gouti et al., 2014, 2015; Tsakiridis and Wilson, 2015). By contrast, elevated levels of Bra/T along with upregulated *msgn1* and *tbx6* drive cells towards mesodermal fate (Chalamalasetty et al., 2011). In cell culture experiments, inputs from Wnt and FGF signaling has produced NMP-like cell population while abolishment of Wnt or FGF signaling, as well as loss of *cdx*, resulted in abrogation of axis elongation (Metzis et al., 2018). Therefore, there seems to be an agreement on the role of Brachyury, Wnt, FGF signals in the NMP formation however the role of Sox2 is less clear. Several studies suggested that Brachyury, induced by Wnt, FGF pathways, can promote the NMP state (Mugele et al., 2018; Tsakiridis and Wilson, 2015).

Therefore, the role of Sox2 in bringing the cells to NMP state is questioned as NMP state could be achieved by Brachyury with Wnt, FGF pathway. For instance, expression of *brachyury* in mouse epiblast stem cells induced them to become NMPs (Tsakiridis and Wilson, 2015). Therefore, the expression of *Sox2* may not be essential to initiate the NMP state but *Sox2* expression may play a crucial role in leading NMPs to neuronal fate. On the contrary, loss of Wnt signaling in NMPs, which activates *brachyury*, gives rise to spinal cord instead of mesoderm (Goto et al., 2017).

3.1.6 Approach to Study Brachyury Function

As described earlier, brachyury shows diversity in its expression patterns in various species depending on the lineage, along with diversity of brachyury function in these species. It has also been established in various species, with loss of function experiments that the brachyury is essential for the respective functions (Conlon et al., 1996; Gross and McClay, 2001; Kispert et al., 1994; Kusch and Reuter, 1999; Satou et al., 2001; Schulte-Merker et al., 1994; Wilkinson et al., 1990). Therefore, it could be argued that the brachyury protein has evolved and acquired novel functions in various species across animal kingdom ranging from single celled filastereans to mammals, however studies using brachyury cDNA from various phyla to induce brachyury overexpression in xenopus animal cap show that the brachyury function is remarkably conserved (Bielen et al., 2007; Marcellini et al., 2003; Sebe-Pedros et al., 2013) and the differences we observe in the role of brachyury in different species must therefore be attributed to differences in the target genes in these species

From the existing comparative literature on Brachyury expression, it remained unclear, what was the ancestral function of brachyury, and which changes led to its mesoderm-specific roles found in vertebrates?
To address this, we identified the Brachyury target genes in two invertebrate species, *Nematostella vectensis* and *Strongylocentrotus purpuratus*, by genome-wide ChIP-seq screens and compared them to similar datasets from the unicellular eukaryote *Capsospora owczarzaki*, one chordate, *Ciona intestinalis* and two vertebrate species *Xenopus tropicalis* and *Mus musculus*.

ChIP-seq (Chromatin Immunoprecipitation assays with high throughput sequencing) has proven to be quite a powerful method to identify DNA binding sites of transcription factors on a genome-wide scale. Chromatin immunoprecipitation coupled with either chip or sequencing approaches have been used to study triploblastic mesoderm developmental process and transcription factors involved. In *D. melanogaster* ChIP-chip experiments have revealed many key mesodermal transcription factors for instance Bagpipe (Bap), Biniou (Bin), Dorsocross 1, 2 and 3 (Doc), Myocyte Enhancer Factor 2 (Mef2), Tinman (Tin), Snail and Twist (Twi) (Mbodj et al., 2016; Wilczyński and Furlong, 2010; Zinzen et al., 2009)(Rembold et al., 2014). ChIP experiments were also used in mouse models to study another important transcription factor in mesoderm development, MyoD. With ChIP experiments it was shown that MyoD was bound significantly to the regulatory regions of other transcription factors like *pax3*, *pax7*, *six1*, *myogenin*, and *meox1* and genes like *eya2* (Gianakopoulos et al., 2011).
Additionally, RNA-seq from WT versus knockdown embryos can help us understand the effect of such binding on genome-scale gene expression in those species. Therefore, by a comparative analysis of binding sites of brachyury and its effect on target gene expression, we identified key differences in Brachyury mediated gene regulation in diploblastic and triploblastic animals.

3.2: Manuscript Brachyury

The first draft of the manuscript was prepared after cross-species comparison of targets of transcription factor brachyury. In this project I identified the targets for In-house generated ChIP-seq peaks of *N. vectensis* and *S. purpuratus*. I also performed the RNA-seq analysis for *N. vectensis, M. musculus, X. tropicalis*.

An ancestral Wnt-Brachyury feedback loop and recruitment of mesoderm-determining target genes revealed by comparative Brachyury target screens

Michaela Schwaiger[1,4#], Carmen Andrikou[2,5#], Rohit Dnyansagar[1#], Patricio Ferrer Murguia[1], Periklis Paganos[2], Danila Voronov[2], Bob Zimmermann[1], Tatiana Lebedeva[1], Heiko A. Schmidt[3], Grigory Genikhovich[1], Giovanna Benvenuto[2], Maria Ina Arnone[2*], Ulrich Technau[1,6*]

# these authors contributed equally

* corresponding authors: ulrich.technau@univie.ac.at; miarnone@szn.it

[1] Dept. of Neurosciences and Developmental Biology, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, 1090, Vienna

[2] Stazione Zoologica Anton Dohrn, Villa Comunale, 80121, Naples, Italy

[3] Center for Integrative Bioinformatics Vienna, Max Perutz Labs, University of Vienna, Medical University of Vienna, Campus Vienna Biocenter 5, Vienna, Austria

[4] current address: Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland and Swiss Institute of Bioinformatics, Basel, Switzerland

[5] current address: University of Bergen, Department of Biological Sciences, Thormøhlensgate 55, 5006 Bergen, Norway

[6] Max Perutz labs, University of Vienna, Dr. Bohrgasse 5, 1030 Vienna, Austria

Abstract

Transcription factors are crucial drivers of cellular differentiation during animal development and often share ancient evolutionary origins. The T-box transcription factor Brachyury plays a pivotal role as an early mesoderm determinant and neural repressor in vertebrates; yet, the ancestral function and key evolutionary transitions of the role of this transcription factor remain obscure. Here, we present a genome-wide target gene screen using ChIP-seq in the sea anemone *Nematostella vectensis*, an early branching non-bilaterian, and the sea urchin *Strongylocentrotus purpuratus*, a representative of the sister lineage of chordates. Our analysis reveals an ancestral gene regulatory feedback loop connecting Brachyury, FoxA and canonical Wnt signaling involved in axial patterning that predates the cnidarian-bilaterian split about 700 million years ago. Surprisingly, we also found that part of the gene regulatory network controlling the fate of neuromesodermal progenitors in vertebrates was already present in the common ancestor of cnidarians and bilaterians. However, while several endodermal and neuronal Brachyury target genes are ancestrally shared, hardly any of the key mesodermal downstream targets in vertebrates are found in the sea anemone or the sea urchin. Our study suggests that a limited number of target genes involved in mesoderm formation were newly acquired in the vertebrate lineage, leading to a dramatic shift in the function of this ancestral developmental regulator.

Introduction

T-box genes are a class of transcription factors that evolved in the common ancestor of protists and animals and have been found to play important roles in the development of all animals. The founding member, Brachyury, also called T in mice (and more recently TbxT), is a key determinant of mesoderm in vertebrates [1–5]. It is expressed in the presumptive mesoderm and continues to be expressed after gastrulation in the notochord, a defining structure of all chordates [1,6–8]. Heterozygous mouse mutants of the T gene locus show a shortened tail (hence the Greek name Brachy = short, ury = tail), while homozygous mutants die *in utero* lacking most of the body posterior to the forelimbs [9,10]. Knockdown of *brachyury* in *Xenopus* also impairs the posterior development of the embryo, while overexpression of *brachyury* mRNA in explanted animal caps leads to mesoderm differentiation [1,2,4,11]. Similar roles have been found in other vertebrates [6,12], demonstrating the conserved role for Brachyury as a pivotal mesoderm regulator across vertebrates.

Brachyury exhibits a dynamic expression pattern in vertebrate embryos. At gastrula stage it is expressed around the blastopore and during anterior-posterior axis elongation, at the tailbud stage, *brachyury* expression is mainly restricted to the most caudal part of the embryo and the notochord [1,13]. This caudal part of the developing embryo contains a population of bipotent cells called neuromesodermal progenitor cells (NMPs) [14], which exhibit temporal differences in gene expression over developmental time[15,16] and contribute to the spinal cord [17–19] paraxial mesoderm [17,18] and notochord development [17,20].
Some studies define these NMP cells by the expression of *brachyury* and *sox2*, a member of the *soxB1* gene family [15,18,21], while others argue that *brachyury* expression along with Wnt/FGF signaling is sufficient for the cells to become NMPs even in the absence of Sox2 protein [22]. Regardless of the definition of NMPs, the mesoderm-defining role of Brachyury has been established within these cell populations. The notochord-specific function of Brachyury also appears to be conserved in non-vertebrate chordates. In the ascidian *Ciona intestinalis*, representing the sister group to the vertebrates, *brachyury* expression and function is restricted to the notochord and *brachyury* knockdown abolishes notochord development [23]. In amphioxus, there are two *brachyury* paralogs. While to date no functional data are available for amphioxus, the expression pattern of *bra2* is very similar to that in vertebrates, suggesting that it likely plays a similar role in this cephalochordate [24], while *bra1* appears to be more restricted to the notochord [25]. Thus, Brachyury is key for mesoderm

formation and in particular for notochord development in vertebrates, and, possibly, in chordates. However, whether Brachyury has a mesodermal function outside chordates is less clear.

In numerous investigated invertebrate species, *brachyury* is expressed around the blastopore, similar to the early *brachyury* expression in vertebrates [26]. For instance, in Echinodermata (e.g. sea urchins, sea stars), members of the non-chordate deuterostome group Ambulacraria, *brachyury* is expressed at the vegetal pole marking the margin of the blastopore [27–29]. Later, *brachyury* expression is confined to the most vegetal part of the hindgut surrounding the anus (proctodeum) and the stomodeum [30,31]. Sea urchin Brachyury has been functionally implicated in the regulation of gastrulation movements [32]. In protostomes, the role of Brachyury shows remarkable similarities to deuterostomes during gastrulation, for instance in the annelid [33,34] and mollusc embryos [35,36].

Brachyury has also been studied in non-bilaterian animals, including cnidarians. For instance, in the freshwater polyp *Hydra*, the *brachyury* homolog *Hybra1* is expressed at the hypostome and acts as an early marker of head formation during budding, regeneration and embryonic development [37–39]. While hydrozoan cnidarians do not form a blastopore during gastrulation, the area giving rise to the hypostome is equivalent to the blastoporal region in other cnidarian clades. In the anthozoan sea anemone *Nematostella vectensis, brachyury* is expressed at the blastopore margin [40]. Functional studies in *Nematostella vectensis* and the coral *Acropora digitifera* suggest that *brachyury* may be involved in the formation of the pharynx and the endo-ectoderm boundary [41,42]. Strikingly, when *Hybra1* mRNA is injected into blastomeres of the animal pole of *Xenopus* embryos, it is capable of inducing mesoderm formation with high efficiency, showing that in the context of the vertebrate, the cnidarian transcription factor behaves like its cognate vertebrate homolog [43,44]. Thus, the protein is conserved enough to mimic the function of the vertebrate Brachyury and to act as a mesoderm determinant. These observations taken together raise the question: How were the genomic targets and the gene regulatory network downstream of Brachyury altered in the course of animal evolution, in order for this transcription factor to acquire such diverse developmental roles? Therefore, to shed light on the evolution of Brachyury function, genome-wide surveys of its downstream genes in taxa with evolutionarily informative phylogenetic positions are needed (Fig. 1A).

To reconstruct the evolutionary changes in the Brachyury gene regulatory network, we carried out genome-wide ChIP-seq and identified Brachyury target genes in two invertebrate species with key phylogenetic placements: the non-bilaterian, diploblastic sea anemone *Nematostella vectensis* and the deuterostome sea urchin *Strongylocentrotus purpuratus*, belonging to the sister group of chordates. We revealed an ancestral developmental kernel conserved between the cnidarians, echinoderms and vertebrates that includes a feedback loop between Brachyury, FoxA and canonical Wnt signaling. This circuit is involved in the axial patterning of the primary body axis. By contrast, only very few homologous genes that are crucially involved in vertebrate mesoderm formation and differentiation are also Brachyury targets in *Nematostella* or *Strongylocentrotus*. Notably, we did find a significant number of neuronal target genes shared between *Nematostella, Strongylocentrotus* and vertebrates, many of which are repressed as revealed by *brachyury* knockdowns. This suggests that axial patterning, endoderm specification and suppression of neuronal targets may be the ancestral function of Brachyury and the acquisition of target genes conveying the role of Brachyury in mesoderm formation emerged within chordates.

Materials and methods

Animal culture

Nematostella polyps were kept as previously described [45,46] at 18°C and largely in the dark. Spawning was induced by a combination of light and elevated temperature [45]. Embryos were raised at 21°C. Adult *Strongylocentrotus* individuals were kept in circulating seawater aquaria at Stazione Zoologica Anton Dohrn in Naples. Spawning was induced by vigorous shaking of gravid animals. Embryos were raised at 15°C in filtered Mediterranean Sea water (FSW) diluted 9:1 with deionized water [47–50]

Antibody generation

A 6xHis-tagged full-length clone of *Nematostella Brachyury* was expressed in *E. coli* strain BL21. After transformation, bacteria were grown overnight in LB supermedium + Ampicillin. Then, 10 ml LB medium / Ampicillin were inoculated with 500 µl of overnight culture, grown for 3 h (O.D. 600 = 0,5-0,6), IPTG was added to a final concentration of 1 mM, and bacteria culture was incubated for 21 h at 4°C. Brachyury protein was purified using Ni-NTA agarose (Qiagen), followed by polyacrylamide gel electrophoresis. The Brachyury band at ~50 kDa was excised and used to inject two rabbits (PRIMM Biotech).

For *Strongylocentrotus* Brachyury, two polyclonal antibodies from 2 different rabbits were raised against the recombinantly produced C-terminal domain excluding the T-box domain. The antibodies were affinity purified by PRIMM Biotech prior to their use.
For both *Nematostella* and *Strongylocentrotus* the reactivity of polyclonal antibodies from each rabbit were assessed by both Western blot and immunohistochemistry, confirming their specificity (Supplemental Fig. 1). For the ChIP replicates, antibodies from the two different rabbits was used for each.

Chromatin immunoprecipitation and library preparation

ChIP-Seq was carried out essentially as previously published [51]. In brief, embryos at gastrula stage were fixed with 2% formaldehyde for 12 min. Nuclei were collected by Dounce homogenizing the embryos. Chromatin was fragmented to an average size of 100-200bp using a Covaris S1 ultrasonicator with the following settings: Duty Cycle: 20%, Intensity: 5, Cycles/Burst: 200, Duration: 240s, Mode: Frequency Sweeping. To carry out the immunoprecipitation, 300 µg protein (as quantified by Bradford assay) of the fragmented chromatin was incubated with 5 µg of *Nematostella* or *Strongylocentrotus* anti-Brachyury antibodies. Chromatin from ~20 immunoprecipitations was pooled to obtain ~2ng of DNA, as measured by a Pico Analyzer. This was then used for library preparation according to Illumina ChIP-seq DNA Sample Prep Kit instructions (Catalog number IP-102-1001). The quantity and quality (size distribution) of the libraries was confirmed using an Agilent Bioanalyzer. Deep sequencing was performed at the Vienna BioCenter Core Facilities (VBCF) (https://www.viennabiocenter.org/facilities/) with 50 bp SE HiSeq 2500 reads. We obtained 39065464, 33397307 and 43959250 reads for Nematostella Brachyury ChIP replicate 1, replicate 2, and Input, respectively. Of these, 21889877, 17420963, and 25679308 were uniquely mapped to the genome (see below).

Ortholog detection

For ortholog detection we used the OMA database [52–54]. Before we subjected our in silico translated protein sequences to OMA, we excluded peptide sequences smaller than 100 amino acids. To detect the orthologs over the wide evolutionary distance of our investigated organisms, it was necessary to use a less stringent 'Length tolerance ratio' = 0.2 parameter of OMA. The other OMA parameters used default values.

ChIP-seq analysis

In addition to our ChIP-seq datasets for Brachyury in *Nematostella* and *Strongylocentrotus* we acquired raw reads of published Brachyury ChIP-seq datasets of mouse [21] (GEO: GSE93524) and *Xenopus* [55] (GEO: GSE48560). Reads from *M. musculus, X. tropicalis, S. purpuratus, N. vectensis* were mapped to genome versions GRCm38 (mm10), v9.1, v5.0, v1.0, respectively, with BWA using the BWA-MEM algorithm [56]ChIP-seq peaks were called using Peakzilla [57] using Input sequencing as control. Depending on the bimodal distribution of the reads, the Peakzilla algorithm assigns a score to each peak. We examined the relationship between Peakzilla peaks and their scores, and the coverage of ChIP-seq reads across the genomes, thereby determining the optimal score cutoff to filter high-quality peaks in each dataset, where the lowest scoring peaks that pass the cutoff are still visually detected as peaks in the coverage tracks. Consequently, we removed peaks with a score less than 2 for mouse and *Xenopus*, less than 0.5 for *Strongylocentrotus* and less than 5 for *Nematostella*. Of note, the number of detected target genes in *Strongylocentrotus* is noticeably lower than in the other species, which could be either due to biological (i.e. fewer target genes) or technical reasons (i.e. less sensitive antibody) or a combination of both. This may lead to an underestimation of the number of conserved targets in our comparative analyses.

Target gene selection

To identify target genes, a commonly used procedure is to select the closest gene to an identified peak. However, in more compact genomes like *Nematostella* or *Strongylocentrotus* (5-10 genes per 100 kb) peaks in the intergenic space are often only marginally less distant to a gene in the other direction. Therefore, we identified the two closest genes to the binding site. Next, using ortholog information obtained earlier with OMA [52,53], we determined whether either of the two closest genes has an ortholog that is a target in any other species considered in this study. In about 20% of *Nematostella*, 24% of *Strongylocentrotus*, 3% of *Xenopus* and 28% of mouse peaks, no ortholog information was found, thus, the closest gene to the binding site was assumed to be the target. If both closest genes had orthologous targets, we selected the one with larger number of orthologous target genes (Supplemental Fig. 2D). In case an equal number of the orthologous target genes was found, we kept both putative targets and marked them with ++ (Supplementary Table 1).

Motif analysis

To find and compare the binding motifs of Brachyury we used the MEME-ChIP (4.11.2) suite [58]. For the number of motifs to be predicted in a sequence we selected 'anr ' (Any number of repetitions) mode and 1e-10 for the e-value cut-off criterion. The discovered motifs were scanned against the non-redundant sequence profiles obtained from the JASPAR database [59]. The other parameters were set to default (A detailed list of parameters can be found on https://github.com/dnyansagar/gene_regulatory_network).

Co-occupancy of motifs

To investigate what other motifs are present in the vicinity of Brachyury binding motif and their spatial arrangement, we made use of MEME-ChIP with an in-house python script, which finds all the motifs in a peak and their absolute distances from the center of the peak. It also groups the motifs into motif groups based on motif families in JASPAR [59]. In case of overlapping motifs, it selects the motif with better FDR adjusted p-value assigned by Fimo from the MEME-ChIP suite [58].

Brachyury knockdowns and RNA-seq

To knockdown Brachyury function in *Nematostella*, we used two non-overlapping antisense morpholinos against *brachyury*: the translation-blocking morpholino BraATG-MO TCGTCCGAGTGCATGTCCGACTATG, and the splice-blocking morpholino BraSpliceMO TCCCTGGTTGTCAACCATACCGTCC. 3-5 pl of both morpholinos were injected at the concentration of 500 µM. 50 µg/ml Dextran-Alexa Fluor 488 MW 10000 (ThermoFisher) was co-injected as tracer to ensure proper delivery of MO [60]. Standard morpholino (Genetools) StdMO CCTCTTACCTCAGTTACAATTTATA was injected as a control at the same concentration. PolyA-enriched RNA samples were collected and processed for library production using the Lexogen kit and sequencing (50bp single-end HiSeqV4, ~ 30 Million reads per sample, for more details check https://github.com/dnyansagar/gene_regulatory_network/tree/master/rnaseq).

To knockdown *Strongylocentrotus brachyury*, we injected 2-4 pl of a 200 µM solution of Morpholino previously characterized with the sequence: CGCTCATTGCAGGCATAGTGGCG [31]. We quantified the mortality at 24 hr after fertilization and discarded all experiments with <80% survival. Embryos were either fixed for WMISH or used for isolation of RNA extraction at 24 hr. PolyA-enriched RNA samples were collected and processed for library production using the Illumina Truseq RNA sample

preparation kit (unstranded) and sequencing (100bp paired-end HiSeq, ~ 50 Million reads per sample).

Differential gene expression analyses

For *Nematostella*, we generated 6 biological replicates of each translational and splice morpholino. Principal component analysis (PCA) showed that the replicates cluster together (Supplemental Figure 3A). For *Strongylocentrotus*, we generated three biological replicates (Supplemental Figure 3B). This was compared to three replicates of Morpholino knockdown in *Xenopus* [55] (GSE48663) (Supplemental Figure 3C) and two replicates of mouse knockout mutants [21](GSE93524) (Supplemental Figure 3D).

RNA-seq reads for *Mus* (mouse), *Xenopus, Strongylocentrotus* and *Nematostella* were subjected to initial quality control with FastQC (v0.11.5) [61] and a quality control report was generated with MultiQC (v1.6) [62]. The index for *Mus, Xenopus* and *Nematostella* species was generated for STAR aligner (v2.5.3a) [63] and the reads were aligned to the respective genomes. Reads aligned to genomic features were counted with featureCounts (Subread v1.6.2) [64,65]. The count files generated were used as input to SARTools (v1.6.0) [66], which is a wrapper R script around two widely used differential expression analysis tools, namely DESeq2 [67] and edgeR [68]. Genes with an FDR corrected p-value (q-value) of 0.05 are considered as differentially expressed. The RNA-seq reads for *Strongylocentrotus* were quasi-mapped to the *Strongylocentrotus purpuratus* genome v.5 associated transcriptome using Salmon (v0.11.3) [69]. The resulting count files were used in DESeq2 [67] for differential expression analysis, differentially expressed genes with p-adjusted value of less than 0.05 were considered significant. A quantitative analysis of shared orthologous direct or indirect targets that are differentially expressed in the respective organisms is found in Supplemental Fig. 3F.

Whole mount *in situ* hybridisation (WMISH)

WMISH in *Nematostella* was carried out as described before [70] with a few minor modifications. Briefly, embryos were fixed in 4% PFA in Phosphate-Buffered Saline (PBS) for 1 h at RT, and then washed 5x with 0.2% Tween in PBS at RT. The embryos were then washed a single time with 50% methanol in PBS with 0.2% Tween and lastly transferred into 100% methanol and stored at -20 °C until further processing. Hybridization was performed at 1 ng riboprobe/µl, overnight at 63°C, and detection with alkaline phosphatase conjugated anti-digoxigenin (1:2000) antibody (Roche). If no strong signal and mostly background began to develop in the NBT/BCIP solution, the embryos were washed 5x in PTw and then left to stand in PBS with 2% w/v triton-x100 until the background was at acceptable levels. This was repeated until a strong signal was observed.

In *Strongylocentrotus*, WMISH was done as previously described [71]. Briefly, embryos were fixed in 4% PFA in MOPS Buffer overnight at 4°C, then gradually dehydrated to 70% ethanol and kept at -20°C until use. Fixed embryos were gradually rehydrated in MOPS buffer, washed several times with the same medium and were pre-hybridized for 3 hours in the hybridisation buffer. Hybridization was carried out with 0.1 ng/µl of probes and incubated for 1 week. The signal for colorimetric in situ hybridisation was developed using anti-digoxigenin AP (Alkaline Phosphatase) conjugated antibody and the AP substrate, while fluorescent in situ hybridization was developed with fluorophore-conjugated tyramide (1:400 reagent diluents, Perkin Elmer).

Immunohistochemistry

Immunohistochemistry in *Nematostella* was essentially performed as described in [70] with the following changes. Briefly, embryos were fixed at 4°C in 4% PFA in PTwTxD (1xPBS with 0.2% Tween, 0.2% Triton X-100 and 0.2% DMSO). All further washing steps were performed with PTwTxD. They were then washed 1x with PTwTxD and ice-cold acetone and put on ice. Once the embryos had settled down the acetone was removed, and they were washed 10x times with the cold PTwTxD. The embryos were then blocked 2 h at RT in the blocking solution containing 20% v/v of heat-inactivated sheep serum and 80% of 1% w/v BSA in PTwTxD. The blocking solution from the embryos was substituted with the rabbit anti-Bra preadsorbed in the blocking solution (1:500) for the time of the blocking and incubated overnight at 4°C on a table rocker. The next day, the embryos were washed 10x with cold PTwTxD and blocked once again for 2 h at RT in the blocking solution. Blocking

solution was then replaced with preabsorbed goat anti-rabbit Alexa Flour 568, 1:2000, Invitrogen) mixed with DAPI (final concentration 5 µg/ml) and Alexa Fluor 488 phalloidin (final concentration 5 U/ml, Invitrogen) and incubated overnight at 4°C. Next day, the embryos were washed 10x in PTwTxD, infiltrated overnight with Vectashield (Vector labs), and imaged.

Immunohistochemical detection of Brachyury in *Strongylocentrotus* was performed as described in [71]. Briefly, the embryos were fixed in 4% PFA in filtered sea water for 15 min at room temperature (RT), dehydrated in 100% ice cold methanol for 1 min and washed several times with 1x PBS. PBS was replaced with a blocking solution (4% Sheep Serum, 1% BSA in 1x PBS) and the samples incubated for 1h at RT. The blocking solution was replaced with the solution containing the primary antibodies diluted in blocking solution (Brachyury 1:100, Nkx2.1 1:600, SoxB2 1:500) and incubated either overnight at 4°C or 90 min at 37°C. Embryos were washed several times with PBS, the secondary antibody goat anti-rabbit AlexaFluor 488 (Invitrogen) diluted in PBS was added and then washed again several times.

Phylogenetic analysis

Phylogenetic analysis of T-Box, Sox, Zic, TFAP, NR2, RNF, Rfx family members was performed using maximum likelihood methods. Orthologs and paralogs were obtained from the respective proteomes using BLASTP with the protein sequences from *Nematostella* and mouse as query. The peptide sequence alignments were generated using MAFFT with the parameters –maxiterate 1000 –localpair [72] and cut to the unambiguously aligned core region. From the alignments maximum likelihood trees were reconstructed using IQ-TREE 1.6.12 Substitution models of LG+I+G4 (T-Box and Sox) and LG+F+G4 (Zic, TFAP, NR2, RNF, Rfx) were chosen after running model selection in IQ-TREE. Support values were generated using UFboot [73] as implemented in IQ-TREE with 1000 (T-Box and Sox) and 10000 (Zic, TFAP, NR2, RNF, Rfx) bootstrap samples, respectively.

Target gene comparison between species

We used OMA (**O**rthologous **MA**trix) for its ability to report strict orthologs by verifying pairs of genes [52,53,74]. All the translated peptide sequences were filtered against proteins shorter than 100 amino acids. This filtering resulted in 8381 (*C. owczarzaki)*, 25729 (*N. vectensis)*, 28658 (*S. purpuratus)*, 25425 (*C. intestinalis)*, 39662 (*X. tropicalis)* and 21317 (*M. musculus*) proteins. Our orthology analysis resulted in 1882 orthologous groups shared by all species (4791 orthologous groups where at least 5 species have orthologs, 7922

orthologous groups where at least 4 species have orthologs, 13149 orthologous groups where at least 3 species have orthologs, and 28918 orthologous groups where at least 2 species have orthologs).

We then selected the target genes coding for transcription factors using InterProScan [75] annotation based on the presence of the DNA binding domain from the Pfam database, then sorted target genes into 5 nodes based on evolutionary lineages representing Metazoa, Bilateria, Deuterostomia, Chordata and Vertebrata. Genes were annotated according to the mouse gene annotations. Homologous target genes were determined by OMA, afterwards the OMA annotation was manually improved for several genes by more detailed phylogenetic analyses (see sections above). Homologous target genes that are shared between at least two organisms were placed at the node where the two lineages diverged, regardless of how many other members of the ingroup also shared this target gene. For instance, a target gene detected in *Nematostella* and at least one of the bilaterian species would be placed at node II (common ancestor of cnidarians and bilaterians). Vertebrates often have multiple paralogs that have only one homolog in the earlier diverging species, which can lead to double assignments in the analysis. For instance, *foxA*, which is a conserved target in *Nematostella*, sea urchin and vertebrates, has duplicated into several paralogs in vertebrates [76]. OMA finds *foxA2* as a shared target of *Nematostella* and mouse and *foxA3* as a shared target of sea urchin and mouse. In such cases we only kept the homolog at the lower node.
To compare the expression of ancestral target genes in different nodes in mouse gastrulation stage cells of neuronal, endodermal, and mesodermal origin, we extracted the mouse ENSEMBL gene IDs for all genes in each node. In the case where we extended the OMA annotation by phylogenetic analysis, and one ancestral gene was present in node II, for example, FoxA, we used all mouse paralogues, for example, Foxa1, Foxa2, and Foxa3. We then used the Mouse Gastrulation Data R package (Version 1.8.0) to extract single cell RNAseq data for mouse embryos at stages E8.0 and E8.5 for all node target genes described above. We used the gene counts to calculate log normalized expression values (logNormCounts function of R package scuttle) to categorize genes as neuronal, endodermal, or mesodermal (and others). Then we generated a pseudo-bulk expression matrix per mesodermal, endodermal, and neuronal cell category for all genes, and normalized the resulting counts to counts per million reads (cpm). Node genes were selected from this pseudo-bulk matrix and the log-fold-change of neuronal vs mesodermal expression or

endodermal vs mesodermal expression was calculated as log2 of neuronal (or endodermal) cpm +1 / mesodermal cpm +1.

Results

Genome wide detection of Brachyury binding sites in *Nematostella* and *Strongylocentrotus*

In order to reconstruct the evolution of Brachyury function, we analyzed its genomic targets in representatives of phylogenetically informative groups (Fig. 1A), the sea anemone *Nematostella* and the sea urchin *Strongylocentrotus*. We first generated antibodies against *Nematostella* and *Strongylocentrotus* Brachyury proteins (see Materials and Methods, Supplemental Fig. 1A-D). Antibody staining of gastrula stage embryos of *Nematostella* using confocal microscopy confirmed that the antibodies detect nuclei at the ectodermal margin of the blastopore, where mRNA expression is also detected by whole mount in situ hybridisation (WMISH) (Fig. 1B; Suppl. Fig. 1). Similarly, in *Strongylocentrotus* gastrulae, Brachyury protein is detected at the endodermal margin of the blastopore as well as at the future stomodaeum, where the archenteron will break through, reflecting the mRNA expression (Fig. 1C). Loss of function experiments by morpholino knockdowns and localized gain of function experiments confirmed the specificity of the antibodies in *Nematostella* (Supplemental Fig. 1A-E) and *Strongylocentrotus* (Supplemental Fig. 1D)[31].

We next used these antibodies to perform ChIP-seq in early gastrula stage embryos. For *Nematostella*, we identified 2389 putative binding sites in two highly reproducible replicates, which mapped to 1543 putative target genes (Fig. 2A, 2B, for details see Materials and Methods; Supplemental Fig. 1E). For *Strongylocentrotus*, we identified 490 binding sites, which corresponded to 391 target genes (Fig. 2B; Supplemental Fig. 1F). The lower number of detected binding sites in the sea urchin might be due to less effective ChIP enrichments or reflect a biological difference. Notably, a large fraction of the Brachyury peaks in *Nematostella* (850/2389) overlapped with the previously 2559 identified enhancers at the gastrula stage [51] as defined by the combination of p300, H3K27ac, and H3K4me1, suggesting that about 1/3 of the identified enhancers are bound by Brachyury at the gastrula stage (Supplemental Fig. 2A,B). Similarly, a large fraction (338/490) of the Brachyury binding sites detected by ChIP-seq in *Strongylocentrotus* are found in open chromatin as detected by ATAC-seq (Supplemental Fig. 2C). However, we also found an even larger fraction in both species that are not associated with active enhancers or open chromatin (see Discussion). To

compare our findings with other species, we re-analysed previously published ChIP-seq data [21,55] using our peak calling thresholds and found for mouse and *Xenopus* 4000 and 2497 peaks, respectively. Corroborating the published results [21,55], our called mouse peaks corresponded to 3060 putative target genes, while in *Xenopus* 1376 target genes were detected (Fig. 2B).

When analysing the distribution of the Brachyury binding sites with respect to target genes, we found that the distribution reflects the genome sizes of the species. In species with larger genome size the peaks are distributed over a larger range, sometimes beyond 100 kb while in species with smaller genome size the peaks are much more in the vicinity of transcription start site (TSS) (Fig. 2H-I). This is particularly obvious in the case of *Nematostella*, where about 50% of all Brachyury binding sites are located within 1kb upstream or downstream of the TSS, compared to all three deuterostome species (5-12%) (Fig. 2D-F). The smaller genome size also makes the gene density in *Nematostella* higher (about 1 gene / 10 kb) than other species concerned [77,78]. Different transcription factors tend to have strikingly different positional specificity within the enhancer regions [79]. From the binding profile of Brachyury (Fig. 2D-G), Brachyury seems to be strongly concentrated around the TSS site in all the species under consideration.

