Maternal gene evolution from a comparative perspective

Ferenc Tibor Kagan

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2022

UNIVERSITY OF BERGEN

Maternal gene evolution from a comparative perspective

Ferenc Tibor Kagan

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 27.09.2022

Year: 2022

Title: Maternal gene evolution from a comparative perspective

Name: Ferenc Tibor Kagan

Print: Skipnes Kommunikasjon / University of Bergen

Contents

Acknowledgements

This thesis would have not seen the daylight without the contribution of many.

Abstract in English

Parental care is a widespread phenomenon across the tree of life. Species use various means of caring for their progeny. One such care is through provision supplied with the unfertilized eggs. Such provisions are sufficient to cover the energetic requirements of early development or even orchestrate them. Regulating said events is achieved through the supply of transcripts and proteins synthesized by the maternal organism and stored in the eggs. These regulators are then exchanged to their zygotic counterparts when transcription is resumed during development in a process known as the maternal to zygotic transition.

Past research efforts had uncovered many intriguing mechanisms underlying this transition. In commonly employed model species the specific factors, their spatial and temporal dynamics are well known. Nevertheless, a gap still remains in lesser studied species. By filling in such gaps a better understanding of the evolutionary processes shaping this group of genes can be achieved. Their biological importance, both from a developmental perspective and from a disease perspective, warrants an in-depth analysis of their evolution.

In my thesis my aim is to contribute to a better understanding of the evolution of maternal genes by identifying and describing the maternal gene repertoire of lesser known invertebrate species. Furthermore, by utilizing a comparative framework I provide an insight into the molecular evolution of maternal genes. My results suggests that maternal genes, despite being characterized with high sequence divergences, show a conservation in their expression profile.

Abstract in Norwegian

Foreldreomsorg er et utbredt fenomen i livets tre. Arter bruker ulike midler på å ta vare på avkommet sitt. En metode er å tilføre proviant til de ubefruktede eggene. Slike tilførte midler dekker kravene til energi gjennom første del av utviklingen og kan til og med orkestrere utviklingen. Regulering oppnås ved å tilføre og lagre transkripsjonsfaktorer og proteiner fra mor-organismen, i eggene. Disse faktorene blir etter en stund byttet med sine zygotiske motstykker, i en prosess som kalles den maternelle til zygotiske overgangen.

Tidligere forskning har avdekket mange spennende mekanismer som ligger til grunn for den maternelle til zygotiske overgangen. I velstuderte modellarter er de spesifikke faktorene, samt deres romlige og tidsmessige dynamikk godt kjent. Men det gjenstår fortsatt kunnskapshull i mindre studerte arter. Ved å fylle disse hullene får man en bedre forståelse av de evolusjonære prosessene som har formet denne gruppen gener. Den biologiske betydningen av maternelle gener både fra et utviklingsperspektiv og et sykdomsperspektiv, rettferdiggjør en grundig analyse av evolusjonen deres.

I oppgaven min er målet å bedre forståelsen av maternelle gener ved å identifisere og beskrive gen-repertoaret i virvelløse dyr der dette ikke har vært undersøkt før. Videre, ved hjelp av et komparativt rammeverk, gir jeg et innblikk i den molekylære utviklingen av maternelle gener. Resultatene tyder på at disse genene, til tross for store endringer på sekvensnivå, viser en bevaring av ekspresjonsprofilene.

1.Introduction

1.1 Maternal genes

1.1.1 A general introduction of oocytes

Metazoan life begins with the fusion of two distinct gametes, a sperm cell which contains genomic information and the molecular machinery required for transporting it to the target cell. A radically different cell, the oocyte, contains genomic information, nutrients fueling development, the molecular machineries required for fertilization and orchestrating early embryonic development. All components are preassembled during the maturation process of the oocyte, a process called oogenesis.

Generally, oocytes can be classified based on different features such as yolk content (alecithal – f.ex. eutherian mammals, microlecithal – f.ex. echinoderms, mesolecithal – f.ex. frogs and macrolecithal – birds, insects) and distribution of the yolk (telolecithal – f.ex. frogs or birds, centrolecithal – f.ex. insects, hololecithal – f.ex. echinoderms and isolecithal – f.ex. amphioxus).
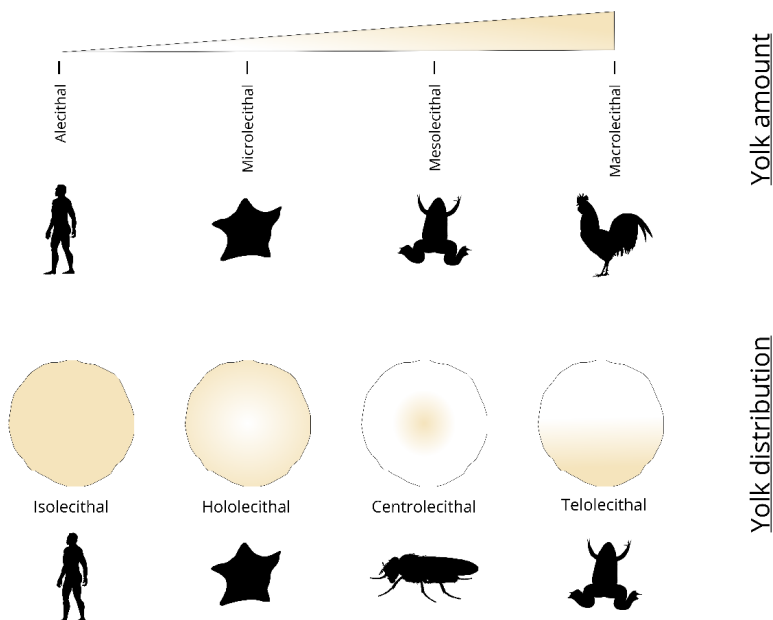
Figure 1. Categories of oocytes based on yolk amount and distribution.

Some oocytes possess an inherent visible asymmetry dividing them roughly into 2 opposite ends: an animal pole and a vegetal pole (AV axis), which is also known as primary axis (Conklin, 1905a, 1905b). According to the definition of Robert G. Edwards and Helen K. Beard this animal-vegetal axis can be identified as: "the first polar body marks the animal pole, and the vegetal pole lies diametrically opposite in the oocyte" (Edwards & Beard, 1997). The asymmetry is sometimes visible in the oocyte morphology, as e.g. in *Drosophila melanogaster* or through the distribution of cytoplasmic components such as the nucleus, pigment granules, yolk granules or mitochondria (e.g. yellow crescent in ascidians - Conklin, 1905b, 1905a). Further asymmetries can be discovered at a molecular level, where RNAs or proteins are unevenly distributed in macromolecular structures in the cytoplasm (Driever et al., 1990).

This setup does not hold completely true to all animals as seen in eutherian mammals. Some evidence was found to support a slight asymmetry in the eggs of this group (f.ex. Dalcq, 1955) where the authors have observed a bilateral symmetry of RNA and an RNA-free vacuolated cytoplasm which was interpreted as a possible dorsal-vegetal axis. Additionally, following maturation, mitochondria in mouse eggs relocate to one hemisphere of the cortex (Calarco, 1995). These results suggest an asymmetry of lesser functional relevance in eutherian mammals, possibly because the embryos are more reliant on the mother organism. To further complicate the picture the asymmetry can be highly stage specific within the process of oogenesis. A well conserved structure in oocytes, the Balbiani body (Figure 2), is only found transiently during the oogenesis in several studied species. It is in fact an amalgamation of different organelles (mostly mitochondria, from where it derives an alternative naming of mitochondrial cloud), cytoplasmic RNAs and proteins (Kloc et al., 2014).

In the past researchers have observed that the AV axis can potentially predict the later polarity of the embryo, most commonly the anterior-posterior axis (Frohnhöfer & Nüsslein-Volhard, 1986) or dorso-ventral axis (Nüsslein-Volhard et al., 1980) but this

can be modified as in for example *Caenorhabditis elegans* (Goldstein et al., 1993) where the sperm entry point alters the predefined axis.
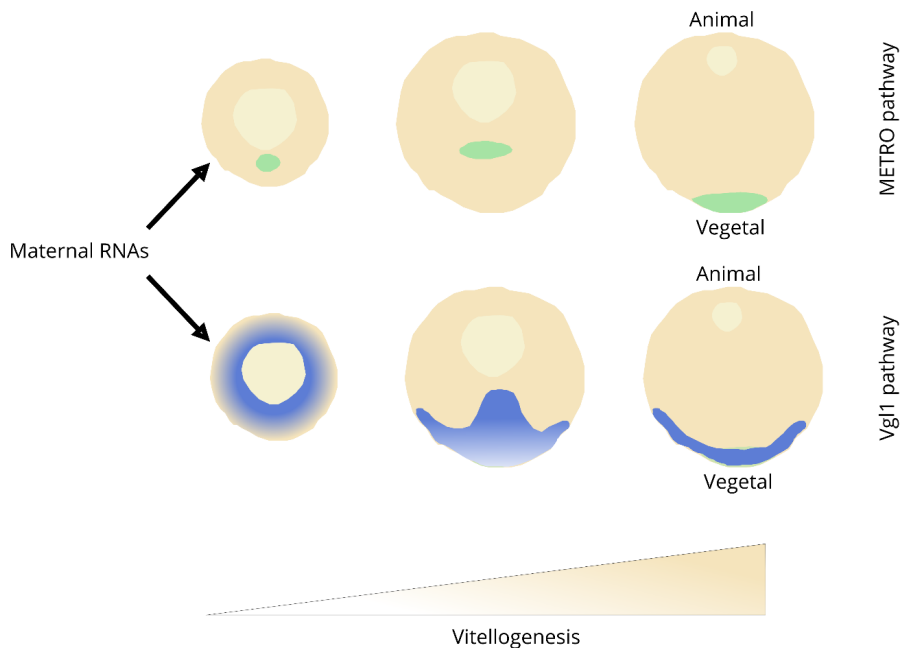
Figure 2. Distribution of maternal RNAs during *Xenopus laevis* oogenesis. Two major pathways have been identified, the METRO pathway and the Vgl1 pathway. Both lead to an asymmetric RNA localization to the vegetal pole of the mature oocytes (adapted from Kloc & Etkin, 1995).

1.1.2 Maternal genes as drivers of early development

The maternal contribution of regulators is insufficient to drive the whole development, instead the zygotic genome will become transcriptionally active and will govern development in conjunction with maternal contribution degradation (Frohnhöfer & Nüsslein-Volhard, 1986; Harvey, 1936; Nüsslein-Volhard, 1977; Nüsslein-Volhard et al., 1980; Nüsslein-Volhard & Wieschaus, 1980). The period of this molecular switch has been termed the maternal-to-zygotic transition (MZT). The MZT has been defined most recently by Nadine L. Vastenhouw, Wen Xi Cao and Howard D. Lipshitz as: "a coordinated series of molecular events, starting with the degradation of maternally deposited transcripts and ending with global activation of the zygotic genome"

(Vastenhouw et al., 2019). During the maternal-to-zygotic transition two major regulatory phases can be separated. In the first phase the maternal genes are regulated on a post-transcriptional (Stoeckius et al., 2014; Thomsen et al., 2010) and post-translational (Krauchunas et al., 2012) level. In a second phase the zygotic genome is activated, whereby the major regulatory events will shift towards a transcriptional one (de Iaco et al., 2017; M. T. Lee et al., 2014).

A demonstration of maternal factor requirement for early development has been performed as early as 1936. Experimentally enucleated sea urchin zygotes go through cleavage cycles despite the lack of a cell nucleus. To explain this, it was suggested that cytoplasmic constituents are responsible for governing the cleavage cycles. These were shown later to be maternally inherited RNAs and (Harvey, 1936). Not to be confused with zygotic genome activation (ZGA) or mid-blastula transition (MBT). ZGA being defined as the onset of *de novo* transcription from the zygotic genome and MBT being a developmental transition characterized by cell cycle lengthening, asynchronous cell cycles and appearance of cell motility in teleost and amphibian early embryogenesis. All three events can overlap regarding timing, hence the confusion around the interchangeability of the terms. To study the timing of the MZT earliest approaches used pharmacological inhibition of *de novo* transcription with the use of α-amanitin. In the presence of this blocking agent the development will halt at the stage from where the transcription of zygotic regulators is necessary for progression (seee.g. Stroband et al., 1992). Initially, to identify key players during the MZT conventional mutagenesis approaches were employed in *Drosophila melanogaster* (Nüsslein-Volhard, 1977; Nüsslein-Volhard et al., 1980; Nüsslein-Volhard & Wieschaus, 1980)**.** Resulting phenotypes were characterized by the female progeny bearing mutations reaching adulthood but being sterile. The loci linked to these phenotypes were homozygous recessive and their lethal phenotype would be noticeable in the offspring generation. Mutations of this kind are more prominently affecting early embryogenetic events. Initial explorations of these mutants yielded 15 genes regulating segmentation of early

Drosophila embryos, which was subdivided into three categories based on the mutant phenotypes: segment polarity genes, pair-rule genes and finally gap genes (Nüsslein-Volhard & Wieschaus, 1980) . Mutagenizing the whole 2nd chromosome of *Drosophila melanogaster* yielded similar results, with 67 loci linked to sterile females with normal egg morphology, some of which was identified from earlier as being a maternal effect gene (Schüpbach & Wieschaus, 1986).

1.1.3 Clearance of maternally provided transcripts

The post-transcriptional regulation of maternally deposited RNA molecules is affecting the half-life of them and their accessibility towards the translational machinery. The maternal genes can vary in their proportion of the total gene pool ranging from about one-third in the mouse *Mus musculus* (Q. T. Wang et al., 2004) or *Caenorhabditis elegans* (Stoeckius et al., 2014) to three-quarters in *Drosophila melanogaster* (Thomsen et al., 2010) or *Strongylocentrotus purpuratus* (Wei et al., 2006). A subset of these loaded transcripts will be first destabilized then degraded during the MZT (Thomsen et al., 2010, Stoeckius et al., 2014). The timing of the MZT is not a punctual event, rather it is characterized as process where the absolute timeframe or relative developmental stage of which is dependent on the studied species. In a well-studied model, *Danio rerio*, the first cohort of gene transcribed from the zygotic genome are detected at around 3.5 hours post fertilization (or mid-blastula stage according to Kimmel et al., 1995). These early transcripts mark the MZT by which point the majority of maternally contributed genes have gone through clearance (Aanes et al., 2011). A stark contrast to this is the nematode *Ascaris suum* where maternal RNAs are already degraded at around the time of fertilization (Wang et al., 2014) Initial clearance of these maternal RNAs is under the control of maternally provided degradation machinery and following zygotic genome activation its degradation is replaced by the newly transcribed zygotically encoded genes. Despite the inhibitory presence of α-amanitin there is a degradation of maternal transcripts detectable in cleavage stage *Danio rerio* embryos (Mishima & Tomari, 2016b). The transcripts which are stable in the presence of a transcriptional inhibitor but degrade in its absence are under the control of zygotically transcribed machinery, a

striking example is the micro-RNA *mir-430* control over the maternal transcripts (Giraldez et al., 2008, Mishima & Tomari, 2016).

Several mechanisms have been found to be involved in regulating the stability of maternal RNAs: RNA binding proteins can both stabilize or destabilize them, small non-coding RNA molecules (acting on their own or in combination with RBPs) destabilize them, so does some RNA modifications and a more obscure mechanism of codon optimality. Control through RBP is a universal process (D'Agostino et al., 2006; Detivaud et al., 2003; Rabani et al., 2017; Ramos et al., 2004).

In several organisms a presence of Y-box proteins has been described, in some of these a direct link between these RBPs and stability of maternal RNAs has been uncovered. Such is the case for the frog *Xenopus laevis* (Matsumoto et al., 1996) or the worm *Caenorhabditis elegans* (Arnold et al. 2014). Conversely purely destabilizing RBPs can act in two known ways: firstly, the RBPs Smaug, BRAT and PUM exert their effect through recruiting a well conserved CCR4-NOT-deadenylase complex in *Drosophila melanogaster* (Semotok et al., 2005, Laver et al., 2015). Secondly the AU-rich element binding proteins (ARE-BP) in *Caenorhabditis elegans* (D'Agostino et al., 2006D'Agostino et al., 2006), *Danio rerio* (Rabani et al., 2017), in conjunction with CUGBP1 destabilize transcripts by recruiting the poly(A) ribonuclease (PARN) (Graindorge et al., 2008).

Apart from RBPs a major regulatory mechanism in most species is through small non-coding RNAs such as miRNAs, endogenous small-interfering RNAs or PIWI interacting RNAs. The *miR-430* miRNA family, as mentioned before, has been reported as regulating maternal transcript degradation in *Danio rerio* (Bazzini et al., 2016) and *Xenopus laevis* (Lund et al., 2009). Analogously in *Drosophila melanogaster* the *miR-309* cluster has been connected to the clearance of a substantial proportion of maternal transcripts (Bushati et al., 2008). Interestingly the *Drosophila melanogaster* clearance has been linked also to the piRNA degradation pathway (Barckmann et al., 2015). Furthermore endo-siRNA presence has been observed in *Caenorhabditis elegans* and in

Mus musculus oocytes where it regulates transcript clearance although in the latter its functional significance remains to be elucidated (Ohnishi et al., 2010, Stoeckius et al., 2014).

The epitranscriptome has also proven to be of relevance to maternal transcript clearance in several species. Specifically, N^6-methylation of adenosine residues (m^6A) has an impact in RBP recognition of the maternal transcripts (Liu et al., 2015, Zhao et al. 2017). The epitranscriptome reader, YTHDF2, is a key player in regulating clearance in several species (Zhao et al., 2017, Ivanova et al., 2017, Deng et al., 2020).

Another layer of RNA level regulation is achieved by codon optimality of maternal transcripts. An optimal codon in this context is one which has high cognate tRNA abundance associated to it. This in turn will affect the translation speed and thus enable selecting for RNAs with slow translation throughput (Presnyak et al., 2015). Such a selective degradation of maternal RNAs has been observed in *Danio rerio* (Bazzini et al., 2016). Apart from codon optimality, translation efficiency is heavily influenced by the poly(A) length of the maternal transcripts (Eichhorn et al., 2016, Subtelny et al., 2014). Naturally the machineries generally responsible for polyadenylation, such as for example the cytoplasmic polyadenylation element binding protein (CPEB) or GLD2 are key players in it. Interestingly the polyadenylation is occurring after transcription in the cytoplasm, furthermore not all maternal transcripts are targeted for polyadenylation (Winata et al., 2018).

Because of regulation at the transcript level, one would expect that signs of this regulation would be reflected in regulatory regions, such as 3' untranslated regions. Indeed, there is evidence suggesting that lengths of 3' UTRs correlate with abundance of cis regulatory motifs (Stark et al., 2005). Conversely genes with emphasized post-transcriptional regulation show longer than average 3' UTR regions (Tadros et al., 2007). Cross-species comparison were made for 3' UTR lengths of maternal genes and it was found that maternal genes have significantly overrepresented longer 3' UTRs compared to genes excluded from the maternal transcriptome (Shen-Orr et al., 2010).

Post-translational regulation has also been proven essential during MZT. This encompasses protein modifications such as phosphorylation or ubiquitination. In turn these will result in downstream effects by altering protein functions or stability. During egg activation in *Drosophila melanogaster* it is estimated that 30% of the proteome is affected by phosphorylation (Krauchunas et al., 2012). Key regulators can also be ubiquitinated and subsequently degraded by proteasomes, as in the case of the OMA-1 and OMA-2 whereby clearance of these proteins releases TAF-4, an initiator protein for zygotic genome activation (Guven-Ozkan et al., 2008).

1.1.4 Proposed functions of the MZT

Clearance of maternal transcripts is a highly complex transition, yet it appears to be present across all metazoan model systems (Vastenhouw et al., 2019) and even found in plants (Baroux et al., 2008; P. Zhao et al., 2022) suggesting an important role during multicellular embryogenesis. The functions of the clearance can be manifold in a context dependent manner. Degrading evenly distributed maternal transcripts could make way for the zygotic counterparts to be expressed in spatially restricted fashion, thereby having a permissive role (Tadros & Lipshitz, 2009). On the other hand, it can have an instructive function whereby clearing the maternal transcripts acts as an internal clock. Reaching a threshold would enable the dividing embryos to time their cell cycle lengthening (Tadros & Lipshitz, 2009). Degradation of transcripts involved strictly in oogenesis (e.g. transcripts coding DNA repair genes) has also been proposed as a role of the MZT (Pan et al., 2005). Another possibility for the function of degradation might be a reprogramming one. Starting from the mature egg, the early embryos need to transition through stages of totipotency, pluripotency and lastly redifferentiation (Lee et al., 2014). A possibility to regulate these cell state transitions while lacking transcription would be one of having all necessary genes present as transcripts and regulating their translation or stability. It is worth noting that the above proposed functions have varying levels of empirical support (de Renzis et al., 2007; Edgar & Datar, 1996; Edgar & O'Farrell, 1990) and should be handled accordingly. Despite the varied suggested roles, we see a clear

evolutionary conservation of the phenomena, the constraint might be present due to some or all the mentioned roles as they are not mutually exclusive.

1.1.5 Maternal effects through maternal genes

Maternal genes exert maternal effects on the progeny, which according to Timothy A. Mousseau and Charles W. Fox can be defined as: "the phenotype of an individual is determined not only by its own genotype and the environmental conditions it experiences during development, but also by the phenotype or environment of its mother" (Mousseau & Fox, 1998, page 5). Maternal effects can range from behavioral ones (such as oviposition determines sex of the progeny in reptiles) to molecular ones (such as the maternal genes). Maternal effects can be observed at three different stages of maternal care: before fertilization, after fertilization and before giving birth or hatching and finally after fertilization and giving birth or hatching. Given that maternal genes have an impact on the fitness of the progeny (see maternal lethal mutations e.g. Kemphues et al., 1988) they exert a maternal effect. Furthermore, as this group of genes is stored in oocytes before fertilization it can be categorized as a prezygotic maternal effect, where the maternal phenotype affects the gamete quality.