Brachyury binding motifs are highly conserved

The Brachyury DNA binding motif was originally found to be a palindromic sequence by an in vitro Selex approach [80,81], which was supported by ChIP-seq in *Xenopus* embryos [55]. We detected 700 and 238 palindromic Bra binding motifs in the sea anemone and sea urchin Brachyury ChIP-seq peaks, respectively. This indicates that the ability for dimerization is an ancestral feature of Brachyury (Fig. 2B-C). We detected significantly higher Brachyury peaks with palindromes compared to peaks with half-palindromes (Supplemental Fig. 4A), suggesting a higher binding affinity of the dimer compared to the half palindromes. However, we found no significant correlation between the distribution of half-palindromes or palindromes with the expression level or GO category of the target genes (data not shown). Thus, the biological significance of the distribution of half-palindromes versus palindromes remains obscure at this point. However, the detailed comparison of the binding motifs of Brachyury in *Nematostella, Strongylocentrotus, Xenopus* and mouse showed that the canonical binding motifs are highly conserved throughout metazoans (Fig. 2C). Similar motifs have also been found among the ATAC-seq peaks in the protist *Capsaspora* [82]. This suggests that the DNA-binding domain and the corresponding binding motif have been

conserved from protists to humans, suggesting a strong positive selection. This might explain why the overexpression of the cnidarian or even *Capsaspora brachyury* mRNA in frogs induces the formation of mesoderm [43,44,83].

In order to detect potential co-factors or competitors of Brachyury in the four species, we searched the peaks for an enrichment of other common motifs, also considering their position with respect to the peak summit. As expected, Brachyury motifs locate centrally, while other transcription factor binding motifs are enriched within ~50bp from the peak center, suggesting possible co-binding (Supplemental Fig. 4B). When scanning the peak sequences with FIMO for known motifs matching to the enriched motifs identified in any species, we found a similar distribution of putative binding sites for homeodomain, bHLH, Pax, HMG (Sox) and Fox proteins in all species (Supplemental Fig. 4B). Especially in *Nematostella*, a substantial number of putative Sox, Fox and Hox motif sites are found together with Brachyury (Supplemental Fig. 4B). This is of particular interest, as the expression of *foxA*, *foxB,* as well as *soxB1* and *soxB2* is partially overlapping with *brachyury* in *Nematostella* (see below).

Brachyury can act as activator or repressor of developmental regulator genes in both the sea anemone and sea urchin

Next, we wished to study the role of *Nematostella* and *Strongylocentrotus* Brachyury on gene expression by knockdown of Brachyury function (Supplemental Fig. 3). In *Nematostella*, we generated 4, 5 and 6 replicate transcriptomes of gastrula stage embryos each injected with control, splicing or translation blocking morpholinos, respectively. A principal component analysis showed that splice morpholino and translation morpholino transcriptomes cluster together (Supplemental Fig. 3A), distinct from the control Morpholino experiments, without batch effects. In *Strongylocentrotus*, we used a previously published and validated translation blocking morpholino [31]. In *Strongylocentrotus,* differential gene expression analysis revealed 90 differentially expressed ChIP target genes (corresponding to 23 % of ChIP targets) in *bra* knockdowns (44 down-regulated and 46 up-regulated) (Supplemental Fig. 3B). In *Nematostella,* the overlap between the ChIP target and differentially expressed genes is 90 genes (6% of ChIP-targets), of which 38 are down-regulated and 52 are up-regulated (Supplemental Fig. 3A,E,F; see also the summary of all data in Supplemental Table 1). We compared these data with transcriptomes comparing knockouts or knockdowns from mouse and *Xenopus*, respectively (Supplemental Fig. 3C-E). Of 3060 mouse putative target genes,

108 (3.5% of ChIP-targets) are differentially expressed in *brachyury* mutants, among them 81 downregulated and 27 are upregulated [21]. By comparison, in *Xenopus,* of 1376 Brachyury target genes 73 (5.3% of ChIP-targets) are differentially expressed in Brachyury knockdowns (40 are down-regulated and 33 up-regulated) [55]. Thus, the proportion of genes regulated in relation to ChIP target genes is in line with earlier published reports of 1% to 10% [84–86]. Gene Ontology (GO) analyses of the Brachyury ChIP targets found in *Nematostella* and *Strongylocentrotus* shows an enrichment in categories of transcriptional regulation, Wnt signaling, multicellular organism development, signal transduction and biological regulation (Supplemental Fig 4C; Supplemental Table 3), which is similar to the GO categories of Brachyury ChIP targets reported in vertebrates [21,55,87]. This suggests that in these species Brachyury is a developmental regulator that mainly acts positively or negatively to control other developmental genes.

Brachyury forms an ancestral gene regulatory loop with Wnt signaling and activates or represses key blastoporal genes

When analysing the target genes in *Nematostella* and in *Strongylocentrotus*, we found that many members of Wnt, FGF, Notch and BMP signaling pathways as well as several transcription factors were among the direct target genes (Fig. 3; Fig.6, Supplemental Table 2). Interestingly, at the tailbud stage of vertebrates Brachyury forms a feedback loop with canonical Wnt3 signaling as well as with FGF signaling in the neuromesodermal progenitors (NMPs) [88–90]. In *Nematostella* and *Strongylocentrotus*, Wnt signaling is active in the blastopore region where *brachyury* is expressed [91–94]. Moreover, functional manipulation of the Wnt signaling pathway has demonstrated that *brachyury* is activated by Wnt/beta-catenin signaling [95–97]. To functionally investigate the regulation of key target blastoporal transcription factors and signaling pathway members by Brachyury, we carried out WMISH in wild-type and morphant embryos in *Nematostella* and *Strongylocentrotus.* We focused on Brachyury ChIP targets that were regulated in our RNA-seq experiments or were previously shown to have an overlapping or mutually exclusive expression pattern with *brachyury*. In both sea anemone and sea urchin, *brachyury* expression is upregulated upon Morpholino mediated knockdown, suggesting a negative feedback loop of Brachyury on its own gene expression (Figs. 4A,5A).

In *Nematostella,* other genes with an overlapping expression with *brachyury* are either abolished in the overlapping domains (e.g., *wnt1, wnt3, foxA, foxB, fgf8a, ephrinB2, myc2*) or strongly downregulated (*wnt4, wntA*) upon Brachyury knockdown (Fig. 4; Supplemental Fig.

5). By contrast, *wnt2*, which shows a complementary expression pattern to *brachyury*, expands into the blastopore lip domain in *brachyury* morphants (Fig. 4), indicating that Brachyury represses *wnt2* from invading the oral territory in wild-type embryos.

In *Strongylocentrotus*, zygotic *brachyury* expression initiates at the blastopore just before the onset of gastrulation (18 hpf). Shortly before the archenteron breaks through (24 hpf), the future stomodaeum (oral ectoderm) of the late gastrula also starts to express *brachyury* (Fig. 1C). Therefore, Brachyury is expected to have different target genes in these two domains. For instance, different genes are co-expressed with *brachyury* in these two domains (e.g., *hox11/13b* is co-expressed with *brachyury* in the blastoporal region, while *gsc* is co-expressed with *brachyury* in the oral ectoderm (Fig. 5A and Supplemental Fig. 5B)). We validated several ChIP target genes that showed a change of expression in the differential gene expression analyses in *brachyury* morphants. Among the *brachyury* co-expressed genes of the presumptive endoderm, *foxA* [98] and *wnt16* [99] are downregulated, while *hox11/13b* [100] and *otx* [101] are upregulated in Brachyury knockdowns. Moreover, the mesodermal markers *ets1* and *ese* [102] are ectopically expressed in the presumptive endoderm, suggesting a repressive function of Brachyury on these genes in this domain of wild type embryos (Fig. 5A and Supplemental Fig. 6B). These results show that, like in vertebrates, *Nematostella* and *Strongylocentrotus* Brachyury can act both as direct activator and repressor on different target genes. By contrast, some target genes, which are also co-expressed with *brachyury* in the blastoporal region (e.g., *soxC, wnt8, eve)* [101,103,104] remain unaffected by the *brachyury* knockdown (Supplemental Fig. 5A), suggesting that other factors than *brachyury* may play a decisive role in their regulation.

Taken together, in both the sea urchin and the sea anemone Brachyury activates numerous members of the Wnt-beta catenin and the Wnt-PCP pathway in the blastoporal region of the gastrulating embryo, as well as *foxA*, which in sea urchin is expressed both in the blastopore and the future stomodeum. Since *brachyury* in turn is downstream of canonical Wnt signaling both in *Strongylocentrotus* and in *Nematostella* (as well as other cnidarians), we conclude that in cnidarians, sea urchins and vertebrates, there is a positive feedback loop of Brachyury and Wnt signaling at the blastopore and its derivative tissues.

Nematostella Brachyury represses neuronal genes in the oral domain

One of the surprising recent findings in vertebrates was that Brachyury has a dual role in the differentiation of "neuromesodermal progenitors": it activates mesodermal genes but directly

also represses neuronal genes, thereby antagonizing the function of Sox2 in the promotion of neural fate [21]. Surprisingly, in *Nematostella*, Brachyury also regulates many genes that are expressed at the aboral half, thus not overlapping with *brachyury* expression. During gastrulation, the aboral half is the domain of early neurogenesis [105]. Notably, many of these aborally expressed target genes are involved in neurogenesis, e.g. the *achaete-scute* homolog *ash-A, islet-1, tbx2/3, masterblind (mbnl), rfx4, noc* and *lhx-1*, which are expressed in single scattered cells, typical for early neurogenic factors [105–108]. Interestingly, in Brachyury morphants, the expression domain of many of these genes expands to the oral domain, suggesting that Brachyury might be directly inhibiting early neurogenesis in the oral domain of the blastopore (Fig. 4B, Supplemental Fig. 5). Of note, the putative neuronal markers that show no significant change of expression do also not show a single cell pattern, but rather a global gastrodermal expression pattern, suggesting that the function of Brachyury on the expression of neuronal genes is restricted to the ectoderm (compare Fig. 4 and Supplemental Fig. 5).

Sea urchin Brachyury is involved in proper ectodermal patterning and fate establishment of the different ectodermal domains

We also investigated the effect of *Strongylocentrotus* Brachyury on genes of the oral ectodermal territory, where *brachyury* is expressed during gastrulation. The oral ectoderm domain is adjacent to the anterior neuroectoderm (ANE) of the embryo that will later give rise to the apical organ of the pluteus larva. The expression of the oral ectoderm (future stomodaeum) gene *goosecoid* [109], which shows an overlapping expression with *brachyury* in this region in wild-type embryos, is severely reduced in Brachyury morphants (Supplemental Fig. 6B,D). Similarly, the anterior neuroectoderm marker Nkx2.1 is downregulated in Brachyury morphants as revealed both by differential RNA sequencing and immunohistochemistry (Supplemental Fig. 6C,D). Notably, several other genes that are also expressed in the ANE (e.g. *fgf9/16/20, fzd5/8, six3/6, soxB2*) [49,110,111] are downregulated in Brachyury morphants (Fig. 5B, Supplemental Fig. 6B,C,D), suggesting that Brachyury might have an indirect role in patterning the anterior neural ectoderm. Moreover, genes involved in the specification of the oral-aboral (ventral-dorsal) axis such as *nodal* and *bmp2/4* [112] that partially co-localize with *brachyury* seem to be severely affected, with the expression of *nodal* reduced and the expression of *bmp2/4* abolished (Fig. 5B). *Fgf9/16/20* that is expressed in the oral side [113] is also abolished (Fig. 5B). This suggests that the proper formation of the oral-aboral axis is compromised in Brachyury morphants. Notably, the

expression of the aboral ectoderm marker *spec1* [114] is also reduced in Brachyury morphants, confirming the general aberrant patterning of the ectoderm. Such a role of sea urchin Brachyury at the O/A organizer is in line with what recently suggested by single cell transcriptomics data [115].

To summarize, similar to vertebrates, in *Nematostella* Brachyury appears to contribute to the inhibition of neuronal differentiation in the blastoporal domain, thereby restricting it to the aboral part. In *Strongylocentrotus* neuronal markers (e.g., *hbn*, *nkx2.1*, *nkx3.2*; Supplemental Fig. 5C and D) appear to be downregulated upon Bra knockdown, however, mostly indirectly due to the fact that in the absence of Brachyury the anterior neuroectoderm is generally mis-patterned. This ectodermal mispatterning results not only in loss of distinct ectodermal territories (oral-aboral-ANE), but also in the inability of the ANE to promote neurogenesis and, thus, proper neuronal differentiation.

Phylogenetic comparisons reveal ancestral targets of Brachyury

In order to reconstruct the evolutionary changes of Brachyury function that led to its mesoderm determination role, conserved among chordates, we aimed to compare the direct Brachyury target genes in organisms with available information, ranging from protists to vertebrates. To this end, we first generated a reliable set of orthologs in the species under consideration using OMA (see Methods). We then compared cnidarian and sea urchin Brachyury targets identified by ChIP-seq with the published target gene sets from the urochordate *Ciona intestinalis* [116,117] and two vertebrates, the mouse *Mus musculus* [21] and the frog *Xenopus tropicalis* [55]. We added to this comparison the putative Brachyury targets from the protist *Capsaspora owczarzaki* determined by the search of Brachyury binding motifs in ATAC-seq peaks [82]. The OMA matrix allowed us to assign pair-wise homologs between these species and thereby reconstruct the ancestral target genes at each of the phylogenetic branching nodes (Figure 6; Supplemental Table 1). Since the mouse has the best-annotated genome, we used it as a reference for the functional annotations in the analyses. As in all investigated metazoans regulation of transcription, multicellular organism development and signal transduction are among the most dominant GO categories of the Brachyury targets, we first focused our analysis on transcription factors. Notably, *brachyury* is the only target gene that is shared among all organisms, suggesting that self-regulation is an ancestral feature and under strong selection (Supplemental Fig. 7).

There is ample evidence that transcription factor binding sites and hence the corresponding target genes can diverge rapidly [118]. However, important target genes might be under higher selection pressure and maintained over longer evolutionary distances. Thus, by comparison of lineages of various phylogenetic distances, we should be able to identify ancestral target genes of distinct phylogenetic groups. Therefore, in order to reconstruct how the Brachyury target gene set was conserved or was changed during the course of animal evolution, we sought to determine by pairwise comparisons, which target genes evolved early and have been maintained in several lineages and which genes were recruited only in more recent lineages. A gene was considered an ancestral target for a given ancestor node when a homolog is shared between the earliest branching organism (outgroup) and at least one of the other ingroup species (for details see Materials and Methods). To unravel shared ancestral and lineage-specific developmental regulators, we focused the analyses on target genes coding for transcription factors (Fig. 6), while the whole set of target genes is found in Supplemental Table 1. The pairwise analysis suggested that six genes were TF coding targets in the last common ancestor of *Capsaspora* and Metazoa (Fig. 6). Node II is of particular interest, as it reflects the common ancestor of the diploblastic sea anemone *Nematostella* with the deuterostomes. Notably, we found numerous homologous target TFs shared between *Nematostella* and at least one of the deuterostome species (*Strongylocentrotus, Ciona, Xenopus*, *Mus*) (Fig. 6).

To understand the functional consequences of shared or species-specific Brachyury target genes, we annotated the target genes encoding transcription factors by their differential expression in different germ layers and tissues during mouse gastrulation. As a reference, we used a recently published single cell transcriptome dataset from mouse embryos corresponding to the gastrula stage (stage E8.0 and 8.5, see Materials and Methods) [119]. When we calculated the log2 normalized endodermal versus mesodermal as well as neuronal versus mesodermal gene expression levels, we found that ancestral node II genes (*Nematostella* plus deuterostomes) tend to have significantly higher expression levels in endodermal and neuronal cell types compared to mesodermal cell types (Fig. 6). A similar trend of more endodermal targets compared to mesodermal genes was detected at nodes III (deuterostomes) and node IV (chordates), albeit less pronounced than at node II. Only at node V (genes exclusively shared among the two vertebrates), we detected more mesodermal genes than neuronal or endodermal genes, in line with the conserved function of Brachyury in mesoderm specification in vertebrates (Fig. 6). While we could confirm the enrichment of neuronal

expression for node II genes, and the enrichment for mesodermal expression for node V genes in a different mouse single cell gene expression dataset (stage E8.5, see Supplementary Figure 9) [120] we did not see an enrichment in endodermal gene expression for node 2- 4 genes. We hypothesize that this is due to the fact that in the dataset used in Figure 6 [119], cells expressing both Brachyury as well as Noto were annotated as endoderm instead of notochord. Thus, among the shared target genes coding for transcription factors, genes with a role in neuronal cell types are more ancestral than the mesodermal target genes, indicative of the shift of target genes of Brachyury in the chordate and vertebrate lineage.

Next, we were interested, whether there are some crucial TFs that would make the difference in the evolutionary change of Brachyury function or whether it is rather a step-wise and gradual shift. We found 16 TFs shared between *Strongylocentrotus* and the chordates, of which only few have a defined mesodermal function in vertebrates. When *Ciona* was compared with the vertebrates (node IV), we detected 21 TFs as shared Brachyury targets. Given *brachyury* is expressed in the notochord in all chordates, the number of shared target genes between *C. intestinalis* and vertebrates is relatively modest and many of them are not exclusively, if at all, expressed in the notochord, in line with previous studies [121]. Although *Ciona* shares with vertebrates the expression of Brachyury in the notochord, only five (*six1/2, tbx18, tbx6, cdx1, lmx1*) have assigned functions in mesoderm formation in vertebrates. One interesting target gene is *tbx6,* which has crucial functions downstream and in conjunction with Brachyury in the formation of paraxial mesoderm in vertebrates [122–125]. There is no *tbx6* homolog in Cnidaria, nor in any other non-bilaterian phylum. By contrast, there are putative homologs in the deuterostome *Strongylocentrotus* and representatives of the protostomes (e.g., the mollusc *Lottia gigantea*; (Supplemental Fig.8A), however, our data suggest that in the sea urchin *tbx6* does not seem to be a target of Brachyury. We conclude that *tbx6* likely evolved by a gene duplication event in the bilaterian ancestor (although our phylogenetic analysis (Supplemental Fig. 8A) did not provide strong support for a monophyletic group of vertebrate and invertebrate Tbx6). In *Ciona,* there are three *Ci-tbx6* paralogous genes, which show an expression in the developing paraxial muscles, i.e., complementary to *brachyury* [126]. This suggests that *tbx6* is likely negatively regulated by Brachyury in *C. intestinalis*. Thus, *tbx6* was recruited as a novel target only in the chordates.

Node V target genes (32 of which are TFs) are shared between *Xenopus* and mouse. Since homologs of these genes are not found as targets in the other organisms, they supposedly

have evolved newly or have been recruited in vertebrates. Unlike node II genes, we find node V genes to be more highly expressed in mesodermal cell types (Fig. 6 and Supplementary Fig. 9). In addition, several of these genes are known as crucial regulators involved in mesoderm development. Among others, this includes *mesp1/2, mesogenin, twist, mef2c* and *smad6*. Thus, these TFs are likely to be involved in conveying the role of Brachyury as a mesoderm determination factor during early gastrulation and a pioneering factor in neuromesodermal progenitor differentiation.

In summary, this phylogenetic analysis of the target genes shows that numerous ancestral target genes that predate the split of cnidarians and bilaterians have a neuronal function at least in vertebrates, whereas a large fraction of key mesodermal target genes was only acquired at the level of the vertebrates.

Discussion

In this study, we revealed the genome-wide targets of Brachyury in two invertebrates, the diploblast sea anemone *Nematostella vectensis* and a basally branching deuterostome, the sea urchin *Strongylocentrotus purpuratus* and compared them with known sets of target genes from similar ChIP experiments in other organisms.

Limitations of this study

We took particular care to use comparable stages of *Nematostella* and sea urchin embryos for our ChIP-seq experiments at early gastrulation, however, as with any such comparative analyses, we cannot fully rule out that both missing or detected target genes are due to specific differences in the developmental stage, tissue origin and the experimental design. Also, each ChIP-seq experiment has a specific sensitivity and therefore leads to more or less detected target genes, which is probably the main reason, why we detect fewer target genes in the sea urchin. Therefore, as a cautionary note, we cannot make statements and draw conclusions on individual target genes that might be missing in a given dataset. On the other hand, mouse and frog datasets share a large number of target genes, many of which have also been validated experimentally at other stages, suggesting that a large fraction of the conserved target genes can be captured.

Conservation of binding motifs of Brachyury across metazoans

Our genome-wide ChIP-seq experiments in the sea anemone *Nematostella* and the sea urchin *Strongylocentrotus* confirmed that the binding motif of Brachyury is deeply conserved

throughout the animal kingdom and likely even extending to the protist *C. owczarzaki.* This is consistent with the extreme conservation of the T-domain of Brachyury over >700 Million years and the induction of mesoderm formation by the ectopic expression of *Capsaspora* or cnidarian *brachyury* mRNA in the animal hemisphere of the frog embryo [43,44,82]. Notably, in all organisms we find both single binding sites as well as palindromes, which were first predicted by *in vitro* selex studies [80,81] . Palindromic motifs bind dimers of Brachyury, and the binding of dimers is expected to be stronger than the binding of monomers to single motifs. This may have a functional consequence on the effect on gene expression. Indeed, we detected significantly higher peaks with palindrome motifs in *Nematostella* (Supplemental Fig. 4A). However, we could not detect any significant correlation of the distribution of single and palindromic binding sites with respect to the distance to the TSS nor to the category of target genes, nor to the suppression or activation of the target gene (Supplemental Fig. 4B-D, data not shown). A large fraction of the detected Brachyury binding sites coincide with promoters and enhancers defined by specific combinations of chromatin modifications in *Nematostella* and by open chromatin (detected by ATAC-seq) in *Strongylocentrotus*, suggesting that Brachyury binds primarily to predicted cis-regulatory elements (Supplemental Fig. 2A-C). However, it is noteworthy that a certain fraction of the Brachyury peaks is also found in regions of closed chromatin. This raises the possibility that in *Nematostella* as well as in *Strongylocentrotus* Brachyury can act as a pioneer transcription factor as it has recently been described in vertebrates [87,127].

Interestingly, a large fraction (about 23% in *Nematostella* and 27% in *Strongylocentrotus*) of the Brachyury binding peaks do not harbor a Brachyury motif (Supplemental Fig. 4C,D). While some of these could be false positives, we assume that many are events of indirect DNA binding through another transcription factor. This suggests that Brachyury might act in concert with a variety of other transcription factors. Indeed, we do find motifs of Sox, Fox and homeodomain TFs enriched in the vicinity of the Brachyury motifs, suggesting either competitive or cooperative binding (Supplemental Fig. 4B). Interestingly, in *Nematostella, brachyury* is co-expressed with *soxB1, soxB2a* (also termed *sox1*)*, foxA,* and *foxB* and notably, all of them are also direct targets of Brachyury, suggesting that they are interconnected in a gene regulatory network (see below). Similarly, in *Strongylocentrotus,* homeodomain genes such as *hox11/13b, otx, six3/6,* are not only co-expressed, but also direct targets of Brachyury. Future work should focus on investigating the physical interactions of these transcription factors.

A conserved blastoporal feedback loop of Brachyury, FoxA and Wnt signaling

Outside metazoans, *brachyury* and several other T-box genes are found in the protist *Capsaspora owzcarzaki*, as well as in certain fungi, indicating that it evolved before the emergence of the metazoans (Fig. 7C) [83,128,129]. The function of *brachyury* in these organisms is not known but the T-domain of *Capsaspora* is sufficiently conserved to induce mesoderm when expressed in the frog embryo [83].

Among non-bilaterian animals, *brachyury* has also been detected in sponges [83,130], ctenophores and placozoans [131–134]. While the correspondence of the body plans of sponges and placozoans to cnidarian or bilaterian body plans appears less clear at first glance, there is evidence for a conservation of the main body axes and a co-expression with Wnt genes [135,136]. Furthermore, functional studies in the ctenophore *Mnemiopsis leydii* suggests that Brachyury is expressed at the blastopore and is required for the invagination and formation of the stomodeum [133].

Within cnidarians, expression of *brachyury* at the blastopore or its derivative tissue is widely conserved, e.g. in *Hydra* [38] *Hydractinia* [137], *Clytia* [138], and *Acropora* [42,139]. Similar to *Nematostella*, in all these cnidarian species, *brachyury* is coexpressed with members of the Wnt pathway. Functional studies in *Acropora* and *Nematostella* suggest a role of *brachyury* in pharynx formation, in line with our results [41,42].

The blastoporal expression is also conserved in a broad range of bilaterian species, for instance the polychaete *Platynereis dumerilii* [33], the mollusc *Patella vulgata* [36], the hemichordate *Ptychodera flava* [140,141] and the ecdysozoan phylum Priapulida [142]. However, in different species of brachiopods (Spiralia, Lophophorata), potentially divergent functions of Brachyury are observed with respect to the formation of the gut openings [143144145]. In *Drosophila melanogaster*, the *brachyury* ortholog *brachyenteron* is required for the ectodermal hindgut formation, where it is co-expressed with the *foxA* ortholog *forkhead* [146–149]. Similar roles have been reported in short germ insects, e.g. *Tribolium castaneum* [146] and the intermediate germ insect, the cricket *Gryllus bimaculatus* [150]. Thus, despite their divergent mechanisms of posterior elongation, these insects have retained a conserved role for *brachyury* in hindgut development and posterior visceral muscles. Taken together, these comparative data strongly support an ancestral blastoporal expression of *brachyury* and Wnt signaling [26,151], and where investigated, often in conjunction with *foxA* expression in the

developing gut. Functional studies in some of these species support a role of *brachyury* in axial patterning as well as fore/hindgut development.

In vertebrates, several Wnt pathway members like *dkk1, wnt5b* , *wnt8a, wnt9a, wnt11b* (PCP) are targets of Brachyury [21,55], and *brachyury* is also regulated by Wnt3 and by eFGF [152] in a feedback loop. It was recently shown that Wnt signaling acts upstream of *brachyury* in *Nematostella* [95,96,153]. In this study, our ChIP-seq data in *Nematostella* have further revealed that numerous *wnt* ligands, all four *frizzled* receptors*, dishevelled,* and □*-catenin* are direct targets of Brachyury, demonstrating the direct regulation of the canonical as well as the putative PCP pathway by Brachyury. Similarly, in *Strongylocentrotus*, *brachyury* is activated by the canonical Wnt pathway and, in turn, Brachyury targets several genes of the Wnt pathway, highlighting a deeply conserved feedback loop of Wnt and Brachyury (Fig. 7A) [27,99,154]. Since *brachyury* expression and Wnt signaling is co-localized at the blastopore lip in most bilaterian species, we postulate that *brachyury* forms an ancestral gene regulatory feedback loop at the blastopore. Furthermore, because Wnt signaling is known to play a crucial role in establishing the primary body axis in vertebrates, ambulacrarians (echinoderms and hemichordates) and cnidarians [155], we propose that one of the ancestral roles of Brachyury in metazoans was in axial patterning (Fig. 7B). In support of this idea, knockdown of *brachyury* abolishes the axis induction capacity of the blastopore lip in *Nematostella* [95]. While we do not have evidence for a control of *brachyury* by FGF signaling at present, we also find FGF ligands (*fgf8* in *Nematostella* and *fgf9/16/20* in *Strongylocentrotus*)*, as well as foxA* (and *foxB* in *Nematostella*) as deeply conserved target genes of Brachyury, suggesting that these genes might also belong to the axis patterning kernel (Fig. 7A,B).

The evolution of the mesoderm inducing role of Brachyury in chordates

Since Brachyury has a conserved function in mesoderm development in vertebrates and is required for notochord development in urochordates (and possibly in cephalochordates), we looked for conserved downstream mesodermal genes in *Nematostella* and *Strongylocentrotus*. Our analysis showed that only few genes with a role in mesoderm formation in vertebrates are also Brachyury target genes in *Nematostella* (e.g., *tbx2/3, tbx20, mbnl1, msx, gata1/2/3*; see Fig. 6; Supplemental Table 1; Supplemental Fig. 6). Instead, a surprisingly large number of mostly negatively regulated TFs that are involved in neurogenesis are shared between *Nematostella* and vertebrates, suggesting that the negative regulation of neuronal regulators is more ancestral. Surprisingly, in *Strongylocentrotus*, the few mesodermal Brachyury target

genes (e.g., *ets1, gcm, foxn2/3*, see Supplemental Table 1, Fig. 6A) are repressed by Brachyury in the presumptive endoderm, in stark contrast to what is observed in chordates. This difference highlights the plasticity of Brachyury in acquiring new functions and targeting new genes.

Some of the key factors driving mesoderm differentiation in chordates, such as *tbx6* and *myoD*, appear to have evolved only in the deuterostome or bilaterian lineage, respectively, since they do not exist in cnidarian genomes. In *Strongylocentrotus*, two *myoD* paralogous genes are present (*myoD1* and *myoD2*), both with distinct mesodermal functions in skeletogenesis and myogenesis [48], but there is no evidence that they are target genes of Brachyury. Quite the contrary, our functional studies suggest that, in the sea urchin, *brachyury* represses mesodermal fate (through repression of the mesoderm specification gene *ese*) and rather promotes endoderm formation.

Tbx6 is a key downstream gene in vertebrates, conveying the mesoderm determination function of Brachyury in feed-forward loops. It is a crucial factor in the development of the paraxial mesoderm, and mutation of *tbx6* leads to the formation of two supernumerary neural tubes instead of somites [122,124]. *Tbx6* is also a Brachyury target gene in the ascidian *C. intestinalis* [117], expressed in the future paraxial muscle, hence mutually exclusive to the *brachyury* expressing notochord and therefore likely to be negatively regulated by Brachyury. The evolution of chordate *tbx6* (arisen through duplication of a member of the *tbx2/3/4/5* families) was probably a crucial step in the recruitment of Brachyury to a mesodermal function.

Besides *tbx6*, another crucial evolutionary step was the acquisition of other key mesodermal downstream differentiation determinants in vertebrates, such as *vegT, mesp1/2, twist, mesogenin, mef2, myf5, myf6, myod* as direct target genes of Brachyury, linking early mesoderm formation to muscle differentiation. All these genes have been shown to play decisive roles in the specification of the mesoderm in vertebrates. While the Myogenic Regulatory Factors (MRFs) (*myf5, myf6, mrf4, myod*) are not present in the cnidarian genomes, *twist* exists, but is neither a direct target of Brachyury in *Nematostella* nor in *Strongylocentrotus,* nor does it seem to play a major role in axis formation or germ layer formation [156,157]. Thus, the evolution of several key mesodermal determinants as targets of Brachyury includes both adoption of conserved genes to mesoderm formation as well as the evolution of novel genes in the vertebrate lineage.

Brachyury and the evolution of neuromesodermal progenitors

The tailbud stage of vertebrates has been viewed as secondary axis formation [158] or even as a continuation of the gastrulation process itself [159]. The tailbud harbors a population of neuromesodermal progenitors (NMPs), which has the potential to differentiate into neural tube or into mesoderm, depending on the concentration of Wnt3a. NMPs express both *brachyury* and *sox2*, a marker for pluripotency and early neural fate [15,160,161]. Several studies support the view that *sox2* expression drives the cells towards neural fate, and a high level of *brachyury* expression promotes mesodermal fate [21,162,163]. Functional as well as ChIP-seq studies have shown that Brachyury and Sox2 mutually inhibit each other, which eventually leads to a segregation of neuronal and mesodermal cell fates [21]. Brachyury also represses a number of other neuronal target genes, in line with its role in suppressing neuronal fate [55]. Sox2 belongs to the subfamily of SoxB1, together with its vertebrate-specific paralogs Sox1 and Sox3. There is no one-to-one ortholog of *sox2* in cnidarians and sea urchin, but there is a cnidarian and sea urchin *soxB1* homolog [164]. Interestingly, while there is currently no evidence that *soxB1* is a target in *Strongylocentrotus* (although both *soxB1* and *soxB2* are differentially expressed in Bra knockdowns, see Supplemental Fig. 5B-C), at least in *Nematostella, soxB1* is a direct target of Brachyury (Fig. 3D). But unlike in vertebrate NMPs, there to date there is no evidence for mutual inhibition of SoxB1 and Brachyury in *Nematostella*. Furthermore, although *soxB1* is expressed in the domain of aboral neurogenesis, functional perturbation of soxB1 does not argue for a direct role of *soxB1* in promoting neural development in *Nematostella* (Supplemental Fig. 4C). However, Brachyury contributes to the repression of neuronal differentiation during embryogenesis. This effect might be in concert with the role of Wnt signaling. Since a significant number of these neuronal genes are also negatively regulated by Brachyury in the NMPs of vertebrates, we conclude that repression of neuronal differentiation is an (unexpected) ancestral role of Brachyury, regardless of its axial position. This is not to say that NMPs have their origin in the common ancestor of cnidarians and bilaterians, but that part of the NMP gene regulatory network and the role of Brachyury therein has a deeper origin.

Concluding remarks

Our comparative approach of genome-wide target genes of a key developmental regulator, Brachyury, has revealed both deeply conserved kernels as well as key acquisitions of novel target genes, that were either recruited or evolved newly by gene duplications. Brachyury has a surprisingly conserved expression pattern around the blastopore in most animal phyla. Since this blastoporal expression is part of the ancestral kernel, we postulate that such

feedback loops can evolutionarily stabilize a given expression pattern and its ancestral role. Nevertheless, the function of the gene can still drastically evolve, by the acquisition of only a few, but crucial, new target genes, which in the case of Brachyury, convey the new function in feed-forward loops. At present, the number of comparative studies of genome-wide target-gene screens of conserved transcription factors is very limited, but future approaches over a wide range of distantly related organisms may offer insights into the general mechanisms of robustness and evolvability of gene regulatory networks in the evolution of animal body plans.

Acknowledgements

We are grateful to Ira Blitz for critically reading the manuscript. We would like to thank the members of the Technau and the Arnone lab for discussions. This work was supported by the EU-ITN EVONET, to U.T. and M.I.A., and by grants of the Austrian Science Fund FWF to U.T. (P34404, P31018) and by EU-ITN EVOCELL (Grant no. 766053) to M.I.A.

Data Availability Statement

The raw files from the ChIP-seq and RNA-seq experiments have been deposited to the NCBI GEO database under the accession number GSE182573 and GSE198320 for *N. vectensis* and *S. purpuratus* respectively.
Additionally, all scripts developed for this study are available at GitHub at https://github.com/dnyansagar/gene_regulatory_network and https://github.com/xxxmichixxx/MouseEmbryoSingleCell as well as https://github.com/fmi-basel/gbuehler-MiniChip.

References

1. Smith, J. C., Price, B. M. J., Green, J. B. A., Weigel, D. & Herrmann, B. G. Expression of a xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67**, 79–87 (1991).
2. Cunliffe, V. & Smith, J. C. Ectopic mesoderm formation in Xenopus embryos caused by widespread expression of a Brachyury homologue. *Nature* **358**, 427–430 (1992).
3. Wilson, V., Manson, L., Skarnes, W. C. & Beddington, R. S. The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* **121**, 877–86 (1995).
4. O'Reilly, M. A. J., Smith, J. C. & Cunliffe, V. Patterning of the mesoderm in Xenopus: Dose-dependent and synergistic effects of Brachyury and Pintallavis. *Development* **121**, 1351–1359 (1995).
5. Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617–622 (1990).
6. Kispert, A., Ortner, H., Cooke, J. & Herrmann, B. G. The chick Brachyury gene: Developmental expression pattern and response to axial induction by localized activin. *Developmental Biology* **168**, 406–415 (1995).
7. Stemple, D. L. Structure and function of the notochord: An essential organ for chordate development. *Development* vol. 132 2503–2512 Preprint at https://doi.org/10.1242/dev.01812 (2005).

8. Wilkinson, D. G., Bhatt, S. & Herrmann, B. G. Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* **343**, 657–659 (1990).
9. Beddington, R. S., Rashbass, P. & Wilson, V. Brachyury--a gene affecting mouse gastrulation and early organogenesis. *Dev Suppl* **165**, 157–65 (1992).
10. Stott, D., Kispert, A. & Herrmann, B. G. Rescue of the tail defect of Brachyury mice. *Genes and Development* **7**, 197–203 (1993).
11. Conlon, F. L. & Smith, J. C. Interference with Brachyury function inhibits convergent extension, causes apoptosis, and reveals separate requirements in the FGF and activin signalling pathways. *Developmental Biology* **213**, 85–100 (1999).
12. Schulte-Merker, S., Van Eeden, F. J. M., Halpern, M. E., Kimmel, C. B. & Nüsslein-Volhard, C. no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* **120**, 1009–1015 (1994).
13. Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. & McMahon, A. P. T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes and Development* **13**, 3185–3190 (1999).
14. Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. & Nicolas, J. F. Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. *Developmental Cell* **17**, 365–376 (2009).
15. Wymeersch, F. J. *et al.* Position-dependent plasticity of distinct progenitor types in the primitive streak. *Elife* **5**, 1–28 (2016).
16. Gouti, M. *et al.* A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Developmental Cell* **41**, 243-261.e7 (2017).
17. Davis, R. L. & Kirschner, M. W. The fate of cells in the tailbud of Xenopus laevis. *Development* **127**, 255–267 (2000).
18. Gouti, M. *et al.* In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS Biology* **12**, (2014).
19. Henrique, D., Abranches, E., Verrier, L. & Storey, K. G. Neuromesodermal progenitors and the making of the spinal cord. *Development (Cambridge)* **142**, 2864–2875 (2015).
20. Selleck, M. A. & Stern, C. D. Fate mapping and cell lineage analysis of Hensen's node in the chick embryo. *Development* **112**, (1991).
21. Koch, F. *et al.* Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Developmental Cell* **42**, 514-526.e7 (2017).
22. Mugele, D. *et al.* Genetic approaches in mice demonstrate that neuro-mesodermal progenitors express T/Brachyury but not Sox2. *bioRxiv* 503854 (2018) doi:10.1101/503854.
23. Takahashi, H. *et al.* Brachyury downstream notochord differentiation in the ascidian embryo. *Genes Dev* **13**, 1519–23 (1999).
24. Holland, P. W. H., Koschorz, B., Holland, L. Z. & Herrmann, B. G. Conservation of Brachyury (T) genes in amphioxus and vertebrates: Developmental and evolutionary implications. *Development* **121**, 4283–4291 (1995).
25. Satoh, N. *et al.* A Preliminary Single-Cell RNA-Seq Analysis of Embryonic Cells That Express Brachyury in the Amphioxus, Branchiostoma japonicum. *Frontiers in Cell and Developmental Biology* **0**, 1731 (2021).
26. Technau, U. Brachyury, the blastopore and the evolution of the mesoderm. *BioEssays* **23**, 788–794 (2001).
27. Croce, J. *et al.* Wnt6 activates endoderm in the sea urchin gene regulatory network. *Development* **138**, 3297–3306 (2011).
28. Harada, Y., Yasuo, H. & Satoh, N. A sea urchin homologue of the chordate Brachyury (T) gene is expressed in the secondary mesenchyme founder cells. *Development* **121**, (1995).

29. Shoguchi, E., Satoh, N. & Maruyama, Y. K. Pattern of Brachyury gene expression in starfish embryos resembles that of hemichordate embryos but not of sea urchin embryos. *Mechanisms of Development* **82**, 185–189 (1999).
30. Croce, J., Lhomond, G. & Gache, C. Expression pattern of Brachyury in the embryo of the sea urchin Paracentrotus lividus. *Development Genes and Evolution* **211**, 617–619 (2001).
31. Annunziata, R. & Arnone, M. I. A dynamic regulatory network explains ParaHox gene control of gut patterning in the sea urchin. *Development* **141**, 2462–2472 (2014).
32. Gross, J. M. & McClay, D. R. The role of Brachyury (T) during gastrulation movements in the sea urchin Lytechinus variegatus. *Developmental Biology* **239**, 132–147 (2001).
33. Arendt, D., Technau, U. & Wittbrodt, J. Evolution of the bilaterian larval foregut. *Nature* **409**, 81–85 (2001).
34. Boyle, M. J., Yamaguchi, E. & Seaver, E. C. Molecular conservation of metazoan gut formation: Evidence from expression of endomesoderm genes in Capitella teleta (Annelida). *Evodevo* **5**, 1–19 (2014).
35. Kin, K., Kakoi, S. & Wada, H. A novel role for dpp in the shaping of bivalve shells revealed in a conserved molluscan developmental program. *Developmental Biology* **329**, 152–166 (2009).
36. Lartillot, N., Lespinet, O., Vervoort, M. & Adoutte, A. Expression pattern of Brachyury in the mollusc Patella vulgata suggests a conserved role in the establishment of the AP axis in Bilateria. *Development* **129**, 1411–1421 (2002).
37. Technau, U. & Scholz, C. B. Origin and evolution of endoderm and mesoderm. *International Journal of Developmental Biology* **47**, 531–539 (2003).
38. Technau, U. & Bode, H. R. HyBra1, a Brachyury homologue, acts during head formation in Hydra. *Development* **126**, 999–1010 (1999).
39. Technau, U. *et al.* Parameters of self-organization in Hydra aggregates. *Proc Natl Acad Sci U S A* **97**, 12127–12131 (2000).
40. Scholz, C. B. & Technau, U. The ancestral role of Brachyury : expression of NemBra1 in the basal cnidarian Nematostella vectensis (Anthozoa). *Development Genes and Evolution* 563–570 (2003) doi:10.1007/s00427-002-0272-x.
41. Servetnick, M. D. *et al.* Cas9-mediated excision of nematostella brachyury disrupts endoderm development, pharynx formation and oral-aboral patterning. *Development (Cambridge)* **144**, 2951–2960 (2017).
42. Yasuoka, Y., Shinzato, C. & Satoh, N. The Mesoderm-Forming Gene brachyury Regulates Ectoderm-Endoderm Demarcation in the Coral Acropora digitifera. *Current Biology* **26**, 2885–2892 (2016).
43. Marcellini, S., Technau, U., Smith, J. C. & Lemaire, P. Evolution of Brachyury proteins: Identification of a novel regulatory domain conserved within Bilateria. *Developmental Biology* **260**, 352–361 (2003).
44. Bielen, H. *et al.* Divergent functions of two ancient Hydra Brachyury paralogues suggest specific roles for their C-terminal domains in tissue fate induction. *Development* **134**, 4187–4197 (2007).
45. Fritzenwanker, J. H. & Technau, U. Induction of gametogenesis in the basal cnidarian Nematostella vectensis (Anthozoa). *Development Genes and Evolution* **212**, 99–103 (2002).
46. Genikhovich, G. & Technau, U. Induction of spawning in the starlet sea anemone nematostella vectensis, in vitro fertilization of gametes, and dejellying of zygotes. *Cold Spring Harbor Protocols* **4**, pdb.prot5281 (2009).
47. Leahy, P. S. Laboratory Culture of Strongylocentrotus purpuratus Adults, Embryos, and Larvae. *Methods in Cell Biology* **27**, 1–13 (1986).
48. Andrikou, C., Iovene, E., Rizzo, F., Oliveri, P. & Arnone, M. I. Myogenesis in the sea urchin embryo: The molecular fingerprint of the myoblast precursors. *Evodevo* **4**, (2013).

49. Anishchenko, E., Arnone, M. I. & D'Aniello, S. SoxB2 in sea urchin development: Implications in neurogenesis, ciliogenesis and skeletal patterning. *Evodevo* **9**, 1–10 (2018).
50. Perillo, M. *et al.* New Neuronal Subtypes With a "Pre-Pancreatic" Signature in the Sea Urchin Stongylocentrotus purpuratus. *Frontiers in Endocrinology* **9**, 650 (2018).
51. Schwaiger, M. *et al.* Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Research* **24**, 639–650 (2014).
52. Altenhoff, A. M., Gil, M., Gonnet, G. H. & Dessimoz, C. Inferring Hierarchical Orthologous Groups from Orthologous Gene Pairs. *PLoS ONE* **8**, e53786 (2013).
53. Altenhoff, A. M. *et al.* The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Research* **46**, D477–D485 (2018).
54. Altenhoff, A. M. *et al.* OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Research* **49**, D373–D379 (2021).
55. Gentsch, G. E. *et al.* InVivo T-Box Transcription Factor Profiling Reveals Joint Regulation of Embryonic Neuromesodermal Bipotency. *Cell Reports* **4**, 1185–1196 (2013).
56. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
57. Bardet, A. F. *et al.* Identification of transcription factor binding sites from ChIP-seq data at high resolution. *Bioinformatics* **29**, 2705–2713 (2013).
58. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
59. Khan, A. *et al.* JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research* **46**, D260–D266 (2018).
60. Renfer, E. & Technau, U. Meganuclease-assisted generation of stable transgenics in the sea anemone Nematostella vectensis. *Nature Protocols* **12**, 1844–1854 (2017).
61. Andrews S. FastQC: a quality control tool for high throughput sequence data. 11%0AHow does this look?%0A%0AAndrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc%0A%0A (2010).
62. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
63. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
64. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
65. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research* **41**, (2013).
66. Varet, H., Brillet-Guéguen, L., Coppée, J.-Y. & Dillies, M.-A. SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLOS ONE* **11**, e0157022 (2016).
67. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, (2014).
68. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**, 4288–4297 (2012).
69. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417–419 (2017).
70. Genikhovich, G. & Technau, U. In situ hybridization of starlet sea anemone (Nematostella vectensis) embryos, larvae, and polyps. *Cold Spring Harbor Protocols* **4**, pdb.prot5282 (2009).

71. Perillo, M., Paganos, P., Spurrell, M., Arnone, M. I. & Wessel, G. M. Methodology for Whole Mount and Fluorescent RNA In Situ Hybridization in Echinoderms: Single, Double, and Beyond. in *Methods in Molecular Biology* vol. 2219 195–216 (Humana Press Inc., 2021).
72. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
73. Minh, B. Q., Nguyen, M. A. T. & Von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30**, 1188–1195 (2013).
74. Roth, A. C. J., Gonnet, G. H. & Dessimoz, C. Algorithm of OMA for large-scale orthology inference. *BMC Bioinformatics* **9**, 518 (2008).
75. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
76. Carlsson, P. & Mahlapuu, M. Forkhead transcription factors: Key players in development and metabolism. *Developmental Biology* **250**, 1–23 (2002).
77. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science (1979)* **317**, 86–94 (2007).
78. Zimmermann, B. *et al.* Sea anemone genomes reveal ancestral metazoan chromosomal macrosynteny. *bioRxiv* (2021) doi:10.1101/2020.10.30.359448.
79. Grossman, S. R. *et al.* Positional specificity of different transcription factor classes within enhancers. *Proc Natl Acad Sci U S A* **115**, E7222–E7230 (2018).
80. Müller, C. W. & Herrmann, B. G. Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* **389**, 884–888 (1997).
81. Kispert, A. & Herrmann, B. G. The Brachyury gene encodes a novel DNA binding protein. *The EMBO Journal* **12**, 3211–3220 (1993).
82. Sebé-Pedrós, A. *et al.* The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. *Cell* **165**, 1224–1237 (2016).
83. Sebe-Pedros, A. *et al.* Early evolution of the T-box transcription factor family. *Proceedings of the National Academy of Sciences* **110**, 16050–16055 (2013).
84. Farnham, P. J. Insights from genomic profiling of transcription factors. *Nature Reviews Genetics* vol. 10 605–616 Preprint at https://doi.org/10.1038/nrg2636 (2009).
85. Scacheri, P. C. *et al.* Genome-wide analysis of menin binding provides insights into MEN1 tumorigenesis. *PLoS Genetics* **2**, 406–419 (2006).
86. Yang, A. *et al.* Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Molecular Cell* **24**, 593–602 (2006).
87. Tosic, J. *et al.* Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nature Cell Biology* **21**, 1518–1531 (2019).
88. Garriock, R. J. *et al.* Lineage tracing of neuromesodermal progenitors reveals novel wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development (Cambridge)* **142**, 1628–1638 (2015).
89. Goto, H., Kimmey, S. C., Row, R. H., Matus, D. Q. & Martin, B. L. FGF and canonical Wnt signaling cooperate to induce paraxial mesoderm from tailbud neuromesodermal progenitors through regulation of a two-step epithelial to mesenchymal transition. *Development (Cambridge)* **144**, 1412–1421 (2017).
90. Martin, B. L. & Kimelman, D. Brachyury establishes the embryonic mesodermal progenitor niche. *Genes and Development* **24**, 2778–2783 (2010).
91. Davidson, E. H. *et al.* A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Developmental Biology* **246**, 162–190 (2002).
92. Gross, J. M. & McClay, D. R. The role of Brachyury (T) during gastrulation movements in the sea urchin Lytechinus variegatus. *Developmental Biology* **239**, 132–147 (2001).

93. Kusserow, A. *et al.* Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433**, 156–160 (2005).
94. Logan, C. Y., Miller, J. R., Ferkowicz, M. J. & McClay, D. R. Nuclear beta-catenin is required to specify vegetal cell fates in the sea urchin embryo. *Development* **126**, (1999).
95. Lebedeva, T. *et al.* Cnidarian-bilaterian comparison reveals the ancestral regulatory logic of the β-catenin dependent axial patterning. *Nature Communications* **12**, 4032 (2021).
96. Röttinger, E., Dahlin, P. & Martindale, M. Q. A Framework for the Establishment of a Cnidarian Gene Regulatory Network for "Endomesoderm" Specification: The Inputs of ß-Catenin/TCF Signaling. *PLoS Genetics* **8**, (2012).
97. Kirillova, A. *et al.* Germ-layer commitment and axis formation in sea anemone embryonic cell aggregates. *Proc Natl Acad Sci U S A* **115**, 1813–1818 (2018).
98. Tu, Q., Brown, C. T., Davidson, E. H. & Oliveri, P. Sea urchin Forkhead gene family: Phylogeny and embryonic expression. *Developmental Biology* **300**, 49–62 (2006).
99. Cui, M., Siriwon, N., Li, E., Davidson, E. H. & Peter, I. S. Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo. *Proc Natl Acad Sci U S A* **111**, E5029–E5038 (2014).
100. Arenas-Mena, C., Cameron, R. A. & Davidson, E. H. Hindgut specification and cell-adhesion functions of Sphox11/13b in the endoderm of the sea urchin embryo. *Development Growth and Differentiation* **48**, 463–472 (2006).
101. Peter, I. S. & Davidson, E. H. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Developmental Biology* **340**, 188–199 (2010).
102. Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A. & Arnone, M. I. Identification and developmental expression of the ets gene family in the sea urchin (Strongylocentrotus purpuratus). *Developmental Biology* **300**, 35–48 (2006).
103. McClay, D. R., Miranda, E. & Feinberg, S. L. Neurogenesis in the sea urchin embryo is initiated uniquely in three domains. *Development (Cambridge)* **145**, (2018).
104. Wikramanayake, A. H. *et al.* Nuclear β-catenin-dependent Wnt8 signaling in vegetal cells of the early sea urchin embryo regulates gastrulation and differentiation of endoderm and mesodermal cell lineages. *Genesis* **39**, 194–205 (2004).
105. Nakanishi, N., Renfer, E., Technau, U. & Rentzsch, F. Nervous systems of the sea anemone Nematostella vectensis are generated by ectoderm and endoderm and shaped by distinct mechanisms. *Development* **139**, 347–357 (2012).
106. Layden, M. J. & Martindale, M. Q. Non-canonical Notch signaling represents an ancestral mechanism to regulate neural differentiation. *Evodevo* **5**, 1–14 (2014).
107. Layden, M. J., Boekhout, M. & Martindale, M. Q. Nematostella vectensis achaete-scute homolog NvashA regulates embryonic ectodermal neurogenesis and represents an ancient component of the metazoan neural specification pathway. *Development* **139**, 1013–1022 (2012).
108. Rentzsch, F., Layden, M. & Manuel, M. The cellular and molecular basis of cnidarian neurogenesis. *Wiley Interdisciplinary Reviews: Developmental Biology* vol. 6 Preprint at https://doi.org/10.1002/wdev.257 (2017).
109. Angerer, L. M. *et al.* Sea urchin goosecoid function links fate specification along the animal-vegetal and oral-aboral embryonic axes. *Development* **128**, 4393–4404 (2001).
110. Wei, Z., Yaguchi, J., Yaguchi, S., Angerer, R. C. & Angerer, L. M. The sea urchin animal pole domain is a Six3-dependent neurogenic patterning center. *Development* **136**, 1179–1189 (2009).
111. Range, R. C. Canonical and non-canonical Wnt signaling pathways define the expression domains of Frizzled 5/8 and Frizzled 1/2/7 along the early anterior-posterior axis in sea urchin embryos. *Developmental Biology* **444**, 83–92 (2018).

112. Duboc, V. *et al.* Nodal and BMP2/4 pattern the mesoderm and endoderm during development of the sea urchin embryo. *Development* **137**, 223–235 (2010).
113. Röttinger, E. *et al.* FGF signals guide migration of mesenchymal cells, control skeletal morphogenesis and regulate gastrulation during sea urchin development. *Development* **135**, 353–365 (2008).
114. Otim, O., Amore, G., Minokawa, T., McClay, D. R. & Davidson, E. H. SpHnf6, a transcription factor that executes multiple functions in sea urchin embryogenesis. *Developmental Biology* **273**, 226–243 (2004).
115. Satoh, N. *et al.* A single-cell RNA-seq analysis of Brachyury-expressing cell clusters suggests a morphogenesis-associated signal center of oral ectoderm in sea urchin embryos. *Developmental Biology* **483**, 128–142 (2022).
116. Katikala, L. *et al.* Functional Brachyury Binding Sites Establish a Temporal Read-out of Gene Expression in the Ciona Notochord. *PLoS Biology* **11**, (2013).
117. Kubo, A. *et al.* Genomic cis-regulatory networks in the early Ciona intestinalis embryo. *Development* **137**, 1613–1623 (2010).
118. Dermitzakis, E. T. & Clark, A. G. Evolution of transcription factor binding sites in mammalian gene regulatory regions: Conservation and turnover. *Molecular Biology and Evolution* **19**, 1114–1121 (2002).
119. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature 2019 566:7745* **566**, 490–495 (2019).
120. Grosswendt, S. *et al.* Epigenetic regulator function through mouse gastrulation. *Nature* **584**, 102–108 (2020).
121. Reeves, W. M., Wu, Y., Harder, M. J. & Veeman, M. T. Functional and evolutionary insights from the Ciona notochord transcriptome. *Development (Cambridge)* **144**, 3375–3387 (2017).
122. Chapman, D. L. & Papaioannou, V. E. Three neural tubes in mouse embryos with mutations in T-box gene Tbx6. *Nature* **391**, 695–697 (1998).
123. Griffin, K. J., Amacher, S. L., Kimmel, C. B. & Kimelman, D. Molecular identification of spadetail: regulation of zebrafish trunk and tail mesoderm formation by T-box genes. *Development* **125**, 3379–3388 (1998).
124. Chapman, D. L., Cooper-Morgan, A., Harrelson, Z. & Papaioannou, V. E. Critical role for Tbx6 in mesoderm specification in the mouse embryo. *Mechanisms of Development* **120**, 837–847 (2003).
125. Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. & Papaioannou, V. E. Tbx6, a mouse T-box gene implicated in paraxial mesoderm formation at gastrulation. *Developmental Biology* **180**, 534–542 (1996).
126. Takatori, N. *et al.* T-box genes in the ascidian Ciona intestinalis: Characterization of cDNAs and spatial expression. *Developmental Dynamics* **230**, 743–753 (2004).
127. Beisaw, A. *et al.* BRACHYURY directs histone acetylation to target loci during mesoderm development. *EMBO Rep* **19**, 118–134 (2018).
128. Sebé-Pedrós, A. & Ruiz-Trillo, I. Evolution and Classification of the T-Box Transcription Factor Family. *Current Topics in Developmental Biology* **122**, 1–26 (2017).
129. Sebe-Pedros, A., de Mendoza, A., Lang, B. F., Degnan, B. M. & Ruiz-Trillo, I. Unexpected Repertoire of Metazoan Transcription Factors in the Unicellular Holozoan Capsaspora owczarzaki. *Molecular Biology and Evolution* **28**, 1241–1254 (2011).
130. Leininger, S. *et al.* Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. *Nature Communications* **5**, 1–15 (2014).
131. Martinelli, C. & Spring, J. Distinct expression patterns of the two T-box homologues Brachyury and Tbx2/3 in the placozoan Trichoplax adhaerens. *Development Genes and Evolution* **213**, 492–499 (2003).

132. Martinelli, C. & Spring, J. T-box and homeobox genes from the ctenophore Pleurobrachia pileus: Comparison of Brachyury, Tbx2/3 and Tlx in basal metazoans and bilaterians. *FEBS Letters* **579**, 5024–5028 (2005).
133. Yamada, A., Martindale, M. Q., Fukui, A. & Tochinai, S. Highly conserved functions of the Brachyury gene on morphogenetic movements: Insight from the early-diverging phylum Ctenophora. *Developmental Biology* **339**, 212–222 (2010).
134. Yamada, A., Pang, K., Martindale, M. Q. & Tochinai, S. Surprisingly complex T-box gene complement in diploblastic metazoans. *Evolution and Development* **9**, 220–230 (2007).
135. Adamska, M. *et al.* Wnt and TGF-β expression in the sponge Amphimedon queenslandica and the origin of metazoan embryonic patterning. *PLoS ONE* **2**, e1031 (2007).
136. Dubuc, T. Q., Ryan, J. F. & Martindale, M. Q. "Dorsal-Ventral" Genes Are Part of an Ancient Axial Patterning System: Evidence from Trichoplax adhaerens (Placozoa). *Molecular Biology and Evolution* **36**, 966–973 (2019).
137. Duffy, D. J., Plickert, G., Kuenzel, T., Tilmann, W. & Frank, U. Wnt signaling promotes oral but suppresses aboral structures in Hydractinia metamorphosis and regeneration. *Development* **137**, 3057–3066 (2010).
138. Momose, T., Derelle, R. & Houliston, E. A maternally localised Wnt ligand required for axial patterning in the cnidarian Clytia hemisphaerica. *Development* **135**, 2105–2113 (2008).
139. Hayward, D. C., Grasso, L. C., Saint, R., Miller, D. J. & Ball, E. E. The organizer in evolution-gastrulation and organizer gene expression highlight the importance of Brachyury during development of the coral, Acropora millepora. *Developmental Biology* **399**, 337–347 (2015).
140. Peterson, K. J., Cameron, R. A., Tagawa, K., Satoh, N. & Davidson, E. H. A comparative molecular approach to mesodermal patterning in basal deuterostomes: the expression pattern of Brachyury in the enteropneust hemichordate Ptychodera flava. *Development* **126**, (1999).
141. Tagawa, K., Humphreys, T. & Satoh, N. Novel pattern of Brachyury gene expression in hemichordate embryos. *Mechanisms of Development* **75**, 139–143 (1998).
142. Martín-Durán, J. M., Janssen, R., Wennberg, S., Budd, G. E. & Hejnol, A. Deuterostomic development in the protostome Priapulus caudatus. *Current Biology* **22**, 2161–2166 (2012).
143. Martín-Durán, J. M., Passamaneck, Y. J., Martindale, M. Q. & Hejnol, A. The developmental basis for the recurrent evolution of deuterostomy and protostomy. *Nature Ecology & Evolution* **1**, 1–10 (2017).
144. Luo, Y. J. *et al.* Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. *Nature Ecology and Evolution* **2**, 141–151 (2018).
145. Andrikou, C., Passamaneck, Y. J., Lowe, C. J., Martindale, M. Q. & Hejnol, A. Molecular patterning during the development of Phoronopsis harmeri reveals similarities to rhynchonelliform brachiopods. *Evodevo* **10**, 1–15 (2019).
146. Berns, N., Kusch, T., Schröder, R. & Reuter, R. Expression, function and regulation of Brachyenteron in the short germband insect Tribolium castaneum. *Development Genes and Evolution* **218**, 169–179 (2008).
147. Kusch, T. & Reuter, R. Functions for Drosophila brachyenteron and forkhead in mesoderm specification and cell signalling. *Development* **126**, 3991–4003 (1999).
148. Reim, I., Frasch, M. & Schaub, C. T-Box Genes in Drosophila Mesoderm Development. in *Current Topics in Developmental Biology* vol. 122 161–193 (Academic Press Inc., 2017).
149. Singer, J. B., Harbecke, R., Kusch, T., Reuter, R. & Lengyel, J. A. Drosophila brachyenteron regulates gene activity and morphogenesis in the gut. *Development* **122**, 3707–3718 (1996).
150. Shinmyo, Y. *et al.* Brachyenteron is necessary for morphogenesis of theposterior gut but not for anteroposterior axial elongationfrom the posterior growth zone in the intermediate-germbandcricket Gryllus bimaculatus. *Development* **133**, 4539–4547 (2006).

151. Swalla, B. J. Building divergent body plans with similar genetic pathways. *Heredity* vol. 97 235–243 Preprint at https://doi.org/10.1038/sj.hdy.6800872 (2006).
152. Arnold, S. J. *et al.* Brachyury is a target gene of the Wnt/β-catenin signaling pathway. *Mechanisms of Development* **91**, 249–258 (2000).
153. Pukhlyakova, E., Aman, A. J., Elsayad, K. & Technau, U. β-Catenin-dependent mechanotransduction dates back to the common ancestor of Cnidaria and Bilateria. *Proc Natl Acad Sci U S A* **115**, 6231–6236 (2018).
154. Sethi, A. J., Wikramanayake, R. M., Angerer, R. C., Range, R. C. & Angerer, L. M. Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos. *Science (1979)* **335**, 590–593 (2012).
155. Niehrs, C. On growth and form: A Cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. *Development* vol. 137 845–857 Preprint at https://doi.org/10.1242/dev.039651 (2010).
156. Genikhovich, G. & Technau, U. Complex functions of Mef2 splice variants in the differentiation of endoderm and of a neuronal cell type in a sea anemone. *Development* **138**, 4911–4919 (2011).
157. Martindale, M. Q., Pang, K. & Finnerty, J. R. Investigating the origins of triplosblasty: "Mesodermal" gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). *Development* **131**, 2463–2474 (2004).
158. Griffith, C. M., Wiley, M. J. & Sanders, E. J. The vertebrate tail bud: three germ layers from one tissue. *Anatomy and Embryology 1992 185:2* **185**, 101–113 (1992).
159. Gont, L. K., Steinbeisser, H., Blumberg, B. & De Robertis, E. M. Tail formation as a continuation of gastrulation: the multiple cell populations of the Xenopus tailbud derive from the late blastopore lip. *Development* **119**, 991–1004 (1993).
160. Bergsland, M. *et al.* Sequentially acting Sox transcription factors in neural lineage development. *Genes and Development* **25**, 2453–2464 (2011).
161. Wilson, V., Olivera-Martinez, I. & Storey, K. G. Stem cells, signals and vertebrate body axis extension. *Development* vol. 136 1591–1604 Preprint at https://doi.org/10.1242/dev.021246 (2009).
162. Bouldin, C. M. *et al.* Wnt signaling and tbx16 form a bistable switch to commit bipotential progenitors to mesoderm. *Development* **142**, 2499–2507 (2015).
163. Takemoto, T. *et al.* Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394–398 (2011).
164. Magie, C. R., Pang, K. & Martindale, M. Q. Genomic inventory and expression of Sox and Fox genes in the cnidarian Nematostella vectensis. *Development Genes and Evolution* **215**, 618–630 (2005).

Figures:

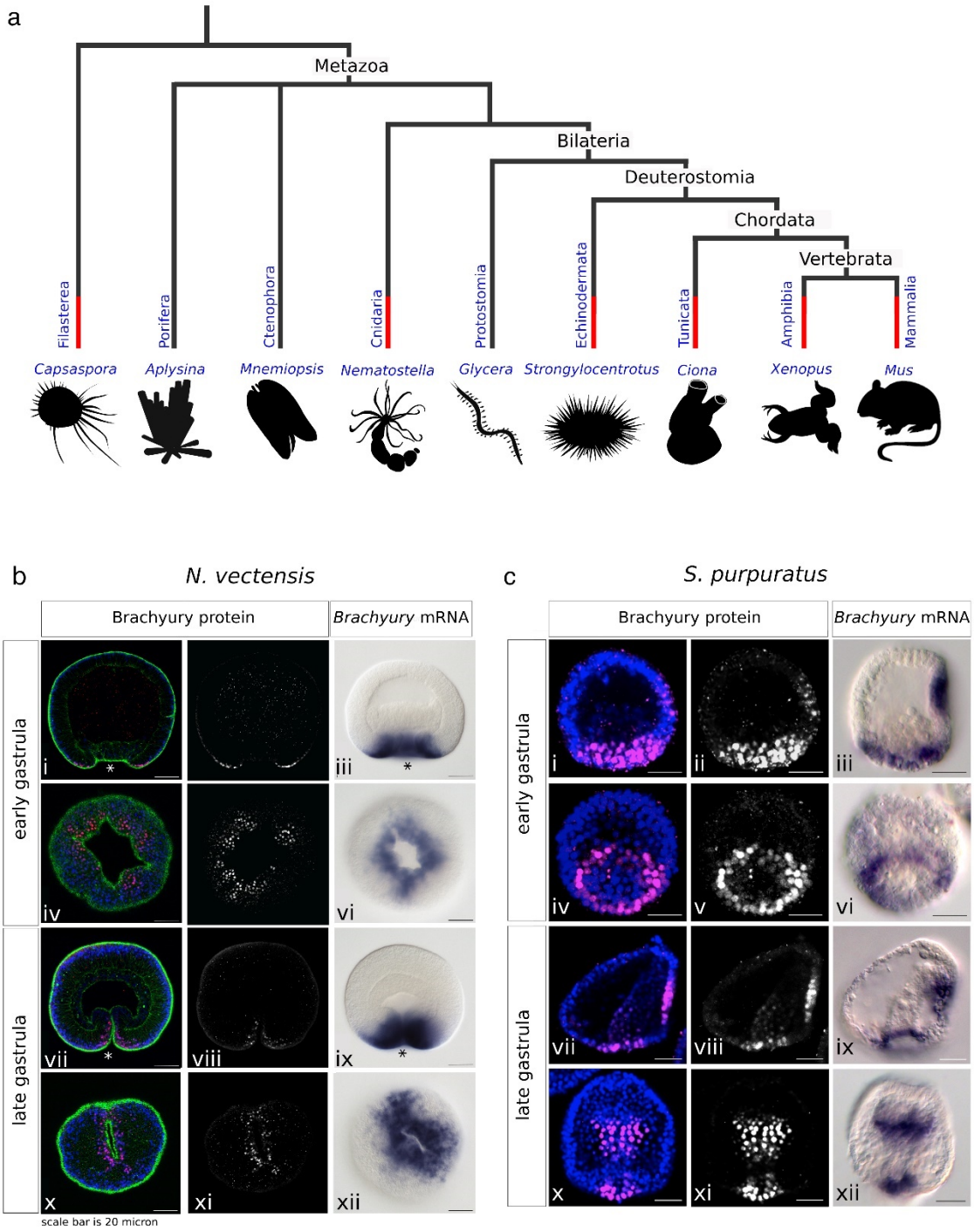

a
Metazoa
Bilateria
Deuterostomia
Chordata
Vertebrata
Filasterea
Porifera
Ctenophora
Cnidaria
Protostomia
Echinodermata
Tunicata
Amphibia
Mammalia
Capsaspora
Aplysina
Mnemiopsis
Nematostella
Glycera
Strongylocentrotus
Ciona
Xenopus
Mus
b
N. vectensis
Brachyury protein
Brachyury mRNA
early gastrula
late gastrula
i
ii
iii
iv
v
vi
vii
viii
ix
x
xi
xii
scale bar is 20 micron
c
S. purpuratus
Brachyury protein
Brachyury mRNA
early gastrula
late gastrula

Figure 1: Lineages sampled in this study and Brachyury protein and mRNA expression. (**A**) Simplified phylogenetic tree depicting the major animal lineages. Phyla analysed in this study are highlighted in red. Genome-scale Brachyury target gene screens from *Capsaspora owczarzaki, Nematostella vectensis, Strongylocentrotus purpuratus, Ciona intestinalis, Xenopus tropicalis,* and *Mus musculus*, representing Filasterea, Cnidaria, Echinodermata, Tunicata, Vertebrata (amphibians and mammals) respectively were analyzed. (**B**) Anti-Brachyury antibody stainings and Whole mount in situ hybridization (WMISH) of *Nematostella* embryos at the early and late gastrula stage. Left column: merged image of Brachyury protein expression (red), acetylated tubulin staining (green) and nuclei (blue). Middle column: Brachyury protein expression only. Right column: WMISH of *brachyury* (dark blue) of corresponding stages. Asterisk indicates blastopore. Rows of Bi and Bvii show lateral views, Biv and Bx show oral views (**C**) Immunofluorescence of *Strongylocentrotus* anti-Brachyury antibody in 24 hpf blastula (Ci, Civ) and 48 hpf gastrula (Cvii-Cx) stage. Left column: Merged image of Brachyury protein expression (red), acetylated tubulin staining (green) and nuclei (blue) in the 24 hpf blastula and 48 hpf gastrula stage. Middle column: Brachyury protein only. Right column: WMISH of *brachyury* at corresponding stages. Orientation: Animal is on top. Rows of Ci and Cvii are lateral views (oral to the right), Civ and Cx are oral views.