While the maternal effect is present, the maternal genotype will override the offspring one. To illustrate this Michael J. Wade described a simple genetic model where a single autosomal locus with two alleles is present in a population with maternal effect in place. In this model the two alleles are *A* and *a*, *A* being dominant over *a*. There are 3 possible maternal genotypes and progeny genotypes: *AA*, *Aa*, *aa*. In a scenario where the offspring is characterized by an *Aa* genotype and the mother having an *AA* genotype, the observable phenotype of the offspring will be one of *AA* (see Table 1

Mother's genotype	Offspring's genotype		
	AA	*Aa*	*aa*
AA	P(AA) G(AA)	P(AA) G(Aa)	P(AA) G(aa)

Aa	P(Aa)	P(Aa)	P(Aa)
	G(AA)	G(Aa)	G(aa)
aa	P(aa)	P(aa)	P(aa)
	G(AA)	G(Aa)	G(aa)

). Even in this very simplistic framework, there will be drastic consequences. More specifically, natural processes of selection and mutations will tend to shift towards higher variance. Differently put, genotypes associated with maternal effects show higher divergences, compared to genes without maternal effects. Theoretical work has even quantified a theoretical magnitude, selection is half as strong on the genes under maternal effects (Mousseau & Fox, 1998, 1st chapter). Apart from the weaker selection, there is a generational uncoupling of the selection, meaning that the population under selection will not respond to selection in the same generation, instead there will be a lag of one generation (Kirkpatrick & Lande, 1989). These theoretical estimates have found empirical support throughout the years. Indeed, genes under this inheritance pattern show high divergence levels as in *Drosophila melanogaster* where genes, such as the maternally expressed *bicoid*, show approximately two fold higher nucleotide diversity compared to its paralog lacking a maternal expression, *zerknüllt* (Demuth & Wade, 2007). With a weaker purifying selection it is also expected to have higher intraspecific variation, an expectation which has found support for maternal gene sequence polymorphisms in *Drosophila melanogaster* populations (Cruickshank & Wade, 2008).

A more recent survey of literature covering genes involved in reproduction, a broader category containing maternal genes as a subset, has uncovered a pattern of high nucleotide divergence between species together with a highly polymorphic sites within species populations. The authors ruled out several alternative hypotheses capable of explaining high divergence of these genes and arrive at the conclusion that for reproductive genes a plausible null hypothesis would that of relaxed selection (Dapper & Wade, 2020). Furthermore, maternal genes show evidence for going through more gene

duplications recently, suggesting genes of an overall younger origin (Antonio Marco, 2017, p 486).

Table 1. Table of genotype relationships among mother and offspring. In each cell P() denotes observable phenotype, G() is representing the underlying genotype. Shaded cells indicate instances impossible in a standard inheritance pattern. Modified after Michael J. Wade 1998.

Mother's genotype	Offspring's genotype		
	AA	*Aa*	*aa*
AA	P(AA) G(AA)	P(AA) G(Aa)	P(AA) G(aa)
Aa	P(Aa) G(AA)	P(Aa) G(Aa)	P(Aa) G(aa)
aa	P(aa) G(AA)	P(aa) G(Aa)	P(aa) G(aa)

1.2 Gene expression evolution

1.2.1 Gene expression as a heritable trait

With the advancement of methodologies, the low-throughput approaches were replaced by high-throughput technologies such as microarrays and RNA sequencing (Tadros et al., 2007, Collart et al., 2014). Due to the scale of high-throughput sequencing, new opportunities for getting a deeper understanding of biology have opened. Among many insights one is of relevance for this introduction, namely the connection between gene expression and phenotype (Nachtomy et al., 2007). Contrary to the classic view of adaptation occurring in the sequences and thereby functionally affecting the resulting protein structure, gene expression is impactful on phenotypic evolution without altering the sequence of the protein itself. Instead, its effect is noticeable by altering the protein abundance, spatial distribution or timing (Harrison et al., 2012). Even with gene expression differences at a single locus, there can be significant phenotypic effects. Beak morphology varies in Darwin's finches according to fluctuations of BMP4 gene expression levels (Arhat et al., 2004) or the blonde phenotype of European people in connection with KITLG expression are well studied examples (Guenther et al., 2014).

1.2.2 Genomic signatures of gene expression level

Through its effects on the phenotype, gene expression can be itself subject to selective forces. Upstream of expression, regulatory steps are encoded in sequences and those can be subjected to mutations. Such mutations can then be fixed in populations if they provide a fitness advantage (Ghanbarian & Hurst, 2015; Signor & Nuzhdin, 2018). Relationships of this sort can be uncovered by finding expression quantitative trait loci (eQTLs) associated with given gene expressions (Majewski & Pastinen, 2011). eQTLs in essence are genomic loci associated with changes in gene expression. A consequence of their sequence information nature is that eQTLs can be passed on to next generations, therefore gene expression variation among individuals can be passed on to the next generations and even fixated if proven advantageous in a given context (Wright et al., 2014)

Phenotypic variability, defined by Wagner and Altenberg as the potential of an organism to vary (Wagner & Altenberg, 1996), is a key property of populations in the light of selection as with higher variability the population can be more robust to environmental fluctuations. Maintaining a high variability is costly from an evolutionary perspective as increased diversity will also introduce disadvantageous phenotypes in each environment. Therefore, a fine balance is maintained in nature between variability and homogeneity of individuals (Willmore et al., 2007). This also holds true to variability in gene expression, where the variability can arise from individual sequence polymorphisms (Hulse & Cai, 2013) or even stochastic processes (Kærn et al., 2005).

1.2.3 Models of trait evolution and their connection with gene expression evolution

As in the case of sequence evolution, where sophisticated models have been proposed to describe their evolutionary dynamics, so does gene expression evolution have proposed models. Generally, two main forces are affecting diversity of gene expression across taxa: genetic drift drives the gene expression to diverge between species and conversely purifying selection pulls expression towards an optimum Hodgins-Davis & Townsend, 2009). Mathematical models formulated for trait evolutions can then be used to describe these processes. The simplest model used for trait evolutions of continuous traits is the

Brownian motion model, where a trait changes through time randomly both in its direction and its magnitude (Cavalli-Sforza & Edwards, 1967). By introducing a novel parameter to this model, the drag towards an optimal value, one can arrive at the Ornstein-Uhlenbeck model (Uhlenbeck & Ornstein, 1930). Here traits are still changing through time around an optimal trait value with departures from this optimum resulting in dragging the trait values back towards it. Apart from these two models, many other have been developed which model different evolutionary processes by modifying covariation matrixes underlying phylogenies or adding new parameters to the two broad models (Freckleton et al., 2002; Pagel, 1997; Pagel, 1999; Rohlfs & Nielsen, 2015).

Most studies focusing on change of gene expression in an evolutionary framework draw conclusions from pairwise comparisons (e.g. Kryuchkova-Mostacci & Robinson-Rechavi, 2016, Levin et al., 2016). First pointed out by Felsenstein in 1985 standard statistical approaches, regressions in his cautionary paper, are problematic as they do not take in account the underlying covariation of traits due to shared ancestry (Felsenstein, 1985). Moreover, he also proposed a solution to the problem in the form of phylogenetic independent contrasts. Under a Brownian motion model then the datapoints associated with the tips can be transformed in a pairwise fashion into independent datapoints. By comparing the trait values at the tips in conjunction with knowing that these values have diverged due to their evolutionary history (represented by their branch lengths from the common node) the resulting standardized contrast is then assigned to the node from which the two tips originate from. This step is repeated iteratively across all the internal nodes. As a result, these contrasts can then be used in standard statistical analyses. Returning to the works of Nadezda Kryuchkova-Mostacci and Marc Robinson-Rechavi or Levin et al. one can correct for the errors arising from pairwise comparisons and find that the results of these papers lose statistical significance (Dunn et al., 2018). Non-independence of traits across a phylogeny has gotten more spotlight with recent works from several research groups. David Brawand et al. has been focusing on gene expression evolution across different organs in a mammalian phylogeny (Brawand et al., 2011),

Kenji Fukushima and David Pollock also studied organ specific gene expression evolution covering a broader range of species within vertebrates (Fukushima & Pollock, 2020). Attention has also been given to exclusively invertebrate gene expression evolution in several studies. In Siphonophores zooid gene expression evolution has been studied by Catriona Munro et al. (Munro et al., 2021), in Drosophiliids ovary specific gene expression evolution has also been investigated by Samuel H. Church et al. (Church et al., 2021). Finally, evolution of maternal gene expression was investigated covering Drosopholiid species, albeit utilizing pairwise comparisons, in the past by Joel Atallah and Susan E. Lott. Their findings suggest that maternal genes which undergo clearance are highly divergent both in expression levels and also in representation, in contrast the maternal genes with zygotic counterpart show a more conserved pattern (Atallah & Lott, 2018).

1.2.4 The hourglass model

Early comparative embryological studies found similarities across a wide range of vertebrate embryos suggesting a connection between evolution and development. A parallel was drawn between development and evolution and the recapitulation model was proposed (Haeckel, 1866; von Baer, 1828). This model predicts that early stages of embryogenesis will develop traits shared by a wide array of lineages and later developmental stages will result in traits shared by specific lineages. The recapitalization model has since been debated and reevaluated. Initial alternative models proposed were inspired by the notion that developmental steps are built in a sequential order and reliant on previous steps (Garstang, 1922; Riedl, 1978). Therefore, it was proposed that earliest developmental steps are the most constrained. Molecular approaches shed a new insight into the relationship between development and evolution and led to the proposal of the hourglass model (Duboule, 1994). This model predicts the most amount of constraint during mid-embryogenetic steps. The early and late developmental steps are therefore divergent in this framework. Transcriptomic studies in the past had found evidence to support the developmental hourglass model in several phyla (Domazet-Lošo & Tautz,

2010; Drost et al., 2015; Gildor et al., 2019; Hu et al., 2017; Irie & Kuratani, 2011; Levin et al., 2012; Simakov et al., 2015). The focus of these studies was centered around vertebrate species with some occasionally focusing on invertebrate species. Some contradictory support also has been found in spiralian species, where there was no observable evidence for the mid-embryogenesis evolutionary constraint (L. Wu et al., 2019)

1.3. Aims

Given their importance in early development, maternal genes are heavily understudied outside standard model systems such as for example *Drosophila melanogaster*. Previous studies also highlighted the overwhelming importance of post-transcriptional regulation for maternal genes and therefore transcript abundance levels are valuable sources of information readily available with high throughput sequencing technologies. With theoretical work laid down by population geneticist suggesting a high evolutionary divergence for the group of genes with maternal inheritance it would be expected to unravel many species-specific attributes. This in turn could provide valuable lessons for evolutionary studies and for species specific biology. There is a clear need for a comparative approach across a wider phylogeny to get a better understanding and as mentioned above, the methodologies capable to deal with phylogenies have been developed, and a great width of datasets is easily accessible (see Material and methods section).

My thesis aims to undertake a comparative analysis of maternal gene expression evolution. After sampling species specific RNA-seq datasets across the metazoan phylogeny it divides maternal genes into two categories using differential gene expression analysis. It characterizes both the functional profiles and gene architecture of these categories separately. Using phylogenetic comparative approaches it then identifies the modes of evolution in the two categories and it contrasts them. It also models the dynamics of gain-loss of expressions and orthogroups for each category. These analyses could provide a rich resource for testable genes in model systems for constraints of expression or divergence of expression.

Finally, three datasets generated for lesser known research organisms are undertaken for a more thorough dissection. Conserved expression is found for *nanos*, novel potential regulators of zygotic genome silencing are identified, and unexpected genes are also identified in the maternal set previously known only as axon guidance regulators.

1.4. Introducing species

Priapulus caudatus is a marine predator, burrowing worm and is one of the about 20 species representing the Priapulida. Species belonging to this group are morphologically characterized with a vermiform morphology, terminal mouth, scalid scales and varying sizes ranging from 1 mm to 40 cm (Shirley et al., 1999). They are dioecious (separate male and female sexes) with only vague sexual dimorphism and reproduce by free-spawning. Oogenesis in this species has been described at an ultrastructural scale (Nørrevang, 1965). Some species of this taxon have available descriptions of their development. In *Priapulus caudatus* the early cleavage is total and equal with the first cleavage (at 9 C^o 15 hours post fertilization) and the second cleavage being meridional to the animal-vegetal axis. The third cleavage will produce an animal hemisphere and a vegetal hemisphere with slightly bigger blastomeres (Wennberg et al., 2008). Gastrulation will commence at the 64 cell stages, 3 days after fertilization with an immigration of future mesodermal and endodermal precursors inside the embryo (Wennberg et al., 2008). Interestingly, despite belonging to the Protostome clade (generally characterized by having the adult mouth forming from the blastopore), they show a deuterostomic (the blastopore will form the anus, while the mouth forming later on on a different site) development (Martín-Durán & Hejnol, 2015). The embryo hatches after 9 days, with the larval stages resembling the adult animal Wennberg et al., 2009). Following hatching the animal will undergo many molting events before it reaches its final size.

Priapulida is one of the three groups within Scalidophora group (with Kinorhyncha and Loricifera), which is positioned basally to Nematoida (comprising of Nematoda and Nematomorpha) and its sister group to Panarthropoda (Howard et al., 2022). These three main lineages together will form the Ecdysozoa clade (Howard et al., 2022). Fossil records from the Early Cambrian for Priapulida are abundant and have a highly similar morphology to present day species. After the Early Cambrian there is a sharp decline in found fossils, possibly due to a massive extinction event, which could also explain the sparse species number of this group in the present (Huang et al., 2004; Vannier et al.,

2010). Molecular evidence also suggests a low evolutionary rate for priapulids as their mitochondrial sequence data show low divergence (Webster et al., 2006). The above information puts this group in the spotlight for evolutionary developmental questions regarding ecdysozoan evolution. Unfortunately, molecular studies covering this group are scarce so far, despite having genomic sequence available. Most studies have evolutionary developmental questions in focus while utilizing *in-situ* hybridization techniques or next generation sequencing (e.g. Martín-Durán & Hejnol, 2015 or Hogvall et al., 2019).

Hypsibius exemplaris is a microscopic invertebrate belonging to the taxon Tardigrada. Along with Onychophora and Arthropoda this clade generally is accepted as part of Panarthropoda, a group nested at a basal position within Ecdysozoa (Howard et al., 2022). Species belonging to this phylum can be broadly characterized by an oval shaped body, which can be broken down into a head segment and four body segments. Their eight unarticulated appendages end in varied number of claws. The head encompasses a tardigrade specific buccopharyngeal structure, which can be used to classify species (Schuster et al., 1980Schuster et al., 1980). Up to date thousands of tardigrade species have been described occupying diverse environments, ranging from marine to limno-terrestrial (Møbjerg et al., 2011). They are most well known for their resilience to extreme conditions in their tun state (Persson et al., 2011). All reproductive modes can be found in tardigrades: dioecious with minor sexual dimorphism, parthenogenetic and hermaphroditic (Bertolani, 2001). *Hypsibius exemplaris* is mostly represented by females reproducing through parthenogenesis, but some males have been found and it has been suggested that due to some unknown factors sexual reproduction can arise in these species (Gabriel et al., 2007). Development in *Hypsibius exemplaris* has been thoroughly described. Adult stage is reached in 4 days, with one to ten eggs of diameter 50–75 μm being laid in the molted cuticle of the mother in order to protect the progeny. Cleavages are total with some cleavage asynchronicity arising as early as 4 cell stage. The blastula stage is lacking a blastocoel and gastrulation starts at 60 cell stage (Gabriel et al., 2007). Information concerning developmental axis formation is scarce, with evidence of a dorso-

ventral axis found via stereotypic nuclear migration in the 4 cell stage (Gabriel et al., 2007). Following hatching the tardigrades are self sufficient with the hatchling having all adult organs already present. Despite having a thorough description of their oogenetic process (Jezierska et al., 2021), molecular players involved in early development are not known.

The third studied species included in this project is *Terebratalia transversa*, a brachiopod. This clade can be characterized as filter feeding animals attached to various surfaces with their pedicles. Generally, their dorsal valves are smaller than the ventral ones, with some exceptions. Shell chemical composition is also variable and so is the connection between the two valves in between some lineages with some having a mineralized hinge, others having muscles or connective tissues bridging the shells (Williams et al., 1996). Great diversity can also be discovered in their feeding and respiratory organ, the lophophore (Rudwick, 1970). Brachiopods generally have two sexes and are free-spawners. In *Terebratalia transversa* the gonad is located in between the coelomic cavity of the double layered mantle epithelium. The oogenetic process has been described in great detail in the work of John Arthur Long (Long, 1964). A few key elements are the synthesis of granules and lipid droplets during early stages of oogenesis located asymmetrically adjacent to the nucleus, as maturation proceeds these will disperse throughout the cytoplasm, the nucleus will take up a central position. Oocytes are also in close contact with a layer of follicle cells. RNA concentration is initially high and decreases with maturation, it was not clarified if this occurs as a consequence of degradation or an artefact of dilution due to increase in oocyte size (Long, 1964).

Paleontological data dates Brachiopoda back to as early Cambrian (Skovsted et al., 2011). No clear consensus has been achieved regarding their relationship with Phoronida, which can be considered as a close sister group or a subgroup of Brachiopoda (Sperling et al., 2011; Cohen, 2013). Molecular data bring Brachiopda close to other spiralian groups such as the annelids, nemerteans, bryozoans or mollusks (Dunn et al., 2008; Hejnol et al., 2009; Mallatt et al., 2012).

Oogenetic processes have been described to all 3 species using histochemical approaches, oogenetic stages have been separated, even RNA content has been followed through for some species throughout this process. Genomic information is also accessible for some species with varying degrees of quality. Despite all of the available resources no information has been found in published literature about maternal genes in these species, given their phylogenetic positions they have the potential to provide valuable lessons about maternal gene evolution in invertebrate species.

2. Material and methods

2.1 Husbandry and collection

Adult *Priapulus caudatus* individuals were collected in Gullmarsfjorden (Fiskebäckskil) in 2018 November by dredging up to 60 meters deep near the coast. The mud was filtered and animals were placed in seawater sampled on the collection spot. As there is no method yet developed for naturally spawning the animals an invasive approach was used. The female gonads were dissected from the animals and placed on top of a mesh (150 μm diameter) in filtered seawater. The oocytes were shaken of the gonads and washed several times. After removing debris from the gonad the eggs were fixed in 4% paraformaldehyde solution (diluted in seawater) overnight on room temperature. Following fixation eggs were washed five times with phosphate buffered saline solution (PBS) and stored at 4 C° until use.

Adult *Terebratalia transversa* individuals were collected by dredging in San Juan Channel, between San Juan Island and Shaw Island (Washington, USA) in January of 2019. Animals were maintained in seawater aquaria at Friday Harbor Laboratories. Here too an invasive approach was used for spawning. Animals were opened and their gonad, which are attached to the inner lining of the valves was stripped. After stripping maceration of the gonads commenced on top of a mesh (150 μm diameter) to filter tissue debris. Eggs were washed repeatedly with seawater until reaching maturation (indicated by germinal vesicle breakdown or GVB and detachment of follicle cells). Male gonads were handled similarly as in *Priapulus caudatus*. Activation of the sperm was achieved by adding buffered seawater of pH 9.8 using Tris. Embryo batches were reared in 1L beakers and at selected developmental stages fixed with 4% paraformaldehyde solution diluted in filtered seawater for 1 hour at room temperature. Following fixation embryos were washed five times using PBS, followed by dehydration with 100% methanol. Samples were stored in -20 C° until use.

Hypsibius exemplaris adult strain was sourced from Sciento in 2018. Cultures were kept in 50 ml – 75 ml Chalkley's media (CM; $CaCl_2$ 0.006 g/L, KCl 0.004 g/L, NaCl 0.1 g/L

in autoclaved ddH$_2$O) at 22 C^o in 12h dark – 12h light cycle. Once a month the medium was changed by filtering animals with a (50 µm) mesh and washing off them with fresh media. Animals were fed once every two weeks with 1 ml *Chlorococcus sp.* algae culture.

2.2 Sampling for RNAseq and cell lysis

Hypsibius exemplaris cultures were monitored for egg laying events. Egg laying is done in conjunction with molting, therefore right after egg laying the adult exits the cuticle leaving behind the eggs in the shed exuvia. At room temperature the first cleavage commences after 2 hours post laying by which time adult have left the exuvia (Gabriel et al., 2007). Before collection surfaces and tools used for collection were cleaned with RNaseZap and filtered tips were used throughout the collection to avoid RNase contamination. Animals where the adult and embryos were present together were selected from the cultures. Staging of embryos was performed under Axioscope Zeiss microscope with differential interference contrast (DIC) settings. Embryos were dissected from the exuvia with Tungsten needles, washed three times in CM. Afterwards there were transferred into 4.4 µl of lysis solution (2.4 µl of 1:19 Triton-X and RNAse inhibitor 40U/µl from MegaScript T7 transcription kit diluted in nuclease free water; 1 µl 10mM each dNTP; 1 µl 5 µM oligo-dT) (Picelli et al., 2014Picelli et al., 2014). Sampled embryos were transferred to PCR strip tubes (one embryo in each tube), labeled and snap frozen in liquid nitrogen. In order to crack the eggshell of the embryos three cycles of snap freezing and thawing were used. Snap freezing was achieved with liquid nitrogen, thawing with a 42 C^o water bath, each step lasted 30 seconds. Following this samples were shipped to GeneCore (EMBL Genomics Core Facilities) where RNA extraction and amplification (18 PCR cycles proved to be the most optimal) was performed according to Smart-Seq2 protocol (Picelli et al., 2014). Sequencing was performed with 75 base pair single end reads on a single lane of a NextSeq machine.

2.3 Cloning of *in situ* hybridization probes

2.3.1 Primer design for probe synthesis

Primer pairs were designed in MacVector 11.1.2. The pairs with longest product without GC bias were selected for amplification PCR from cDNA. The pairs used in this project are listed below.

Table 2. Table containing primer pairs targeting genes of interest with estimated lengths.

Species	Gene name	Forward primer	Reverse primer	Estimated_length
	nanos	TGGTGAGTGTTTACCCGTCTTCC	CGAAAGTCTGGAACGCTATATTGC	467
Priapulus caudatus	*cbx1*	CCTCCTCCTCCATCTTCTTC	GTCTTTCGTCTCTTTCTCATTCATC	524
	emsy	CGCAGGAGAGGATGGTGTATTTG	CGCAGGAGAGGATGGTGTATTTG	1200
	plexinA2	CACCAGAGCCAAACCAGTTACC	GAGTGTTGTCAAAACGCAGTTCG	1329
	plexinB	ACATCTTCTCAACAACACTCTC	AACCACTCAAATCTCCCAAATC	1485
Terebratalia transversa	*roundabout*	TGCTCGTGGTCTGTCTCTGTAACG	GTGGGAGGTTGTATGCTGGAAG	1139
	slit1	AGAACTAACCCAGTCACACC	GACAACTGACAAATCTCCCC	1072
	slit2	TGACGATAAACGCAGCCAG	CGATTTCAGCCCAGCAAAAC	1300
	slit3	GGCTGAACAGTCTTTGCTTTCCC	CTTTGAGGACACTTACACTCCGAAC	1466

2.3.2 RNA extraction

Priapulus caudatus and *Terebratalia transversa* RNA extraction were performed by previous lab members using a standard TRIzol extraction approach. Briefly the steps were: embryo or tissue homogenization with vortexing samples together with 200 µl TRIzol reagent. This was followed up by phase separation with adding 0.1 ml BCP (1-bromo-3-chloro-propane, isopropyl-alcohol 75% ethanol)/ml of samples, then vortexing, letting tubes sit at room temperature for 10 minutes and finally centrifuging for 15 minutes at 4°C with 15115 rcf (relative centrifugal force). After phase separation and centrifugation, the upper clear phase was removed and transferred to a RNase free tube. To this 0.5 ml of RNase free isopropanol was added and stored overnight at -20°C. The next day another round of centrifugation commenced for 8 minutes at 4°C with 13.000 rpm, liquid was carefully removed, without pipetting up the pellet. Pellets were washed with 1 ml 75% RNase free ethanol and centrifuged again for 5 minutes at room temperature with 12.000 rpm. Following this ethanol was removed and the samples were left to air-dry, after which resolubilization followed with 25 µl - 50 µl diethyl-

pyrocarbonate (DEPC) treated water at 55° C for 10 minutes. Yield was measured as a last step on a NanoDrop machine and samples were stored at -80° C until use.