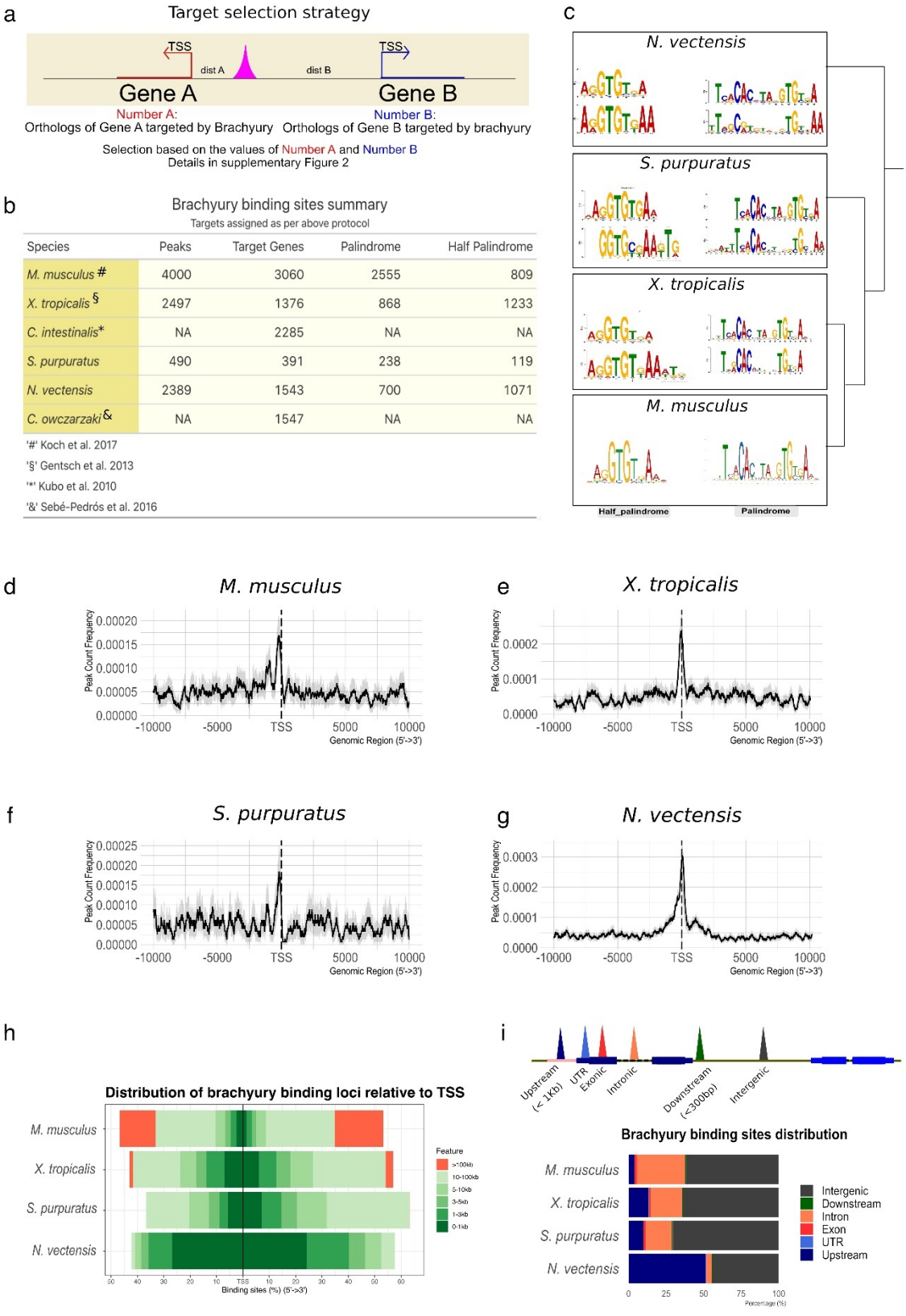


Figure 2: Distribution of Brachyury binding sites and motif analyses. (**A**) Summary of target selection strategy for Brachyury binding. To correctly identify the target of the binding

site, we make use of the ortholog information. From the two closest genes to the binding site, we select the gene that has more orthologs, that are also targets of Brachyury in the other species. The detailed algorithm for the target selection is shown in Supplementary Figure 2D. (**B**) Summary table showing the number of ChIP-seq peaks and their target genes, identified using above mentioned strategy. The table also lists the presence of centrally enriched (Centrimo/ MEME-ChIP) palindrome or half palindrome motifs within the peak region. (**C**) Brachyury motifs (Palindromic and/or half palindromic) are highly enriched in ChIP-seq peaks along with other transcription factor motifs. In each comparison the lower motif is the one found in our respective datasets while a vertebrate motif from the JASPAR database is shown above. For mouse the enriched motifs are the same as the one found in the JASPAR database. (**D-G**) Brachyury binding profile around the TSS (Transcription Start Site) with 95% confidence interval calculated with 1000 bootstrap. x-axis: distance to TSS, y-axis: density of peaks overlapping each bp (as in distance to/from TSS) on the x-axis. The plot was generated using the plotAvgProf function from the ChipSeeker R package. (**H**) Genomic distribution of Brachyury peaks with respect to the TSS over the whole genome. All genomic regions around TSSs were binned into 6 bins depending on their distance to the TSS, and split into upstream (left side of the plot) and downstream (right side of the plot) of the TSS. The percentage of peaks overlapping each bin is plotted on the x-axis. This plot was generated using the plotAnnoBar function of the ChipSeeker R package. (**I**) Distribution of the Brachyury binding sites with respect to genomic features such as promoters (defined here as 1000 bp upstream and 200 bp downstream of the TSS), introns, exons, UTRs and Downstream (Defined here as 300 bp downstream of a gene) and intergenic regions. The percentage of Brachyury peaks (x-axis) overlapping each region as shown in the diagram above is shown for each species (y-axis).

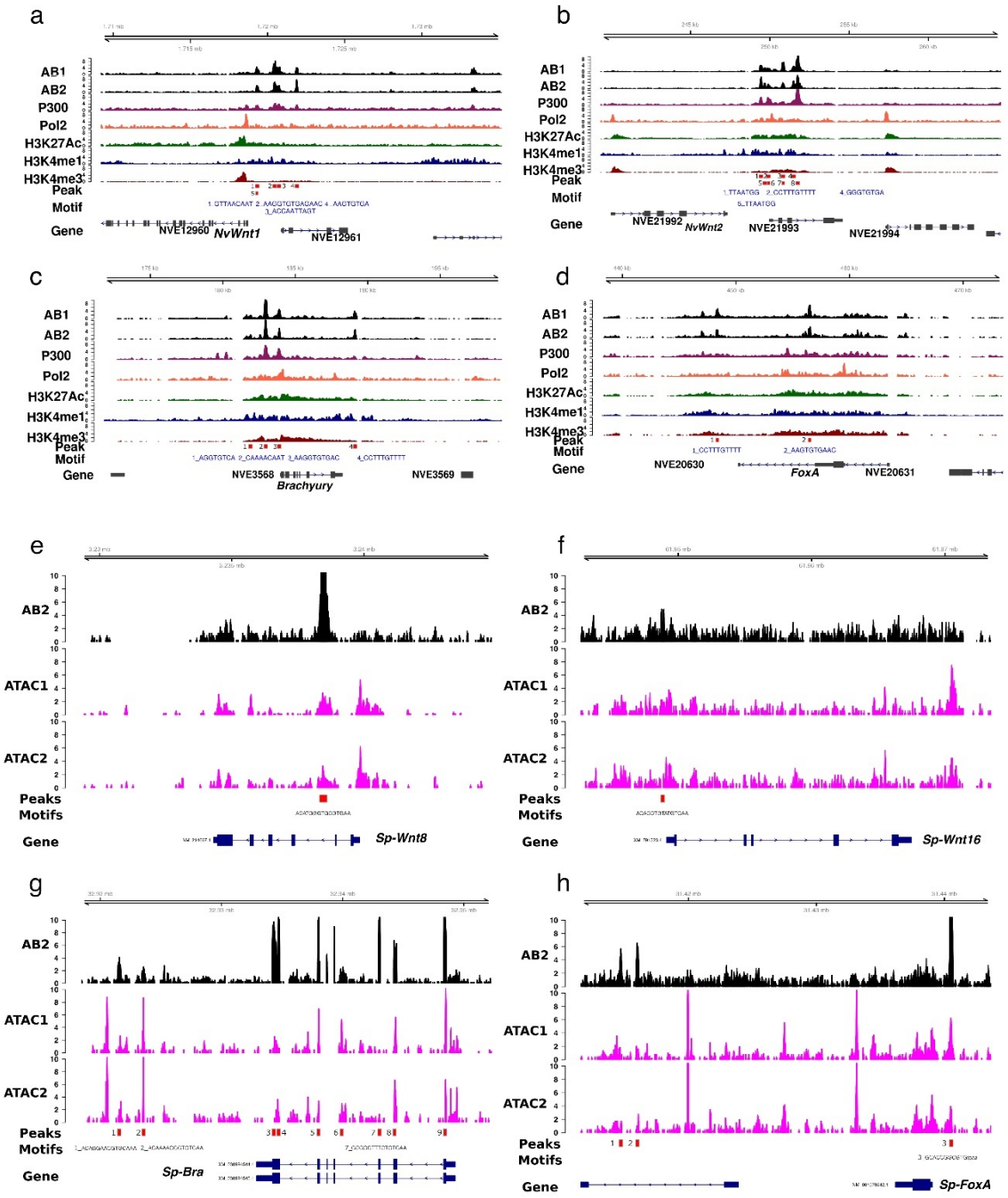


Figure 3: Brachyury binding sites at selected target genes in *Nematostella and Strongylocentrotus*. Genome tracks of Brachyury ChIP reads (AB1 and AB2) in relation to

histone modification or ATACseq reads (x-xis shows the position on the chromosome, y-axis for each track shows the coverage with ChIP reads). (**A-D**) Binding sites (Bra_AB1 and Bra_AB2) in *Nematostella* frequently overlap with previously identified gene regulatory elements in [51] (Histone acetyltransferase p300, RNA polymerase II (Pol2), Histone H3 lysine 27 acetylation (H3K27ac), Histone H3 lysine 4 methylation (H3K4me1 and H3K4me3)). For the details of the overlap, see Supplementary Figure 2A,B. (**E-H**) Most of the binding sites (Bra_AB1 and Bra_AB2) in *Strongylocentrotus* overlap with open chromatin sites identified by ATAC-seq (unpublished). For the details of the overlap, see Supplementary Figure 2C.

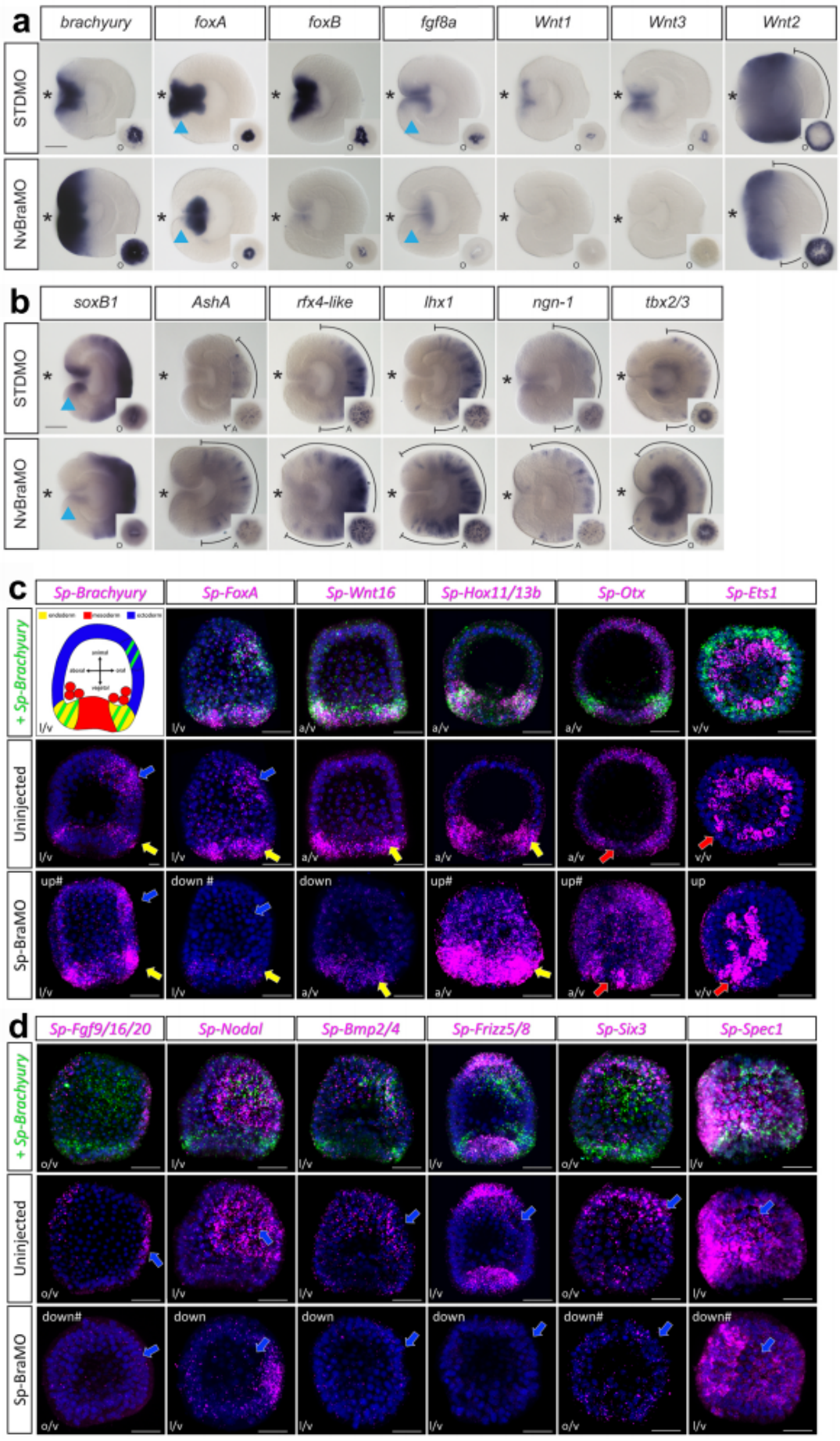

a
brachyury
foxA
foxB
fgf8a
Wnt1
Wnt3
Wnt2
STDMO
NvBraMO
b
soxB1
AshA
rfx4-like
lhx1
ngn-1
tbx2/3
STDMO
NvBraMO
c
Sp-Brachyury
Sp-FoxA
Sp-Wnt16
Sp-Hox11/13b
Sp-Otx
Sp-Ets1
+ Sp-Brachyury
Uninjected
Sp-BraMO
up#
down #
down
up#
up#
up
d
Sp-Fgf9/16/20
Sp-Nodal
Sp-Bmp2/4
Sp-Frizz5/8
Sp-Six3
Sp-Spec1
+ Sp-Brachyury
Uninjected
Sp-BraMO
down#
down
down
down
down#
down#

Figure 4: Validation of Brachyury target genes in *Nematostella* by WMISH after morpholino induced knockdown of Brachyury and in *Strongylocentrotus* by double FISH

Relative spatial expression of (**a**) oral markers *brachyury, foxA, foxB, fgf8a, wnt1, wnt3, and wnt2* in control Morpholino (upper row) and Bra MO injected embryos. Note that except for *brachyury* itself, oral markers are abolished or strongly diminished. Wnt2, a midbody marker, invades the oral territory of brachyury in Bra MO injected embryos. Blue arrowhead points to domain of *foxA* expression overlapping with *bra* expression. (**b**) aboral markers soxB1, *ash1, rfx4, lhx, ngn, tbx2/3*). Aboral neuronal marker gene expression expands orally in Bra knockdown embryos. The asterisks indicate the oral side of the embryo. Quantitative data are found in Supplemental Figure 1D. All embryos shown in lateral view. Scale bar corresponds to 50 μm. (**c**) Schematic representation (Left top corner) of Brachyury expression and regulatory input in the endoderm (yellow), ectoderm (red) and ectoderm (blue). WMISH of *wnt16, foxA, otx, hox11/13b, ets1* (magenta) relative to *Brachyury* expression (green). White arrowheads indicate co-expression. (**d**) WMISH of *fgf9/16/20, nodal, bmp2/4, frizzled, six3, spec1* in Brachyury morphants compared to wild type 24h embryos. Pictures are projections of confocal stacks. Nuclei are labeled blue with DAPI. # indicates that the ChIP target is also present in the differential RNAseq dataset. Arrows show the embryonic domain in which we see an effect (red arrow: mesodermally derived; yellow: endodermally derived; blue ectodermally derived). l/v: lateral view; v/v vegetal view; o/v oral view; a/v aboral view. up: upregulated gene; down; down-regulated gene. The scale bar is 20 μm.

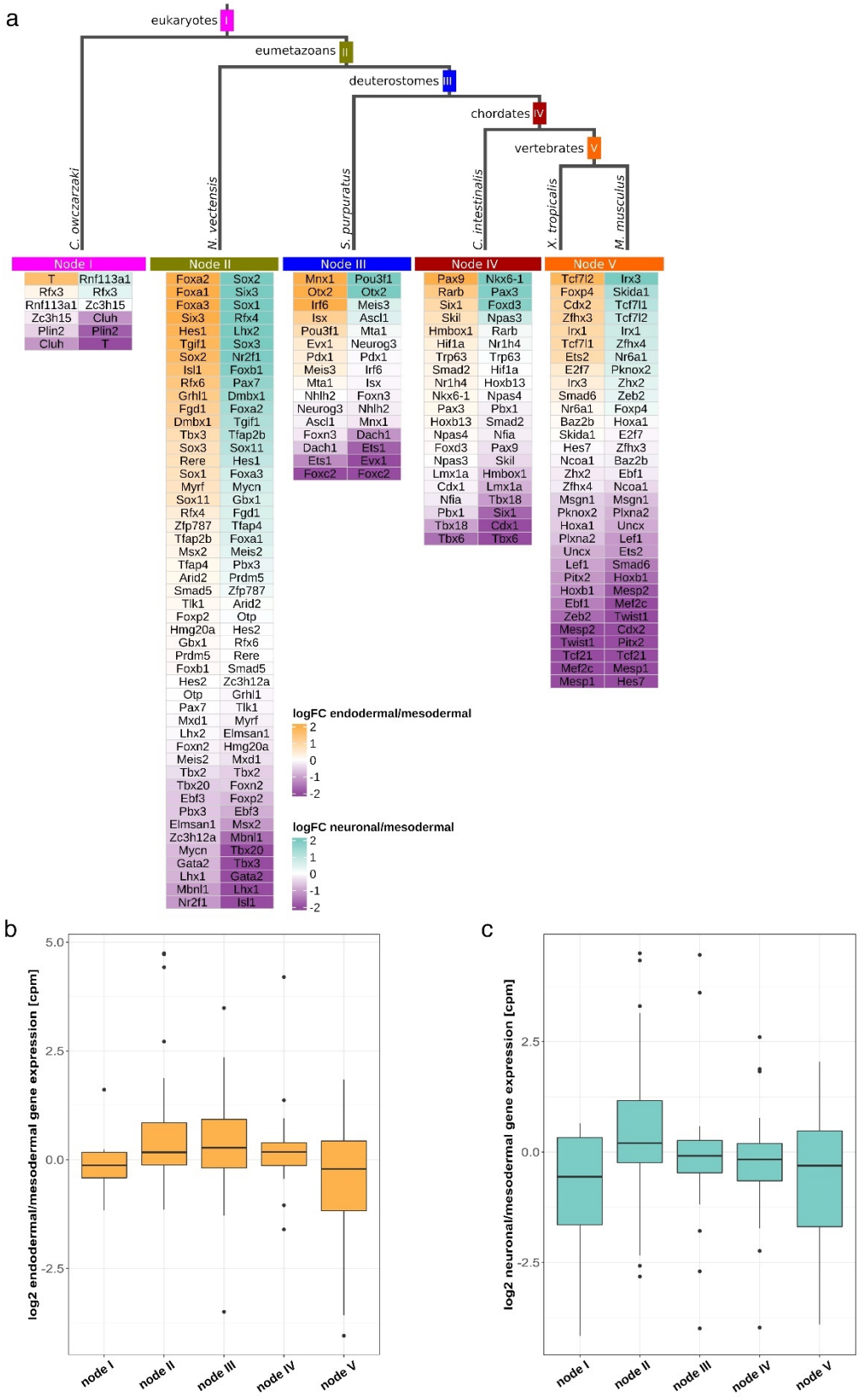


Figure 5: Apomorphic and synapomorphic target genes of Brachyury. To assess the evolutionary origin of vertebrate Brachyury target genes and their function, apomorphic and

synapomorphic target genes were assigned to nodes of common ancestors. Transcription factor coding genes targeted by Brachyury that are shared between an outgroup and at least one member of the ingroup are considered ancestral for each node. The expression of the target genes in neuronal versus mesodermal cell types was annotated using single cell RNAseq data from E8.0 and E8.5 mouse embryos. A) Heatmap of the log2 fold change (logFC) of neuronal (turquoise) or endodermal (yellow) vs mesodermal (purple) gene expression for each gene using the corresponding mouse gene symbols as they appeared in the single cell dataset in each node (see Methods for details). B) Boxplot of the log2 fold change (y-axis) of endodermal vs mesodermal gene expression for all genes per node (x-axis). Note that node2 is enriched in endodermally expressed genes while node 5 is enriched in mesodermally expressed genes (p-value = 0.008, Wilcoxon Rank Sum Test). C) Boxplot of the log2 fold change (y-axis) of neuronal vs mesodermal gene expression for all genes per node (x-axis). Note that node 2 is enriched in neuronal expression while node 5 is enriched in mesodermal expression (p-value = 0.008, Wilcoxon Rank Sum Test).

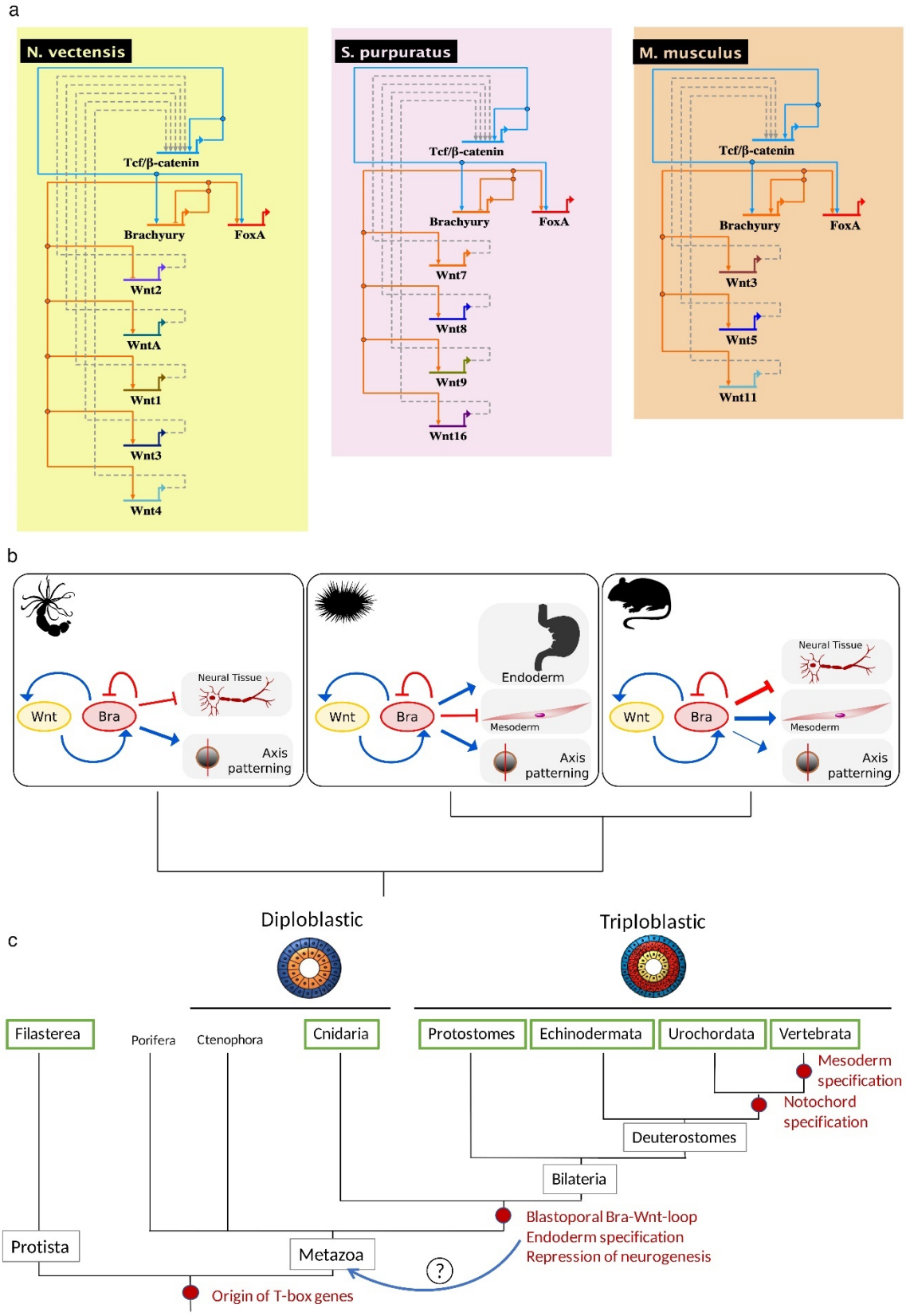
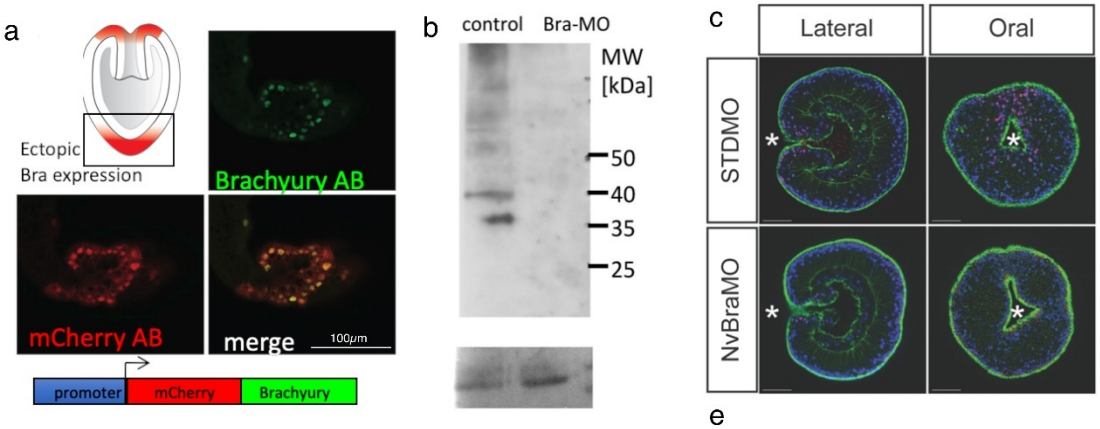


Figure 7: Comparison of regulatory functions of Brachyury in the sea anemone, sea urchin and vertebrates. (**A**) The feedback loop of Brachyury and the Wnt signaling pathway

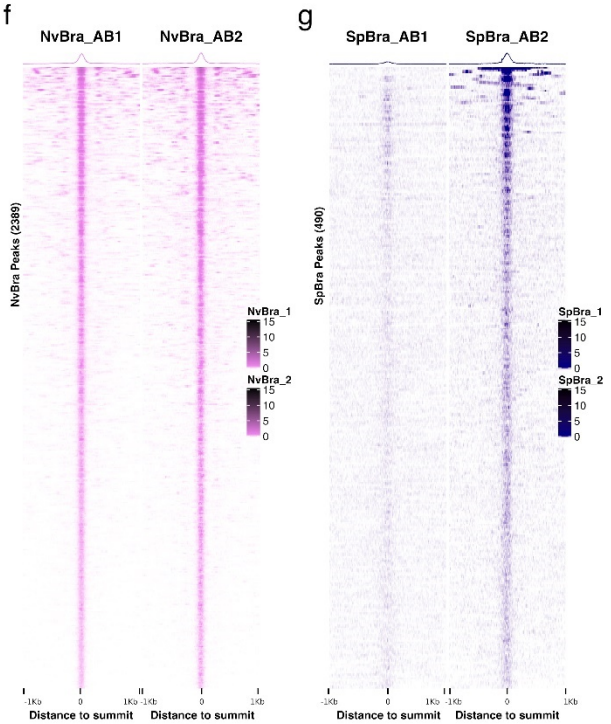
is conserved between the sea anemone, sea urchin and vertebrate, although individual Wnt ligands have been changed as targets of Brachyury. (**B**) Summary of the main functions of Brachyury in axial patterning, neuronal differentiation and germ layer specification in the diploblastic sea anemone, the sea urchin and vertebrates. **(C)** Scenario of stepwise evolution of Brachyury function from protists to vertebrates. See text for explanation.