2.3.2 cDNA synthesis

From purified RNA extraction cDNA was synthesized using SuperScript III First-Strand Synthesis System for RT-PCR from Invitrogen, following the description of the provided handbook guide. In 0.2 ml RNase free PCR tubes 10 µl of a solution was added (1 µl of 10 mM dNTP mix with all four nucleotides homogeneously, 1 µl 50 ng/µl hexamere containing solution and finally 8 µl of the purified RNA extractions from different developmental stages and tissues covering all possible transcript profiles). This solution was heated up to 65°C for 5 minutes, followed by an addition of 10 ul in each tube (2 µl 10X reaction buffer from the kit, 4 µl 25 mM $MgCl_2$, 2 µl 0.1 M DTT, 1 µl RNase OUT and 1 µl SuperScript III polymerase). This was followed by thermal cycling steps, 25°C for 10 minutes, 50°C for 50 minutes and 85°C for 5 minutes. The tubes were placed on ice for a minute, after cooling them down they were incubated with 1 µl of RNase H for 20 minutes at 37°C. The resulting cDNA solutions had their concentrations measured with a NanoDrop machine and were diluted to get 300 ng/µl concentrations. All thermal cycling steps were performed in a standard Eppendorf Thermocycler.

2.3.4 PCR amplification from cDNA

Isolation of gene of interest has been done with a PCR setup. Primers targeting the gene of interest were designed ahead of time and ordered from Sigma Aldrich. For each reaction a final volume of 25 µl was used, this contained a Promega supplied 5 µl of 5X reaction buffer, 3 µl of 25 mM $MgCl_2$, 0.5 µl of 10 mM dNTP mix containing all four nucleotides homogeneously, 0.15 µl of Taq-polymerase and 13.35 µl nuclease free water. Additionally, 1 µl of 10 µM of forward and reverse primers each and 1 µl of cDNA template. Reactions were carried out in an Eppendorf Thermocycler where initial denaturation was achieved with 95°C for 5 minutes. Amplification cycles were repeated 35 times and consisted of 95°C for 30 seconds, annealing was done in a primer specific manner for 30 seconds, elongation similarly was performed with a primer specific

temperature for 2 minutes. A final elongation step done, again at a temperature optimal for given primer pair, for 10 minutes. Following this the products were retrieved, 5 µl of product was loaded on 50 ml of 1% agarose gel with TAE buffer and 5 µl of SybrSafe. Using gel electrophoresis products were separated at a constant 90 volts for 45 minutes next to a lane containing 5 µl of GeneRuler 1kb Plus DNA ladder. Afterwards the product sizes were compared to the estimated product sizes from the PCR primer design step.

Purification of the product was done with MiniElute PCR Purification Kit purchased from Qiagen. One volume of PCR product was diluted to five volumes of PB buffer and thoroughly mixed. The mixture was transferred to a MiniElute Column within a collection tube and centrifuged for a minute at 12000 rcf (relative centrifugal force). Flow through was discarded and to the column 750 µl of PE buffer was added followed by another 1 minute centrifugation at 12000 rcf. Flow through was discarded and columns were centrifuged again to discard any residual PE buffer. Following this the columns were transferred into a clean tube and 10 µl of EB buffer was added to them. After incubation for 2 minutes on room temperature these tubes were centrifuged one last time for 1 minute at 12000 rcf.

2.3.5 Ligation and bacterial transformation

PCR products were ligated with pGEM-T Easy Vector System (Promega kit). For the reaction 1.5 µl of product was used, which was mixed with 3.5 µl of 2X Ligation Buffer, 0.5 µl pGEM-T Easy Vector and 0.5 µl of T4 DNA ligase for an hour at room temperature in 0.5 ml Eppendorf tubes.

The ligated product was then used for transformation of competent *Escherichia coli* cells. In a prechilled tube 2 µl of the product was pipetted and mixed with 50 µl of competent cells. The mixture was left to rest on ice for 20 minutes following which a heat-shock commenced in a water bath for 50 seconds at 42°C. The mixture was left to rest on ice for 2 minutes again, after which 450 µl of prewarmed at 37°C ampicillin containing LB

media was added and together they were incubated at 37ºC on 450 rpm in an Eppendorf Thermomixer machine.

LB plates containing ampicillin were prepared ahead of time (500 ml MiliQ purified water, 10 gr of LB powder, 17.5 gr of agar powder was mixed and autoclaved after which 500 µl of ampicillin was added). On each plate 50 µl of IPTG (isopropyl beta-D-1-thiogalactopyranoside) and 50 µl X-Gal (5-Bromo-4-Chloro-3-Indolyl-beta-D-Galactoside) was added and spread with a sterilized spreader. After drying 150 ml of the incubated bacterial culture was plated in each plate. These were then incubated overnight at 37ºC.

2.3.6 Colony selection

Multiple white colonies were picked for further inspections using a filter containing pipette tip from the plates. First a PCR was setup to verify correct product insertion. In conjunction these potentially good colonies were transferred to an ampicillin containing LB media and incubated overnight.

The PCR setup was the following: 15.85 µl nuclease free water, 5 µl of 5X reaction buffer, 3 µl of 25 mM $MgCl_2$, 0.5 µl of 10 mM dNTP mixture containing all four nucleotides, 0.15 µl of Taq polymerase and 0.25 µl of 100 µM T7 and SP6 primers each was mixed and to this was a picked colony containing pipette tip was submerged for a minute. Following this the tip was then transferred into a tube containing 2 ml of ampicillin containing LB liquid media and incubated overnight at 37ºC at 210 rpm. The PCR mixture was then subjected to the following program in an Eppendorf Thermocycler machine: denaturation was achieved with 95ºC for 5 minutes. Amplification cycles were repeated 35 times and consisted of 95ºC for 30 seconds, annealing was done at 50ºC for 30 seconds, elongation was performed at 72º C for 1 minute and 45 seconds. These steps were cycled through 35 times. A final elongation step was done at 72ºC for 10 minutes. Following this the products were retrieved, 10 µl of product was loaded on 50 ml of 1% agarose gel with TAE buffer and 5 µl of SybrSafe. Using gel electrophoresis products were separated at a constant 90 volts for 45 minutes next to a lane containing 5 µl of

GeneRuler 1kb Plus DNA ladder. Afterwards the product sizes were compared to the original estimated product sizes from the PCR primer design step. If a given colony contained the appropriately sized insert in the vector than the following day these were used for miniprep.

2.3.7 Miniprep

For selected colonies miniprep was performed the following day using NucleoSpin Miniprep kit (purchased from Macherey Nagel). The 2 ml cultures were transferred to a 2 ml Eppendorf tube and centrifuged for 30 seconds at 12000 rcf. Supernatant was removed and discarded, the leftover pellet was resuspended with 250 µl of A1 buffer to this 250 µl of A2 buffer was added and mixed. The mixture was left at room temperature for 5 minutes, after which 300 µl of A3 buffer was added and centrifuged for 5 minutes at 12000 rcf. Supernatant was then transferred to a NucleoSpin column inserted in a collection tube and centrifuged again for 1 minute at 12000 rcf. Flow-through was discarded and 600 µl of A4 buffer was added to the column. After another round of centrifugation at 12000 rcf for 2 minutes the columns were transferred into clean 1.5 ml Eppendorf tubes and eluted with 50 µl of Elution buffer after 1 minute of incubation at room temperature.

2.3.8 Linearization and reverse transcription

Inspection of the miniprep was done by using Sanger sequencing at University of Bergen sequencing facility. For this a PCR reaction was prepared according to their description: 1 µl of 5X reaction buffer, 3.2 µl of 1 µM primer,1 µl of BigDye v3.1, 3.8 µl nuclease free water and 1 µl of extracted plasmid DNA. For each plasmid 2 reactions were set up, one with SP6 primers and the other one with T7 primers to not validate sequence identity, but also its orientation for anti-sense probe synthesis. The reaction itself was carried out with the following setup on a Eppendorf Thermocycler machine: after 5 minutes of denaturation at 95ºC amplification cycles were repeated 25 times and consisted of 95ºC for 10 seconds, annealing was done at 50 Cº for 5 seconds, elongation was performed 60ºC for 4 minutes. The final product was diluted with another 10 µl of nuclease free

water and delivered at the sequencing facility. The resulting sequences and their orientation were confirmed with the help of MacVector v18.1.3.

Linearization of the inserts before probe synthesis was performed with a PCR setup repeated two times for each plasmid to increase the final template DNA concentration required for probe synthesis. 10 µl of 5X reaction buffer, 6 µl of 25 mM MgCl$_2$, 0.5 µl of 100 µM T7 or SP6 primer (in accordance with the anti-sense orientation of the insert),1 µl of 10 mM dNTP mix, 0.3 µl of Taq polymerase, 30.7 µl nuclease free water and 1 µl of extracted plasmid DNA. The reaction itself was carried out with the following setup on an Eppendorf Thermocycler machine: after 5 minutes of denaturation at 95°C amplification cycles were repeated 40 times and consisted of 95°C for 30 seconds, annealing was done at 50°C for 30 seconds, elongation was performed 72°C for 1 minute and 45 seconds. After amplification cycles a last elongation was set at 72°C for 10 minutes. The products were confirmed using gel-electrophoresis approach. 5 µl were aliquoted out of the 50 µl volume and separated on 5 µl SybrSafe containing 1% agarose gel with TAE buffer at a constant 95 volts for 45 minutes.

Probe synthesis was done with the use of MEGAscript Kit (purchased from Ambion). To label the probe for colorimetric *in-situ* hybridization digoxigenin (DIG) labelled UTP (purchased from Roche) was used. The synthesis was set up the following way: 0.5 µl of 50 mM A/T/GTP, 1.2 µl of 50 mM UTP, 2.1 µl of 10 mM, 1 µl kit provided enzyme mix and linearized DNA template until 10 µl final volume. This mixture was assembled in a nuclease free 0.5 ml Eppendorf tube and incubated at 37°C for 5 hours, after which 0.5 µl of RNase free DNase I (kit provided) was added to each tube and incubated for another 5 minutes at 37°C. After DNase treatment 5 µl of lithium-chloride precipitation solution was added to each reaction and stored at -20°C overnight to precipitate synthetized RNA. Following day mixtures were thawed and centrifuged at 15600 rcf for 15 minutes at 4 C^o. Supernatant was removed and remainder pellet was washed with 0.5 ml of 70% RNase free ethanol. The mixture was centrifuged again at 7200 rcf for 15 minutes at 4 C^o. As much as possible ethanol was removed without disturbing the pellet. Rest of the ethanol

was air dried. To the dry pellets 30 µl of preheated RNAsecure (Thermofischer) was added and the mixture was incubated at 60 C^o for 10 minutes. Samples were cooled to room temperature and their RNA concentration was measured with a NanoDrop machine. Samples were diluted in hybridization buffer to 50 ng/µl final concentration and stored at -20°C until further use.

2.4 *In-situ* hybridization

2.4.1 Solution used for *in-situ* hybridization

DEPC water was prepared by adding 750 µl of DEPC to 1 liter of distilled water, incubated overnight at 37 C^o and the next day autoclaved. Afterwards it was ready to be used in steps where RNase contamination would be detrimental.

Final volume of hybridization buffer consisted of final proportions of 50% formamide, 5X SSC (diluted from 20X SSC pH 4.5 from Sigma Aldrich), 50 µg/ml heparin, 0.1% Tween-20, 1% sodium-dodecyl-sulfate and 100 µg/ml salmon sperm DNA (Sigma Aldrich) and DEPC treated water for achieving the required concentrations. In the case of hybridization washing solution the heparin and salmon sperm DNA volumes were replaced by DEPC treated water. 20X SSC was composed of final concentrations of 0.3 M of Na-citrate and 3 M of NaCl in DEPC treated water.

Phosphate buffered saline (PBS) solution stocks were prepared with 18.6 mM NaH_2PO_4, 84.1 mM Na_2HPO_4 and 1750 mM NaCl final concentrations. The pH of it was adjusted to 7.4. This stock was then diluted 1:10 to get working concentrations set. For PTw to PBS was added Tween-20 with a final proportion of 0.1% for the detergent. In the case of the PBT blocking solution 0.2% Triton X-100 and 0.1% bovine serum albumin final proportions were mixed into 1X PBS.

Boehringer-Mannheim blocking solution was diluted in maleic acid buffer, which consists of 100 mM maleic acid and 150 mM NaCl final concentrations in distilled water.

For the alkaline phosphatase buffer 100 mM NaCl, 50mM $MgCl_2$, 100 mM Tris of pH 9.5 and 0.5% Tween-20 final concentrations were mixed in DEPC treated water. In the

case of alkaline phosphatase buffer without MgCl$_2$ the MgCl$_2$ volume was replaced by DEPC treated water.

2.4.2 Colorimetric *in-situ* hybridization

In-situ hybridizations (ISH) were performed on paraformaldehyde fixed oocytes in a 24 well dish, where each well was hybridized with one probe. Washes were performed with 500 µl for 5 minutes on a rocker table at room temperature, unless stated otherwise. Change of solutions was done under custom made hoods with use of a stereomicroscope. Prehybridization and hybridization steps were done in rocking incubators with adjustable temperatures.

If fixed samples were dehydrated in methanol, rehydration was achieved by two washes of decreasing methanol concentrations (60% methanol/40% PTw and 30% methanol/70% PTw). After rehydration 4 washes with PTw were done to get rid of residual methanol. Permeabilization of samples was achieved through a 0.01 mg/ml proteinase-K (Ambion) treatment in PTw for 10 minutes at room temperature on a rocker table. To stop digestion two consecutive washes of PTw with 2 mg/ml glycine was used. Following these two washes of PTw with 1% triethanolamine was done. To the second wash 1.5 µl of acetic anhydride was added and gently mixed into solution by swirling the plate manually first and then on the rocker table. After 5 minutes another 3 µl of acetic anhydride was added and swirled manually and rocked on the table for another 5 minutes. This was followed up with two PTw washes and a refixation step with 3.7% paraformaldehyde in PTw for 1 hour at room temperature. Refixed samples were washed five times with PTw to remove residual paraformaldehyde, the last of the washes was done at 80 C^o for 10 minutes. Prehybridization was done in two washes of hybridization buffer, first at room temperature for 10 minutes, second one overnight at 67°C.

The following day probes were diluted to 5 ng/µl in hybridization buffer and were denatured at 80°C for 10 minutes. Blank hybridization buffer was replaced with buffer containing probes in each well and left to be incubated for 3 day at 67°C in the incubators.

After hybridization washes of 900 µl volume were the following in the incubators at 67°C: two washes of blank hybridization buffer, first 10 minutes and second 40 minutes, a decreasing ladder of a mixture of hybridization wash solution and 2X saline-sodium citrate (SSC) buffer of pH 7 for 30 minutes each (75 % hybridization wash/25% 2X SSC, 50% hybridization wash/50% 2X SSC, 25% hybridization wash/75% 2X SSC and finally 100% 2X SSC), lastly three washes of 0.5X SSC for 20 minutes each. Following these steps samples were washed at room temperature. Next a decreasing ladder of SSC and PTw was used, each wash was done for 10 minutes on a rocker table. The ladder consisted of 75% 0.5X SSC/25% PTw, 50% 0.5X SSC/50% PTw, 25% 0.5X SSC/75% PTw and finally 100% PTw.

Next the samples were prepared for antibody tagging. Five washes of blocking solution (PBT) were followed by a wash of Boehringer-Mannheim blocking solution for 1 hour at room temperature. This was replaced afterwards by the same blocking solution containing anti-DIG antibodies (Roche) diluted by 1:5000 and incubated with the samples overnight at 4°C. This was followed up with five washes of PBT for 15 minutes each at room temperature and five washes of PTw solution for 30 minutes at room temperature, last of which was done overnight at 4°C.

Development of signal was started the next day. Preparations for this included two washes of alkaline-phosphatase (AP) buffer without $MgCl_2$ for 10 minutes each at room temperature. Next followed two washes of AP with $MgCl_2$. Development was done in this buffer with added 3.3 µl/ml nitroblue tetrazolium chloride (NBT, Roche) and 3.3 µl/ml 5-bromo-4-chloro-3-indolyl-phosphate (BCIP, Roche) at room temperature. Signal development was followed under a stereomicroscope and solutions were changed twice each day if development stretched the span of days always prepared freshly on that day. When signal has developed the reaction was stopped with two washes of AP buffer without $MgCl_2$ and five washes of PTw. For storage of the samples the PTw was replaced by 70% glycerol in PTw and stored at 4°C.

2.5 Data collection and processing

All datasets, software (with citations where applicable) and codes used in this project can be found at: https://github.com/fka21/Maternal_gene_expression_evolution

2.5.1 RNAseq and associated datasets

Collection of RNAseq datasets was done manually through SRA-explorer (https://sra-explorer.info/#). NCBI's sequence read archive (SRA) was searched for keywords covering topics related to this project (e.g. "oocyte", "egg", "early development"). Filtering criteria for datasets were the following: reads had to originate from RNAseq experiments, species of origin had to be within the Metazoa lineage, biological replicates had to be included, sampling had to cover early development up until gastrulation. If the criteria were met, then the raw fastq files were downloaded.

Where possible, genomic resources were also downloaded from Ensembl, NCBI or ParaSite databases.

2.5.2 Fastq processing

Sequences generated during this project were quality inspected using FastQC. Raw fastq files were trimmed using fastp. Adapter removal, low quality nucleotide trimming, low complexity filtering and base correction with paired end data was applied. Upon completion fastq files were inspected again with FastQC to confirm quality. All subsequent applications were done with trimmed fastq files.

2.5.3 Sequence retrieval for gene trees

Coding sequences were acquired through the use of Ensembl Rest-API with a custom R script. For each gene of interest orthologs of model species were queried from both Ensembl and Ensembl-Metazoa and only longest isoforms were kept for further analysis. Perfect duplicate sequences were removed using seqkit.

2.6 Transcriptome assembly

Where genomic resources were unavailable transcriptome assemblies were generated. For this project, despite having a published genome, *Hypsibius exemplaris* transcriptome was

assembled *de novo* for use. This was done for two reasons: the current annotation of the genome lacks UTR sequence annotations and the *de novo* transcriptome assembly improved mappability of reads considerably during quantification. Quality of this transcriptome was of comparable quality to the genome based on BUSCO analysis (Simão et al., 2015).

An over-assembly approach was chosen as it improves the quality of the assembly (Hölzer & Marz, 2019; Surget-Groba & Montoya-Burgos, 2010). With this approach multiple assemby software, multiple studies and multiple k-mer sizes produce a diverse set of contigs. These contigs then can be filtered using various approaches to achieve a non-redundant transcriptome. The three main assemblers used were: Trinity, TransAbyss and RNASpades. K-mer sizes were chosen according to read sizes. After trimming reads had a distribution of varying lengths, the shortest one was used as reference length for k-mer choice. The k-mer had to be of a minimum length of 50% read length and could not exceed 80% read length size. In this range several k-mer values were chosen for different assemblies. A notable exception to this is Trinity, which has a fixed k-mer value of 32.

Assemblies were then concatenated for each species and using EvidentialGene pipeline further processed (Gilbert, 2019). Non-redundant transcripts are retained and from them open reading frames predicted, from which coding sequences and subsequently protein sequences are retrieved. Homology information is also added through aligning coding sequences to uniprot-swissprot database (Bateman et al., 2021) with the ultrafast diamond aligner. The quality of each assembly was evaluated with BUSCO scores. This was run in proteome mode and in all cases it was used against the metazoa_odb10 database.

2.7 Quantification and categorization

Quantification step was performed with salmon's pseudoaligner algorithm. If genomic sequence was available, then that was used as decoy during indexing of the cDNA sequences, otherwise cDNA sequences were indexed directly. Quantifications were done with mapping validation, bias awareness enabled and bootstrapping.

As presented in the introduction a subset of maternal genes will degrade throughout MZT. On the other hand, some transcripts persist throughout the MZT by being *de novo* transcribed from the zygotic genome. Because of this abundance dynamic the two maternal gene categories could reflect different functions during early development. Based on theories proposed to explain degradation of some maternal transcripts one should expect a mixture of genes with specific regulatory roles (e.g. asymmetrically localized germ cell markers) and genes without functions for early development (e.g. the transcripts required for oogenesis). While the persistently expressed ones could potentially have a housekeeping role as their presence is required throughout the whole early development. This categorization in turn can help in getting a better understanding in the evolution of maternal genes as, if the above assumptions are met, different selective forces should shape the dynamics of the subcategories.

Following the logic from above one can then categorize genes present in the maternal transcriptome. Here the maternal transcriptome is defined as: all transcripts present in oocyte stage samples with an expression level of 2 transcripts per million (TPM) as suggested previously (Wagner et al., 2013). A separate approach using the zFPKM normalization approach was also tested (Hart et al., 2013). The list of maternal genes proved to be highly similar in the two cases. Due to its simplicity with similar results the cut-off value was favored for downstream analyses. Following the dynamics of these maternal transcripts up until the MZT then one can define the cleared maternal transcripts with robust statistical test as done during differential gene expression (DGE) analysis. In this context down-regulated genes correspond to the degraded genes, whilst up-regulated genes have more obscure origins. These can originate either from *de novo* transcription or from cytoplasmic polyadenylation as the RNAseq datasets commonly use polyA tail capture approaches. Because of ambiguous origins the up-regulated genes were not further investigated. Genes without statistically significant fold changes while being present in the maternal transcriptome are thus categorized as having persistent

expression. Genes significantly downregulated from the maternal transcriptome compared to stages following the MZT are categorized then as degraded genes.

Above analyses were done in the R programming language using custom scripts and widely used R packages. For DGE a standard pipeline of tximport and DESeq2 pipeline was used. Not all species have a reported timeframe for MZT, therefore contrasting points had to be determined. For these Euclidean distances were quantified across the variance stabilized samples and these were clustered using hcclust and visualized with pheatmap. The heatmap provided visual information for shifts in the transcriptome and in conjunction with the known developmental stages the earliest major transcriptional shift was set as anchoring point for the MZT. This was further validated with PCA analysis of variance stabilized samples. To account for unknown variables during data collection a surrogate variable analysis was performed using the sva package and these variables were added to the design formula. Where variables apart from the developmental stages were known the design formulas were setup with these accounted for.