Supplementary Figures



d

Quantitative analysis of Nematostella Bra knockdown experiments			
	control	BraMO	STDMO
Symmetric expression	171	11	149
Weak/Asymmetric expression	14	38	9
No Staining	21	236	12
Total	206	285	170



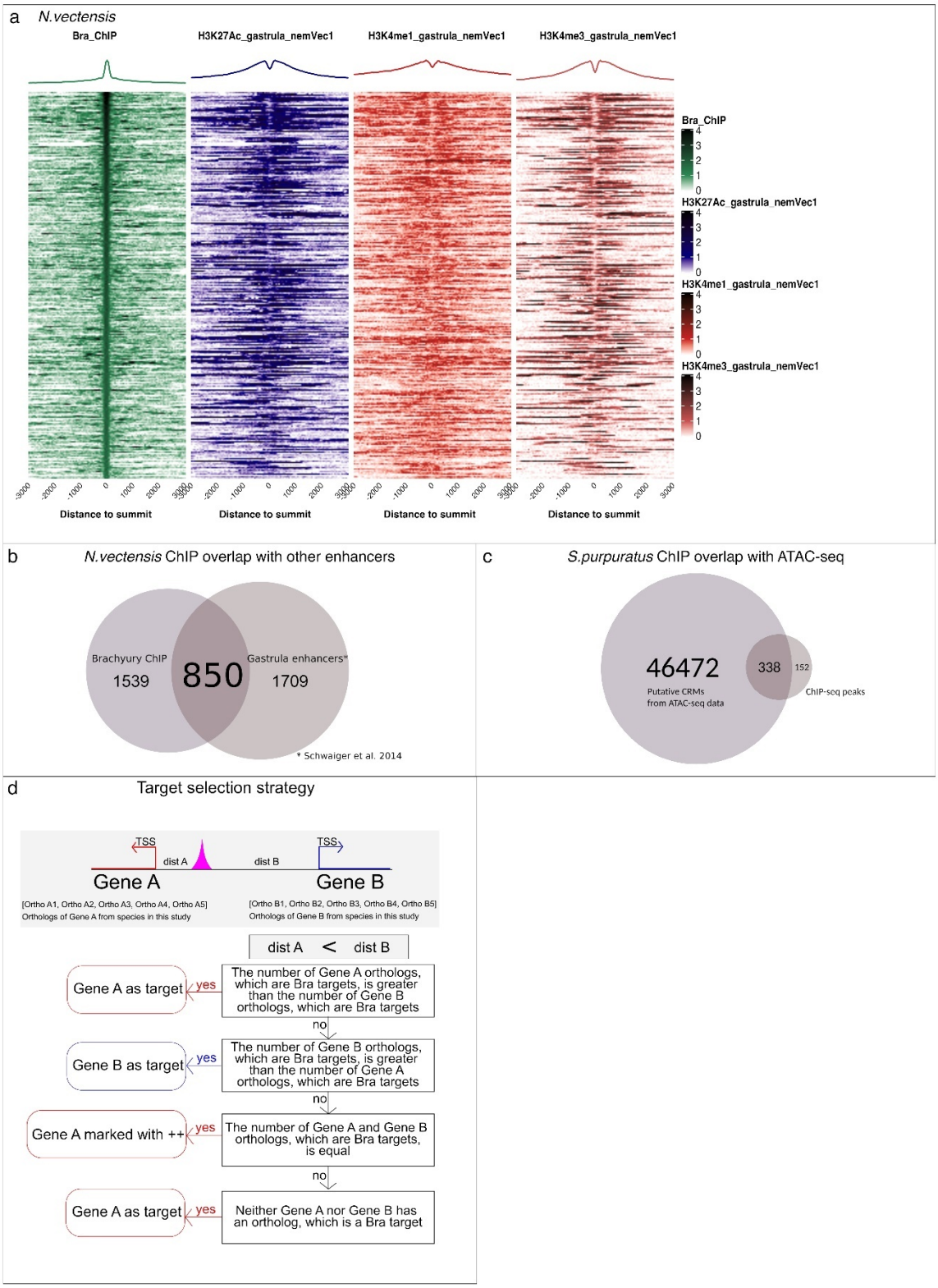
h

Dataset Comparisons

Table X. Datasets used for comparative analyses in this study

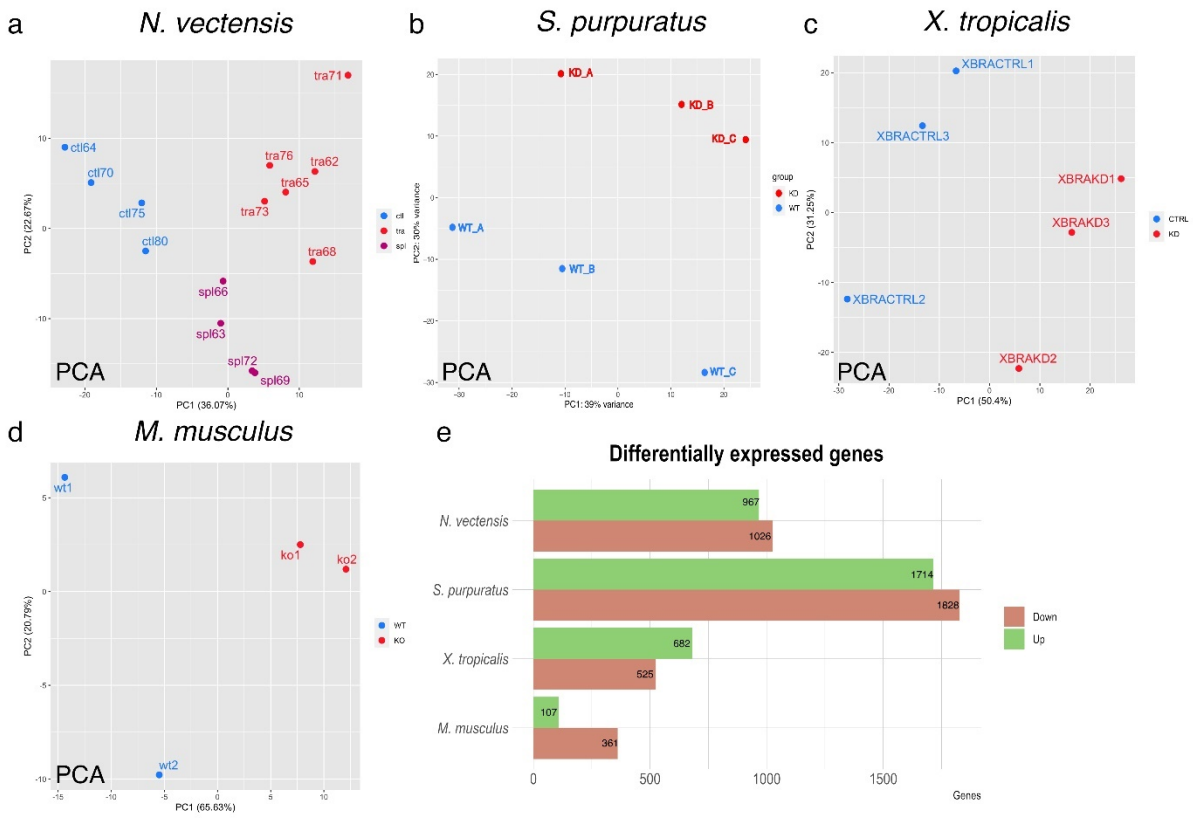
Organism	Method	embryonic Stage	Source of cells	RNAseq	Reference
Mus musculus	ChIP-seq	E8.5	In-vitro differentiated NMPs	WT vs. Mutant	Koch et al., 2017
	scRNA-seq	E8.5	whole embryo	WT	Pijuan-Sala et al. 2019
	scRNA-seq	E8.5	whole embryo	WT	Grosswendt et al. 2020
Xenopus tropicalis	ChIP-seq	gastrula 11-12.5	whole embryo	WT vs. MO KD	Gentsch et al., 2013
Ciona intestinalis	ChIP-on-chip	gastrula	whole embryo	WT vs. MO KD	Kubo et al., 2010
Strongylocentrotus purpuratus	ChIP-seq	gastrula 24h	whole embryo	WT vs. MO KD	this study
Nematostella vectensis	ChIP-seq	gastrula 26h	whole embryo	WT vs. MO KD	this study
Capsaspora owczarzarki	ATAC-seq	n.a.	cell culture	n.d.	Sébé-Pedros et al., 2016

Supplementary Figure 1: Specificity of the Brachyury antibodies and target gene detection strategy. (**A**) Ectopic expression of Brachyury at aboral pole of *Nematostella* confirms the specificity of the antibody. Zygotes were injected with a plasmid of EF1a::mCherry-p2A-Brachyury. Embryos with mosaic expression were stained for mCherry and Brachyury antibodies. **(B)** Western Blot of Brachyury antibody in control and Morpholino mediated knock-down of *brachyury in Nematostella* (**C**) Immunocytochemistry of gastrula stage embryos of *Nematostella* (oral views) in controls and after morpholino-mediated knockdown of Brachyury. (**D**) Quantitative summary of Brachyury knockdown with morpholino oligonucleotides. (**E**) Western blot of anti-Brachyury in *Strongylocentrotus*. The estimated protein size is approx. 50kD. Developmental stages tested: 6h, 12h, 48h, 72h. Recombinant protein (RP) size: 9 kD. **(F-G)** Heatmap of Brachyury binding sites in Bra_AB1 and Bra_AB2 of *Nematostella* and *Strongylocentrotus* related to the TSS of all target genes. **(H)** Summary of the sources of the datasets used in this study.



Supplementary Figure 2: **Brachyury binding sites and relationship to chromatin modifications.** (**A**) Heatmap of *Nematostella* Brachyury binding sites from this study with

chromatin modification sites earlier identified in Schwaiger et al. 2014. (**B**) Venn diagram showing the overlap of Brachyury binding sites in *Nematostella* identified in this study overlaps with the enhancer/promoter sites previously identified in [51]. (**C**) Venn diagram of Brachyury binding sites identified in this study in *Strongylocentrotus* showing the overlap with the previously identified open chromatin sites identified with ATAC-seq data**.** (D) Brachyury target selection strategy. Two closest genes on either side of the binding site were considered and their respective orthologs in the species under study were identified. A gene was prioritized as a target, if it was also a target gene in one or more species.



f Orthologous direct/indirect targets

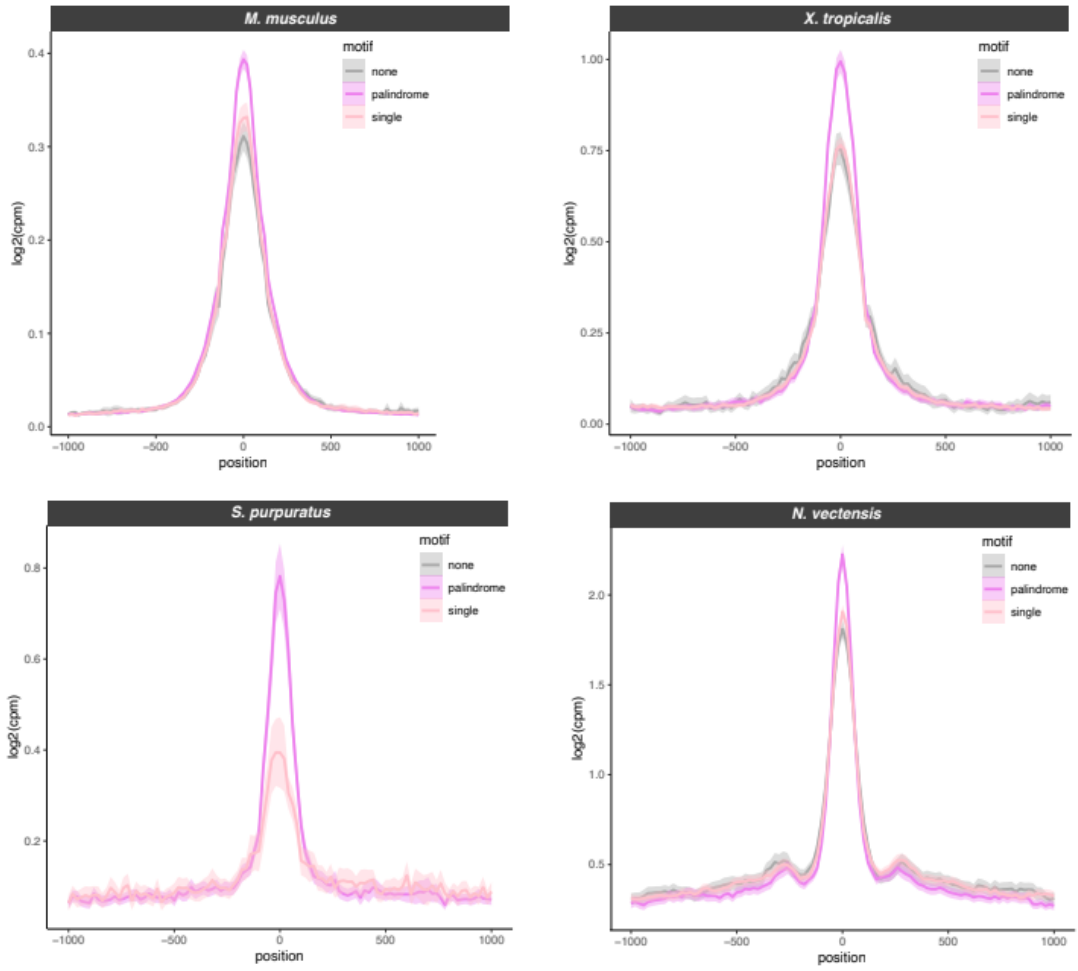
Query dataset	Direct target in query			Indirect target in query		
M. musculus	*X. tropicalis*	*S.purpuratus*	*N. vectensis*	*X. tropicalis*	*S.purpuratus*	*N. vectensis*
Regulated	16	9	4	259	188	205
Not Regulated	31	23	14	796	760	554
Not detected	27	5	8	737	105	138
X. tropicalis	*M. musculus*	*S.purpuratus*	*N. vectensis*	*M. musculus*	*S.purpuratus*	*N. vectensis*
Regulated	23	4	14	234	142	172
Not Regulated	49	26	16	1023	564	376
Not detected	13	7	9	71	53	82
S.purpuratus	*M. musculus*	*X. tropicalis*	*N. vectensis*	*M. musculus*	*X. tropicalis*	*N. vectensis*
Regulated	13	6	9	23	19	31
Not Regulated	33	18	24	106	68	68
Not detected	5	28	4	14	54	24
N. vectensis	*M. musculus*	*X. tropicalis*	*S.purpuratus*	*M. musculus*	*X. tropicalis*	*S.purpuratus*
Regulated	2	3	9	87	95	142
Not Regulated	18	13	15	608	380	561
Not detected	9	10	13	74	295	85

Regulated :- >= 10 baseMean score and adj_pval < 0.05 and
Not Regulated :- >= 10 baseMean score but adj_pval >0.05
Not detected:- DEG not detected

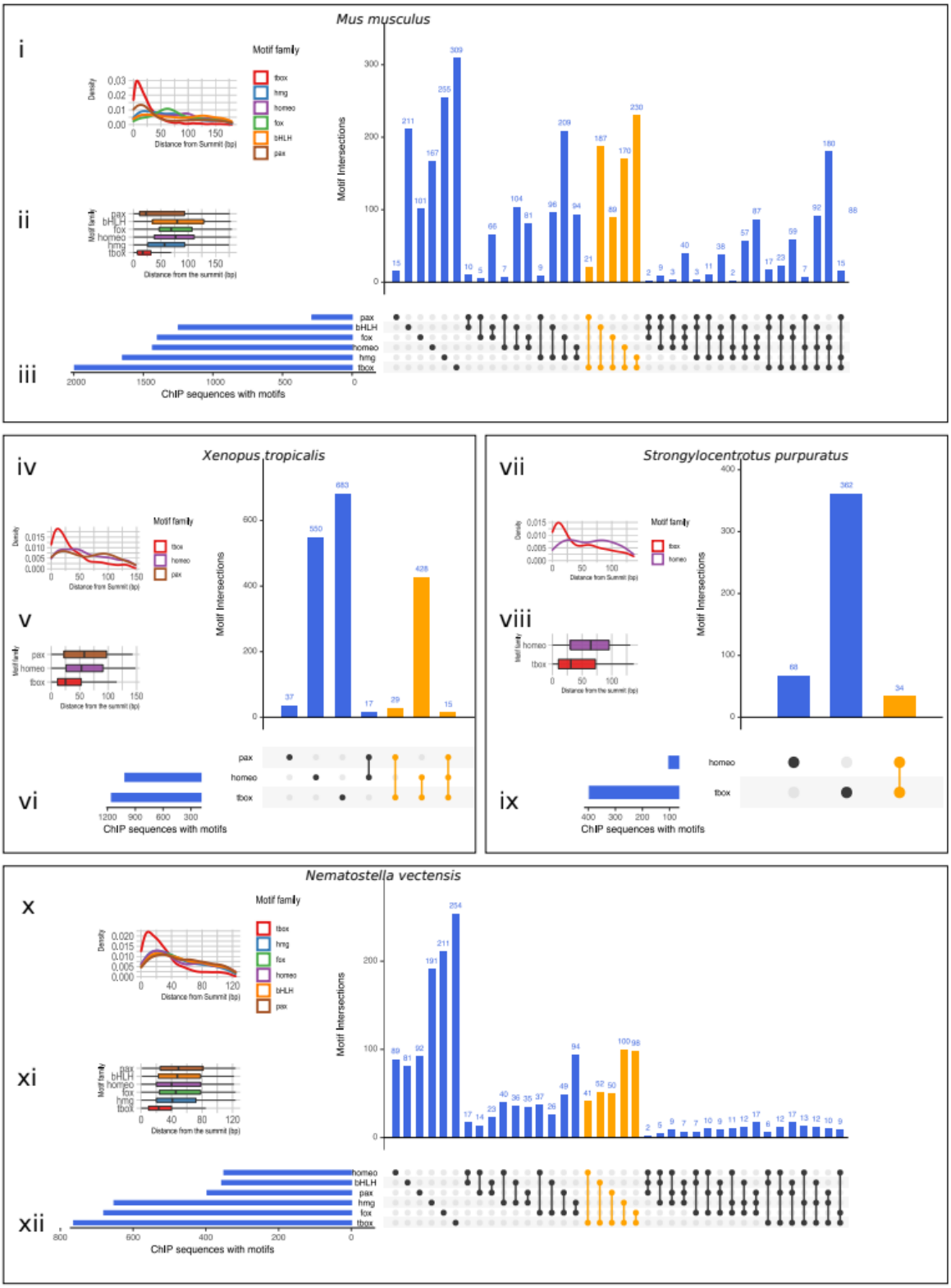
Supplementary Figure 3: **PCA analyses and summary of DEG analyses of RNAseq experiments. (A-D)** Principal Component Analysis (PCA) of RNA-seq datasets in

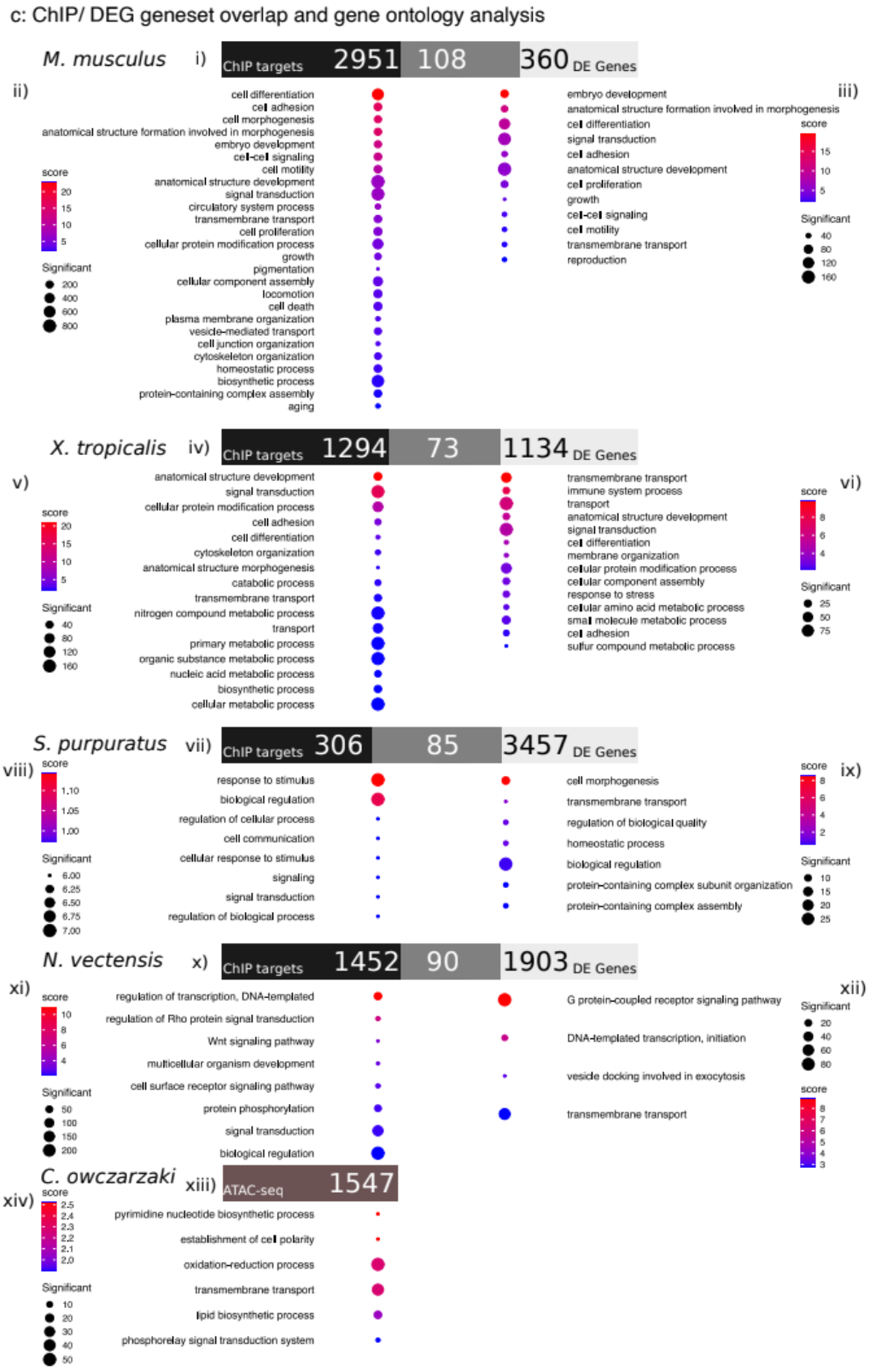
Nematostella, *Strongylocentrotus*, *Xenopus* and *Mouse* respectively. The blue dots represent the control/WT samples while as red dots indicate KD/KO samples/ In case of *Nematostella* both pre-mRNA splicing (spl) and translation blocking (tra) morpholinos were used. (**E**) Summary of differentially expressed genes after Bra KD/KO. Differential expression analysis was done using DEseq2 R package with 0.05 alpha value. Data for *Nematostella*, *Strongylocentrotus* and *Xenopus* is from morpholino induced knock down of Bra transcripts while mouse data is a result of Bra knock-out. (**F**) Overlap of direct (ChIP-seq detected) and indirect targets (DEGs) across different species. Each species was used as a query species (query dataset) and the genes determined that are differentially expressed, and also ChIP targets ("column "direct targets in query") or not ChIP targets (column "indirect targets in query"). The numbers refer to the number of orthologs in each one of the other species. The numbers are low, since the overlap between ChIP-seq targets and DEGs from RNAseq experiments is only 1-10% within a given species.

a: Brachyury peak metaplots split palindrome/single



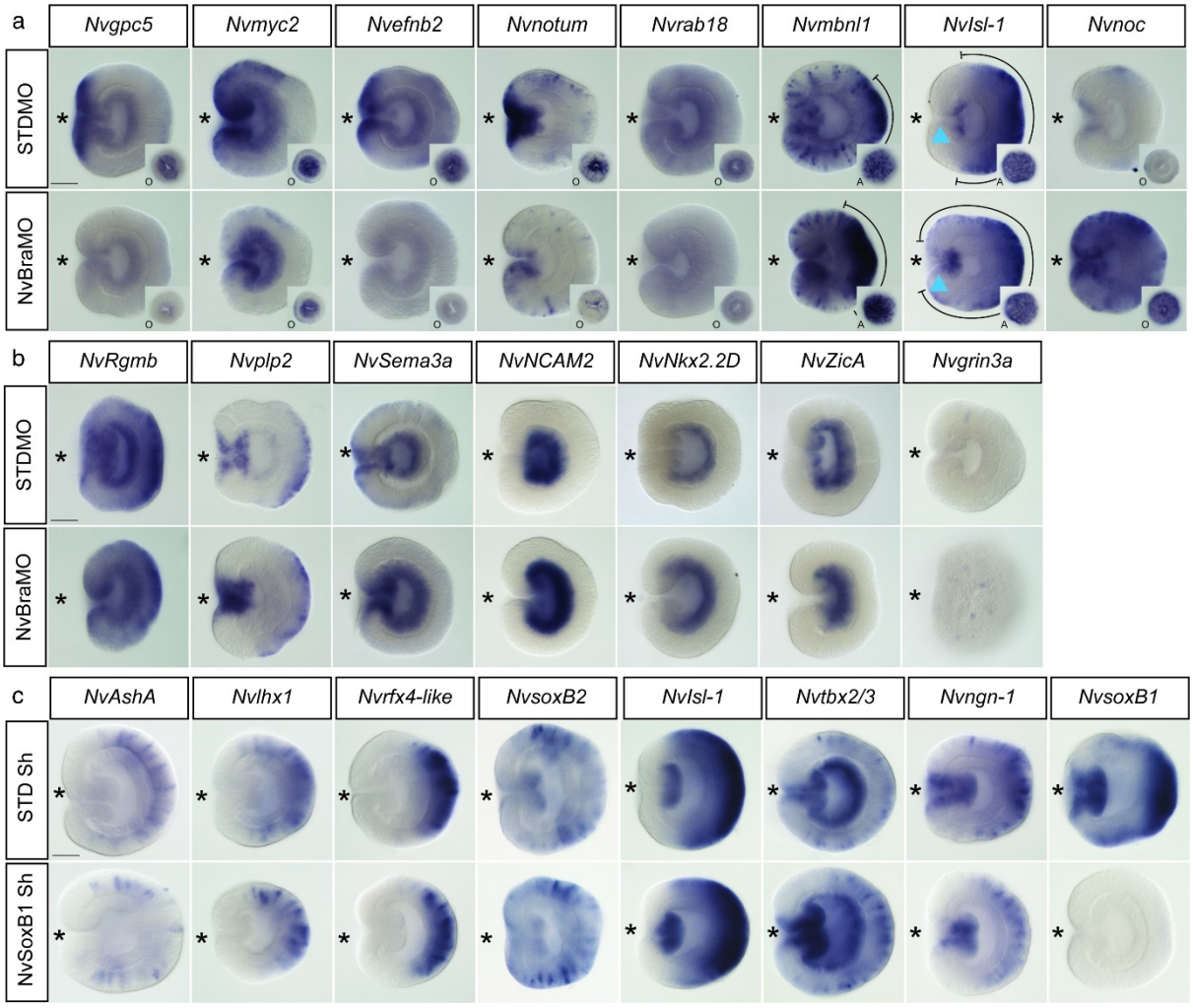
b: Intersection of motifs within the peak region



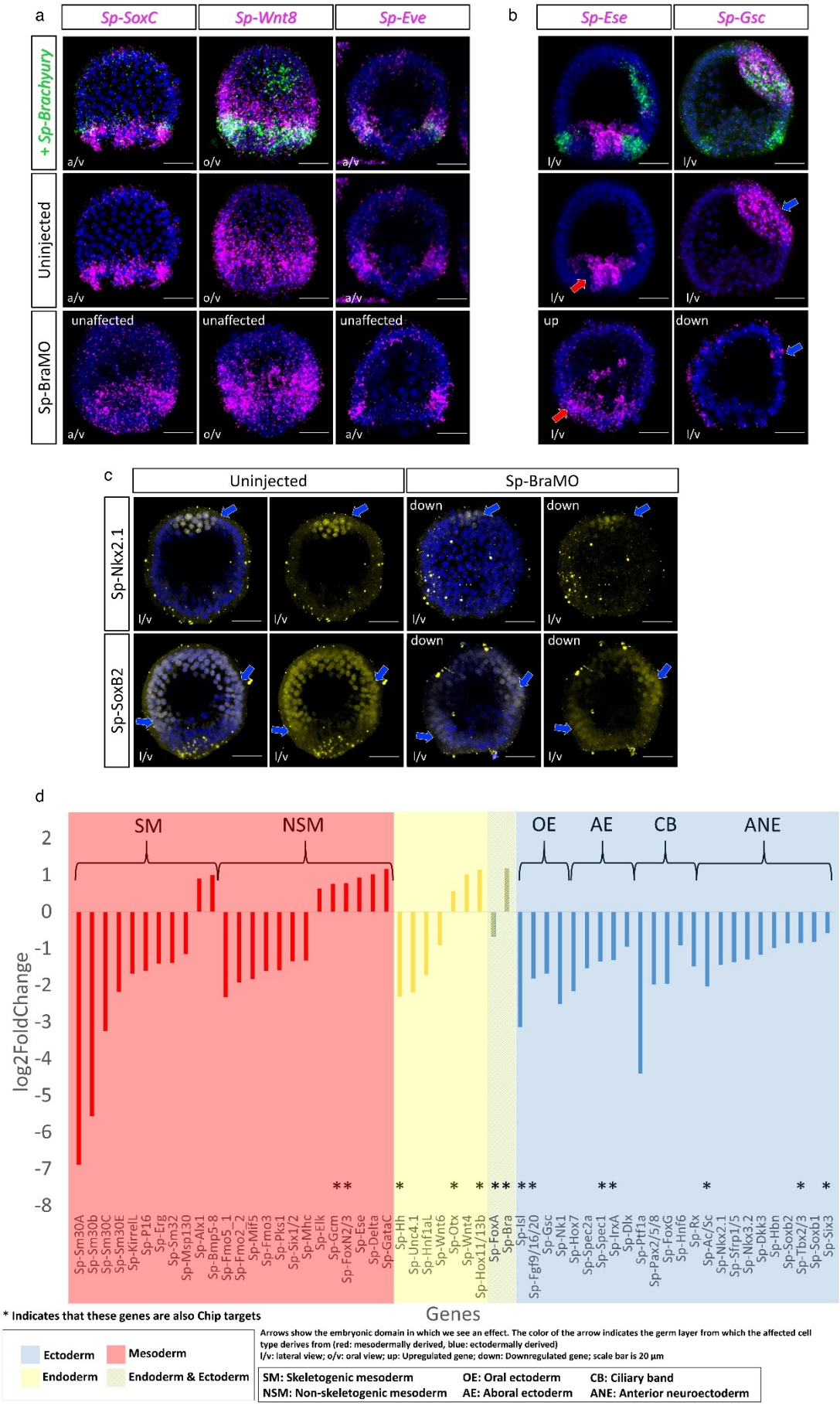

c: ChIP/ DEG geneset overlap and gene ontology analysis
M. musculus
i)
ChIP targets
2951
108
360
DE Genes
ii)
cell differentiation
cell adhesion
cell morphogenesis
anatomical structure formation involved in morphogenesis
embryo development
cell-cell signaling
cell motility
anatomical structure development
signal transduction
circulatory system process
transmembrane transport
cell proliferation
cellular protein modification process
growth
pigmentation
cellular component assembly
locomotion
cell death
plasma membrane organization
vesicle-mediated transport
cell junction organization
cytoskeleton organization
homeostatic process
biosynthetic process
protein-containing complex assembly
aging
score
Significant
iii)
embryo development
anatomical structure formation involved in morphogenesis
cell differentiation
signal transduction
cell adhesion
anatomical structure development
cell proliferation
growth
cell-cell signaling
cell motility
transmembrane transport
reproduction
score
Significant
X. tropicalis
iv)
ChIP targets
1294
73
1134
DE Genes
v)
anatomical structure development
signal transduction
cellular protein modification process
cell adhesion
cell differentiation
cytoskeleton organization
anatomical structure morphogenesis
catabolic process
transmembrane transport
nitrogen compound metabolic process
transport
primary metabolic process
organic substance metabolic process
nucleic acid metabolic process
biosynthetic process
cellular metabolic process
score
Significant
vi)
transmembrane transport
immune system process
transport
anatomical structure development
signal transduction
cell differentiation
membrane organization
cellular protein modification process
cellular component assembly
response to stress
cellular amino acid metabolic process
small molecule metabolic process
cell adhesion
sulfur compound metabolic process
score
Significant
S. purpuratus
vii)
ChIP targets
306
85
3457
DE Genes
viii)
response to stimulus
biological regulation
regulation of cellular process
cell communication
cellular response to stimulus
signaling
signal transduction
regulation of biological process
score
Significant
ix)
cell morphogenesis
transmembrane transport
regulation of biological quality
homeostatic process
biological regulation
protein-containing complex subunit organization
protein-containing complex assembly
score
Significant
N. vectensis
x)
ChIP targets
1452
90
1903
DE Genes
xi)
regulation of transcription, DNA-templated
regulation of Rho protein signal transduction
Wnt signaling pathway
multicellular organism development
cell surface receptor signaling pathway
protein phosphorylation
signal transduction
biological regulation
score
Significant
xii)
G protein-coupled receptor signaling pathway
DNA-templated transcription, initiation
vesicle docking involved in exocytosis
transmembrane transport
Significant
score
C. owczarzaki
xiii)
ATAC-seq
1547
xiv)
pyrimidine nucleotide biosynthetic process
establishment of cell polarity
oxidation-reduction process
transmembrane transport
lipid biosynthetic process
phosphorelay signal transduction system
score
Significant

Supplementary Figure 4: **Motif analyses of Brachyury ChIP-peaks and Gene Ontology analyses of target genes.**

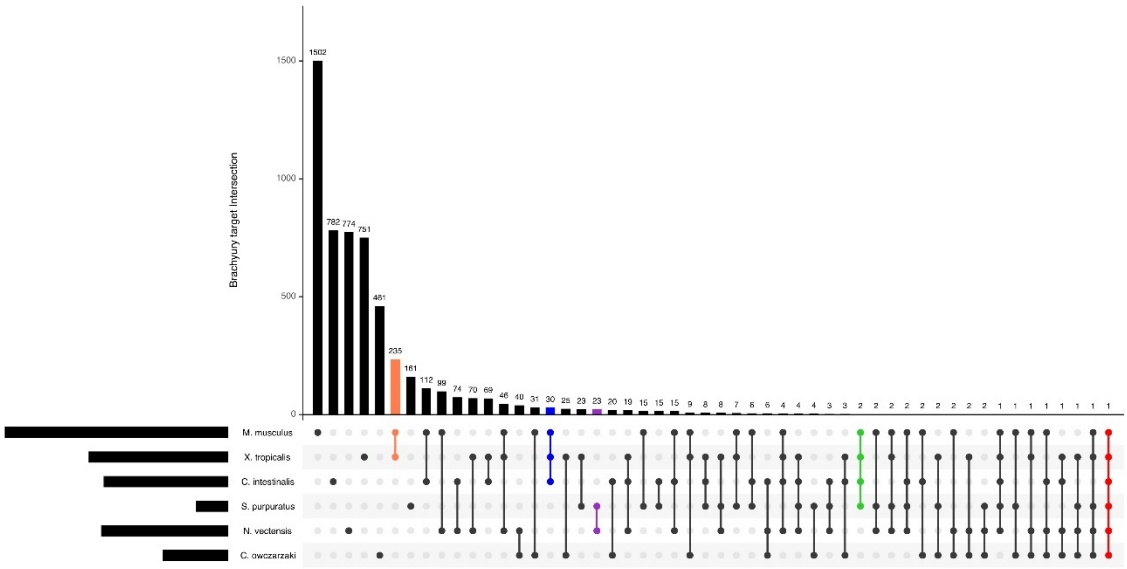
(A) Brachyury ChIP metaplots around Brachyury peaks containing palindrome or half palindrome (single) or no Brachyury binding motifs. The average read count (normalized to a million reads) was calculated for Brachyury ChIP-seq reads for regions around peak summits spanning 2kb in 20bp windows. The shaded area around the lines represents the 95% confidence interval across peaks in a category. Note that in all four species, peaks with palindromes show significantly higher ChIP-seq read counts, which may serve as a proxy for the strength of Brachyury binding. (**B**) Intersection of motifs with the peak region. The Brachyury peak region (as identified using the Peakzilla algorithm) was scanned for presence of other transcription factor binding sites using Fimo (MEME-ChIP suite) with default settings. The resulting binding motifs were grouped by the transcription factor families Paired box (Pax), basic helix–loop–helix (bHLH), Forkhead box (Fox), homeobox (homeo), High mobility group (hmg), T-box (tbox). Presence of these motifs together with the Brachyury motif is highlighted in orange in the upset plots (iii, vi, ix, xii). Distance of these motifs from the peak center was also tracked and is shown in adjoining plots (I, ii, iv, v, vii, viii, x, xi). **(C)** ChIP / DE gene set overlap and GO analysis. Overlap (dark grey) between ChIP targets (black) and differentially expressed (light grey) genes in *M. musculus* (i), *X. tropicalis* (iv), *S. purpuratus* (vi), *N. vectensis* (x). In *C. owczarzaki* (xiii) only the number of ATAC-seq peaks with Bra motifs is shown. Gene ontology analysis of Brachyury ChIP targets for *M. musculus* (ii), *X. tropicalis* (v), *S. purpuratus* (vii), *N. vectensis* (xi) and *C. owczarzaki* (xiv) and of differentially expressed genes after Bra KO/KD (iii, vi, ix, xii). Only gene ontology terms for biological process and molecular function were reported. The color of the dot represents the score (-log(p-value)) assigned by topGO while the size of the dot represents the number of genes associated with the term. **(D)** GO analyses of target genes of peaks with or without a Bra consensus motif. Note that no significant difference can be detected.



Supplementary Figure 5: **Expression analysis of Brachyury ChIP targets by WMISH after knockdown in *Nematostella*.** (**A**) Morpholino-mediated knockdown of target genes with complex expression pattern show partial down or upregulation in the ectodermal layer. (**B**) Neuronal target genes that are not affected by Brachyury knockdown. Note that all unaffected genes are expressed in the inner layer of the embryo. (**C**) Knockdown of SoxB1 (a homolog of vertebrate Sox2) shows no effect on neuronal target genes regulated by Brachyury.


a
Sp-SoxC
Sp-Wnt8
Sp-Eve
+ Sp-Brachyury
Uninjected
Sp-BraMO
unaffected
b
Sp-Ese
Sp-Gsc
up
down
c
Uninjected
Sp-BraMO
Sp-Nkx2.1
Sp-SoxB2
d
SM
NSM
OE
AE
CB
ANE
log2FoldChange
Genes
* Indicates that these genes are also Chip targets
Ectoderm
Mesoderm
Endoderm
Endoderm & Ectoderm
Arrows show the embryonic domain in which we see an effect. The color of the arrow indicates the germ layer from which the affected cell type derives from (red: mesodermally derived, blue: ectodermally derived)
l/v: lateral view; o/v: oral view; up: Upregulated gene; down: Downregulated gene; scale bar is 20 µm
SM: Skeletogenic mesoderm
NSM: Non-skeletogenic mesoderm
OE: Oral ectoderm
AE: Aboral ectoderm
CB: Ciliary band
ANE: Anterior neuroectoderm

Supplementary Figure 6: Expression analyses of Brachyury target genes in *Strongylocentrotus* upon knockdown. (**A**) ChIP target genes that are not affected by Brachyury knockdown. (**B**) Expression analysis of Brachyury RNA seq targets by WMISH after morpholino induced knockdown (**C**) Expression analysis of Brachyury RNA seq targets by immunohistochemistry after morpholino induced knockdown. Arrows show the embryonic domain in which we see an effect (red arrow: mesodermally derived; blue ectodermally derived). (**D**) Differentially expressed genes after Morpholino induced knockdown that are also ChIP targets at 24h. Key genes playing a crucial role in endoderm development are highlighted in yellow, key genes playing a role in mesoderm development are highlighted in red while key genes playing a role in ectoderm development are highlighted in blue. Asterisks indicate genes that are also ChIP targets. l/v: lateral view; v/v vegetal view; o/v oral view; a/v aboral view. up: upregulated gene; down; down-regulated gene. The scale bar is 20 μm.

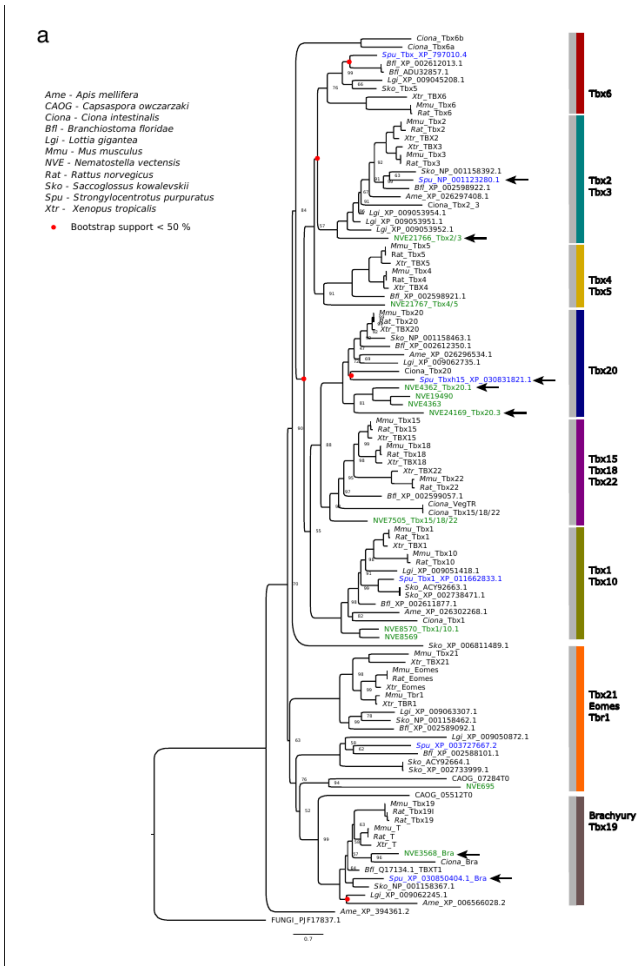


Vertebrata							Chordata	Non-chordates Metazoa
A830080D01	Ccno	Elmo1	Iqca	Neurl1a	Rhob	Tle1	Adgra3	LOC764665
Acta2	Cd209c	Ephb1	Iqgap2	Nin	Rnf165	Tlr5	Angpt1	Arhgap20
Aen	Cd209f	Ephb3	Irf2bp2	Ninj2	Ror2	Tmem72	Bcas3	Atp2b1
Aimp2	Cdh2	Erbb4	Irx1	Nipal2	Rspo3	Trrap	Cldn19	B3GALT5
Akap2	Cdon	Erc2	Irx3	Nkain2	Sall3	Ttc39a	Col11a1	BP10
Amotl2	Cdx1	Ercc4	Itgbl1	Nr6a1	Scaf8	Twist1	Col13a1	C1qtnf12
Ano5	Cdx2	Ets1	Jph2	Nrp1	Selenon	Ube2v1	Crim1	CHRNA7
Apln	Celf2	Ets2	Kirrel	Olfr364-ps1	Sema4c	Uncx	Dgkh	Cog1
Aplnr	Chd7	Exoc6	Klf6	Otud7b	Sestd1	Urm1	Epha4	Cyp2u1
App	Chrna9	Exoc7	Krt18	Pacrg	Sfrp1	Usp13	Ext1	Dmbx1
Arf1	Cksl1b	F2rl1	Krt19	Pam	Sgpp2	Vgll4	Farp1	Ephb2
Arhgap15	Clybl	Fam120b	Lef1	Pappa	Shb	Vps13b	Gm21949	Gbx1
Arhgef28	Cnih1	Farp2	Lin9	Pcdh7	Sipa1l3	Vwa8	Gnai1	HES-1
Aridl a	Col5a1	Fbxo25	Litaf	Pdgfra	Skap1	Xpo6	Gtdc1	hlh-6
Arl4a	Cox4i2	Fgfr2	Lmf1	Pdx1	Skida1	Zbtb10	Hmbox1	Ipo4
Arrdc4	Csnk1e	Flnb	Lrig1	Peak1	Slc12a2	Zdhhc21	Kcnmb2	Kif12
Arsi	Csnk1g3	Fmn1 2	Ltbp1	Peli2	Slc7a6	Zeb2	Kirrel3	Ngdn
Astn2	D630003M21	Fmod	Macrod2	Pfkfb3	Slit3	Zfhx3	Mid1	Notum
Atg7	Dapl1	Foxp4	Magi1	Pitx2	Smad6	Zfhx4	Mmp21	Nrarp
Atic	Denndl b	Frmd4b	Maml3	Pknox2	Smim15	Zfp217	Nbea	Spata18
Atl2	Dennd2c	Fto	Mcph1	Pla2g12b	Snrpc	Zfp521	Ntng1	Thoc1
Atp8a2	Dgcr6	Gli3	Med13l	Plekhg1	Sntb1	Zfp703	Pax3	Wee1
Auts2	Dkk1	Gm15262	Med27	Plpp1	Sp5	Zhx2	Plcl2	Wnt8b
Bahcc1	Dmd	Gmnc	Mef2c	Plxna2	Sp9	Zic2	Prkg1	
Bambi	Dna2	Gramd1c	Mesp1	Pnlip	Spred2	Zmat4	Qk	
Baz2b	Dnajc5b	Has2	Mesp2	Ppp1cb	Spry2		Sall1	Bilateria
Bcl7a	Dnmt3a	Hes7	Mex3b	Ppt1	Srek1		Sema6a	Rgma
Cachd1	Dock4	Hoxa1	Mipol1	Prr5l	St3gal1		Six1	Tmcc1
Cacna2d1	Dock8	Hoxb1	Mmp7	Ptpn14	Sulf1		Slit2	
Cacnb4	Dusp5	Hsd17b3	Mreg	R3hdm2	Taf3		Sndl	Metazoa
Cadps2	Dusp6	Iars	Msgn1	Rbfox1	Tanc1			Brachyury
Cbx3	E2f7	Ier5	Mycbp2	Rbm20	Tcf21			
Ccdc148	Ebf1	Il17rd	Ncoa1	Rbms1	Tcf711			
Ccdc88c	Echdc2	Immp1l	Ndfip2	Rcbtb2	Tcf712			
Ccnh	Eif2b5	Ints7	Ndufs4	Rftn1	Tfb1m			

Supplementary Figure 7: Shared Brachyury targets between lineages of Metazoa, Bilateria, Chordata and Vertebrata. Upset plot of shared orthologous genes as detected by

OMA between different lineage combinations. Note that *brachyury* is the only target gene found in all investigated organisms. The large number of shared target genes between mouse and Xenopus indicates that this screen is robust against slight differences in developmental staging, source of cells, experimental design and sensitivity.

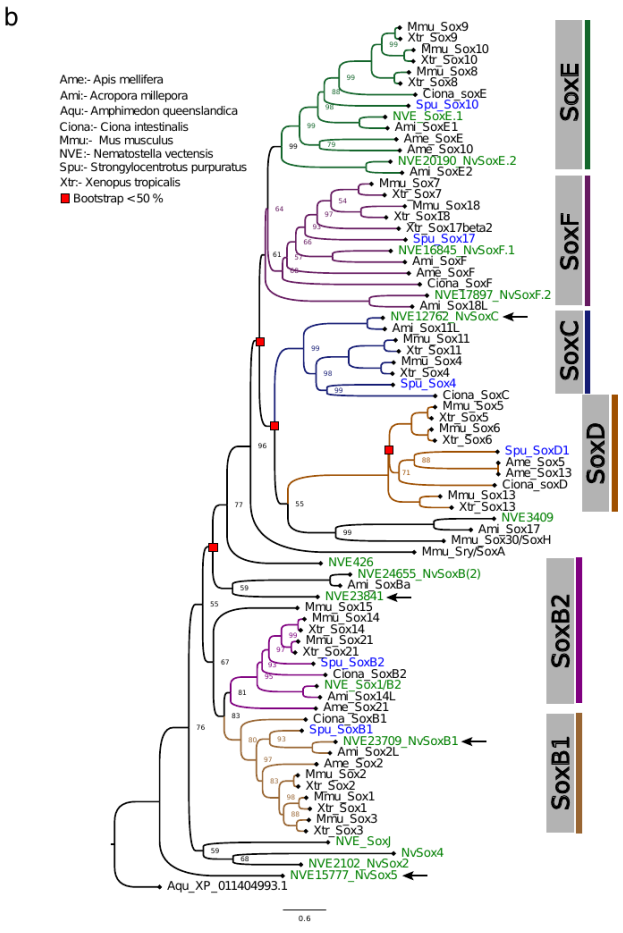
Supplementary Figure 8: Protein phylogenies of selected target genes (attached separately)



(a) T-box family phylogenetic tree

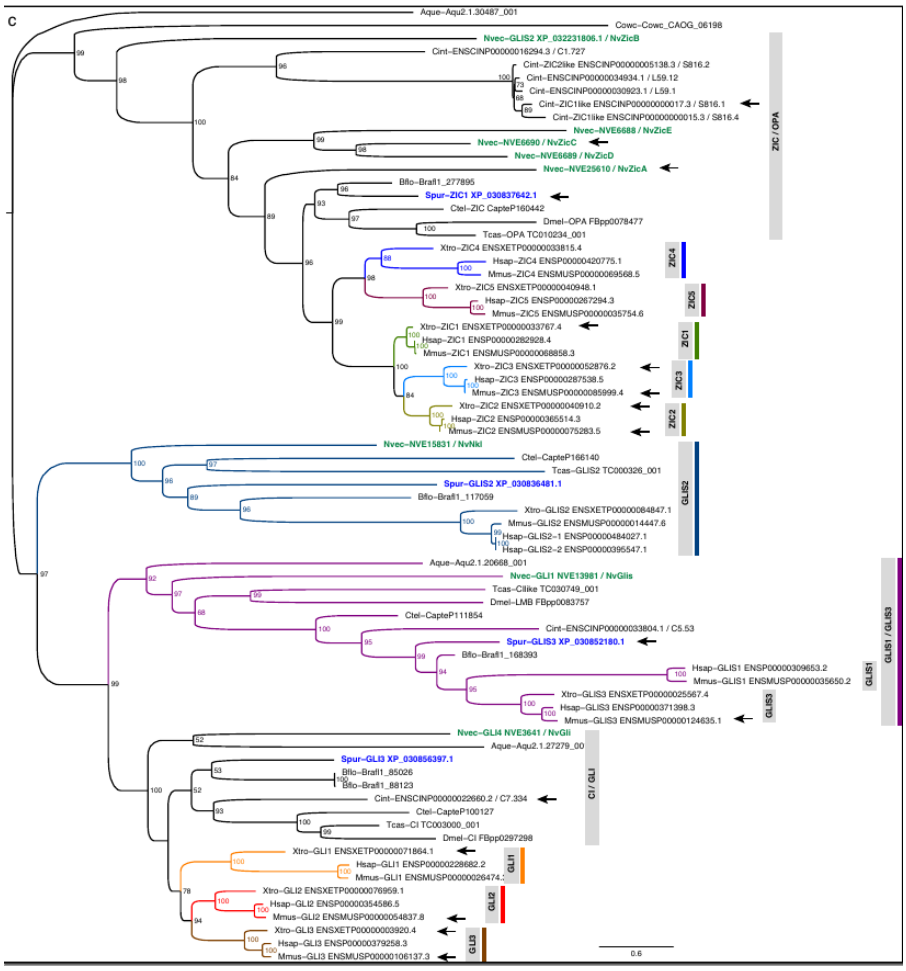
T-Box family tree constructed with T-box genes from *Apis mellifera* (Ame), *Branchiostoma floridae* (Bfl), Capsaspora owczarzaki (CAOG), *Ciona intestinalis* (Ciona), *Lottia gigantea* (Lgi), *Mus musculus* (mmus), *Nematostella vectensis* (NVE), *Strongylocentrotus purpuratus* (Spu), *Xenopus tropicalis* (xtro), *Rattus norvegicus* (Rat), *Saccoglossus kowalevskii* (Sko). Tree was constructed using maximum likelihood method with 1000 UFboot samples, the values at the nodes represent the support values. UFboot values below 50% are not shown and the nodes are marked with a red circle. The values at nodes with 100% support are also

not shown. The tree was rooted on a T-box gene from the fungus *Paramicrosporidium saccamoebae*. Brachyury target genes in *Nematostella* and *Strongylocentrotus* are indicated by arrows.



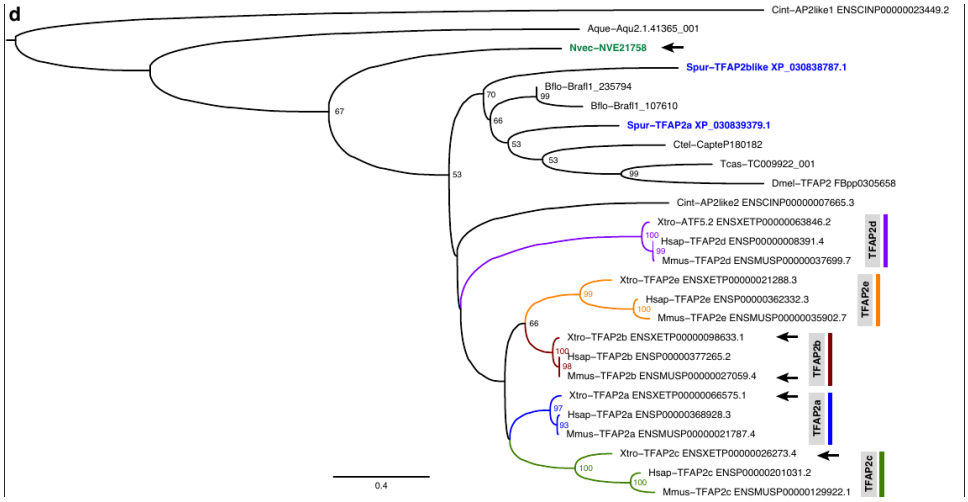
(b) Sox family phylogenetic tree

Sox family tree constructed with Sox genes from *Apis mellifera* (ame), *Acropora millepora* (Ami), *Amphimedon queenslandica* (Aq), *Ciona intestinalis* (ci), *Mus musculus* (mmu), *Nematostella vectensis* (NVE), *Strongylocentrotus purpuratus* (Spu), *Xenopus tropicalis* (xtro). Tree was constructed with a maximum likelihood method with 1000 UFboots samples, the values at the nodes represent the support values. UFboots values below 50% are not shown and the nodes are marked with a red circle. The values at nodes with 100% support are also not shown. The tree was rooted with a sponge *A. queenslandica* Sox protein. Brachyury target genes in *Nematostella* and *Strongylocentrotus* are indicated by arrows.



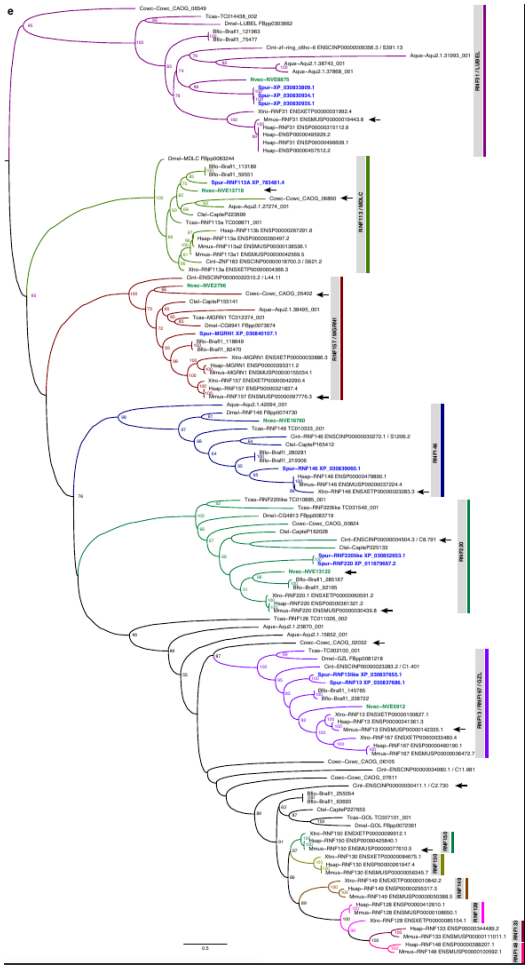
(c) Zic family phylogeny

ZIC family tree constructed from ZIC genes from *Amphimedon queenslandica* (Aque), *Branchiostoma floridae* (Bflo), *Capitella teleta* (Ctel), *Capsaspora owczarzaki* (Cowc), *Ciona intestinalis* (Cint), *Drosophila melanogaster* (Dmel), *Homo sapiens* (Hsap), *Mus musculus* (Mmus), *Nematostella vectensis* (Nvec), *Strongylocentrotus purpuratus* (Spur), *Tribolium castaneum* (Tcas), *Xenopus tropicalis* (Xtro). This protein maximum likelihood tree was constructed using IQ-Tree with 10000 UFboot samples. The values at the nodes represent the UFboot support, where values below 50 % are not shown. The tree is rooted between the ZIC and GLI/GLIS subfamilies. Sequences from *Nematostella* are marked bold in green, those from *Strongylocentrotus* in bold and blue. Brachyury target genes in *Nematostella* and *Strongylocentrotus* are indicated by arrows.



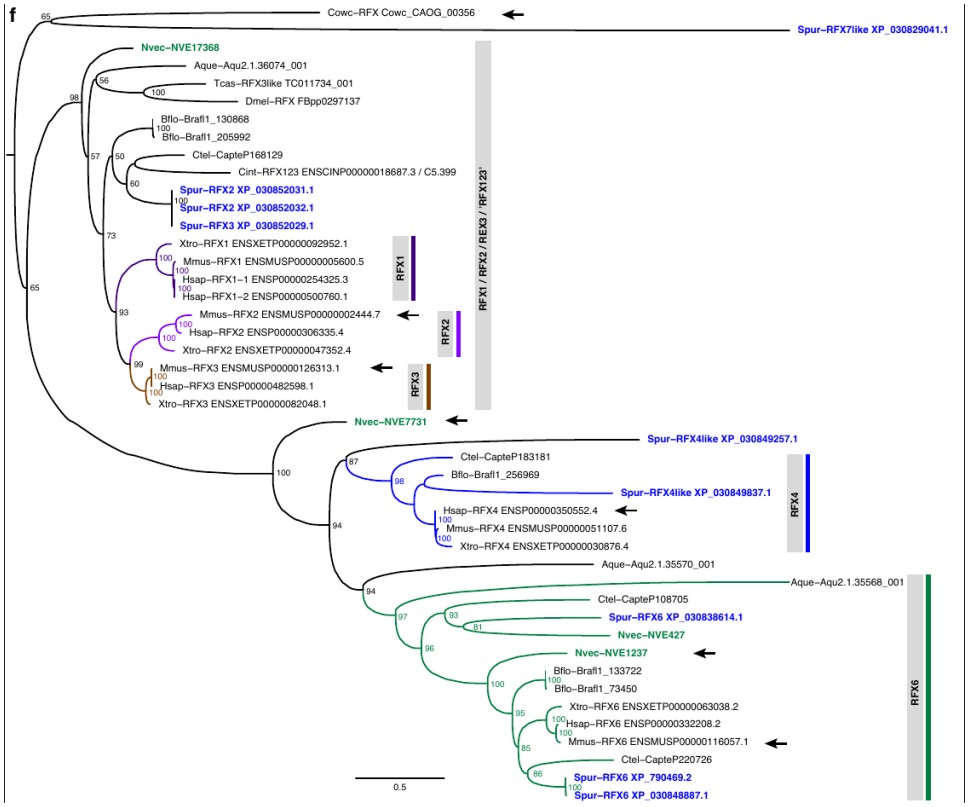
(d) TFAP2 family phylogeny

TFAP2 family tree constructed from TFAP2 genes from *Amphimedon queenslandica* (Aque), *Branchiostoma floridae* (Bflo), *Capitella teleta* (Ctel), *Ciona intestinalis* (Cint), *Drosophila melanogaster* (Dmel), *Homo sapiens* (Hsap), *Mus musculus* (Mmus), *Nematostella vectensis* (Nvec), *Strongylocentrotus purpuratus* (Spur), *Tribolium castaneum* (Tcas), *Xenopus tropicalis* (Xtro). This protein maximum likelihood tree was constructed using IQ-Tree with 10000 UFboot samples. The values at the nodes represent the UFboot support, where values below 50 % are not shown. The tree is shown midpoint-rooted with Figtree. Sequences from *Nematostella* are marked bold in green, those from *Strongylocentrotus* in bold and blue. Brachyury target genes in *Nematostella* and *Strongylocentrotus* are indicated by arrows.



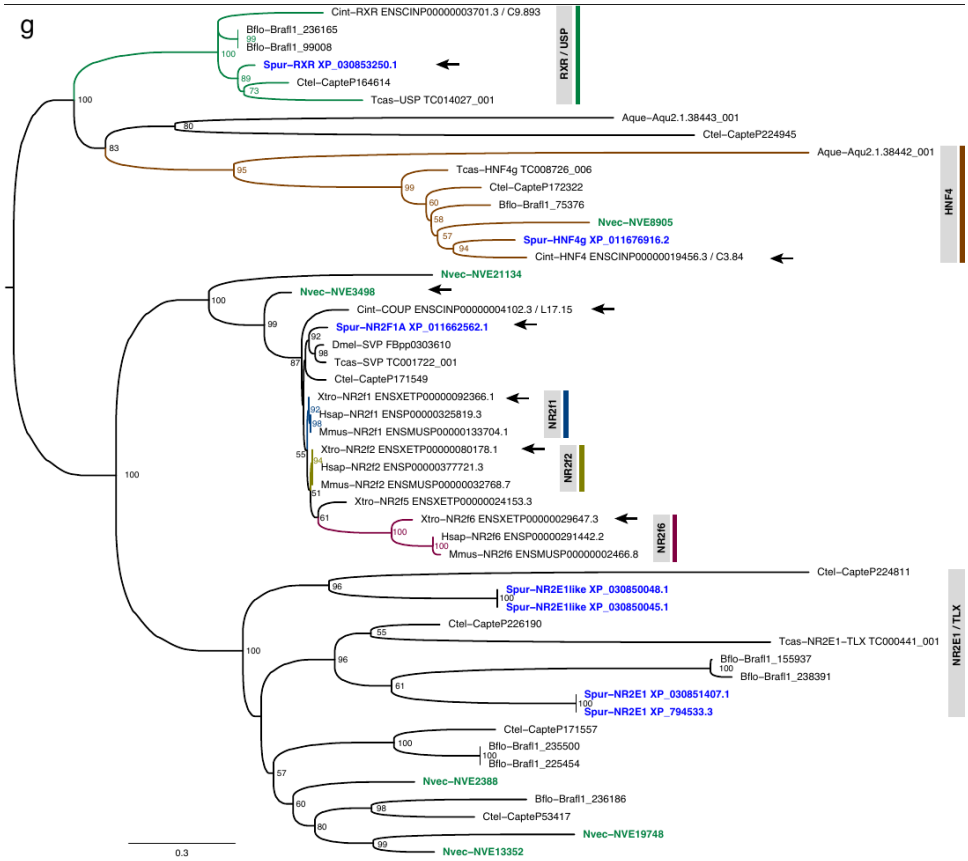
(e) RNF family phylogeny

RNF family tree constructed from RNF genes from *Amphimedon queenslandica* (Aque), *Branchiostoma floridae* (Bflo), *Capitella teleta* (Ctel), *Capsaspora owczarzaki* (Cowc), *Ciona intestinalis* (Cint), *Drosophila melanogaster* (Dmel), *Homo sapiens* (Hsap), *Mus musculus* (Mmus), *Nematostella vectensis* (Nvec), *Strongylocentrotus purpuratus* (Spur), *Tribolium castaneum* (Tcas), *Xenopus tropicalis* (Xtro). This protein maximum likelihood tree was constructed using IQ-Tree with 10000 UFboot samples. The values at the nodes represent the UFboot support, where values below 50 % are not shown. The tree is shown midpoint-rooted with Figtree. Sequences from *Nematostella* are marked bold in green, those from *Strongylocentrotus* in bold and blue. Brachyury target genes in *Nematostella* and *Strongylocentrotus* are indicated by arrows.