Following the above preparations the degraded maternal genes were determined with cutoff values for adjust p-values of 0.05 and $\log_2$ fold change -2. To validate DGE results *in-situ* hybridization based categorizations from the Fly-FISH database (Lécuyer et al., 2007, Wilk et al., 2016) were retrieved and compared to the *Drosophila melanogaster* list of degraded genes. Also, for *Priapulus caudatus* and *Terebratalia transversa* selected candidates were validated with *in-situ* hybridization.

2.8 Functional enrichment

Functional enrichment analysis was done in R programming language environment. Functional annotations were either imported from available genomic resources or assigned *de novo*. For the latter two strategies were used: the transcriptome assembly provided homology information was used for inferring gene ontology annotations or a web-service tool Pannzer2 assigned high probability annotations. For the enrichment itself clusterProfiler package's enricher() and enrichGO() functions were utilized. The former was used in cases with custom gene ontological annotation databases built *de*

novo, the latter for available annotations. If custom annotations were provided to enricher() as background set all GO annotations retrieved for all genes per each species were used. All ontological categories were tested and considered enriched with a cut-off value of < 0.05 for the adjusted p-values. Both categories of maternal genes were tested this way separately, ordering of the genes was done by the absolute value of the $\log_2$ fold changes in degraded category and the TPM values for the non-degraded genes.

For the three studied species different biological pathways were retrieved from the Kyoto encyclopedia of genes and genomes (KEGG, Kanehisa & Goto, 2000). To map genes from these organisms first their gene identifications were converted to a universal entrez identification code.

2.9 Gene architectural feature analysis

Information on the lengths of different architectural characteristics of genes were extracted from their genome annotations. Species where only coding sequence and intron lengths were available were omitted. Where multiple feature sizes were annotated due to alternative splicing events the means of these lengths were taken. For plotting the mean lengths or proportions were used.

Due to their significant role in regulating transcript stability the 3' untranslated regions (UTR) from the three focal species were inferred from their transcripts with the aid of protein sequences in the ExUTR pipeline. A minimum length threshold of 200 bp was applied and duplicates were removed. Following this the UTRs were searched for any enriched RNA binding protein motifs present within using the SEA algorithm of the MEME suite. Experimentally validated RNA binding motifs were retrieved from ATtRACT database (Giudice et al., 2016) and RBPDB (Cook et al., 2011) combined withthe default database of the MEME webserver (Ray, Kazan, Cook, Weirauch, Najafabadi, Li, Gueroussov, Albu, Zheng, Yang, et al., 2013). Only motifs with an E value less than 0.05 were considered as enriched.

The opposite side of this interaction was also queried in the three species. Pfam database (v35; Mistry et al., 2020) was searched against the translated CDSs using HMMER with default settings. Results were filtered with keyword searches ("RNA", "binding") in the description of the domains. Enrichment was tested using a hypergeometric test in R.

2.10 Orthology assignment and species tree inference

From the translated CDSs only the longest isoforms were retained for orthology assignments. For the assignments OrthoFinder2 was used. Initial alignments were done with diamond ultra-sensitive mode, clustering was done with default inflation parameter and gene trees were estimated in a multiple sequence alignment mode with MAFFT. A species tree was provided which was retrieved from the Open Tree of Life database with the rotl API R package. A separate run of OrthoFinder2 was also performed without any species tree provided, this way the pipeline clustered orthogroups for estimating a maximum likelihood species tree. These selected orthogroups were realigned with MAFFT's L-INS-I setup, trimmed to retain only parsimoniously informative sites and finally concatenated with AMAS before the species tree estimation. For the species tree estimation IQ-TREE2 was used. The partitioned multiple sequence alignment (224 partitions with 10.18% missing data) was provided with the Open Tree of Life provided species tree used as a starting tree.

As phylogenetic comparative methods require a dated tree the generated tree was calibrated using geiger's congruify approach (Eastman et al., 2013). Internal branching event timings were accessed from the TimeTree database (S. Kumar et al., 2017). For scaling itself TreePL's approach was used. An initial optimization run was followed up by a second run where different smoothing parameter values were tested. In a third and final run with all parameters set to optimal values dating was performed.

For maternal and degraded categories orthogroups species were pruned. This resulted in a presence/absence matrix. Additionally, the species tree was pruned for the presence of maternal or degraded genes for each orthogroup. This was followed up by quantifying the longest distance between all pair-wise tip distances for each orthogroup. This metric was

used to quantify how distantly related species does each category cover. Distances were calculated with distTips function from adephylo package. Both number of nodes separating tips and branch length sums were inspected.

A data matrix was constructed combining expressions with the orthology relationships. Raw TPM values across a wide variety of species, each with differing number of genes and variation in gene lengths, are not directly comparable without normalizing first. Therefore, within each species a normalization was performed using edgeR's trimmed mean of M values (TMM) normalization method across replicates. In the next step another normalization across species was done according to a recently proposed approach (Munro et al., 2021.). Following this, replicates were collapsed into a single value by calculating their means and this single value was assigned to each orthogroup for that species. If multiple genes were assigned to one orthogroup, then these paralogs were summarized into a single value represented by their mean. Finally, the expression values were log transformed with an added pseudocount of 0.01. Batch effects across different experiments were accounted for using metadata provided for each sequencing run and limma's removeBatchEffect() function. A separate data matrix was constructed containing information about the categories of maternal genes within each orthogroup. This table was later used to prune the species tree to retain only degraded or persistently expressed maternal genes respectively.

2.11 ML tree estimation from gene expression dataset

The matrix containing gene expression datasets was used for downstream analyses, one of which involved maximum likelihood phylogenetic tree estimation from it. The matrix was transformed with two approaches. In one approach a binary transformation was used, whereby if expression was present (normalized TPM ≥ 2) it would be assigned a 1 and if expression was absent (normalized TPM < 2) it would be assigned a 0. This was performed separately for the matrix containing pruned for persistently expressed genes and degraded genes. A second transformation discretized the gene expression values into 5 categories based on the normalized TPM values. The cut-off thresholds used were the

following: category 1 (0 > TPM < 2), category 2 (2 ≥ TPM < 10), category 3 (10 ≥ TPM < 100), category 4 (100 ≥ TPM < 1000) and category 5 (TPM ≥ 1000).

In all three cases the ML tree was estimated using IQ-TREE2 with the automatic best fitting model selection and ultra- fast bootstrap calculations (1000 replicates). Cophylogenetic plots were generated using phytools's cophylo function. The distances from the species phylogeny were calculated using phangorn's treedist function.

2.12 Gene trees and sequence evolution

For the three studied species selected maternal gene sequences were analyzed in a phylogenetic framework. Gene trees were estimated using a maximum likelihood approach with IQ-TREE2. For this CDS of orthologs were retrieved (see 2.5.3 Sequence retrieval for gene trees) and duplicated ones removed. To this, sequences from the three species were added and the whole dataset was translated into amino acid sequences using transeq from the EMBOSS Tool suite. For each gene of interest a domain characterizing that gene was identified in those sequences. Sequences were discarded if given domain could not be identified (Table *3*).

Table 3. Table providing information of the Pfam domians used to filter sequences.

Gene name	Domain	Pfam ID	Description
Nanos	Zf-nanos	PF05741.15	Nanos RNA binding domain
Cbx1	Chromo	PF00385.26	Chromo (CHRromatin Organisation MOdifier) domain
	Chromo_shadow	PF01393.21	Chromo shadow domain
Emsy	ENT	PF03735.16	Emsy N-Terminus

The domain hit sequences were then extracted and aligned using MAFFT in G-INS-I mode. In sequences where domains were duplicated only one of the domains was retained for downstream analysis. A maximum likelihood gene tree was then estimated. Sequence evolution models were chosen by best fit based on ModelFinder, which is implemented in the IQ-TREE2 algorithm. All branches were resampled for bootstrap values using IQ-TREE2's ultrafast bootstrapping method. As outgroup for *emsy* three homologs of TDRKH (Tudor and KH domain containing protein) were chosen, for *cbx1* three

homologs of PHD19 (PHD finger protein 19) were chosen and finally for *nanos* 3 homologs for ZNF3 (zinc finger protein 3) were chosen. The outgroups separated well from the tree, although they showed long branch attraction effects, therefore were left out in a second round of multiple sequence alignment, which placed long branches away from basal positions.

In parallel whole translated CDS was also aligned using MAFFT with the same settings. These were then backtranslated to CDSs with EMBOSS Tools Suite and with the estimated gene trees based on the specific domains codon based selection analyses were performed with Hyphy. Nonsynonymous to synonymous site ratios for each gene tree and alignment were approximated with fixed effect likelihood (FEL) approach (Kosakovsky Pond & Frost, 2005), while relaxation of selection for *a priori* defined branches was done with RELAX (Wertheim et al., 2015). For the former the branches to test for relaxation of selection were the ones containing invertebrate lineages. These same branches were subject to contrast-FEL, also part of Hyphy package, which tests for sites experiencing significant differences in nonsynonymous to synonymous site ratios in *a priori* selected lineages.

2.13 Model fittings

In the R programming framework models of character evolution were fit to each orthogroup. Before fitting the species tree was pruned for the two maternal gene categories separately based on the presence of a gene in the orthogroup being part of a category. Furthermore, a filtering step for the minimum tree size was added before fittings partly to save computational power and partly because model fittings can be sensitive to tree size (Cooper et al., 2016). Following pruning and filtering models were fitted using geiger. Standard errors for each species were also provided during model fitting to explain a portion of variation outside phylogeny in the case of continuous characters. For continuous characters, such as the normalized expression values, models of Brownian motion or Ornstein-Uhlenbeck processes were fitted in an initial round (Figure 3). Alternatively, as a null-model a white noise model was also fitted, which

leaves out phylogenetic relationships from explaining the data distribution. The best fitting model was decided based on second-order Akaike Information Criterion (AICc). Following this a parametric bootstrapping step was inserted. Data was simulated based on the best fitting model, parameters used for the simulation were extracted directly from the fitted model estimates. On this simulated dataset another round of fitting was performed and the model confidence was evaluated based on the proportion of correct model identification. In parallel it was inspected if the parameters estimated from the original dataset were significantly different from the parameter estimates for the simulated dataset using a Kolmogorov-Smirnov test. Orthogroups where the best fitting model could not be replicated in 50% of the cases were discarded from further analysis.
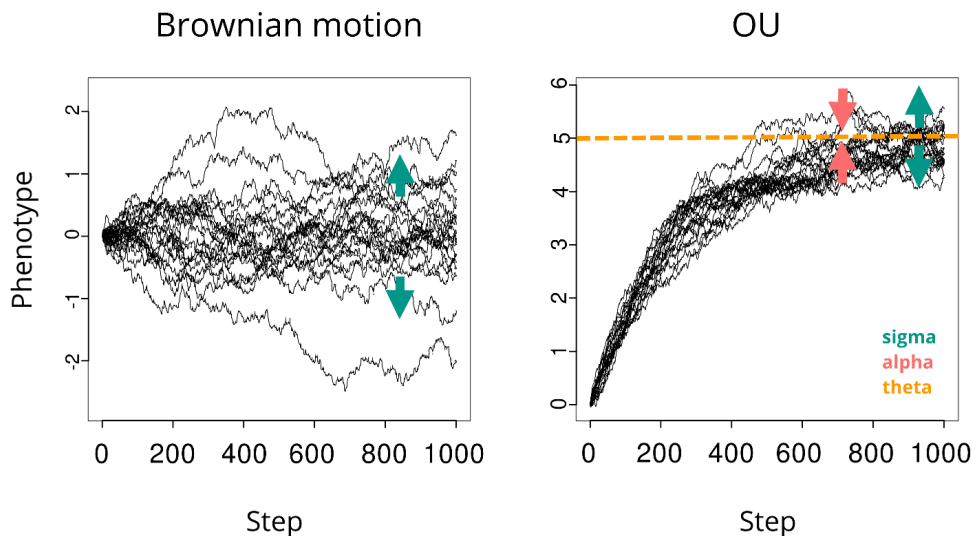
Figure 3. 1000 step simulation of trait evolution across a tree under a Brownian motion model and a OU model. Highlighted on the plots are the parameters and their effect on trait evolution.

A second round of model fittings were performed separately in orthogroups with Brownian motion models or OU models. In this round more complicated models of trait evolution were considered in order to explain more intricate patterns of expression evolution across clades. For the Brownian motion models heterogeneous rate models were considered, where instead of a single σ^2 parameter multiple σ^2 parameter are tested

across all nodes. This model has been implemented using the R package motmot. If including multiple σ^2 parameters significantly increased the likelihood compared to a single rate Brownian motion the heterogeneous model was kept as a better alternative. Up to 4 parameters were allowed to be tested across each orthogroup following Brownian motion models. In the case of orthogroups where OU model was the best fitting one an alternative model was considered where multiple Θ could be present in the data. This was implemented using the surface R package. Here a likelihood ratio test was performed in order to evaluate improvement in model fittings compared to the single Θ OU model. As both models work in an iterative fashion on the species tree, nodes where a shift occurred can be then identified. Finally, for some orthogroups out of this second round of model fittings a third model type was considered. The subject orthogroups were chosen upon manual inspection of the results from the second round of model fittings. If high dispersion of expression values was common in a tree with a clade centering around a strict value an orthogroup was chosen for further testing. Here a nested model of trait evolution was considered in which in most lineages evolve according to Brownian motion model, but in a clade there is a shift towards a different model (in this case OU model was considered). The clade subject for testing for shifts was chosen from the output of the second round of model fittings. The model fittings were chosen according to AICc's. The nested models fittings were also performed using motmot package.

A similar approach was used for discrete characters, the difference lying in the input data, the fitted models and simulation of the models. Input data was discretized based on either the expression value (if TPM <= 2, then not expressed, otherwise considered as expressed) or the orthogroup occupancy (no expression found because of representative lacking in the orthogroup or not). The fitted models were either the equal rate model (ER), all rates different model (ARD) or as a null model an ER model with a branch transformation step eliminating underlying phylogeny.

3. Results

3.1 Transcriptome assembly

BUSCO scores evaluate the completeness of a predefined gene set characteristic of a lineage against the transcriptome. Based on the BUSCO scores the assembly pipeline resulted in consistently complete transcriptomes (Figure 4, Table 4). A fixed species tree topology was retrieved from Open Tree of Life (Redelings & Holder, 2017). Branch lengths for this tree were adjusted according to the orthogroups used by OrthoFinder's algorithm to estimate a species phylogeny. These orthogroups were separately realigned, concatenated and used for estimation of branch lengths for the fixed topology species tree (Figure 5).

Table 4. Summary statistics for de novo transcriptome assemblies (ORF – open reading frame). *-in number of nucleotides.

| Species	Transcript number	N50	Longest transcript*	ORF	Coding number	Annotated number
Heliocidaris erythrogramma	102,483	1,530	19,737	105,432	21,403	11,354
Heliocidaris tuberculata	70,765	1,785	20,480	73,370	22,823	12,707
Hypsibius exemplaris	40,568	1,117	11,326	41,790	31,326	16,659
Lytechnius variegatus	105,487	1,565	21,158	108,263	23,742	12,552
Mesocentrotus franciscanus	67,922	2,819	132,566	31,366	30,952	14,508
Paracentrotus lividus	113,651	721	15,574	114,215	25,809	15,534
Platynereis dumerilii	118,615	982	18,083	119,836	36,494	17,758
Terebratalia transversa	32,446	1,501	13,394	32,817	12,565	11,604
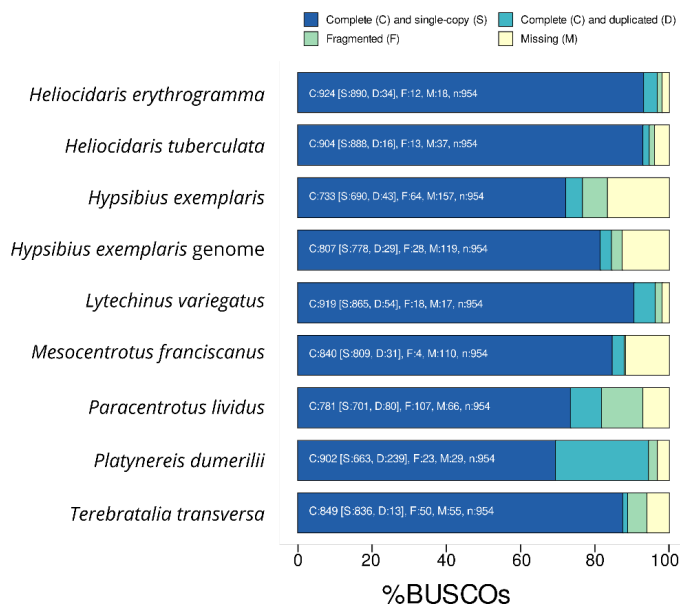
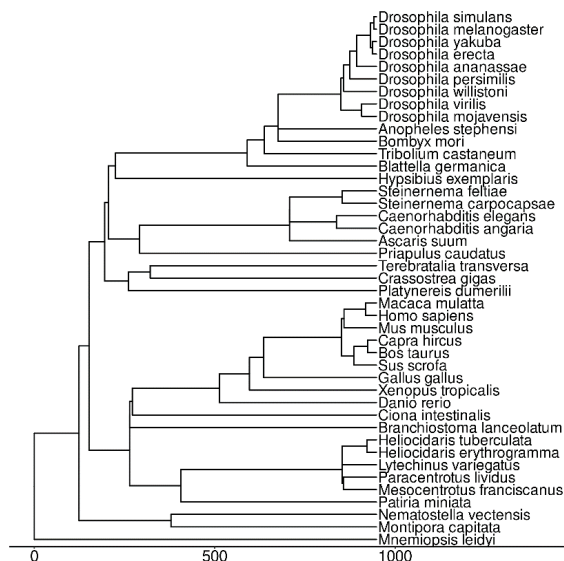Figure 4. BUSCO scores distribution of the de novo assembled transcriptomes used in this project

Figure 5. Species phylogeny according to ROTL with branch lengths estimated from OrthoFinder and scaled to fossil data from TimeTree. Axis denotes millions of years.

3.2 *Priapulus caudatus* maternal transcriptome analysis

In the case of *P.caudatus* I found that a third of the genes are expressed in the maternal transcriptome. Most of the gene expression values are skewed towards the threshold TPM value of 2, with some outliers such as *cyclin B* or *srr55* having high TPM values. Oocyte and cleavage stage samples are tightly clustered in both principal component analysis (PCA) plots and heatmaps of sample Poisson distances, in contrast the later developmental stages are clearly distinguishable (Figure 6 A, B). On the PCA plot the first principal component, which explained 70% of variance across samples, corresponded to the developmental stages of the samples. I found a separation of the oocyte and cleavage stages from the gastrula stage, indicating a major transcriptome overhaul in between cleavage and gastrula stages. The separation was consistent in the PCA and sample Poisson distances. This shift in the expression patterns is reminiscent of MZT, therefore in differential gene expression analyses the latest stage I compared the downregulation of maternal genes to is the gastrula stage. Oocyte stages were contrasted to gastrula and cleavage stages, the latter one was also contrasted to gastrula stage. In total 3115 genes were found to be significantly downregulated during the MZT period. The GO analysis results show that the maternal transcriptome includes genes involved in translation, protein degradation through ubiquitination, mRNA processing and chromosomal function enriched. The downregulated genes are enriched for similar functions with the addition of terms such as translation or rRNA binding.

I have retrieved several signaling pathways from KEGG database and investigated their temporal dynamics. On average 40%-60% of the genes involved in the inspected pathways were not found in the annotations of the genome. A possible cause for this could be that KEGG database is populated with human gene names and the *P.caudatus* genome is annotated with data from several different species with different naming conventions. Nevertheless, genes covering the canonical Wnt signaling pathway are present in the oocyte (Figure 6 C). Many components of this pathway show a strong expression from gastrulation onwards, as previously shown (Hogvall et al., 2019). Other

signaling pathways of the early development were partially recovered in the maternal transcriptome, such as the calcium signaling pathway (Figure 6 D) which has been reported to be of great importance in several studied species (Fluck et al., 1991; Muto et al., 1996; Stein et al., 2022).
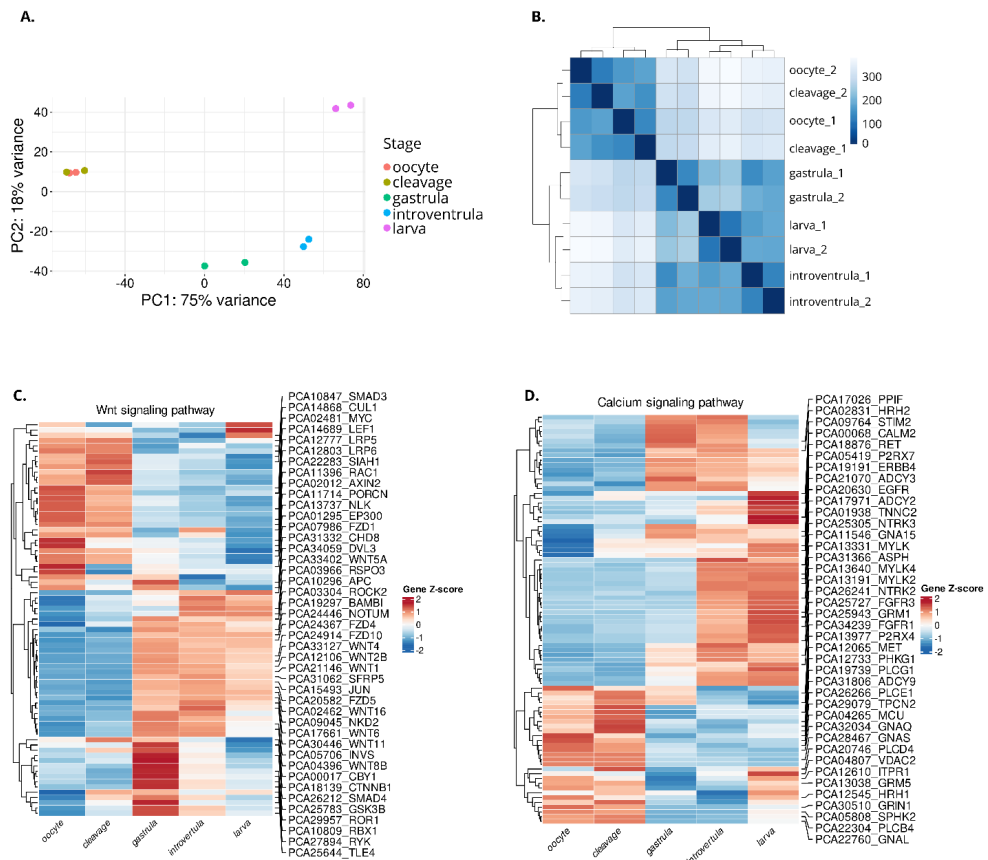
Figure 6. MZT identification in *P.caudatus*. Results of clustering approaches used on samples to distinguish major transcriptome overhauls, an indication of the MZT, are shown in panel A and B. The distance matrix was used as a quantitative approach to determine the contrasting points for differential gene expression analysis (B). Heatmap of developmental stage wise relative gene expression across developmental stages in *P.caudatus* for two major pathways (C, D).

Using a fuzzy clustering approach (L. Kumar & Futschik, 2007) on the gene expression dataset covering the development I have found three major patterns (Figure 7). One

pattern shows stronger expression later in development as in the case of clusters 4 (characterized by enrichment of protein translation and transport related GO terms), cluster 10 (characterized by enrichment in GO terms connected to ribosomal functions and synapse formation) and cluster 2 (with enriched terms centered around rRNA processing and mitochondrial functions). The second pattern can be characterized with expression peak during gastrulation such as clusters 6 (enriched for RNA splicing and protein ubiquitination and degradation) or cluster 7 (enriched terms of mRNA splicing and metabolism, signaling pathways and mitosis regulation). The third pattern is that of genes having gene expression peaks during the MZT period followed by a sharp decline afterwards. Clusters 1, 5 and 8 (enriched for GO terms regarding protein functions such as binding regulation or protein chemical modifications, mRNA splicing, cell cycle regulation) have a peak during cleavage stages and a sharp decline by gastrulation. Cluster 3 and 9 contain genes with peak expression in the oocyte stages and a decline by gastrulation. These clusters contain enriched GO terms such as histone modifications, protein chemical modifications and cell cycle regulations.
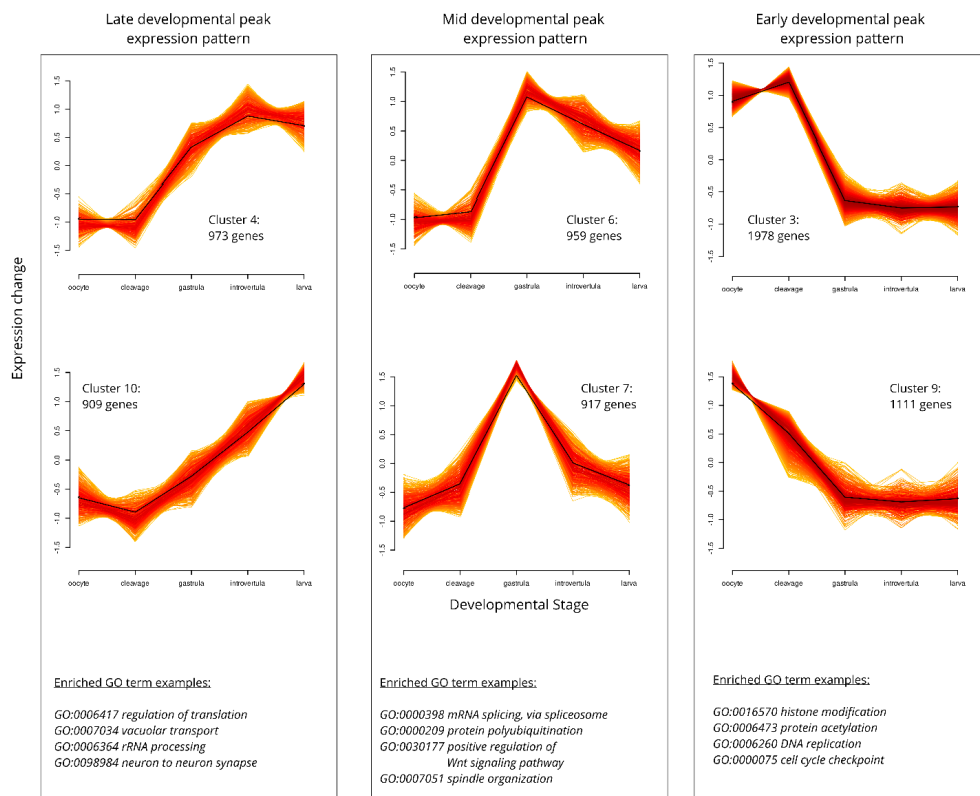
Figure 7 Results of fuzzy clustering of gene expression data throughout embryonic development. Three major patterns were recovered, all 3 represented here. Maternal genes fall in the early developmental peak pattern. Below the graphs examples of significantly enriched GO terms are displayed.

RNA binding proteins (RBP) play a crucial role in regulating the stability of transcripts during the MZT (Matsumoto et al., 1996; Semotok et al., 2005). Therefore, I scanned the maternal transcriptome for protein domains and I evaluated the abundance of RNA binding domains (RBD) (Table 5). I found that both the non-degraded and the degraded subset of maternal genes are significantly enriched for RBDs. This enrichment was even more apparent if inspecting maternal genes without annotations (2.24 fold non-degraded genes and 2.02 fold for degraded genes hypergeometric test p-value 3.9e-15 and 6.8e-4 respectively). Furthermore, I also found that the expression of genes in the maternal transcriptome with RBDs are generally higher compared to the rest of the maternal genes,

regardless of them being degraded or not (Figure 8). The higher abundance and higher expression profile of RBD containing genes suggests that these genes are important factors during the MZT in *P.caudatus*. Interestingly genes with significant differential expression between later developmental stages do not show this pattern of RBD enrichment. Instead, they show a depletion of RBDs (1.68% or 1.01% with p-values from hypergeometric test of 0.63 and 0.97).

I have scanned 3'-UTR sequences of maternal genes for cis-acting motifs that could impact their stability (Figure 8). Most of the results were motifs recognized by splicing factors such as SRSF1/2/5/7 or motifs recognized by ELAV-like proteins. Genes with motifs recognized by splicing factors were found to be enriched in GTPase related functions and protein chemical modification functions. The ELAV-like protein target genes did not have any GO terms enriched associated with them.

Table 5. Table summarizing RBD enrichment results. Bold fonts highlight significant results.

Species	Category	Proportion (%)	Enrichment p-value	Ratio
Priapulus caudatus	Reference	1.9	-	1.00
	Maternal	2.96	**9.23e-25**	1.55
	Degraded	2.47	**1.87e-02**	1.30
Hypsibius exemplaris	Reference	1.46	-	1.00
	Maternal	2.4	**1.81e-09**	1.64
	Degraded	1.76	**2.03e-02**	1.21
Terebratalia transversa	Reference	2.4	-	1.00
	Maternal	2.54	**2.06e-02**	1.06
	Degraded	0.977	9.39e-01	0.41

Because ELAV-like proteins are known for regulating degradation through their 3'-UTR binding (Beckel-Mitchener et al., 2002) I analyzed the expression dynamics for genes that contain these motifs (Figure 8). Firstly, in the priapulid genome 4 ELAV protein homologs were annotated, out which one is comprised of only 1 exon therefore being labelled as unlikely functional and was discarded for further analysis. Of the remaining 3

ELAV homologs one showed a prominent expression overlapping with the MZT. Screening the expression profiles of the genes having ELAV binding motifs in their 3'-UTRs I found 2 mutually exclusive clusters throughout development. One of the clusters is characterized by prominent early expression with weak signal following MZT, conversely the second cluster is more prevalent during later stages of development. The motif numbers in 3`-UTRs range from 1 to several hundreds (e.g. motif recognized by YBX1 which is found 184 times in the 3'-UTR of the *nurim* gene). In downregulated maternal genes I detected lesser enriched motif diversity. Motifs recognized by splicing factors were the only enriched ones I have detected.
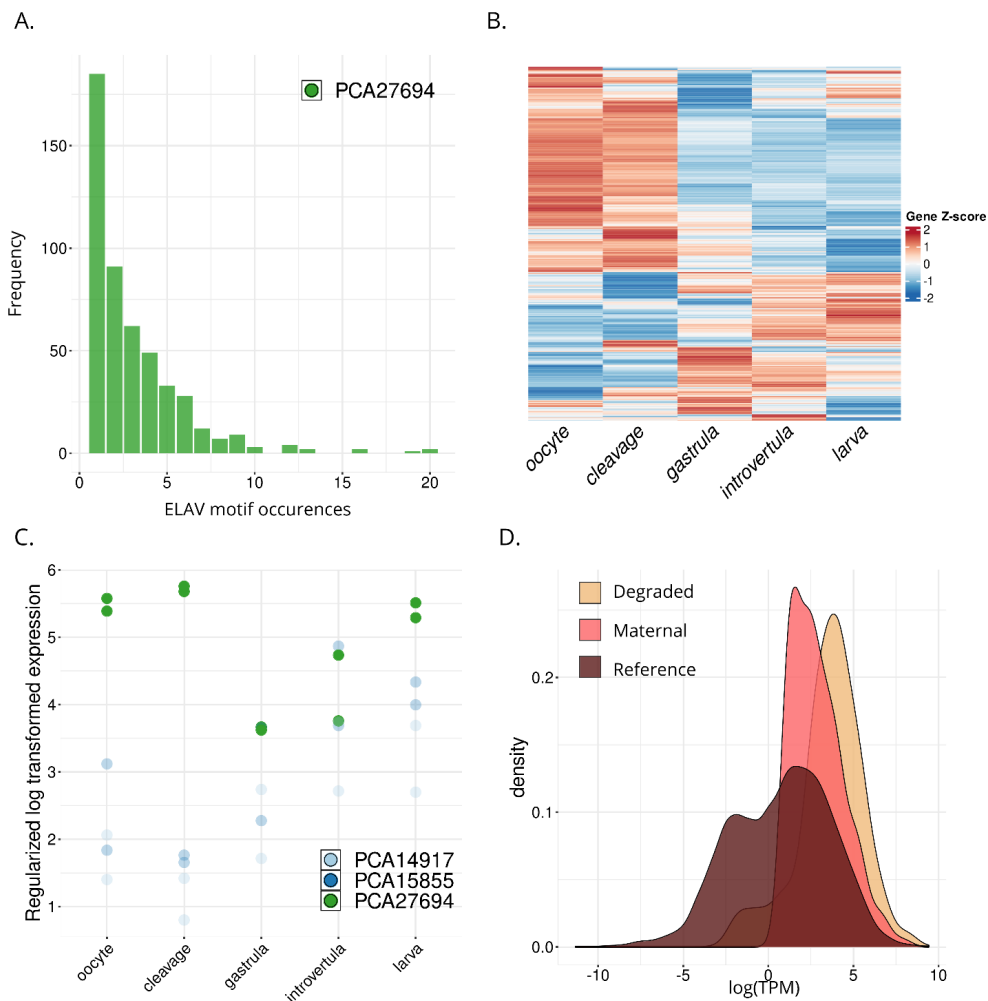
Figure 8. 3`-UTR analysis in the maternal transcriptome. ELAV binding motifs are enriched in the 3`-UTRs of maternal genes (A). A big proportion of maternal genes containing the ELAV binding motifs have peak expressions during early embryogenesis (B). Only one of the 3 ELAV-like proteins are expressed early on in *P.caudatus* (C). Gene expression distribution of genes containing RBDs (D). Expression for maternal and degraded subset from oocyte stage embryos in contrast with RBD containing genes from later developmental stages.

3.3. *Terebratalia transversa* maternal transcriptome analysis

In the *T. transversa* maternal transcriptome I found 71% of all assembled genes as expressed. Because of the cut-off value the average expression of the maternal genes is skewed to TPM 2. Most abundant transcripts are those coding *polyubiquitin* or *histone-1*.

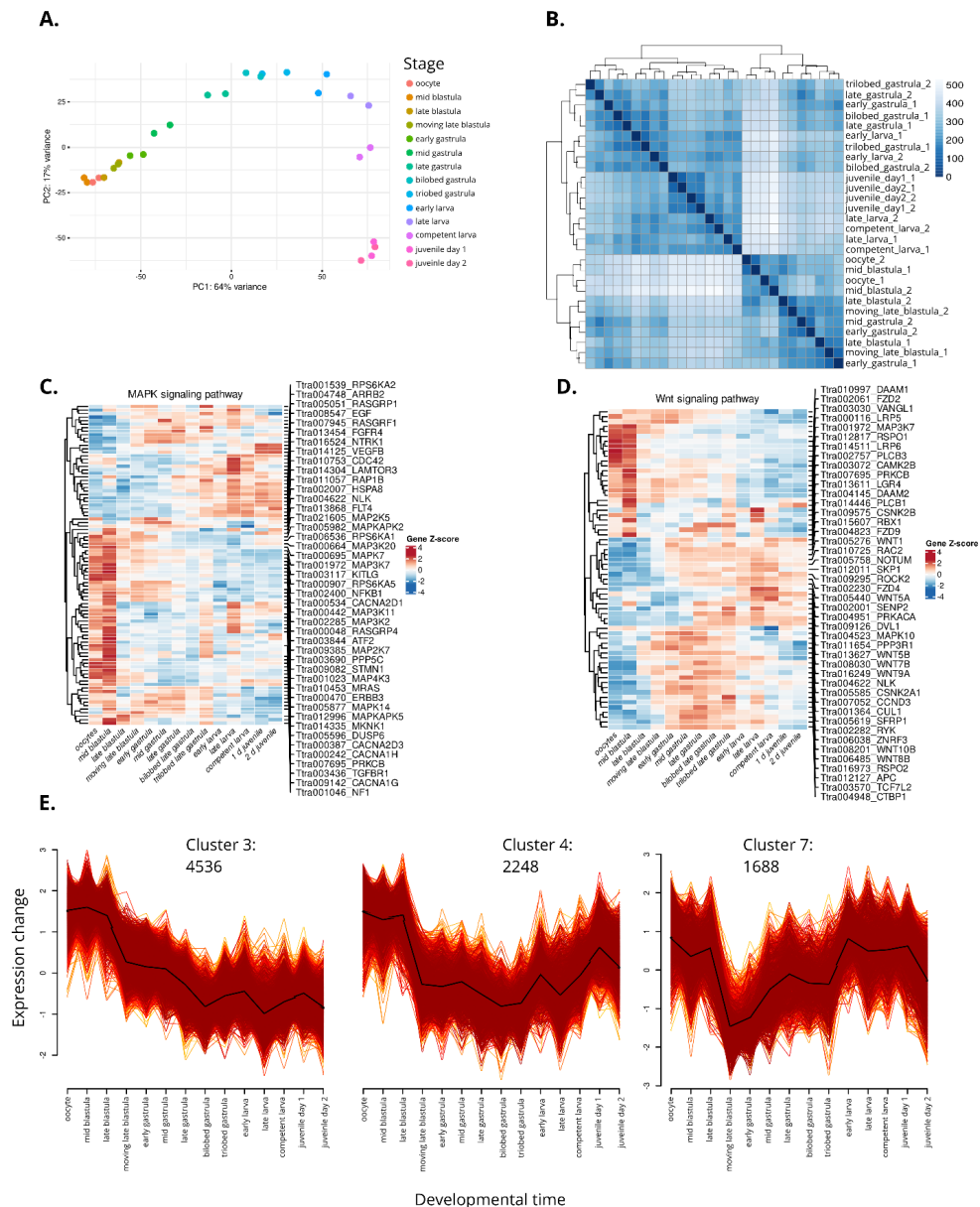Because the *T. transversa* transcriptome contains more timepoints, a more accurate determination of the MZT was possible. In the PCA the samples are mainly separated by the developmental stages on the first principal component. According to the Poisson distance of the samples oocyte stages and mid blastula stages are grouped together. Based on these results the MZT endpoint in *T. transversa* is late blastula stage. Therefore, late blastula stage was used as a contrasting point to explore the degraded maternal genes Figure 9 A, B).

My analysis resulted in 501 downregulated genes under the MZT. GO enrichment analysis yielded few enriched terms for both the maternal transcriptome and the downregulated genes. In the maternal transcriptome mitochondrial processes were enriched. In the degraded genes terms related to protein dimerization, negative regulation of transcription and multicellular organism development were enriched.

I also mapped signaling pathways to the transcriptome from the KEGG database. The retrieval of these pathways varied (46% - 85%). The Wnt pathway is present in the stages leading up to the MZT with no strong Wnt expression being detected and a peak expression in oocytes for Fzr2 receptor. Other noteworthy signaling pathways with strong peaks during early development are that of FoxO signaling pathway and the MAPK signaling pathway. Genes regulating axon guidance during development also have a strong peak during early development (Figure 9 C, D).

Using fuzzy clustering I have found 12 optimal patterns for the *T. transversa* developmental gene expression profile. I narrowed down the analysis to clusters with strong expression patterns during early development. Two clusters showed a steady strong expression until the late blastula stage from which a sharp decline in expression

was detected. Only one of the clusters had enriched GO terms which grouped around nuclear processes (such as DNA unwinding, DNA repair, nucleocytoplasmic transport or methylation). A second pattern represented peaks at the mid blastula stage and drops in expression by late blastula stage. Enriched terms for this cluster are related to protein modifications (protein phosphorylation, protein modification process) and GTPase functions (small GTPase mediated signal transduction, guanyl-nucleotide exchange factor activity) (Figure 9 E).

A. PCA plot showing PC2: 17% variance against PC1: 64% variance, colored by Stage (oocyte, mid blastula, late blastula, moving late blastula, early gastrula, mid gastrula, late gastrula, bilobed gastrula, triobed gastrula, early larva, late larva, competent larva, juvenile day 1, juvenile day 2).

B. Sample correlation heatmap with dendrograms.

C. MAPK signaling pathway heatmap.

D. Wnt signaling pathway heatmap.

E. Clusters 3, 4, and 7 expression change over developmental time.

Cluster 3: 4536

Cluster 4: 2248

Cluster 7: 1688

Enriched GO term examples:

GO:0008168 methyltransferase activity
GO:0003678 DNA helicase activity

Enriched GO term examples:

-

Enriched GO term examples:

GO:0035556 intracellular
signal transduction

Figure 9. *T. transversa* maternal gene repertoire explorations. Clustering approaches suggest late blastula stage as an end point for MZT (A, B). Based on their expression peaks the most complete signaling pathways found in the maternal transcriptome were that of MAPK (C) and Wnt signaling (D). Three clusters were identified showing an early stage specific expression peak during *T. transversa* development (E). Two of which show an absence of expression outside these early steps. Below each graph representative significantly enriched GO terms are displayed

I scanned the transcriptome for RBD protein domains. Just as previously described RBD enrichment was tested in maternal transcriptome and the downregulated genes. Less abundant RBDs were detected in *T.transversa* maternal transcriptome and degraded genes compared to *P.caudatus* or *H.exemplaris*. Although RBDs are significantly enriched in the maternal transcriptome their enrichment ratio is only marginal.

The lack of RBDs is reflected also in the 3'-UTR sequence motifs. I have not detected any enriched motifs in the subset of downregulated genes. The maternal transcriptome was enriched with motifs bound by ELAV-like proteins. The genes containing motifs recognized by ELAV-like proteins show no GO term enrichment. Additionally, these potential ELAV-like protein targets have two developmental expression profiles. First, an early expression which fades out by gastrula stages and a second group of genes having stronger expression patterns throughout later development. In addition to ELAV-like protein motifs, I have found Mex-5 binding motifs enriched in the maternal transcriptome.

3.4 *H. exemplaris* maternal transcriptome analysis

My analysis showed that of the three species studied in this project *H. exemplaris* has the lowest proportion of genes present in the maternal transcriptome. Here 16% of all transcripts are expressed in the maternal transcriptome. The genes with unusually high expressions do not have annotations, nor do they have any protein domains associated with them. Samples are easily distinguishable on the PCA plot with the highest proportion of variance between samples explained by developmental stages (Figure 10 A, B). For differential gene expression analysis, the segmentation stage was used as a contrasting point. I detected 613 differentially expressed genes. Maternal genes in this species are enriched for GO terms such as histone binding, RNA binding, and protein polyubiquitination. Degraded genes have similar terms with some additional ones such as gene silencing by RNA or P-body. Low proportion of the KEGG signaling pathways were mapped to the transcriptome. Most complete signaling pathway found was that of TGF-beta signaling at 85% of genes involved for this pathway missing annotation in the *H. exemplaris* transcriptome. Compared to the previous two species the Wnt pathway identified players are scarce, especially in the oocyte stage.

Using fuzzy clustering I have estimated 12 gene expression patterns throughout the tardigrade development out of which 2 had oocyte specific peaks. Cluster 1 is characterized by a peak expression in the oocyte followed by a steady drop of expression throughout the rest of the development. Several GO terms are enriched in this cluster, such as mRNA binding, mRNA transport, DNA replication and protein kinase related functions. Cluster number 7 has a sharp drop in expression after oocyte stages which persisted throughout development. Membrane transport functions (amino acid transport, protein secretion) and various protein functions (hydrolase activity, transaminase activity) are enriched in cluster 7 (Figure 10 C).
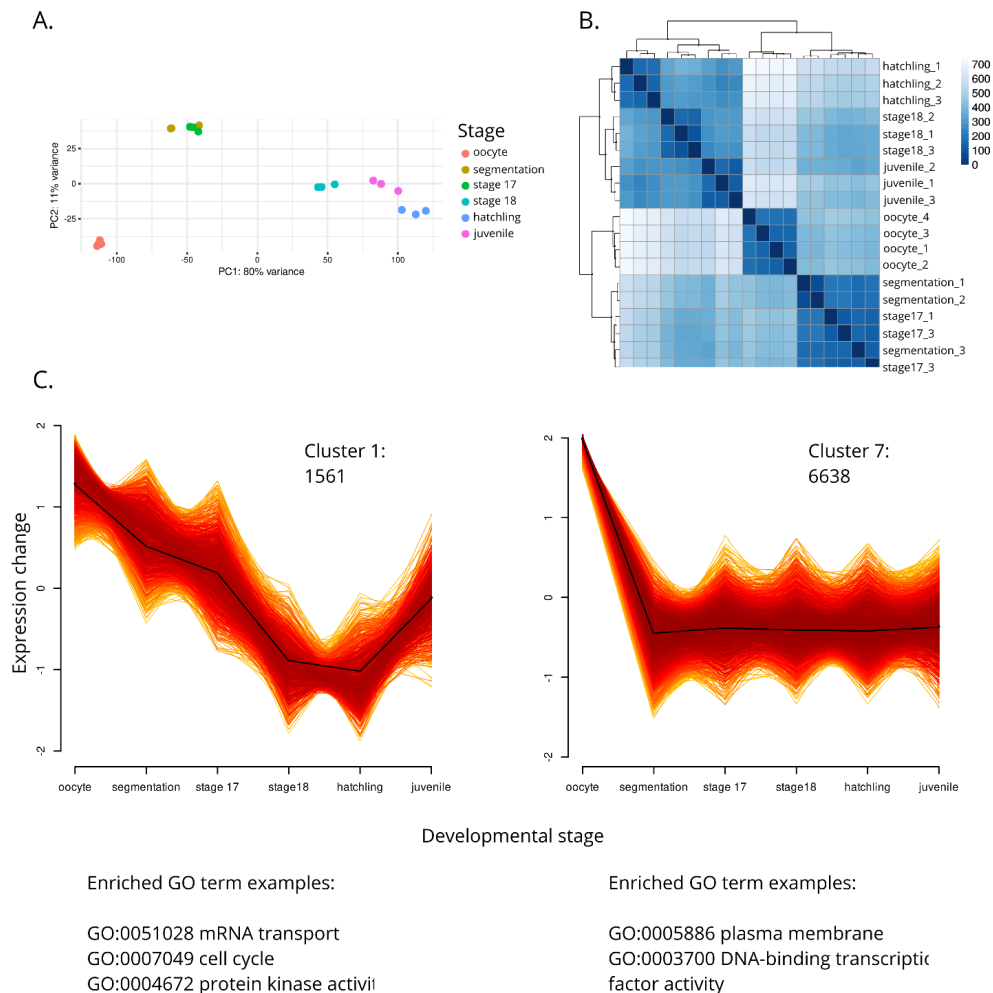
Figure 10. *H. exemplaris* maternal gene complement analysis. PCA plot (A) of samples and Poisson distance based heatmap (B) indicated the MZT timepoint. Fuzzy clustering retrieved 2 clusters indicative of maternal genes (C). Below the expression patterns for these clusters are example GO terms which were significantly enriched.