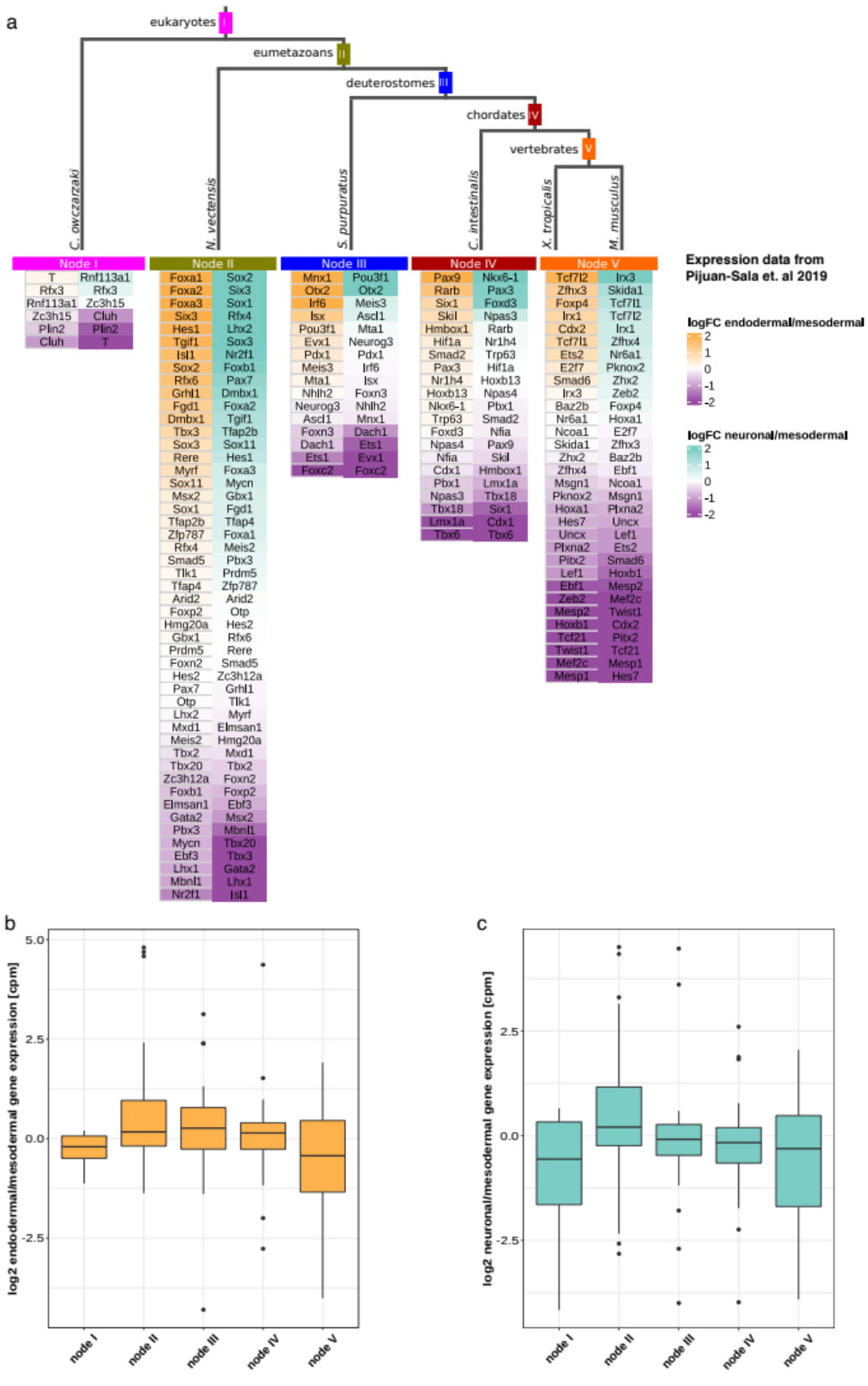
(f) RFX family phylogeny

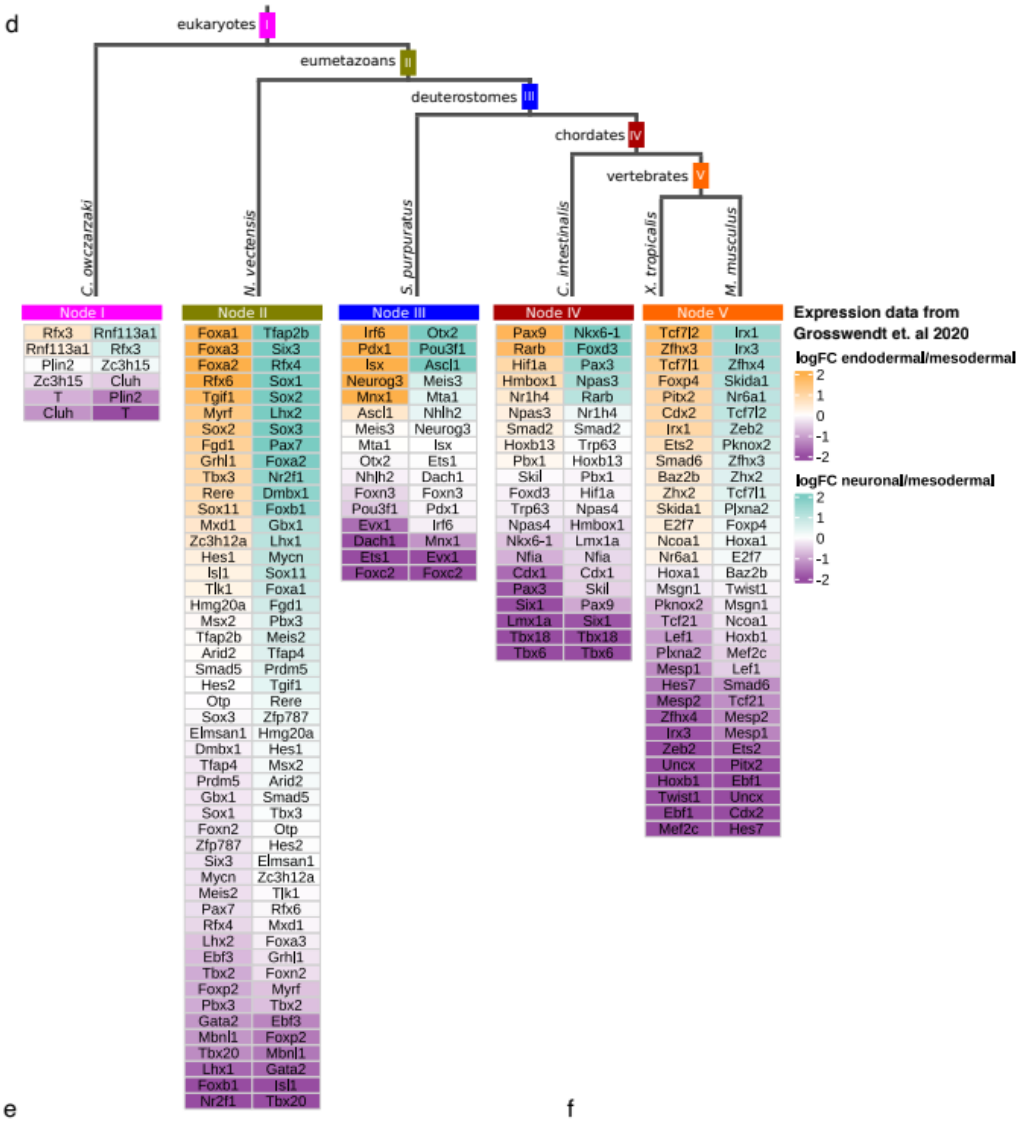
RFX family tree constructed from RFX genes from *Amphimedon queenslandica* (Aque), *Branchiostoma floridae* (Bflo), *Capitella teleta* (Ctel), *Capsaspora owczarzaki* (Cowc), *Ciona intestinalis* (Cint), *Drosophila melanogaster* (Dmel), *Homo sapiens* (Hsap), *Mus musculus* (Mmus), *Nematostella vectensis* (Nvec), *Strongylocentrotus purpuratus* (Spur), *Tribolium castaneum* (Tcas), *Xenopus tropicalis* (Xtro). This protein maximum likelihood tree was constructed using IQ-Tree with 10000 UFboot samples. The values at the nodes represent the UFboot support, where values below 50 % are not shown. The tree is shown midpoint-rooted with Figtree. Sequences from *Nematostella* are marked bold in green, those from *Strongylocentrotus* in bold and blue. Brachyury target genes in *Nematostella* and *Strongylocentrotus* are indicated by arrows.

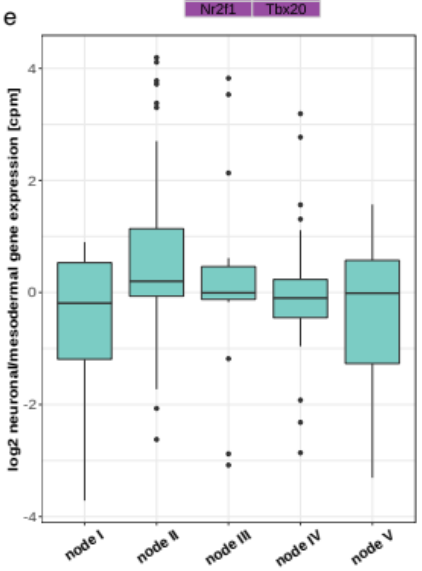


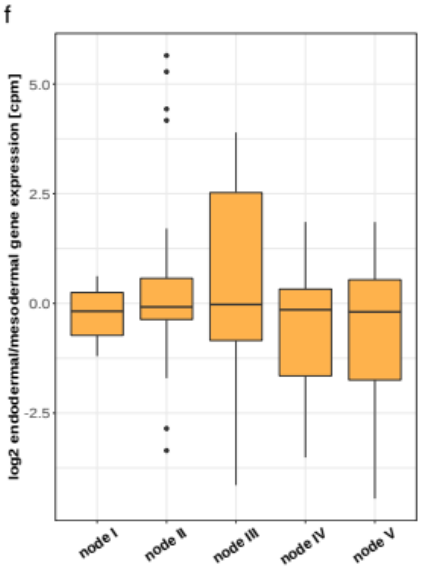
(g) NR2 family phylogeny

NR2f family tree constructed from NR2 genes from *Amphimedon queenslandica* (Aque), *Branchiostoma floridae* (Bflo), *Capitella teleta* (Ctel), *Ciona intestinalis* (Cint), *Drosophila melanogaster* (Dmel), *Homo sapiens* (Hsap), *Mus musculus* (Mmus), *Nematostella vectensis* (Nvec), *Strongylocentrotus purpuratus* (Spur), *Tribolium castaneum* (Tcas), *Xenopus tropicalis* (Xtro). This protein maximum likelihood tree was constructed using IQ-Tree with 10000 UFboot samples. The values at the nodes represent the UFboot support, where values below 50 % are not shown. The tree is shown midpoint-rooted with Figtree. Sequences from *Nematostella* are marked bold in green, those from *Strongylocentrotus* in bold and blue. Brachyury target genes in *Nematostella* and *Strongylocentrotus* are indicated by arrows.


a
eukaryotes I
eumetazoans II
deuterostomes III
chordates IV
vertebrates V
C. owczarzaki
N. vectensis
S. purpuratus
C. intestinalis
X. tropicalis
M. musculus
Node I
T Rnf113a1
Rfx3 Rfx3
Rnf113a1 Zc3h15
Zc3h15 Cluh
Plin2 Plin2
Cluh T
Node II
Foxa1 Sox2
Foxa2 Six3
Foxa3 Sox1
Six3 Rfx4
Hes1 Lhx2
Tgif1 Sox3
Isl1 Nr2f1
Sox2 Foxb1
Rfx6 Pax7
Grhl1 Dmbx1
Fgd1 Foxa2
Dmbx1 Tgif1
Tbx3 Tfap2b
Sox3 Sox11
Rere Hes1
Myrf Foxa3
Sox11 Mycn
Msx2 Gbx1
Sox1 Fgd1
Tfap2b Tfap4
Zfp787 Foxa1
Rfx4 Meis2
Smad5 Pbx3
Tlk1 Prdm5
Tfap4 Zfp787
Arid2 Arid2
Foxp2 Otp
Hmg20a Hes2
Gbx1 Rfx6
Prdm5 Rere
Foxn2 Smad5
Hes2 Zc3h12a
Pax7 Grhl1
Otp Tlk1
Lhx2 Myrf
Mxd1 Elmsan1
Meis2 Hmg20a
Tbx2 Mxd1
Tbx20 Tbx2
Zc3h12a Foxn2
Foxb1 Foxp2
Elmsan1 Ebf3
Gata2 Msx2
Pbx3 Mbnl1
Mycn Tbx20
Ebf3 Tbx3
Lhx1 Gata2
Mbnl1 Lhx1
Nr2f1 Isl1
Node III
Mnx1 Pou3f1
Otx2 Otx2
Irf6 Meis3
Isx Ascl1
Pou3f1 Mta1
Evx1 Neurog3
Pdx1 Pdx1
Meis3 Irf6
Mta1 Isx
Nhlh2 Foxn3
Neurog3 Nhlh2
Ascl1 Mnx1
Foxn3 Dach1
Dach1 Ets1
Ets1 Evx1
Foxc2 Foxc2
Node IV
Pax9 Nkx6-1
Rarb Pax3
Six1 Foxd3
Skil Npas3
Hmbox1 Rarb
Hif1a Nr1h4
Smad2 Trp63
Pax3 Hif1a
Nr1h4 Hoxb13
Hoxb13 Npas4
Nkx6-1 Pbx1
Trp63 Smad2
Foxd3 Nfia
Npas4 Pax9
Nfia Skil
Cdx1 Hmbox1
Pbx1 Lmx1a
Npas3 Tbx18
Tbx18 Six1
Lmx1a Cdx1
Tbx6 Tbx6
Node V
Tcf7l2 Irx3
Zfhx3 Skida1
Foxp4 Tcf7l1
Irx1 Tcf7l2
Cdx2 Irx1
Tcf7l1 Zfhx4
Ets2 Nr6a1
E2f7 Pknox2
Smad6 Zhx2
Irx3 Zeb2
Baz2b Foxp4
Nr6a1 Hoxa1
Ncoa1 E2f7
Skida1 Zfhx3
Zhx2 Baz2b
Zfhx4 Ebf1
Msgn1 Ncoa1
Pknox2 Msgn1
Hoxa1 Plxna2
Hes7 Uncx
Uncx Lef1
Plxna2 Ets2
Pitx2 Smad6
Lef1 Hoxb1
Ebf1 Mesp2
Zeb2 Mef2c
Mesp2 Twist1
Hoxb1 Cdx2
Tcf21 Pitx2
Twist1 Tcf21
Mef2c Mesp1
Mesp1 Hes7
Expression data from Pijuan-Sala et. al 2019
logFC endodermal/mesodermal
2
1
0
-1
-2
logFC neuronal/mesodermal
2
1
0
-1
-2
b
log2 endodermal/mesodermal gene expression [cpm]
5.0
2.5
0.0
-2.5
node I
node II
node III
node IV
node V
c
log2 neuronal/mesodermal gene expression [cpm]
2.5
0.0
-2.5
node I
node II
node III
node IV
node V


d
eukaryotes I
eumetazoans II
deuterostomes III
chordates IV
vertebrates V
C. owczarzaki
N. vectensis
S. purpuratus
C. intestinalis
X. tropicalis
M. musculus
Node I
Rfx3 Rnf113a1
Rnf113a1 Rfx3
Plin2 Zc3h15
Zc3h15 Cluh
T Plin2
Cluh T
Node II
Foxa1 Tfap2b
Foxa3 Six3
Foxa2 Rfx4
Rfx6 Sox1
Tgif1 Sox2
Myrf Lhx2
Sox2 Sox3
Fgd1 Pax7
Grhl1 Foxa2
Tbx3 Nr2f1
Rere Dmbx1
Sox11 Foxb1
Mxd1 Gbx1
Zc3h12a Lhx1
Hes1 Mycn
Isl1 Sox11
Tlk1 Foxa1
Hmg20a Fgd1
Msx2 Pbx3
Tfap2b Meis2
Arid2 Tfap4
Smad5 Prdm5
Hes2 Tgif1
Otp Rere
Sox3 Zfp787
Elmsan1 Hmg20a
Dmbx1 Hes1
Tfap4 Msx2
Prdm5 Arid2
Gbx1 Smad5
Sox1 Tbx3
Foxn2 Otp
Zfp787 Hes2
Six3 Elmsan1
Mycn Zc3h12a
Meis2 Tlk1
Pax7 Rfx6
Rfx4 Mxd1
Lhx2 Foxa3
Ebf3 Grhl1
Tbx2 Foxn2
Foxp2 Myrf
Pbx3 Tbx2
Gata2 Ebf3
Mbnl1 Foxp2
Tbx20 Mbnl1
Lhx1 Gata2
Foxb1 Isl1
Nr2f1 Tbx20
Node III
Irf6 Otx2
Pdx1 Pou3f1
Isx Ascl1
Neurog3 Meis3
Mnx1 Mta1
Ascl1 Nhlh2
Meis3 Neurog3
Mta1 Isx
Otx2 Ets1
Nhlh2 Dach1
Foxn3 Foxn3
Pou3f1 Pdx1
Evx1 Irf6
Dach1 Mnx1
Ets1 Evx1
Foxc2 Foxc2
Node IV
Pax9 Nkx6-1
Rarb Foxd3
Hif1a Pax3
Hmbox1 Npas3
Nr1h4 Rarb
Npas3 Nr1h4
Smad2 Smad2
Hoxb13 Trp63
Pbx1 Hoxb13
Skil Pbx1
Foxd3 Hif1a
Trp63 Npas4
Npas4 Hmbox1
Nkx6-1 Lmx1a
Nfia Nfia
Cdx1 Cdx1
Pax3 Skil
Six1 Pax9
Lmx1a Six1
Tbx18 Tbx18
Tbx6 Tbx6
Node V
Tcf7l2 Irx1
Zfhx3 Irx3
Tcf7l1 Zfhx4
Foxp4 Skida1
Pitx2 Nr6a1
Cdx2 Tcf7l2
Irx1 Zeb2
Ets2 Pknox2
Smad6 Zfhx3
Baz2b Zhx2
Zhx2 Tcf7l1
Skida1 Plxna2
E2f7 Foxp4
Ncoa1 Hoxa1
Nr6a1 E2f7
Hoxa1 Baz2b
Msgn1 Twist1
Pknox2 Msgn1
Tcf21 Ncoa1
Lef1 Hoxb1
Plxna2 Mef2c
Mesp1 Lef1
Hes7 Smad6
Mesp2 Tcf21
Zfhx4 Mesp2
Irx3 Mesp1
Zeb2 Ets2
Uncx Pitx2
Hoxb1 Ebf1
Twist1 Uncx
Ebf1 Cdx2
Mef2c Hes7
Expression data from Grosswendt et. al 2020
logFC endodermal/mesodermal
2
1
0
-1
-2
logFC neuronal/mesodermal
2
1
0
-1
-2


e
log2 neuronal/mesodermal gene expression [cpm]
4
2
0
-2
-4
node I
node II
node III
node IV
node V


f
log2 endodermal/mesodermal gene expression [cpm]
5.0
2.5
0.0
-2.5
node I
node II
node III
node IV
node V

Supplementary Figure 9: Expression of apomorphic and synapomorphic target genes of Brachyury in mouse E8.5 neuronal, endodermal, and mesodermal cell types.

This is the same analysis as shown in Figure 6, except that the single cell gene expression dataset from Grosswendt et al. was used. The expression of the target genes in neuronal versus mesodermal cell types was annotated using single cell RNAseq data from E8.5 mouse embryos (read counts per gene/cell from GEO accession GSE122187). To annotate cells as neuronal, endodermal, or mesodermal, we used the information from Supplementary Table 2 of Grosswendt et al. For endodermal, we used Lineage=Eendo, for mesodermal we used Lineage=Emeso, and for neuronal we used the following cell states: 1,11,24, and 39, which, according to Supplementary Figure 1i of Grosswendt et al corresponds to: neural ectoderm anterior, neural ectoderm posterior, fore/midbrain, and future spinal cord. A) Heatmap of the log2 fold change (logFC) of neuronal (turquoise) or endodermal (yellow) vs mesodermal (purple) gene expression for each gene using the corresponding mouse gene symbols as they appeared in the single cell dataset in each node (see Methods for details). B) Boxplot of the log2 fold change (y-axis) of endodermal vs mesodermal gene expression for all genes per node (x-axis). Note that when, as is the case in this dataset, "gut" is the only annotated endodermal cell type, node 2 is not more enriched in endodermally expressed genes compared to node 5 (p-value = 0.23, Wilcoxon Rank Sum Test). C) Boxplot of the log2 fold change (y-axis) of neuronal vs mesodermal gene expression for all genes per node (x-axis). Note that node 2 is enriched in neuronal expression while node 5 is enriched in mesodermal expression (p-value = 0.037, Wilcoxon Rank Sum Test).

3.3 Discussion Chapter 2

3.3.1 Brachyury Mediated Gene Regulation

The importance of mesoderm in animal evolution cannot be overstated. Diploblastic animals, animals that lack mesoderm, develop open digestive system without body cavity while as triploblastic bilateral animals develop body cavity, coelom (Minot, 1890). Internal organs can grow within this coelom without support from body wall and are protected by fluid cushion. Therefore, by spurring the ability to have internal organs along with its connective tissue derivatives, mesoderm makes it possible to have a complex Bauplan for all bilateral animals. Therefore, it is pivotal to understand the complexity of mesoderm gene regulatory network. One of the most crucial regulators of mesoderm and its derivatives identified so far is the transcription factor Brachyury. It is shown to be a main mesodermal marker in vertebrate mesodermal development (Faial et al., 2015; Gentsch et al., 2013; Gouti et al., 2014, 2017; Koch et al., 2017; Lolas et al., 2014).
Beyond vertebrates, *brachyury* is found in almost all animal lineages and in filastereans and some fungi (Sebé-Pedrós et al., 2018). The role of *brachyury* outside vertebrates is not restricted to mesoderm and its derivatives, in fact, it is often excluded from mesoderm. The diversity of function and spatial expression of *brachyury* outside vertebrate makes it necessary to understand the role of *brachyury* in basal animal lineages. Therefore, to understand the evolutionary change in the role Brachyury, we compared the genome-wide targets of Brachyury in six species, namely the protist *C. owczarzaki*, the sea anemone *N. vectensis*, the sea urchin *S. purpuratus*, tunicate *C. intestinalis* the amphibian *X. tropicalis*, and the mammal *M. musculus*.

To study gene regulation of a transcription factor, it is a prerequisite that we first identify its binding sites. There are many computational tools that can predict putative binding sites of known transcription factors in the genome (Jayaram et al., 2016). However, it is known that in mammals there are many more binding sites than there are actual binding events. To capture true binding sites, Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) has become a standard method. With our ChIP-seq analysis (in-house dataset for *N. vectensis, S. purpuratus* and public datasets for *M. musculus, X. tropicalis*) we found that Brachyury targets developmental genes in all studied animals, indicating that it is located high in the cascade of developmental differentiation. We also found that the distribution of

Brachyury binding sites correlates with the genome size of the species. For instance, in smaller genomes like *N. vectensis*, about half of the binding sites are within 1kb of TSS and no binding sites beyond 100kb. By comparison, in the 10x larger genome of *M. musculus*, about 25-30% of the binding sites are farther than 100kb and less than 10% within 1kb of TSS. Notably about 1/3 Brachyury peaks identified in *N. vectensis* also overlapped with previously identified enhancers (Schwaiger et al., 2014). These enhancers in *N. vectensis* were defined using combination of p300, H3K27ac, and H3K4me1. This suggests that many of the previously identified enhancers are occupied by Brachyury at the gastrula stage.

One striking observation is, among the considerable number of Brachyury target genes, some are upregulated while some are downregulated upon knockdown, in line with recent finding in vertebrates (Koch et al., 2017; Lolas et al., 2014) that Brachyury can act as an activator as well as a repressor. However, many of the target genes show little to no effect on their gene expression. This observation points towards more subtle and synergetic mode of action of Brachyury, which require input from other regulators to affect expression of many of its target genes.

Since the Brachyury protein can form a dimer (Kispert and Herrmann, 1993), it was expected that the binding motif of Brachyury is palindromic. However, this palindromic motif was first identified by a Selex system without the knowledge of the in-vivo Brachyury binding sites. In our study, we found considerable number of both palindromic as well as non-palindromic binding sites of Brachyury, showing not only the dimer can act in vivo, but monomer can also bind at the binding site.
An additional aspect of the transcription factor binding motif analysis was the identification of other transcription factor binding sites that are significantly enriched in the vicinity of Brachyury binding site. An enhancer/promoter region normally contain binding sites for multiple transcription factors. Therefore, we identified enriched motifs within fragment length calculated by Peakzilla (Bardet et al., 2013) around the Brachyury binding. We found large number of binding motifs from important developmental transcription factor families like HMG-domain, Forkhead-domain, Homeo-domain, etc. which occur at a specific distance from the centrally located Brachyury binding site. Future work will show whether these transcription factors may have a co-operative function and may even act as cofactors of Brachyury.

3.3.2 Brachyury Mode of Action

Although Brachyury binding is sequence specific, our results indicate that the target recognition seems more complex with involvement of additional mechanism along with sequence motif. And as mentioned above we find motifs from many other motif families in the vicinity of the brachyury binding motifs. This observation suggests several mechanisms by which Brachyury could be affecting its target gene expression.

1. Brachyury could be interacting with other transcription factors in a context-dependent manner and together with this transcription factor provides stability to the binding of both or one of the transcription factor bindings. In previous studies it has been shown that Brachyury and another member of T-Box family Eomes act by controlling the accessibility of the chromatin of its target mesodermal genes (Tosic et al., 2019).
2. Brachyury facilitates acetylation at H3K27 at its target genes and thereby facilitate / inhibits binding of another transcription factor. In this regard, it has been shown that Brachyury controls H3K27ac in cooperation with histone modifying enzymes, e.g., with acetylation enzyme P300 (Beisaw et al., 2018).
3. Brachyury binding could facilitate H3K4 methylation and affect chromatin access for another transcription factor. Contrarily Brachyury binding may be facilitated by other transcription factor by providing open chromatin.

A combination of these mechanisms could explain how brachyury is associated with activation of some genes while repression of others.

3.3.3 Brachyury Targets in Vertebrates and its Significance

To compare the gene regulatory network between diploblastic and triploblastic species it is very essential to first summarize currently known Brachyury mediated gene regulation in vertebrates.

- After the activation of *brachyury* by Wnt/FGF signaling, *brachyury* first creates necessary conditions in vertebrates for the mesoderm development by controlling the retinoic acid levels. In zebrafish, Brachyury is known to activate *cyp26a1* gene which is involved in RA degradation (Martin and Kimelman, 2010). Low levels of RA are crucial for the mesodermal development of zebrafish. Therefore, activation of *cyp26a1* by Brachyury leads to more degradation of RA, thus creating necessary conditions for mesodermal development.

- Furthermore, Brachyury activates *wnt11,* which promotes the morphogenetic movements in chicken development via the Wnt/PCP pathway (Hardy et al., 2008).
- Brachyury gets additional support from another T-Box family member *tbx6*, which is also a target of Brachyury in mouse and zebrafish and ciona. Tbx6 promotes the mesodermal fate from NMP state and is crucial for paraxial mesoderm formation (Gilchrist et al., 2009; Koch et al., 2017; Kugler et al., 2010). In the absence of *tbx6*, formation of ectopic neural tubes in mouse model has been shown (Chapman and Papaioannou, 1998).
- *Tbx6*, along with Brachyury is essential to repress *sox2* to favor mesodermal fate of the cells (Bouldin et al., 2015; Nowotschin et al., 2012; Ruvinsky et al., 1998; Takemoto et al., 2011).
- *Tbx6* targets the signaling molecule Delta (*dll1)* which activates Notch signaling (Jung et al., 2011). A function of Delta-Notch signaling is shown to play a vital role in germ layer segregation during gastrulation of *Xenopus laevis* (Favarolo and López, 2018; Revinski et al., 2010).
- Another function of *tbx6* is to activate the bHLH transcription factor gene *mesp1* in mouse which is involved in cardiac development (Sadahiro et al., 2018).
- *Tbx6* also activates *msgn1*, which is involved in muscle development, which in turn activates RA-synthesizing enzyme gene *aldh1a2*, thus increasing the levels of RA. In mouse models, embryos devoid of *aldh1a2* and RA have problems in axis elongation and are thus truncated. RA signaling is also shown to be involved in neural differentiation, patterning of neurons and axon outgrowth. Brachyury via *cyp26a1* regulates RA (Martin and Kimelman, 2010), however, the role of RA at the stage of NMPs remains obscure (Gouti et al., 2017).

3.3.4 Brachyury Targets in Diploblasts

Comparing the Brachyury mediated gene regulation in diploblastic and triploblastic animals can help us to speculate about the evolutionary change to a crucial mesodermal determinant from diploblastic to triploblastic animals (Table 1). In the absence of mesodermal inputs, the neural fate is the default state of the tissue (Muñoz-Sanjuán and Brivanlou, 2002; Tosic et al., 2019; Tropepe et al., 2001). Notably, many of the Brachyury targets such as *Tbx6*, *Mesp1*, *Msgn1* in vertebrate promoting mesodermal fate, are either lacking an ortholog in *N. vectensis* or are not Brachyury targets in our dataset (e.g., *twist1*). Surprisingly, while most of

these crucial mesodermal genes are not targets in invertebrates and diploblasts, in these lineages Brachyury instead targets numerous genes which are identifies as neural fate markers in vertebrates. Additionally, genes that would regulate the RA levels in vertebrates are either absent in the genome or not targets of Brachyury in *N. vectensis*. Although there is a debate about the existence of the RA pathway in *N. vectensis,* we did find one putative member of the RA pathway, *RXR-alpha-B*-like as a target of Brachyury in *N. vectensis*. Without the RA receptors, the RA pathway cannot function in *N. vectensis* as it does in vertebrates. Finally*, Tbx6*, one of the most crucial Brachyury targets acting as a key mesodermal determinant in vertebrates, is also absent from diploblastic and even most bilaterian genomes and indeed appears to be a chordate innovation that occurred by gene duplication from an ancestor gene of *tbx2/3, tbx4/5* and *tbx6*.

These differences in the gene regulatory networks help us understand the key evolutionary steps that led to the evolution of Brachyury as a key mesodermal determinant in vertebrates. Indeed, *msgn1* and *mesp1* which are involved in paraxial mesoderm and cardiac development, respectively, have evolved in vertebrates by gene duplication from an ancestor bHLH gene. Even *tbx6* is chordate specific and does not have an ortholog in non-chordates.

The ancestral role of Brachyury.

Our analysis finds that numerous Wnt ligands along with four frizzled receptors, dishevelled, ß-catenin are direct Brachyury targets in the cnidarian model organism *N. vectensis*. Functional studies indicate the existence of a regulatory feedback loop involving Wnt pathway (regular as well as putative PCP) in *N. vectensis*. Existence of a similar feedback loop in sea urchin (Croce et al., 2011; Cui et al., 2014; Sethi et al., 2012) and vertebrates has been reported before. Wnt genes are known to establish primary body axis in vertebrates, echinoderms (and hemichordates) and cnidarians (Lebedeva et al., 2021; Niehrs, 2010) and are also signal to *brachyury* expression, therefore we postulate that *brachyury* forms a part of ancestral feedback loop involving *foxA* and *wnt* genes. Apart from primary body axis establishment this ancestral feedback loop also has additional function in varied species. In *N. vectensis* this feedback loop also suppresses neural marker genes. In sea urchin it also activates endodermal genes and anterior neural fate. Interestingly, in vertebrates it suppresses neural fate genes, activates mesodermal genes while also still playing role in body axis establishment. In summary, our work suggests a scenario that the axis patterning feedback loop of Brachyury and Wnt signaling, together with several other transcription factors (e.g.,

FoxA, FoxB) is the ancestral function of Brachyury in animals, maintained in slightly modified way in several lineages, while the acquisition of several important target genes (e.g., *tbx6, mesp, msgn, twist*) in the vertebrates led to a key shift in function towards its role in mesoderm formation.

Bibliography

Adamska, M., Degnan, S.M., Green, K.M., Adamski, M., Craigie, A., Larroux, C., and Degnan, B.M. (2007). Wnt and TGF-β expression in the sponge Amphimedon queenslandica and the origin of metazoan embryonic patterning. PLoS One *2*, e1031. https://doi.org/10.1371/journal.pone.0001031.

Arendt, D., Technau, U., and Wittbrodt, J. (2001). Evolution of the bilaterian larval foregut. Nature *409*, 81–85. https://doi.org/10.1038/35051075.

Bacon, C.D., Silvestro, D., Jaramillo, C., Smith, B.T., Chakrabarty, P., and Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proc Natl Acad Sci U S A *112*, 6110–6115. https://doi.org/10.1073/pnas.1423853112.

Bagaeva, T.S., Kupaeva, D.M., Vetrova, A.A., Kosevich, I.A., Kraus, Y.A., and Kremnyov, S. V (2019). cWnt signaling modulation results in a change of the colony architecture in a hydrozoan. Dev Biol *456*, 145–153. https://doi.org/10.1016/j.ydbio.2019.08.019.

Bardet, A.F., Steinmann, J., Bafna, S., Knoblich, J.A., Zeitlinger, J., and Stark, A. (2013). Identification of transcription factor binding sites from ChIP-seq data at high resolution. Bioinformatics *29*, 2705–2713. https://doi.org/10.1093/bioinformatics/btt470.

Beddington, R.S., Rashbass, P., and Wilson, V. (1992). Brachyury--a gene affecting mouse gastrulation and early organogenesis. Dev Suppl *165*, 157–165. .

Beheregaray, L.B., Cooke, G.M., Chao, N.L., and Landguth, E.L. (2015). Ecological speciation in the tropics: Insights from comparative genetic studies in Amazonia. Front Genet *5*. https://doi.org/10.3389/fgene.2014.00477.

Beisaw, A., Tsaytler, P., Koch, F., Schmitz, S.U., Melissari, M., Senft, A.D., Wittler, L., Pennimpede, T., Macura, K., Herrmann, B.G., et al. (2018). BRACHYURY directs histone acetylation to target loci during mesoderm development. EMBO Rep *19*, 118–134. https://doi.org/10.15252/embr.201744201.

Berlocher, S.H., and Feder, J.L. (2002). Sympatric Speciation in Phytophagous Insects: Moving Beyond Controversy? Annu Rev Entomol *47*, 773–815. https://doi.org/10.1146/annurev.ento.47.091201.145312.

Bielen, H., Oberleitner, S., Marcellini, S., Gee, L., Lemaire, P., Bode, H.R., Rupp, R., and Technau, U. (2007). Divergent functions of two ancient Hydra Brachyury paralogues suggest specific roles for their C-terminal domains in tissue fate induction. Development *134*, 4187–4197. https://doi.org/10.1242/dev.010173.

Bouldin, C.M., Manning, A.J., Peng, Y.-H., Farr, G.H., Hung, K.L., Dong, A., and Kimelman, D. (2015). Wnt signaling and tbx16 form a bistable switch to commit bipotential progenitors to mesoderm. Development *142*, 2499–2507. https://doi.org/10.1242/dev.124024.

Bowen, B.W., Rocha, L.A., Toonen, R.J., and Karl, S.A. (2013). The origins of tropical marine biodiversity. Trends Ecol Evol *28*, 359–366. https://doi.org/10.1016/j.tree.2013.01.018.

Brauckmann, S. (2012). Karl Ernst von Baer (1792-1876) and evolution. International Journal of Developmental Biology *56*, 653–660. https://doi.org/10.1387/ijdb.120018sb.

Burton, P.M. (2008). Insights from diploblasts; the evolution of mesoderm and muscle. J Exp Zool B Mol Dev Evol *310*, 5–14. https://doi.org/10.1002/jez.b.21150.

Chalamalasetty, R.B., Dunty, W.C., Biris, K.K., Ajima, R., Iacovino, M., Beisaw, A., Feigenbaum, L., Chapman, D.L., Yoon, J.K., Kyba, M., et al. (2011). The Wnt3a/β-catenin target gene Mesogenin1 controls the segmentation clock by activating a Notch signalling program. Nat Commun *2*. https://doi.org/10.1038/ncomms1381.

Chapman, D.L., and Papaioannou, V.E. (1998). Three neural tubes in mouse embryos with mutations in T-box gene Tbx6. Nature *391*, 695–697. https://doi.org/10.1038/35624.