I have detected that RBDs are significantly more abundant in the maternal transcriptome and the degraded genes. Most common RBDs detected were domains involved in nucleolar transport such as JAZ-type zinc finger domain (Yang et al., 1999) or RNA

polyadenylation such as Nab2-type zinc finger domain (Kelly et al., 2010). Binding motifs of ELAV-like proteins, splicing factors and argonaute (AGO) proteins were enriched in the 3`-UTR sequences of the maternal genes.

3.5 Maternal gene expression validation and sequence evolution

3.5.1 Nanos

I have found the ortholog of *nanos* in the maternal transcriptome of *P.caudatus*. Furthermore, it showed a significant down-regulation during MZT. In *T. transversa* one of the two *nanos* paralogs was present in the maternal transcriptome. This paralog was degraded only after gastrulation. No nanos ortholog could be recovered from the *H.exemplaris* transcriptome. In both species with orthologs present an expression peak was noticeable during cleavage stages. This was possibly due to cytoplasmic polyadenylation, a conserved process across Metazoa (Fernando et al., 1994; Paris et al., 1988; L. Wang et al., 2002). In both species the expression peak subsided by gastrulation stages and never peaked again later in development. In *P. caudatus* immature oocytes (i.e. before germinal vesicle breakdown or GVB) *nanos* transcripts were visualized using *in situ* hybridization. The transcripts are gathered around the germinal vesicle, dispersed in a ring like shape (Figure 11 A).
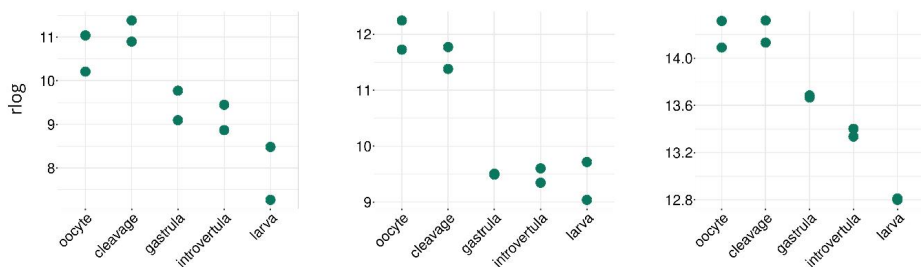
A.
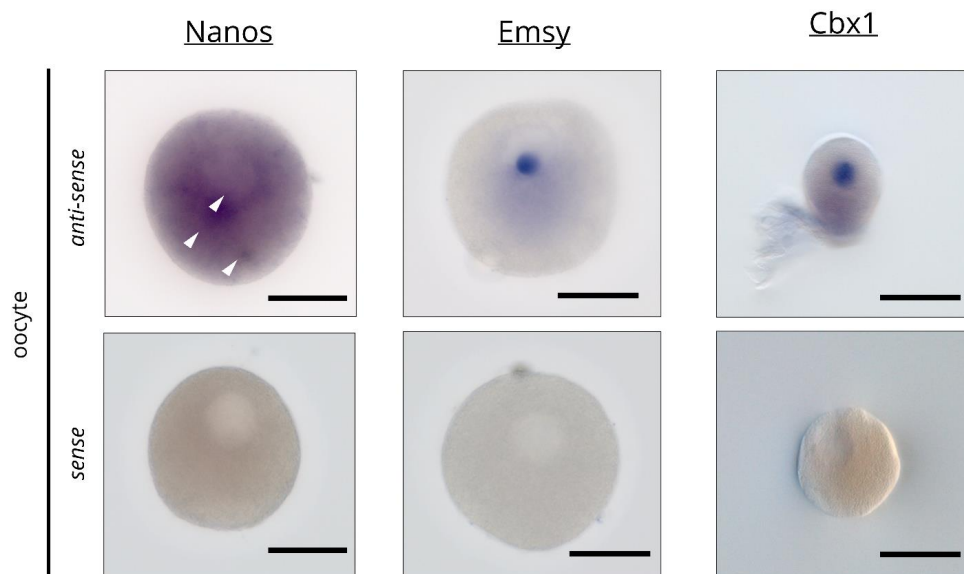
B.

Figure 11. Gene expression of selected maternal genes in *P. caudatus*. *Nanos* shows an early expression pattern with transcript densities found around the cytoplasm (white arrows, A). Both *emsy* (B) and *cbx1* (C) show a restricted expression pattern within the germinal vesicle. For these experiments control experiments included sense probes. Germinal vesicles are visible as a yolk free area. rlog – regularized log transformed counts Scale bar: 0.1 mm

I conducted a phylogenetic analysis of the invertebrate *nanos* orthologs, including the orthologs from *P. caudatus* and *T. transversa*. For constructing the gene tree 973 sequences were retrieved from Ensembl Metazoa database. Only 122 sequences contained the zf-nanos domain with standard alignment parameters and were therefore

retained for further analyses. The sequences from *P. caudatus* and *T.transversa* were also retained as they contained the zf-nanos domain. Two separate genes trees were estimated. One was estimated based on the translated coding sequence (CDS) and the second one was estimated based on the zf-nanos domain sequence.

The best fitting substitution model for the domain based alignment according to ModelFinder is LG with 4 rate category discrete Gamma distribution with a proportion of invariable sites allowed. Tree search was completed within 300 iterations (Figure 12). In the estimated gene tree nematode species and drosophiliid species formed a monophyletic group, whilst the rest of the species were in paralogous groups. The priapulid ortholog grouped together with non-ecdysozoan species as a sister group to a paralogous clade compromised mostly by lophotrochozoan species. The two *Terebratalia nanos* homologs were recovered on different sites of the tree, with both still grouped within a clade of lophotrochozoan species suggesting that the two homologs have diverged since their duplication. I confirmed the evolutionary conservation of *nanos* orthologs (Jaruzelska et al., 2003; Mochizuki et al., 2000) using the fixed effects likelihood (FEL) approach. Codon sites under significant purifying selection were present throughout the whole alignment. At a significance cut-off of 0.001 8.9% of all sites are under purifying selection. This proportion in increased to 63% within the zf-nanos domain. In contrast codon sites under significant diversifying selection were absent from the alignment.

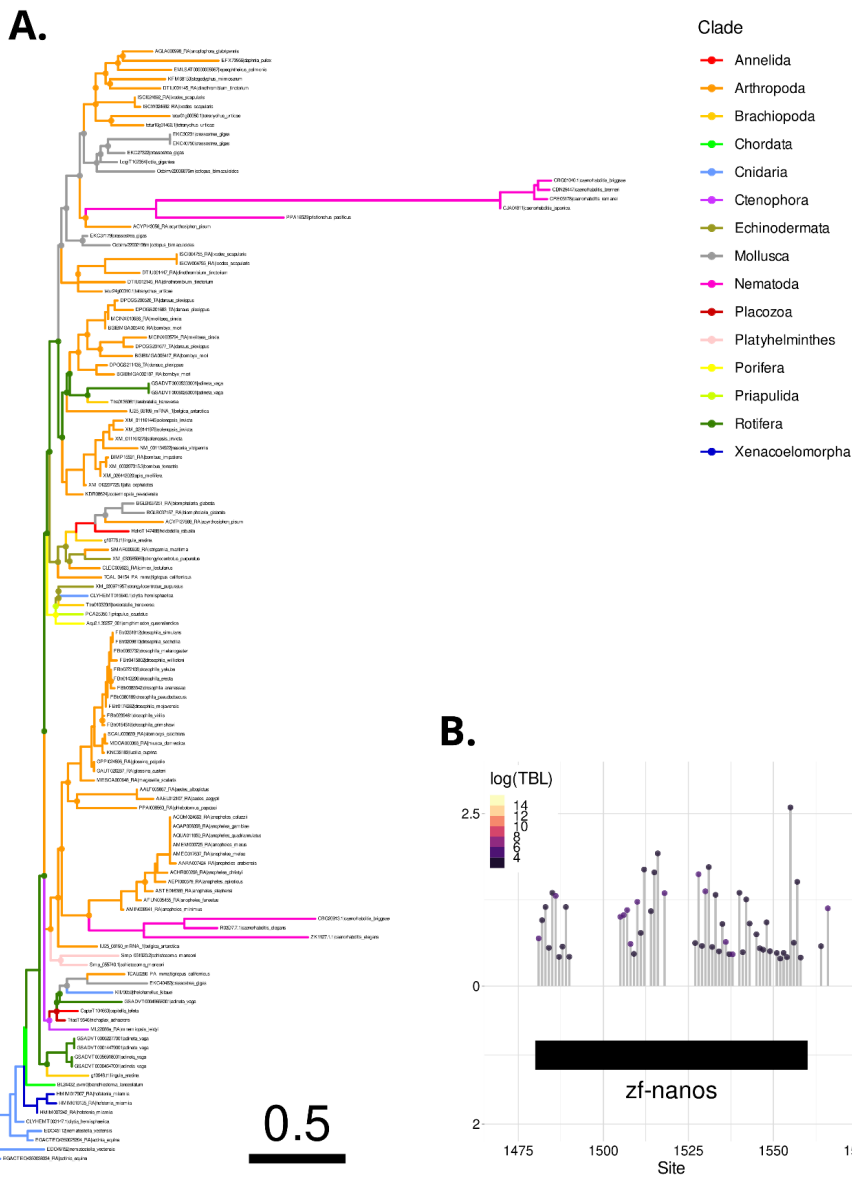Figure 12. ML phylogenetic tree for *nanos* estimated from CDS. The colors highlight the clades. Scale bar denotes nucleotide substitution per site. Bootstrap support with values smaller than 80% are highlighted at nodes as points. Next to the tree is the result of the FEL analysis. Highlighted is the zf-nanos domain. Y-axis denotes estimated alpha and beta values. log(TBL) – log(total branch length contribution to signal)

To get a better insight into duplication and loss events the estimated gene tree was reconciled. For the reconciliation I have retrieved the species tree from NCBI's taxonomy database. The gene tree was rooted based on the species tree topology, therefore *Amphimedon queenslandica nanos* ortholog was used for rooting. Duplications were inferred deep in the species tree with most duplications arising from the Protostomia node (10 events), Eumetazoa node (8 events) and Bilateria node (4 events). Majority of loss events were also inferred deep in the species tree. The algorithm inferred that the *Terebratalia* paralogs are not the result of a recent duplication, as already indicated by the estimated gene tree. The duplication potentially originates from one of the deeper expansion events followed by possibly serial losses in lineages such as Spiralia or Lopchotrochozoa.

3.5.2 Emsy

I chose *emsy* for further analysis from the list of degraded maternal genes in *P. caudatus*. *Emsy* is characterized by the highly conserved ENT domain (Y. Huang et al., 2006; Hughes-Davies et al., 2003), a domain which was also found in *H. exemplaris* and *T.transversa* transcriptomes. Here, I have not found expression of *emsy* in the maternal transcriptomes.

The temporal dynamics of *emsy* in *P. caudatus* can be characterized by strong expression in oocytes and cleavage stage embryos. In later developmental stages a decrease in expression is observable. *In situ* hybridization revealed that *emsy* transcripts show a specific spatial distribution in immature oocytes. Transcripts are localized within the germinal vesicle and distributed asymmetrically (Figure 11).

I have conducted an in-depth analysis of the *emsy* sequence evolution. Smaller scale studies have already indicated sequence conservation (Hughes-Davies et al., 2003) with only a limited number of species included. For a comprehensive analysis I have included all metazoan homologs retrieved from the Ensembl database. To this dataset I have also added sequences from *P.caudatus*. In the analysis 278 ENT domain containing genes were included. Separate tree estimates were done, one for the ENT domain and one for

the translated CDS. ModelFinder inferred different optimal substitution models for the ENT domain (LG with 4 rate category discrete Gamma distribution) and the translated CDS (JTT with amino acid frequencies estimated from the alignment and with FreeRate model for the rate heterogeneity across sites). Successful convergence of the maximum likelihood tree was achieved in 352 iterations for the ENT domain and 199 iterations for the translated CDS.

The topology of the two gene trees revealed the same major pattern. Chordates and Protostome lineages separated well on the tree. Relative positions of species within these two clades change based on the sequence alignment used. Nevertheless, the two separate clades remain constant (Figure 14 A). The chordate clade was characterized by short branch lengths, indicating low substitution rates. A stark exception of this were *Ciona intestinalis* and *Ciona savignyi*. In contrast Protostome lineages had high substitution rates indicated by long branches. This pattern was persistent irrespective of alignment used. The *P.caudatus* ortholog was located in the sister clade of the chordates together with various lineages but separate from other ecdysozoan species.
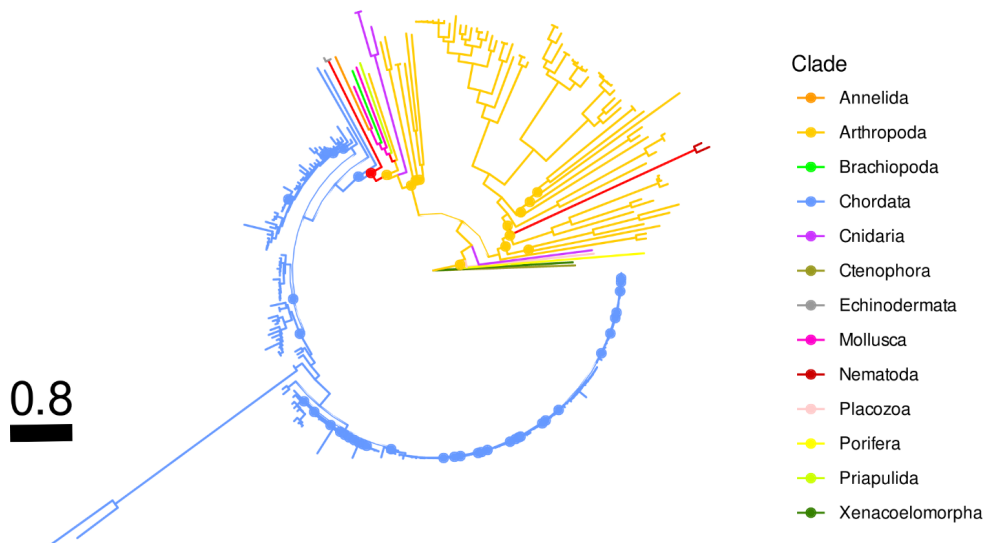
Figure 13. Rooted ML phylogeny of *emsy* gene tree based on CDS. Colors denote metazoan clades. Scale bar displays substitution rate for each codon. Bootstrap support with values smaller than 80% are highlighted at nodes as points.

I used a formal hypothesis testing framework on the CDS alignment to test whether there was sequence conservation and if that conservation varied across lineages. The FEL algorithm identified that 23.3% of the sites are under statistically significant (p-value $\leq$ 0.001) purifying selection. Within the ENT domain 58% of all sites are affected by significant purifying selection (Figure 14 B). The analysis returned no sites under significant diversifying selection. The site-wise distribution of synonymous to non-synonymous substitution rates (ω) is skewed towards zero, indicating the presence of negative selective pressure (Figure 15 B). The ENT domain was identified in the 5` end of the alignment and displayed signs for the presence moderate selective forces.
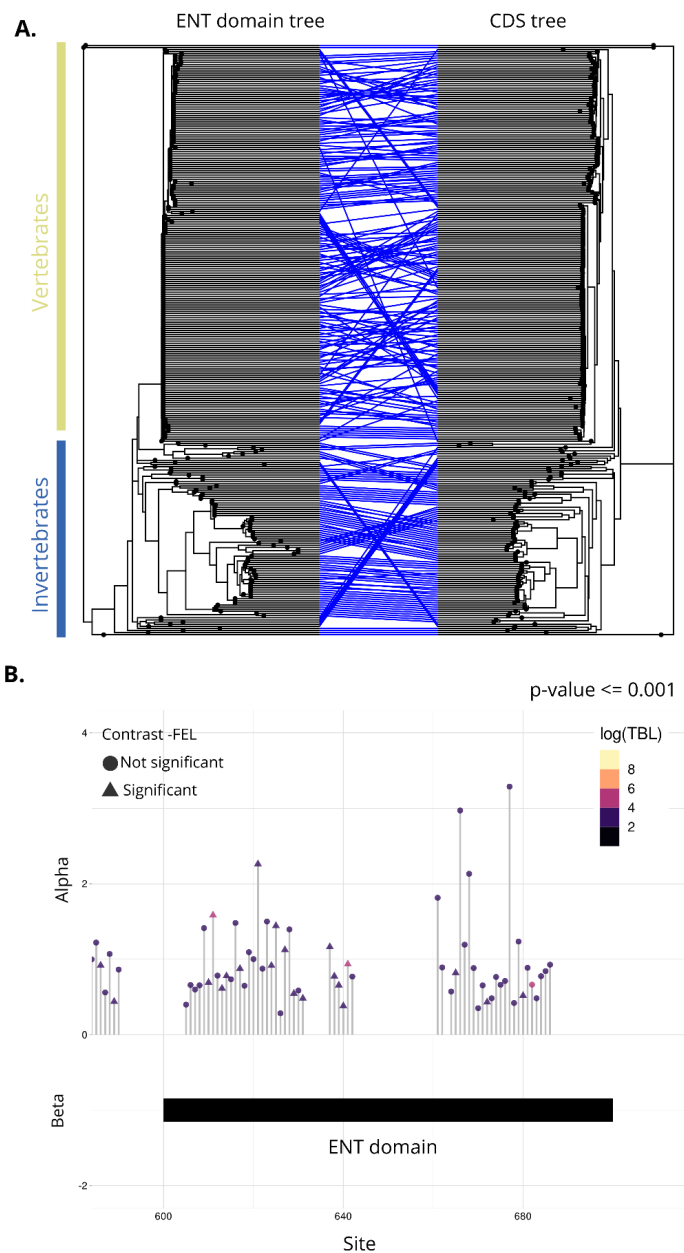
Figure 14. *Emsy* sequence evolution. The gene tree estimated from ENT domain and translated CDS show some differences (A), with the Deuterostome and Protostome lineages maintained. Purifying selection within the ENT domain.

Relative long branches within Protostome lineages suggest a relaxation of selection in these lineages. I have tested the relaxation using the RELAX framework from the Hyphy suite. Chordate lineages were used as reference in this analysis. Protostome lineages show a statistically significant (p-value < 0.001) relaxation of selection with a selection intensity parameter of 0.21. Inferred site-wise ω value proportions indeed shifted towards a neutral regime in the alternative model. I have also tested site-wise differences in selective forces between the test and reference branches using the contrast-FEL method. I found significant (p-value < 0.05) differences in sites which varied according to the alignment section. The ENT domain showed difference in 32.8% of the sites under significant purifying selection. Outside the ENT domain this proportion increased to 52.9% (Figure 15).

I have also inspected the duplication-loss events for the *emsy* gene tree using Notung. I have rooted the gene at the ortholog of *Mnemiopsis leidyi*. This was done based on the species tree retrieved from the NCBI taxonomy database. The reconciliation algorithm identified a surge in duplications at the root of Vertebrates and some of its descending nodes such as Amniota (33 duplication events) or Theria (35 duplication events). Less duplications are estimated deeper in the species tree such as Metazoa node (4 duplication events). In concordance major loss events are estimated in shallower Deuterostome nodes such as Laurasiatheria with 71 loss events or Rodentia with 20 loss events.
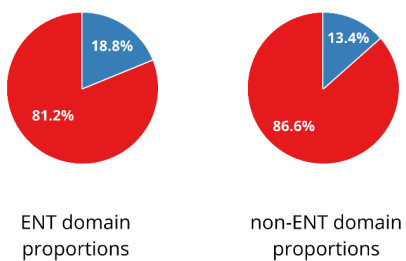
A.
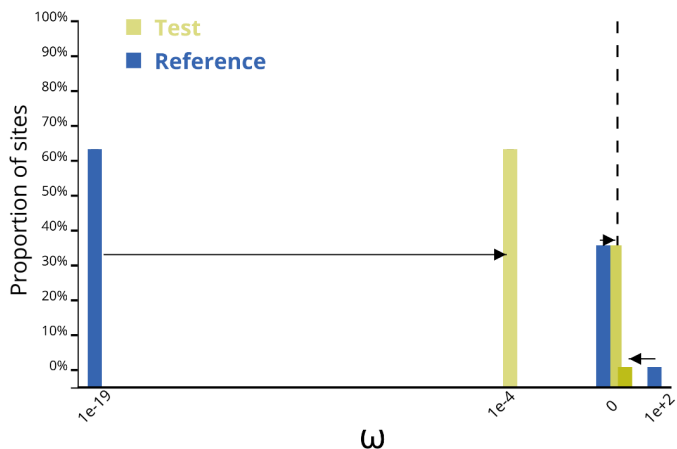
B.
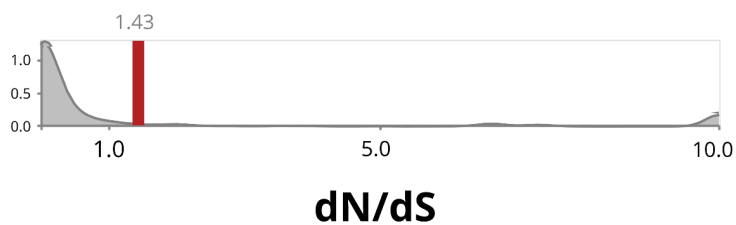
C.

Figure 15. *Emsy* selection analysis. ω density distribution across all sites for emsy. Red bar denotes average ω values (B). Hyphy's Relax method inferred shifts towards neutral sequence evolution in the selected branches (A). Contrast-FEL method found ENT domain sites to have slightly higher significant sites under selection in the reference lineages (C).

3.5.3 Cbx1

I have investigated the orthologs of *cbx1* as evidence suggests it interacts with *emsy* (Y. Huang et al., 2006). In *P.caudatus cbx1* had two paralogs with only one having a maternal expression. The *cbx1* found in the maternal transcriptome is not significantly downregulated during MZT. Both *T.transversa* and *H.exemplaris* transcriptomes contain the *cbx1* ortholog. Only in *T.transversa* does it have a maternal expression without a significant downregulation during MZT.

The temporal dynamics from *P.caudatus* and *T.transversa* show an expression peak in oocytes which declines as development proceeds. The decline is most apparent after gastrulation. The *cbx1* transcripts show a localized distribution in *P.caudatus* oocytes. Their distribution is even and restricted to the germinal vesicle (Figure 11).

The *cbx1* orthology for the 3 species has been confirmed through shoot.bio (Emms & Kelly, 2021). interface. Chromobox protein family evolution has been previously described (Flanagan et al., 2007; X.-Y. Liu et al., 2017; Lomberk et al., 2006). No in-depth investigation is available on *cbx1* sequence evolution, therefore I set out to perform an in-depth analysis. All metazoan homologs from the Ensembl database were retrieved with the addition of *P.caudatus* and *T.transversa* sequences. I used the presence of a chromodomain as a filtering criterion (Jacobs et al., 2001). Following filtering 553 *cbx1* orthologs were retained. The most optimal substitution model for the CDS alignment was that of general variable time matrix model in combination with FreeRate (7 categories) rate heterogeneity model (VT + R7). Convergence for the ML tree was reached after 1001 iterations. The resultant gene tree estimate resembled the e*msy* gene tree in the sense of the dichotomy between branch lengths in chordate and protostome lineages (Figure 16Figure *16*). The orthologs of the chordate *Branchiostoma lancelatum* grouped outside of the chordate clade. Orthologs from species such as *Anolis carolinensis* or *Eptatretus burgeri* showed high divergence compared to the rest of the chordates. Some species showed signs of duplications such as the rotifer *Adineta vaga* and the brachiopod

Lingula anatina. Both *P.caudatus* and *T.transversa* orthologs are part of the sister clade to all chordates.
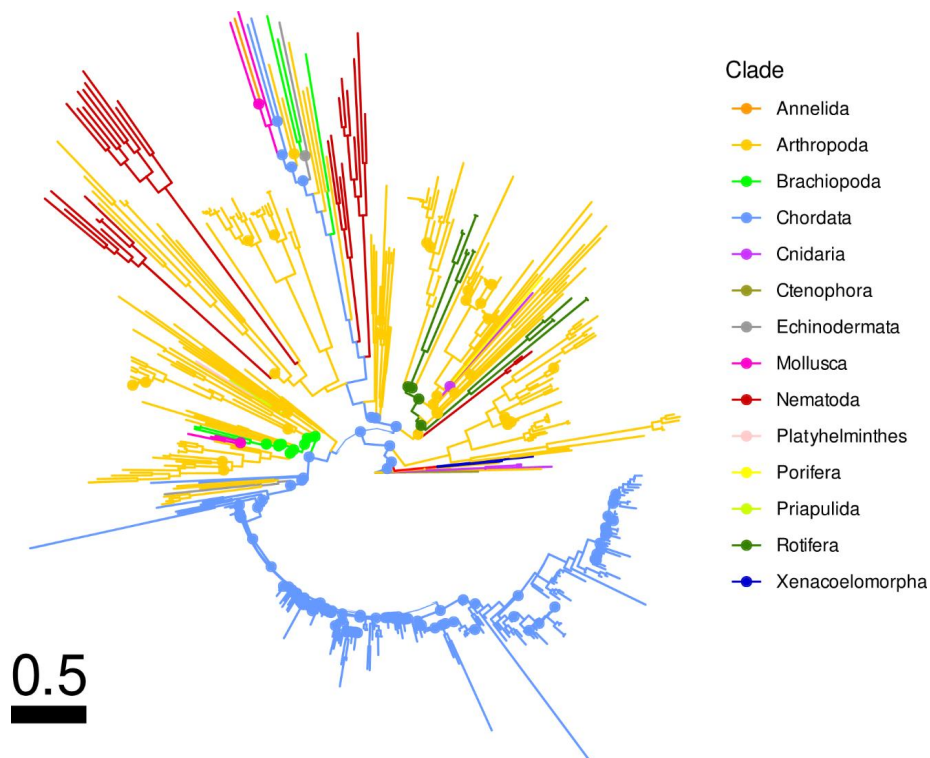
Figure 16. Rooted ML phylogeny of *cbx1* gene tree based on CDS alignment. Colors denote metazoan clades. Bootstrap support with values smaller than 80% are highlighted at nodes as points. Scale bar displays substitution rate for each codon.

I have performed selection analyses for the *cbx1* sequence using the Hyphy software suite (Figure 17). Within the alignment I have located the chromo domain and a fragment of the chromo shadow domain. The chromo domain showed higher proportion (68%) of sites under significant (p-value < 0.001) purifying selection compared to the CDS (8%). The chromo shadow domain had fewer sites under purifying selection (26%) compared to the chromo domain. The branch length patterns indicate a relaxation of selection in protostome lineages. Therefore, I have employed RELAX and contrast-FEL approaches for hypothesis testing. As reference I have chosen the chordate lineages without the

outlier *B. lancelatum*. I found a significant (p-value < 0.001) relaxation of selection in the protostome lineages with a selection intensity parameter of 0.61. The site-wise ω proportions under the alternative hypothesis showed a shift towards a neutral rate value. I have found no signs of differential selection within the chromo shadow domain between the test and reference branches. The chromo domain showed similar proportions (2.6%) of sites under differential selection as the CDS (2.4%).

A.

B.
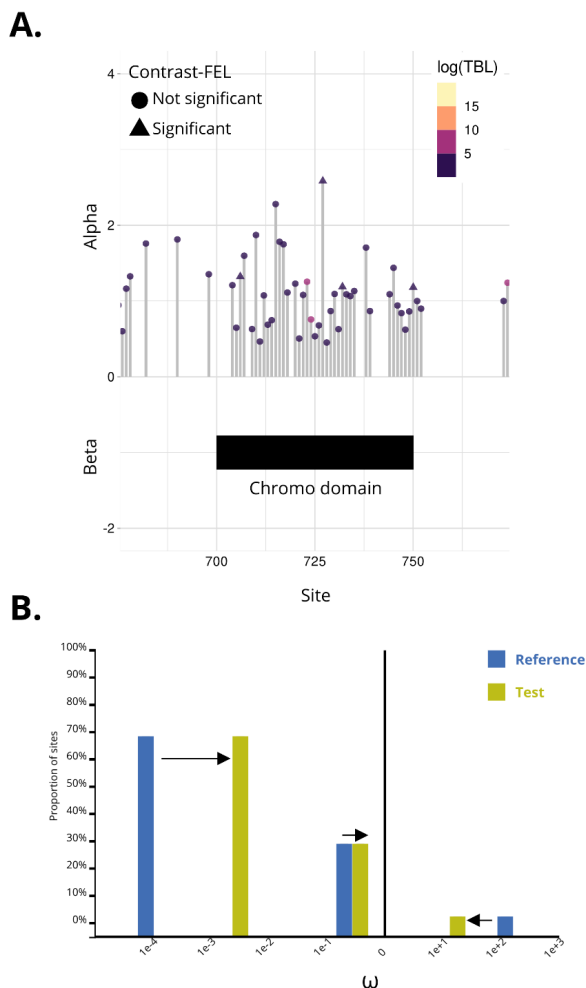
Figure 17. *Cbx1* sequence evolution. FEL analysis inferred selection analysis (A). Highlighted is the chromo domain. The test branches show a significant relaxation of selection (B) with shifts of the site proportions towards a neutral ω.

3.5.4 Axon guidance signaling

Several receptors involved in axon guidance were detected in the maternal *T.transversa* transcriptome (Figure 18). No in-depth phylogenetic analysis was performed as their evolutionary histories were thoroughly investigated before (Junqueira Alves et al., 2019;

Xu et al., 2021; Yu et al., 2014). Nevertheless, their orthology was validated through shoot.bio's web service.

The *robo* receptor (Kidd et al., 1998) is strongly detected in the maternal transcriptome of *T.transversa*. It's expression declines by gastrulation after which a steady increase commences during development. *In situ* hybdridization revealed that the *robo* transcripts are dispersed evenly in the *T.transversa* oocyte cytoplasm. It's ligand (Cecile et al., 2007; Rothberg et al., 1988), *slit* is found in the transcriptome as 3 separate paralogs. Temporal dynamics for the three paralogs are similar with peaks during gastrulation. I have detected weak expression for the ligands in the maternal transcriptome. The lack of expression in oocytes was confirmed with *in situ* hybridization.

Several *plexin* receptor genes (Winberg et al., 1998) were identified in the *T. transversa* maternal transcriptome. Two *plexin* genes (a class B *plexin* and a class A *plexin*) are expressed in the maternal transcriptome. The remaining two class A *plexin* genes were identified as artefacts from the *de novo* assembly pipeline. Both expressed *plexin* genes show a peak in expression in early development. Their expression declines during gastrulation and rises again during larval stages. The presence of *plexin* transcripts in oocytes was confirmed with *in situ* hybridization. Class A *plexin* transcripts are evenly distributed in the cytoplasm of the oocytes. Class B *plexin* transcripts are localized around the germinal vesicle.

A.

Gene Z-score

Ttra000526_Semaphorin-5A
Ttra001605_Semaphorin-1A
Ttra001066_Protein slit
Ttra002432_Sliprotein
Ttra010320_Nanos
Ttra011062_Chromobox protein
Ttra013696_Nanos
Ttra001906_Semaphorin-2A
Ttra000079_Plexin-A
Ttra000102_Roundabout
Ttra014045_Emsy like
Ttra000136_Plexin-B
Ttra002225_Semaphorin-5B

oocytes, mid blastula, late blastula, moving late blastula, early gastrula, mid gastrula, late gastrula, bilobed late gastrula, trilobed late gastrula, early larva, late larva, competent larva, 1 d juvenile, 2 d juvenile

B.

Plexin B | Plexin A | Roundabout

anti-sense

sense

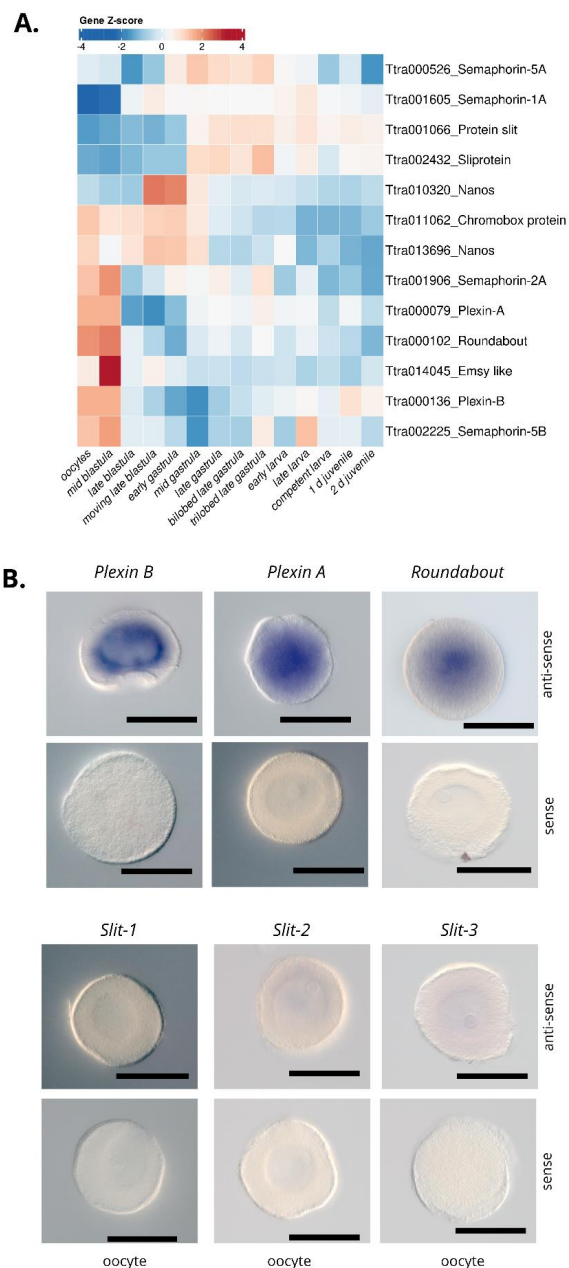Slit-1 | Slit-2 | Slit-3

anti-sense

sense

oocyte | oocyte | oocyte

Figure 18. Axon guidance molecule expression found in oocytes of *T.transversa*. RNAseq data displayed in a heatmap (A). *In situ* hybridization experiments with controls. Scale bar: 0.1 mm

3.6 Patterns of maternal gene expression across Metazoa

I have found that on average the maternal genes comprise 41% of all the genes present in the reference transcriptomes in the surveyed species (Figure 19 A). The highest proportion is found in *T.transversa*, where 71% of all transcripts are expressed in oocytes. I have separated the gene expression distribution into 4 categories: high expression where TPM ≥ 1000, medium expression where $100 \leq$ TPM < 1000, low expression where $2 \leq$ TPM < 100. Genes below 2 TPM are considered as part of baseline noise (Wagner et al., 2013). A shared feature across all species is the lack of genes with high expression (Figure 19 A). Patterns for medium or low expression genes varied across species. In Hexapoda maternal gene expressions are skewed towards medium expression categories. In contrast echinoderm species show an enrichment in low expression categories. Maternal genes undergoing degradation during MZT are abundant in medium expression categories. Outliers to this consensus are *H.exemplaris* and *T.castenum*. Additionally, in the degraded maternal gene set the high expression category profiles are more abundant. Genes with TPM less than 2 can also be found in the degraded maternal genes. A biological source could be that these transcripts are present in oocytes without poly-A tails. Without such tails they would be excluded from libraries based on poly-A capture. These transcripts would then be detected in later developmental stages due to cytoplasmic polyadenylation (Fernando et al., 1994; Radford et al., 2008; Winata et al., 2018). These genes can be degraded before the ending of MZT and therefore included in the degraded subset of maternal genes. A technical source could be that of low count genes with high variance being flagged as differentially expressed.

Fold changes during MZT for maternal genes vary greatly across species. Both decreases and increases of maternal genes were observable (Figure 19 B). Generally, the negative fold changes are associated with degraded genes and genes with high expression values. Positive fold changes on the other hand are associated with low gene expression. Maternal genes without degradation can have both positive and negative fold changes associated with them. Some maternal genes have up to 29 log2 fold change

downregulation such as *Cdc6,* a key factor in oocyte maturation Anger et al., 2005; Lemaître et al., 2002).
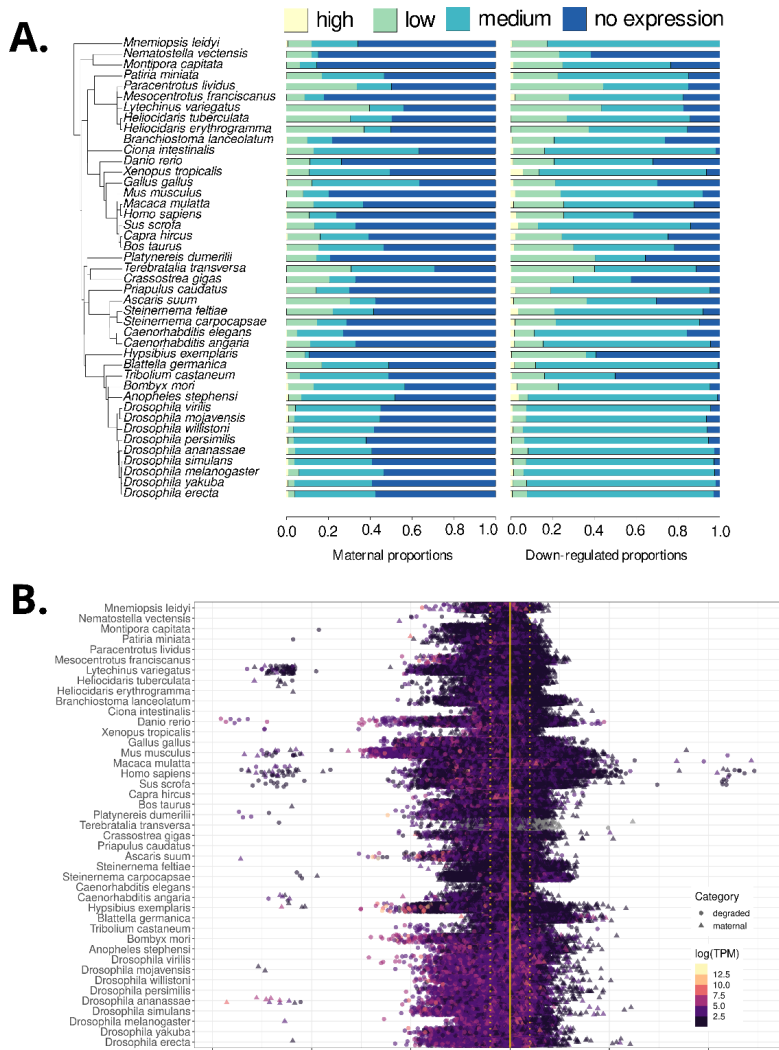

Figure 19. Patterns of maternal gene expression across studied Metazoan species. A variation in the proportion of maternal gene expression magnitudes (A) and fold changes during the MZT (B) is observable across the species tree.

To get a better insight into the general functions of maternal genes I have performed GO enrichment analysis in all species. The results show that degraded genes are less enriched for GO terms, compared to maternal ones (Figure 20 A). Moreover, degraded genes are dissimilar and show sparse overlaps in functions. I have found higher degree of overlaps in functional profile for the maternal genes (Figure 20 B). Overlap distributions across the two subsets showed a significant (Wilcoxon rank sum test p-value < 2.2e-16 with moderate effect size of 0.468) difference (Figure 20 C). Common shared terms across maternal genes were related to RNA splicing, RNA transportation, DNA related functions or protein metabolism (specifically functions involved in degradation). Apart from this piRNA metabolic processes and epidermal growth factor receptor signaling pathways were also enriched.

I have reconstructed species phylogenies using gene expression data. First, I have discretized maternal gene expression data. The resulting matrix was used to reconstruct a phylogeny across species. The best fitting substitution model was that of MK+FQ+R5. A ML tree was found after 132 iterations (Figure 21 E, F). The species phylogeny was not completely recapitulated. The ecdysozoan clade is recovered with the exception of *A.suum* and *H.exemplaris.* These, together with *C.gigas* and *N.vectensis* form the sister clade of Tetrapoda. The misplaced species have relatively long branches, indicating a spurious change in expression values. Echinoderms in conjunction with tetrapods and ecdysozoans retain the internal species phylogeny while having the remaining basal chordates, the ctenophore and one cnidarian as sister groups (Figure 21 C).
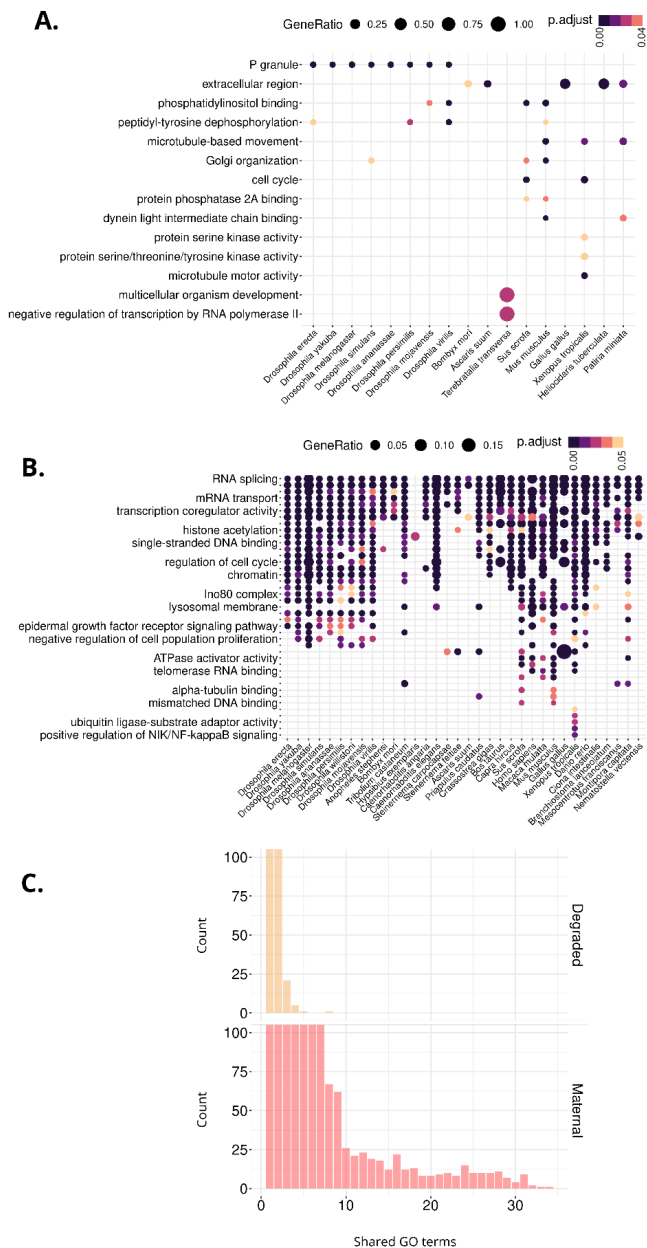
Figure 20. Dotplot of GO enrichment across the two maternal categories. Degraded genes are scarce in GO terms as displayed by the scarce overlaps across species and terms (A). Maternal genes have shared GO terms (B). Counts of common GO terms across categories (C).

Next, I have reconstructed species phylogenies based on presence/absence of gene expression values. The underlying species phylogeny was not recapitulated for either category. For maternal genes the best fitting substitution model was GTR2 + FO + R4. The ML tree has converged within 127 iterations (Figure 21 C, D). The maternal genes reconstructed an approximate of the species tree with a few misplaced species (e.g. *S.scrofa* or *D.melanogaster*). The clade with the misplaced species is separated by a long internal branch from the rest. In this framework this implies more frequent gains or losses of maternal gene expression for this clade. The estimated state frequencies have also been inspected. Absence frequency was estimated at 89.54%.