Chesley, P., and Dunn, L.C. (1936). The Inheritance of Taillessness (Anury) in the House Mouse. Genetics *21*, 525–536. https://doi.org/10.1093/genetics/21.5.525.
Conlon, F.L., Sedgwick, S.G., Weston, K.M., and Smith, J.C. (1996). Inhibition of Xbra transcription activation causes defects in mesodermal patterning and reveals autoregulation of Xbra in dorsal mesoderm. Development *122*, 2427–2435. https://doi.org/10.1242/DEV.122.8.2427.
Croce, J., Range, R., Wu, S.Y., Miranda, E., Lhomond, G., Peng, J.C.F., Lepage, T., and McClay, D.R. (2011). Wnt6 activates endoderm in the sea urchin gene regulatory network. Development *138*, 3297–3306. https://doi.org/10.1242/dev.058792.
Cui, M., Siriwon, N., Li, E., Davidson, E.H., and Peter, I.S. (2014). Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo. Proc Natl Acad Sci U S A *111*, E5029–E5038. https://doi.org/10.1073/pnas.1419141111.
Daly, M. (2005). A systematic revision of Edwardsiidae (Cnidaria, Anthozoa). Invertebrate Biology *121*, 212–225. https://doi.org/10.1111/j.1744-7410.2002.tb00061.x.
Daly, M., Brugler, M.R., Cartwright, P., Collins, A.G., Dawson, M.N., Fautin, D.G., France, S.C., Mcfadden, C.S., Opresko, D.M., Rodriguez, E., et al. (2007). The phylum Cnidaria: A review of phylogenetic patterns and diversity 300 years after Linnaeus*. Zootaxa *1668*, 127–182. .
Daly, M., Rack, F., and Zook, R. (2013). Edwardsiella andrillae, a new species of sea anemone from Antarctic ice. PLoS One *8*. https://doi.org/10.1371/journal.pone.0083476.
Darling, J.A., Reitzel, A.R., Burton, P.M., Mazza, M.E., Ryan, J.F., Sullivan, J.C., and Finnerty, J.R. (2005). Rising starlet: The starlet sea anemone, Nematostella vectensis. BioEssays *27*, 211–221. https://doi.org/10.1002/bies.20181.
Davies, J.A. (2013). Invagination and Evagination: The Making and Shaping of Folds and Tubes. Mechanisms of Morphogenesis 217–233. https://doi.org/10.1016/B978-0-12-391062-2.00018-8.
Dieckmann, U., and Doebeli, M. (1999). On the origin of species by sympatric speciation. Nature *400*, 354–357. https://doi.org/10.1038/22521.
Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., et al. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. Nature *452*, 745–749. https://doi.org/10.1038/nature06614.
Faial, T., Bernardo, A.S., Mendjan, S., Diamanti, E., Ortmann, D., Gentsch, G.E., Mascetti, V.L., Trotter, M.W.B., Smith, J.C., and Pedersen, R.A. (2015). Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. Development *142*, 2121–2135. https://doi.org/10.1242/dev.117838.
Fautin, D.G., and Mariscal, R.N. (1991). Cnidaria: Anthozoa. In '"Microscopic Anatomy of Invertebrates"'(FW Harrison and JA Westfall, Eds.), Vol. 2.
Favarolo, M.B., and López, S.L. (2018). Notch signaling in the division of germ layers in bilaterian embryos. Mech Dev *154*, 122–144. https://doi.org/10.1016/j.mod.2018.06.005.
Frank, U., Leitz, T., and Müller, W.A. (2001). The hydroid Hydractinia: A versatile, informative cnidarian representative. BioEssays *23*, 963–971. https://doi.org/10.1002/bies.1137.
Fritzenwanker, J.H., Genikhovich, G., Kraus, Y., and Technau, U. (2007). Early development and axis specification in the sea anemone Nematostella vectensis. Dev Biol *310*, 264–279. https://doi.org/10.1016/j.ydbio.2007.07.029.
Genikhovich, G., and Technau, U. (2009). The starlet sea anemone nematosteila vectensis: An anthozoan model organism for studies in Comparative genomics and functional

evolutionary developmental biology. Cold Spring Harb Protoc *4*, pdb.emo129. https://doi.org/10.1101/pdb.emo129.
Genikhovich, G., and Technau, U. (2017). On the evolution of bilaterality. Development (Cambridge) *144*, 3392–3404. https://doi.org/10.1242/dev.141507.
Genikhovich, G., Fried, P., Prünster, M.M., Schinko, J.B., Gilles, A.F., Fredman, D., Meier, K., Iber, D., and Technau, U. (2015). Axis Patterning by BMPs: Cnidarian Network Reveals Evolutionary Constraints. Cell Rep *10*, 1646–1654. https://doi.org/10.1016/j.celrep.2015.02.035.
Gentsch, G.E., Owens, N.D.L., Martin, S.R., Piccinelli, P., Faial, T., Trotter, M.W.B., Gilchrist, M.J., and Smith, J.C. (2013). InVivo T-Box Transcription Factor Profiling Reveals Joint Regulation of Embryonic Neuromesodermal Bipotency. Cell Rep *4*, 1185–1196. https://doi.org/10.1016/j.celrep.2013.08.012.
Ghabooli, S., Shiganova, T.A., Briski, E., Piraino, S., Fuentes, V., Thibault-Botha, D., Angel, D.L., Cristescu, M.E., and MacIsaac, H.J. (2013). Invasion Pathway of the Ctenophore Mnemiopsis leidyi in the Mediterranean Sea. PLoS One *8*, e81067. https://doi.org/10.1371/JOURNAL.PONE.0081067.
Gianakopoulos, P., Mehta, V., Voronova, A., Y, C., Z, Y., J, C., X, W., MS, W., SJ, T., and IS, S. (2011). MyoD directly up-regulates premyogenic mesoderm factors during induction of skeletal myogenesis in stem cells. J Biol Chem *286*, 2517–2525. https://doi.org/10.1074/JBC.M110.163709.
Gilchrist, M.J., Lachani, K., Wardle, F.C., Morley, R.H., Flicek, P., Smith, J.C., and Keefe, D. (2009). A gene regulatory network directed by zebrafish No tail accounts for its roles in mesoderm formation.
Gluecksohn-Schoenheimer, S. (1944). The Development of Normal and Homozygous Brachy (T/T) Mouse Embryos in the Extraembryonic Coelom of the Chick. Proceedings of the National Academy of Sciences *30*, 134–140. https://doi.org/10.1073/pnas.30.6.134.
Goto, H., Kimmey, S.C., Row, R.H., Matus, D.Q., and Martin, B.L. (2017). FGF and canonical Wnt signaling cooperate to induce paraxial mesoderm from tailbud neuromesodermal progenitors through regulation of a two-step epithelial to mesenchymal transition. Development (Cambridge) *144*, 1412–1421. https://doi.org/10.1242/dev.143578.
Gouti, M., Tsakiridis, A., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. PLoS Biol *12*. https://doi.org/10.1371/journal.pbio.1001937.
Gouti, M., Metzis, V., and Briscoe, J. (2015). The route to spinal cord cell types: A tale of signals and switches. Trends in Genetics *31*, 282–289. https://doi.org/10.1016/j.tig.2015.03.001.
Gouti, M., Delile, J., Stamataki, D., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2017). A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. Dev Cell *41*, 243-261.e7. https://doi.org/10.1016/j.devcel.2017.04.002.
Gross, J.M., and McClay, D.R. (2001). The role of Brachyury (T) during gastrulation movements in the sea urchin Lytechinus variegatus. Dev Biol *239*, 132–147. https://doi.org/10.1006/dbio.2001.0426.
Gruneberg, H. (1958). Genetical studies on the skeleton of the mouse. XXIII. The development of brachyury and anury. J Embryol Exp Morphol *6*, 424–443. .
Hand, C., and Uhlinger, K.R. (1992). The culture, sexual and asexual reproduction, and growth of the sea anemone Nematostella vectensis. Biological Bulletin *182*, 169–176. https://doi.org/10.2307/1542110.

Hand, C., and Uhlinger, K.R. (1994). The unique, widely distributed, estuarine sea anemone, Nematostella vectensis Stephenson: A review, new facts, and questions. Estuaries *17*, 501–508. https://doi.org/10.2307/1352679.
Hand, C., and Uhlinger, K.R. (1995). Asexual Reproduction by Transverse Fission and Some Anomalies in the Sea Anemone Nematostella vectensis. Invertebrate Biology *114*, 9. https://doi.org/10.2307/3226948.
Harbison, G.R., Madin, L.P., and Swanberg, N.R. (1978). On the natural history and distribution of oceanic ctenophores. Deep Sea Research *25*, 233–256. https://doi.org/10.1016/0146-6291(78)90590-8.
Hardy, K.M., Garriock, R.J., Yatskievych, T.A., D'Agostino, S.L., Antin, P.B., and Krieg, P.A. (2008). Non-canonical Wnt signaling through Wnt5a/b and a novel Wnt11 gene, Wnt11b, regulates cell migration during avian gastrulation. Dev Biol *320*, 391–401. https://doi.org/10.1016/j.ydbio.2008.05.546.
Hays, G.C., Doyle, T.K., and Houghton, J.D.R. (2018). A Paradigm Shift in the Trophic Importance of Jellyfish? Trends Ecol Evol *33*, 874–884. https://doi.org/10.1016/j.tree.2018.09.001.
Herrmann, B.G., Labeit, S., Poustka, A., King, T.R., and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. Nature *343*, 617–622. https://doi.org/10.1038/343617a0.
Hilbish, T.J. (1996). Population genetics of marine species: The interaction of natural selection and historically differentiated populations. J Exp Mar Biol Ecol *200*, 67–83. https://doi.org/10.1016/S0022-0981(96)02645-7.
Izumi, T., and Fujita, T. (2018). Description of three new species of Scolanthus (Cnidaria, anthozoa, actiniaria, edwardsiidae): First records of the genus from Japan. Zookeys *2018*, 1–21. https://doi.org/10.3897/zookeys.794.25243.
Izumi, T., Ise, Y., Yanagi, K., Shibata, D., and Ueshima, R. (2018). First Detailed Record of Symbiosis Between a Sea Anemone and Homoscleromorph Sponge, With a Description of Tempuractis rinkai gen. et sp. nov. (Cnidaria: Anthozoa: Actiniaria: Edwardsiidae) . Zoolog Sci *35*, 188–198. https://doi.org/10.2108/zs170042.
Jaramillo, C., Montes, C., Cardona, A., Silvestro, D., Antonelli, A., and Bacon, C.D. (2017). Comment (1) on "Formation of the Isthmus of Panama" by O'Dea et al. Sci Adv *3*, e1602321. https://doi.org/10.1126/sciadv.1602321.
Jayaram, N., Usvyat, D., and R. Martin, A.C. (2016). Evaluating tools for transcription factor binding site prediction. BMC Bioinformatics *17*, 547. https://doi.org/10.1186/s12859-016-1298-9.
Jung, J., Mo, J.S., Kim, M.Y., Ann, E.J., Yoon, J.H., and Park, H.S. (2011). Regulation of notch1 signaling by delta-like ligand 1 intracellular domain through physical interaction. Mol Cells *32*, 161–165. https://doi.org/10.1007/s10059-011-1046-y.
Kayal, E., Roure, B., Philippe, H., Collins, A.G., and Lavrov, D. V. (2013). Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evol Biol *13*, 5. https://doi.org/10.1186/1471-2148-13-5.
Kiecker, C., Bates, • Thomas, and Bell, E. (2015). Molecular specification of germ layers in vertebrate embryos. Cellular and Molecular Life Sciences *73*. https://doi.org/10.1007/s00018-015-2092-y.
King, T., Beddington, R.S.P., and Brown, N.A. (1998). The role of the brachyury gene in heart development and left-right specification in the mouse. Mech Dev *79*, 29–37. https://doi.org/10.1016/S0925-4773(98)00166-X.
Kispert, A., and Herrmann, B.G. (1993). The Brachyury gene encodes a novel DNA binding protein. EMBO J *12*, 3211–3220. https://doi.org/10.1002/j.1460-2075.1993.tb05990.x.

Kispert, A., Herrmann, B.G., Leptin, M., and Reuter, R. (1994). Homologs of the mouse Brachyury gene are involved in the specification of posterior terminal structures in Drosophila, Tribolium, and Locusta. Genes Dev *8*, 2137–2150. https://doi.org/10.1101/GAD.8.18.2137.
Kitajima, S., Takagi, A., Inoue, T., and Saga, Y. (2000). MesP1 and MesP2 are essential for the development of cardiac mesoderm. Development *127*, 3215–3226. .
Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K., and Herrmann, B.G. (2017). Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. Dev Cell *42*, 514-526.e7. https://doi.org/10.1016/j.devcel.2017.07.021.
Korschelt, E., and Heider, K. (1890). Lehrbuch der vergleichenden Entwicklungsgeschichte der wirbellosen Thiere.
Kugler, J.E., Gazdoiu, S., Oda-Ishii, I., Passamaneck, Y.J., Erives, A.J., and di Gregorio, A. (2010). Temporal regulation of the muscle gene cascade by Macho1 and Tbx6 transcription factors in Ciona intestinalis. J Cell Sci *123*, 2453–2463. https://doi.org/10.1242/JCS.066910.
Kuhn, K., Streit, B., and Schierwater, B. (1996). Homeobox genes in the cnidarian Eleutheria dichotoma: Evolutionary implications for the origin of Antennapedia-class (HOM/Hox) genes. Mol Phylogenet Evol *6*, 30–38. https://doi.org/10.1006/mpev.1996.0055.
Kusch, T., and Reuter, R. (1999). Functions for Drosophila brachyenteron and forkhead in mesoderm specification and cell signalling. Development *126*, 3991–4003. .
Lankester, E.R. (1877). Memoirs: Notes on the Embryology and Classification of the Animal Kingdom: comprising a Revision of Speculations relative to the Origin and Significance of the Germ-layers. J Cell Sci *s2-17*. .
Lapébie, P., Ruggiero, A., Barreau, C., Chevalier, S., Chang, P., Dru, P., Houliston, E., and Momose, T. (2014). Differential Responses to Wnt and PCP Disruption Predict Expression and Developmental Function of Conserved and Novel Genes in a Cnidarian. PLoS Genet *10*, e1004590. https://doi.org/10.1371/journal.pgen.1004590.
Lebedeva, T., Aman, A.J., Graf, T., Niedermoser, I., Zimmermann, B., Kraus, Y., Schatka, M., Demilly, A., Technau, U., and Genikhovich, G. (2021). Cnidarian-bilaterian comparison reveals the ancestral regulatory logic of the β-catenin dependent axial patterning. Nat Commun *12*, 4032. https://doi.org/10.1038/s41467-021-24346-8.
Leininger, S., Adamski, M., Bergum, B., Guder, C., Liu, J., Laplante, M., Bråte, J., Hoffmann, F., Fortunato, S., Jordal, S., et al. (2014). ARTICLE Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. https://doi.org/10.1038/ncomms4905.
Leys, S.P., and Degnan, B.M. (2002). Embryogenesis and metamorphosis in a haplosclerid demosponge: gastrulation and transdifferentiation of larval ciliated cells to choanocytes. Invertebrate Biology *121*, 171–189. https://doi.org/10.1111/J.1744-7410.2002.TB00058.X.
Li, M., Yamada, S., Shi, A., Singh, R.D., Rolland, T.J., Jeon, R., Lopez, N., Shelerud, L., Terzic, A., and Behfar, A. (2021). Brachyury engineers cardiac repair competent stem cells. Stem Cells Transl Med *10*, 385–397. https://doi.org/10.1002/sctm.20-0193.
Lolas, M., Valenzuela, P.D.T., Tjian, R., and Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. Proceedings of the National Academy of Sciences *111*, 4478–4483. https://doi.org/10.1073/pnas.1402612111.
Marcellini, S., Technau, U., Smith, J.C., and Lemaire, P. (2003). Evolution of Brachyury proteins: Identification of a novel regulatory domain conserved within Bilateria. Dev Biol *260*, 352–361. https://doi.org/10.1016/S0012-1606(03)00244-6.
Martin, B.L., and Kimelman, D. (2010). Brachyury establishes the embryonic mesodermal progenitor niche. Genes Dev *24*, 2778–2783. https://doi.org/10.1101/gad.1962910.

Martindale, M.Q., Pang, K., and Finnerty, J.R. (2004). Investigating the origins of triplosblasty: "Mesodermal" gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). Development *131*, 2463–2474. https://doi.org/10.1242/dev.01119.
Martinelli, C., and Spring, J. (2003). Distinct expression patterns of the two T-box homologues Brachyury and Tbx2/3 in the placozoan Trichoplax adhaerens. Dev Genes Evol *213*, 492–499. https://doi.org/10.1007/s00427-003-0353-5.
Mbodj, A., Gustafson, E.H., Ciglar, L., Junion, G., Gonzalez, A., Girardot, C., Perrin, L., Furlong, E.E.M., and Thieffry, D. (2016). Qualitative Dynamical Modelling Can Formally Explain Mesoderm Specification and Predict Novel Developmental Phenotypes. PLoS Comput Biol *12*, e1005073. https://doi.org/10.1371/JOURNAL.PCBI.1005073.
McGuire, J.A., Witt, C.C., Remsen, J. V., Corl, A., Rabosky, D.L., Altshuler, D.L., and Dudley, R. (2014). Molecular phylogenetics and the diversification of hummingbirds. Current Biology *24*, 910–916. https://doi.org/10.1016/j.cub.2014.03.016.
Metzis, V., Steinhauser, S., Pakanavicius, E., Gouti, M., Stamataki, D., Ivanovitch, K., Watson, T., Rayon, T., Mousavy Gharavy, S.N., Lovell-Badge, R., et al. (2018). Nervous System Regionalization Entails Axial Allocation before Neural Differentiation. Cell *175*, 1105-1118.e17. https://doi.org/10.1016/j.cell.2018.09.040.
Miller, D.J., and Ball, E.E. (2000). The coral Acropora: what it can contribute to our knowledge of metazoan evolution and the evolution of developmental processes (John Wiley & Sons, Inc).
Minot, C.-S. (1890). The Mesoderm and the Coelom of Vertebrates.
Mugele, D., Moulding, D., Savery, D., Mole, M., Greene, N., Martinez-Barbera, J.P., and Copp, A. (2018). Genetic approaches in mice demonstrate that neuro-mesodermal progenitors express T/Brachyury but not Sox2. BioRxiv 503854. https://doi.org/10.1101/503854.
Muñoz-Sanjuán, I., and Brivanlou, A.H. (2002). Neural induction, the default model and embryonic stem cells. Nat Rev Neurosci *3*, 271–280. https://doi.org/10.1038/nrn786.
Nakanishi, N., Sogabe, S., and Degnan, B.M. (2014). Evolutionary origin of gastrulation: Insights from sponge development. BMC Biol *12*, 26. https://doi.org/10.1186/1741-7007-12-26.
Niehrs, C. (2010). On growth and form: A Cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. Development *137*, 845–857. https://doi.org/10.1242/dev.039651.
Nosil, P., Crespi, B.J., and Sandoval, C.P. (2002). Host-plant adaptation drives the parallel evolution of reproductive isolation. Nature *417*, 440–443. https://doi.org/10.1038/417440a.
Nowotschin, S., Ferrer-Vaquer, A., Concepcion, D., Papaioannou, V.E., and Hadjantonakis, A.K. (2012). Interaction of Wnt3a, Msgn1 and Tbx6 in neural versus paraxial mesoderm lineage commitment and paraxial mesoderm differentiation in the mouse embryo. Dev Biol *367*, 1–14. https://doi.org/10.1016/j.ydbio.2012.04.012.
O'Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., De Queiroz, A., Farris, D.W., Norris, R.D., et al. (2016). Formation of the Isthmus of Panama. Sci Adv *2*, e1600883. https://doi.org/10.1126/sciadv.1600883.
Orr, M.R., and Smith, T.B. (1998). Ecology and speciation. Trends Ecol Evol *13*, 502–506. https://doi.org/10.1016/S0169-5347(98)01511-0.
Panousopoulou, E., Hobbs, C., Mason, I., Green, J.B.A., and Formstone, C.J. (2016). Epiboly generates the epidermal basal monolayer and spreads the nascent mammalian skin to enclose the embryonic body. J Cell Sci *129*, 1915–1927. https://doi.org/10.1242/JCS.180703/VIDEO-3.
Powell, K.I., Chase, J.M., Knight, T.M., and Johnsen, S.I. and Taugbøl, T. (2010). NOBANIS - Invasive Alien Species Fact Sheet. Am J Bot *98*, 1–11. .

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. v., et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science (1979) *317*, 86–94. https://doi.org/10.1126/science.1139158.
de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences *102*, 6600–6607. https://doi.org/10.1073/pnas.0502030102.
De Queiroz, K. (2007). Species Concepts and Species Delimitation. Syst Biol *56*, 879–886. https://doi.org/10.1080/10635150701701083.
Reeves, W.M., Shimai, K., Winkley, K.M., and Veeman, M.T. (2021). Brachyury controls Ciona notochord fate as part of a feed-forward network. Development (Cambridge) *148*. https://doi.org/10.1242/DEV.195230/223262.
Reitzel, A.M., Daly, M., Sullivan, J.C., and Finnerty, J.R. (2009). Comparative anatomy and histology of developmental and parasitic stages in the life cycle of the lined sea anemone Edwardsiella lineata. J Parasitol *95*, 100–112. https://doi.org/10.1645/GE-1623.1.
Rembold, M., Ciglar, L., Yáñez-Cuna, J.O., Zinzen, R.P., Girardot, C., Jain, A., Welte, M.A., Stark, A., Leptin, M., and Furlong, E.E.M. (2014). A conserved role for Snail as a potentiator of active transcription. Genes Dev *28*, 167–181. https://doi.org/10.1101/GAD.230953.113.
Rentzsch, F., Anton, R., Saina, M., Hammerschmidt, M., Holstein, T.W., and Technau, U. (2006). Asymmetric expression of the BMP antagonists chordin and gremlin in the sea anemone Nematostella vectensis: Implications for the evolution of axial patterning. Dev Biol *296*, 375–387. https://doi.org/10.1016/j.ydbio.2006.06.003.
Revinski, D.R., Paganelli, A.R., Carrasco, A.E., and López, S.L. (2010). Delta-Notch signaling is involved in the segregation of the three germ layers in Xenopus laevis. Dev Biol *339*, 477–492. https://doi.org/10.1016/j.ydbio.2010.01.010.
Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A.A., Farkas, T.E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S.R., et al. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. Nat Ecol Evol *1*. https://doi.org/10.1038/s41559-017-0082.
Ronco, F., Matschiner, M., Böhne, A., Boila, A., Büscher, H.H., El Taher, A., Indermaur, A., Malinsky, M., Ricci, V., Kahmen, A., et al. (2021). Drivers and dynamics of a massive adaptive radiation in cichlid fishes. Nature *589*, 76–81. https://doi.org/10.1038/s41586-020-2930-4.
Ruvinsky, I., Silver, L.M., and Ho, R.K. (1998). Characterization of the zebrafish tbx16 gene and evolution of the vertebrate T-box family. Dev Genes Evol *208*, 94–99. https://doi.org/10.1007/s004270050158.
Sadahiro, T., Isomi, M., Muraoka, N., Kojima, H., Haginiwa, S., Kurotsu, S., Tamura, F., Tani, H., Tohyama, S., Fujita, J., et al. (2018). Tbx6 Induces Nascent Mesoderm from Pluripotent Stem Cells and Temporally Controls Cardiac versus Somite Lineage Diversification. Cell Stem Cell *23*, 382-395.e5. https://doi.org/10.1016/j.stem.2018.07.001.
Saina, M., Genikhovich, G., Renfer, E., and Technau, U. (2009). BMPs and Chordin regulate patterning of the directive axis in a sea anemone. Proc Natl Acad Sci U S A *106*, 18592–18597. https://doi.org/10.1073/pnas.0900151106.
Satou, Y., Imai, K.S., and Satoh, N. (2001). Action of morpholinos in Ciona embryos. Genesis *30*, 103–106. https://doi.org/10.1002/GENE.1040.
Scholz, C.B., and Technau, U. (2003). The ancestral role of Brachyury : expression of NemBra1 in the basal cnidarian Nematostella vectensis (Anthozoa). Dev Genes Evol 563–570. https://doi.org/10.1007/s00427-002-0272-x.
Schulte-Merker, S., and Smith, J.C. (1995). Mesoderm formation in response to Brachyury requires FGF signalling. Curr Biol *5*, 62–67. https://doi.org/10.1016/S0960-9822(95)00017-0.

Schulte-Merker, S., Van Eeden, F.J.M., Halpern, M.E., Kimmel, C.B., and Nüsslein-Volhard, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. Development *120*, 1009–1015. .
Schwaiger, M., Schönauer, A., Rendeiro, A.F., Pribitzer, C., Schauer, A., Gilles, A.F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. Genome Res *24*, 639–650. https://doi.org/10.1101/gr.162529.113.
Sebé-Pedrós, A., and Ruiz-Trillo, I. (2017). Evolution and Classification of the T-Box Transcription Factor Family. Curr Top Dev Biol *122*, 1–26. https://doi.org/10.1016/bs.ctdb.2016.06.004.
Sebe-Pedros, A., Ariza-Cosano, A., Weirauch, M.T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T.R., Gomez-Skarmeta, J.L., et al. (2013). Early evolution of the T-box transcription factor family. Proceedings of the National Academy of Sciences *110*, 16050–16055. https://doi.org/10.1073/pnas.1309748110.
Sebé-Pedrós, A., Chomsky, E., Pang, K., Lara-Astiaso, D., Gaiti, F., Mukamel, Z., Amit, I., Hejnol, A., Degnan, B.M., and Tanay, A. (2018). Early metazoan cell type diversity and the evolution of multicellular gene regulation. Nat Ecol Evol *2*, 1176–1188. https://doi.org/10.1038/s41559-018-0575-6.
Seipel, K., and Schmid, V. (2004). Mesodermal anatomies in cnidarian polyps and medusae. International Journal of Developmental Biology *50*, 589–599. https://doi.org/10.1387/IJDB.062150KS.
Seipel, K., and Schmid, V. (2005). Evolution of striated muscle: Jellyfish and the origin of triploblasty. Dev Biol *282*, 14–26. https://doi.org/10.1016/j.ydbio.2005.03.032.
Selander, E., Møller, L.F., Sundberg, P., and Tiselius, P. (2010). Parasitic anemone infects the invasive ctenophore Mnemiopsis leidyi in the North East Atlantic. Biol Invasions *12*, 1003–1009. https://doi.org/10.1007/s10530-009-9552-y.
Sethi, A.J., Wikramanayake, R.M., Angerer, R.C., Range, R.C., and Angerer, L.M. (2012). Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos. Science (1979) *335*, 590–593. https://doi.org/10.1126/science.1212867.
Shiganova, T., Mirzoyan, Z., Studenikina, E., Volovik, S., Siokou-Frangou, I., Zervoudaki, S., Christou, E., Skirta, A., and Dumont, H. (2001). Population development of the invader ctenophore Mnemiopsis leidyi , in the Black Sea and in other seas of the Mediterranean basin. Marine Biology 2001 139:3 *139*, 431–445. https://doi.org/10.1007/S002270100554.
Singer, J.B., Harbecke, R., Kusch, T., Reuter, R., and Lengyel, J.A. (1996). Drosophila brachyenteron regulates gene activity and morphogenesis in the gut. Development *122*, 3707–3718. .
Smith, J.C., Price, B.M.J., Green, J.B.A., Weigel, D., and Herrmann, B.G. (1991). Expression of a xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. Cell *67*, 79–87. https://doi.org/10.1016/0092-8674(91)90573-H.
Stake, J.L., and Sammarco, P.W. (2003). Effects of pressure on swimming behavior in planula larvae of the coral Porites astreoides (Cnidaria, Scleractinia). J Exp Mar Biol Ecol *288*, 181–201. https://doi.org/10.1016/S0022-0981(03)00018-2.
Stampar, S.N., Morandini, A.C., Branco, L.C., Silveira, F.L. da, and Migotto, A.E. (2015). Drifting in the oceans: Isarachnanthus nocturnus (Cnidaria, Ceriantharia, Arachnactidae), an anthozoan with an extended planktonic stage. Marine Biology 2015 162:11 *162*, 2161–2169. https://doi.org/10.1007/S00227-015-2747-0.
Steele, R.E. (2002). Developmental signaling in Hydra: What does it take to build a “simple” animal? Dev Biol *248*, 199–219. https://doi.org/10.1006/dbio.2002.0744.
Stefanik, D.J., Lubinski, T.J., Granger, B.R., Byrd, A.L., Reitzel, A.M., DeFilippo, L., Lorenc, A., and Finnerty, J.R. (2014). Production of a reference transcriptome and

transcriptomic database (EdwardsiellaBase) for the lined sea anemone, Edwardsiella lineata, a parasitic cnidarian. BMC Genomics *15*, 71. https://doi.org/10.1186/1471-2164-15-71.
Steinmetz, P.R.H., Aman, A., Kraus, J.E.M., and Technau, U. (2017). Gut-like ectodermal tissue in a sea anemone challenges germ layer homology. Nat Ecol Evol *1*, 1535–1542. https://doi.org/10.1038/s41559-017-0285-5.
Stemple, D.L. (2005). Structure and function of the notochord: An essential organ for chordate development. Development *132*, 2503–2512. https://doi.org/10.1242/dev.01812.
Stephenson, T.A. (1935). The British sea anemones. by Stephenson, T. A.
Takahashi, H., Hotta, K., Erives, A., di Gregorio, A., Zeller, R.W., Levine, M., and Satoh, N. (1999). Brachyury downstream notochord differentiation in the ascidian embryo. Genes Dev *13*, 1519–1523. https://doi.org/10.1101/gad.13.12.1519.
Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D.M., Lovell-Badge, R., Papaioannou, V.E., and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. Nature *470*, 394–398. https://doi.org/10.1038/nature09729.
Technau, U., and Bode, H.R. (1999). HyBra1, a Brachyury homologue, acts during head formation in Hydra. Development *126*, 999–1010. .
Technau, U., and Genikhovich, G. (2018). Evolution: Directives from Sea Anemone Hox Genes. Current Biology *28*, R1303–R1305. https://doi.org/10.1016/j.cub.2018.09.040.
Thiebot, J.B., Ito, K., Raclot, T., Poupart, T., Kato, A., Ropert-Coudert, Y., and Takahashi, A. (2016). On the significance of Antarctic jellyfish as food for Adélie penguins, as revealed by video loggers. Mar Biol *163*. https://doi.org/10.1007/s00227-016-2890-2.
Tosic, J., Kim, G.J., Pavlovic, M., Schröder, C.M., Mersiowsky, S.L., Barg, M., Hofherr, A., Probst, S., Köttgen, M., Hein, L., et al. (2019). Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. Nat Cell Biol *21*, 1518–1531. https://doi.org/10.1038/s41556-019-0423-1.
Trewavas, E. (1947). Speciation in Cichlid Fishes of East African Lakes [18]. Nature *160*, 96–97. https://doi.org/10.1038/160096b0.
Tropepe, V., Hitoshi, S., Sirard, C., Mak, T.W., Rossant, J., and Van Der Kooy, D. (2001). Direct neural fate specification from embryonic stem cells: A primitive mammalian neural stem cell stage acquired through a default mechanism. Neuron *30*, 65–78. https://doi.org/10.1016/S0896-6273(01)00263-X.
Tsakiridis, A., and Wilson, V. (2015). Assessing the bipotency of in vitro-derived neuromesodermal progenitors. F1000Res *4*. https://doi.org/10.12688/f1000research.6345.2.
Wake, D.B., and Yanev, K.P. (1986). GEOGRAPHIC VARIATION IN ALLOZYMES IN A "RING SPECIES," THE PLETHODONTID SALAMANDER *ENSATINA ESCHSCHOLTZII* OF WESTERN NORTH AMERICA. Evolution (N Y) *40*, 702–715. https://doi.org/10.1111/j.1558-5646.1986.tb00532.x.
Wilczyński, B., and Furlong, E.E.M. (2010). Dynamic CRM occupancy reflects a temporal map of developmental progression. Mol Syst Biol *6*, 383. https://doi.org/10.1038/MSB.2010.35.
Wilkins, J.S. (2011). Philosophically speaking, how many species concepts are there?
Wilkinson, D.G., Bhatt, S., and Herrmann, B.G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. Nature *343*, 657–659. https://doi.org/10.1038/343657a0.
Wolf, M., Chen, S., Song, J., Ankenbrand, M., and Müller, T. (2013). Compensatory Base Changes in ITS2 Secondary Structures Correlate with the Biological Species Concept Despite Intragenomic Variability in ITS2 Sequences - A Proof of Concept. PLoS One *8*, 66726. https://doi.org/10.1371/journal.pone.0066726.
Yamada, A., Martindale, M.Q., Fukui, A., and Tochinai, S. (2010). Highly conserved functions of the Brachyury gene on morphogenetic movements: Insight from the early-

diverging phylum Ctenophora. Dev Biol *339*, 212–222. https://doi.org/10.1016/j.ydbio.2009.12.019.
Yasuoka, Y., Shinzato, C., and Satoh, N. (2016). The Mesoderm-Forming Gene brachyury Regulates Ectoderm-Endoderm Demarcation in the Coral Acropora digitifera. Current Biology *26*, 2885–2892. https://doi.org/10.1016/j.cub.2016.08.011.
Zaffos, A., Finnegan, S., and Peters, S.E. (2017). Plate tectonic regulation of global marine animal diversity. Proc Natl Acad Sci U S A *114*, 5653–5658. https://doi.org/10.1073/PNAS.1702297114/-/DCSUPPLEMENTAL.
Zapata, F., Goetz, F.E., Smith, S.A., Howison, M., Siebert, S., Church, S.H., Sanders, S.M., Ames, C.L., McFadden, C.S., France, S.C., et al. (2015). Phylogenomic analyses support traditional relationships within Cnidaria. PLoS One *10*. https://doi.org/10.1371/journal.pone.0139068.
Zimmermann, B., Robb, S.M., Genikhovich, G., Fropf, W.J., Weilguny, L., He, S., Chen, S., Lovegrove-Walsh, J., Hill, E.M., Ragkousi, K., et al. (2021). Sea anemone genomes reveal ancestral metazoan chromosomal macrosynteny. BioRxiv https://doi.org/10.1101/2020.10.30.359448.
Zinzen, R., Girardot, C., J, G., M, B., and EE, F. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature *462*, 65–70. https://doi.org/10.1038/NATURE08531.