ModelFinder found GTR2 + FO + R4 + ASC as the best fitting substitution model for the degraded gene binary data. ML convergence was reached after 143 iterations (Figure 21 A, B). For degraded genes more frequent positional shifts were observed. Both lopchotrochozoan and deuterostome clades are split up. Ecdysozoan species retain major clades, such as Nematoda. Additionally, low support nodes are frequent throughout the tree. The maximum likelihood absence frequency was estimated at 94.52%.

To the visual inspection of the phylogenies I have added a quantitative comparison too. Using Robinsons-Foulds (RF) distances I detected the degraded maternal gene expression presence/absence phylogeny to be the most dissimilar to the underlying species phylogeny (RF 56). The maternal presence/absence tree showed higher similarity (RF 54) and the discretized expression based phylogeny was the most similar (RF 44). I found this lineup to be persistent when considering branch length normalized RF distances.
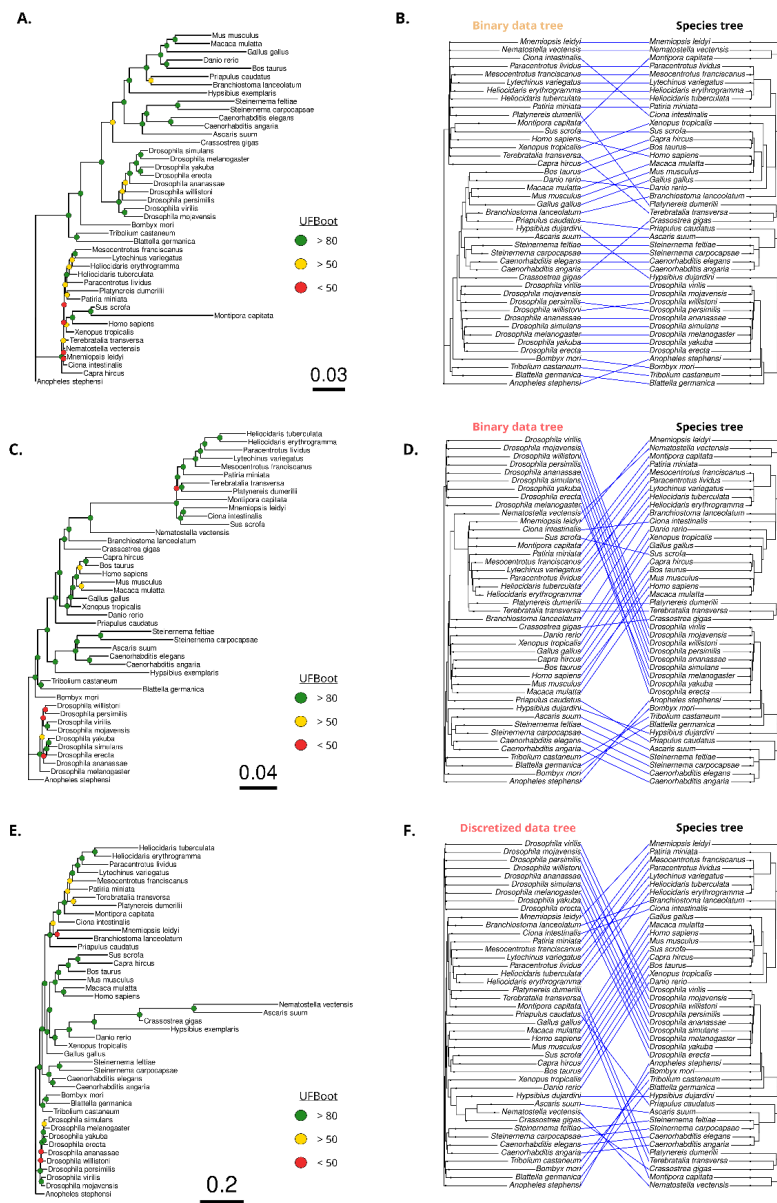
Figure 21. Reconstructed ML phylogenies based on gene expression presence/absence and expression values compared to the species phylogeny. Cophylogenetic plots display the character based ML tree on the left with nodes with UFBoot values < 80 highlighted in red.

As UTR lengths can impact transcript stability (Mishima & Tomari, 2016b) I have set out to compare the structural features of maternal genes (Figure 22). As a reference I have quantified genes not categorized as maternal within each species. The 3`-UTR lengths are imbalanced across species. Ecdysozoan species contain shorter 3`-UTR lengths in comparison to deuterostome species. *C.hircus* is a stark outlier with long 3`-UTR sequence lengths. The 3`-UTR lengths have not been normalized across species therefore caution is warranted when interpreting these comparisons. Differences could also arise due to differing genome annotation qualities. The sequence lengths for reference and maternal genes vary greatly on a species-specific basis. Generally maternal genes and degraded genes are characterized by longer 3`-UTR sequences within each species. This could be an indication for post-transcriptional regulation being more important of a factor in maternal genes as compared to the rest of the genes. However, exceptions to this are present in species *H.sapiens* or *S.carpocapsae* which have equal 3`-UTR lengths. Maternal genes have shorter 3`-UTR sequences compared to the degraded ones in hexapod species. This could be a possible indication that for this clade stability of maternal transcripts is more reliant on 3`-UTR sequences.

I have also compared other architectural features across the categories. Longer exon lengths were detected for the majority of ecdyosozoan species. Exon lengths outside Ecdysozoa were constant with marginal differences across gene categories. When inspecting the CDS the species-specific length differences in exonic sequences diminished. Less variability was noticeable across the tree with maternal and degraded genes having longer CDS. Intron lengths showed an opposite pattern. Deuterostome species could be characterized by longer intronic sequences for maternal categories. Ecdysozoan species were characterized by shorter introns with less difference across categories. A notable outlier is *G.gallus* where degraded genes have long exonic sequences flanked by short intronic sequences. The 5`-UTR sequence lengths showed less variability across species with a few nematode exceptions.
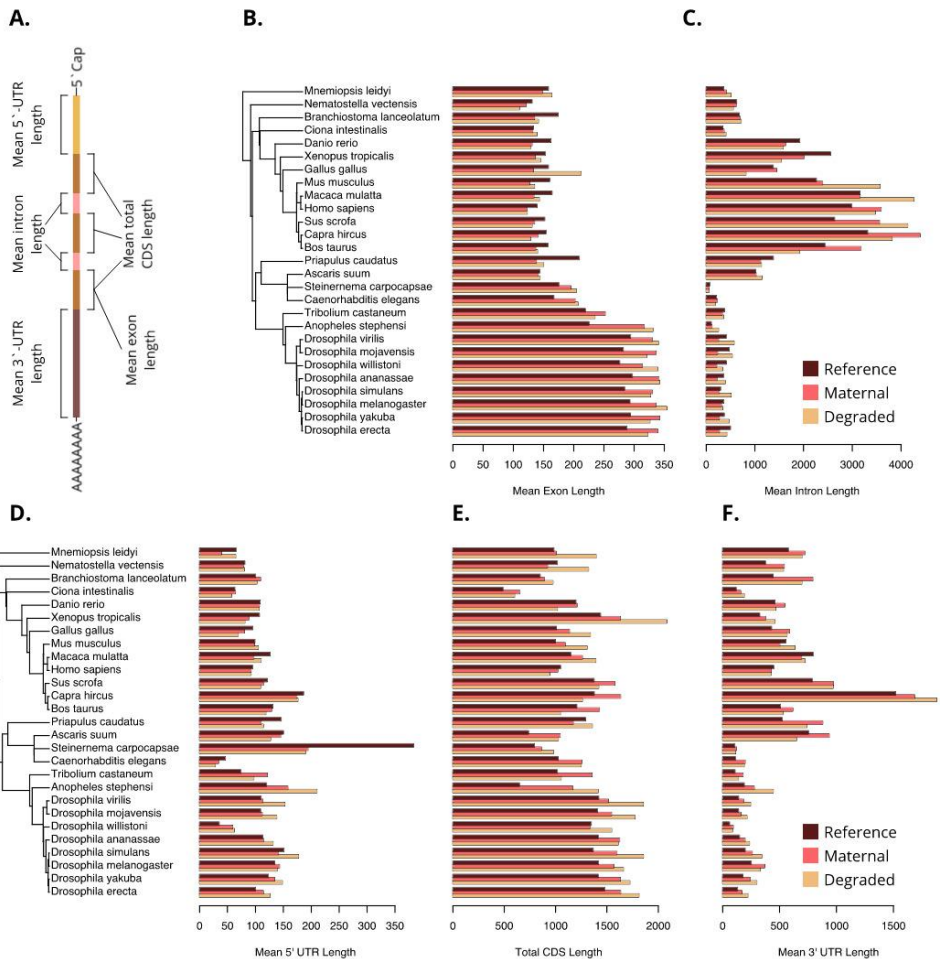
Figure 22. Architectural characteristics of maternal genes across the studied species. Five features were compared for which a visual summary is provided (A). Species lacking this information were discarded from the analysis.

Orthology inference resulted that 87.7% of all genes were grouped into orthogroups. The sizes of orthogroups shows a multimodal distribution with a heavy skew towards smaller orthogroup sizes (Figure 23 A). A bimodality is present when inspecting the maternal genes but is absent from the degraded genes. All species are represented in orthogroups to varying degrees. Generally, the maternal genes are found in more orthogroups

compared to the degraded subset of those without any phylogenetic relationship observable behind this pattern (Figure 23 B).

I have quantified the maximum distances between tips for each pruned orthogroup (Figure 23 D). This has showed a bimodal distribution for maternal genes (p-value < 2.2e-16). The two modes showed differences in fitted OU models (see 3.7 Trait evolution model fittings). Degraded genes showed significantly (Wilcoxon rank sum test p-value < 2.2e-16 with large effect size of 0.536) shorter distances across tips compared to the maternal genes. Together, this suggest that maternal genes are common across a wider array of species separated by more branching events compared to the degraded genes.

I have inspected duplication dynamics for gene categories using OrthoFinder's estimates for gene duplication events (Figure 23 C). Proportionally, the deeper nodes within the species tree experienced duplications within orthogroups containing maternal genes. The degraded genes and reference genes show duplication events in shallower nodes. In the Ecdysozoa node no duplication events were inferred after filtering for their support values. Mean duplication events at the tips for all three categories were more abundant compared to the internal node inferred duplications. In both cases the maternal gene containing orthogroups were inferred to have experienced the highest average duplication events.
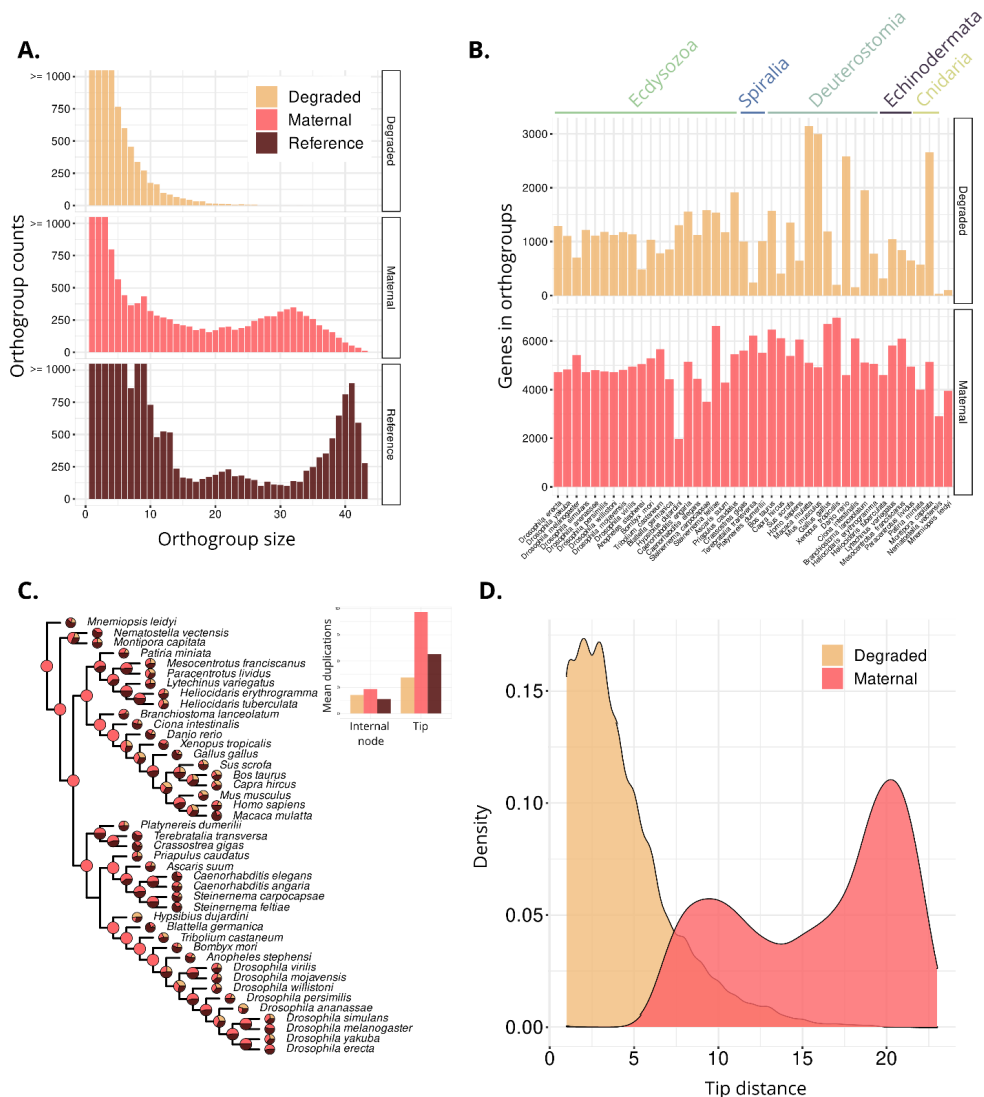
Figure 23. Orthology relationship inference results across different categories of maternal genes. The distribution of orthogroup sizes decreases in degraded subset of genes (A). This is again apparent if looking at species specific genes in orthogroups (B). Relative duplications (displayed as pie charts at each node on the species tree) show that maternal genes experienced duplication events throughout the species tree, whilst degraded ones were more frequent in terminal nodes (C). Tips had more duplications events assigned to them (C). Tip distances between the furthest species in each orthogroup (D)

3.7 Trait evolution model fittings

3.7.1 First round of model fittings

I have fitted evolutionary trait models to expression data from each orthogroup. Model fit was tested by parametric bootstrapping (Garland Jr. et al., 1993; O'Meara et al., 2006). Model reproducibility filtering was passed by 7422 orthogroups for maternal genes and 713 orthogroups for degraded genes. Reproducibility in maternal categories was high for null models, followed by OU model and finally Brownian motion models. For degraded category both null models and Brownian motion models were reproduced with high confidentiality. OU models were difficult to reproduce in this category. Model parameters were also reproduced with high confidence in orthogroups passing reproducibility filtering. A variation in tree sizes was noticed across the two categories. Degraded genes featured smaller tree sizes overall compared to maternal genes. Null models and Brownian models in both categories had smaller tree sizes. OU models were characterized by bigger tree sizes following pruning, as OU models are sensitive to tree sizes (Cooper et al., 2016).

I have detected considerable differences when comparing the proportion models between the two sets of maternal genes. In both categories null models were identified to varying degrees. The degraded genes had higher representation of null models, possibly a consequence of model misspecifications due to the smaller tree sizes. Degraded genes were abundant in Brownian motion models, with OU models being rare. A stark opposite was identified in maternal genes. OU models were detected as the main mode of trait evolution for this category with Brownian motion models also represented. The difference between the maternal genes and degraded genes for Brownian motion models and OU models was found to be significant (χ^2 test p-value < 2.2e-16).
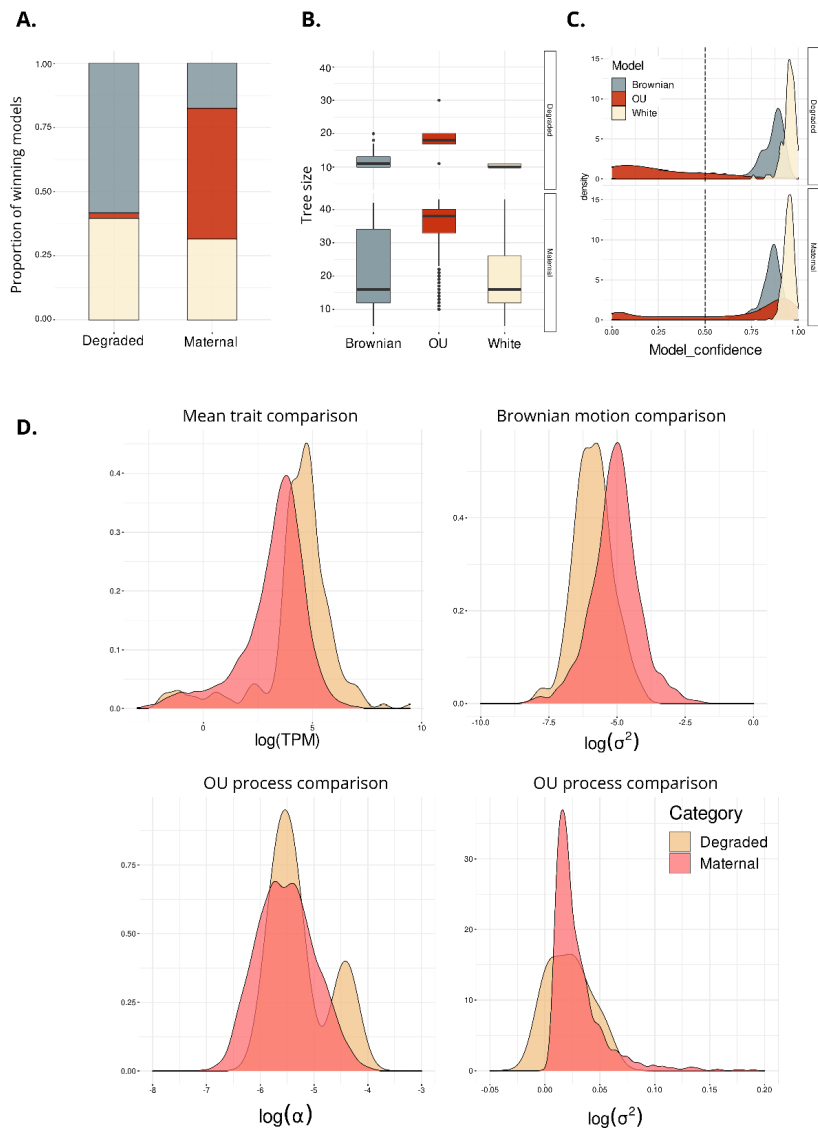
Figure 24. Results of first round for model fittings. Majority of maternal gene expression values evolve according to an OU model, whilst degraded ones show a Brownian motion model enrichment (A). Pruned tree sizes vary across categories, but not across models (B). Model replicabilities are skewed towards higher ends in general (C). Differences in parameter estimates are displayed (D).

Comparing the estimated parameters (Figure 24 C) some significant differences were found across the subsets of maternal genes. The differences were detectable in mean expression values and the σ^2 of the Brownian motion model. Maternal genes have significantly higher expression values and higher σ^2 parameter values (Table 6). This difference could be attributed to differences in pruned tree topology and total branch lengths variables which could impact model inference (Boettiger et al., 2012).

In addition, the maternal pruned tree tip distances (Figure 23 D) showed differences in the most optimal model reproducibility. For both modes the Brownian motion models and null models were similar. Modes with shorter distances between tips had OU models with low reproducibility, whilst the mode with longer distances was characterized by high confidentiality OU models. This reflects the sensitivity of OU models to tree sizes (Cooper et al., 2016).

I have also fitted binary trait evolutionary models (Mk model) to the gene expression data. I considered expression (TPM $\geq$ 2) or no expression (TPM < 2) as the two states. The fitting resulted 2803 maternal and 219 degraded genes passing reproducibility filtering steps. Null models were prevalent for both categories. Apart from the null models only equal rate (ER) models passed filtering steps. Both model types had high reproducibility values (both centered at 80%). Parameters were also reproduced with high confidence. Pruned tree sizes varied between the winning models. Null model tree sizes were bigger than the ER model tree sizes. For the equal rate models, I found a significant (Wilcoxon rank sum test p-value < 2.2e-16 with a moderate effect size of 0.358) difference between the maternal and degraded categories. Degraded genes showed more dynamic transitions between losing or gaining gene expressions compared to the maternal genes. Thus, degraded genes display more dynamic gene expression gain or loss. A notion in line with expanded orthogroups with Brownian motion models. The high prevalence of null models and lack of orthogroups where trait models have varying transition rates between expression gain or loss events is a reflection of low sample sizes. Increasing the dataset would result in better understanding of such transition rates.

Table 6. Table summarizing parameter comparisons. Bold font highlights parameter being significantly different and of relevant effect size.

Model	Compared parameter	P-value	Effect size	Magnitude
All	log(TPM)	1.67e-44	1.43e-01	small
Brownian		2.56e-28	2.82e-01	
	σ	2.68e-36	3.21e-01	**moderate**
OU	log(TPM)	2.76e-01	1.44e-02	small
	σ	4.08e-01	1.09e-02	
	α	1.46e-01	1.92e-02	

I have detected *nanos* having a maternal expression after filtering, it has expression in the maternal transcriptomes of 39 species. The best fitting model of trait evolution was that of OU with 93% reproducibility. Simulated parameters replicated the estimated parameters (Kolmogorov-Smirnov goodness of fit test p-value 0.7921). For *nanos* a multi-regime OU model also showed a better fit (likelihood ratio test p-value 2.3e-07). Two new regimes were added where expression values are close to 0. I found that *emsy* has only maternal expression profile after filtering for tree size (Figure 25 A). It has expression in 36 species out of 43 and the best fitting model for this orthogroup was that of the OU model with high reproducibility (98%). The parameters were also reproduced successfully in simulations (Kolmogorov-Smirnov goodness of fit test p-value 0.831). More complex OU model with multiple regimes proved to be a better fit (likelihood ratio test p-value 0.024). The two added regimes were displayed only in single species such as *C.hircus* where a loss of *emsy* expression had been detected The other regime was exemplified by *M.musculus* with a slightly higher expression compared to the rest of the species. *Cbx1* was expressed in 40 species but found to be degraded only in 18 species. For both categories the best fitting model was the Brownian motion model (92% and 93% reproducibility respectively). The axonal guidance receptor gene expressions generally had the null model as best fitting model (*plexin-B*, *robo*). *Slit* was split into 2 orthogroups, one of these had the Brownian motion model as best fitting model with 95%

reproducibility, while the other one had OU as the best fitting model with low reproducibility (30%). In all cases the simulated parameters were concise with estimated parameters (Kolmogorov-Smirnov goodness of fit test p-value > 0.05) and the pruned tree sizes used for model fittings varied from 21 to 27 species.
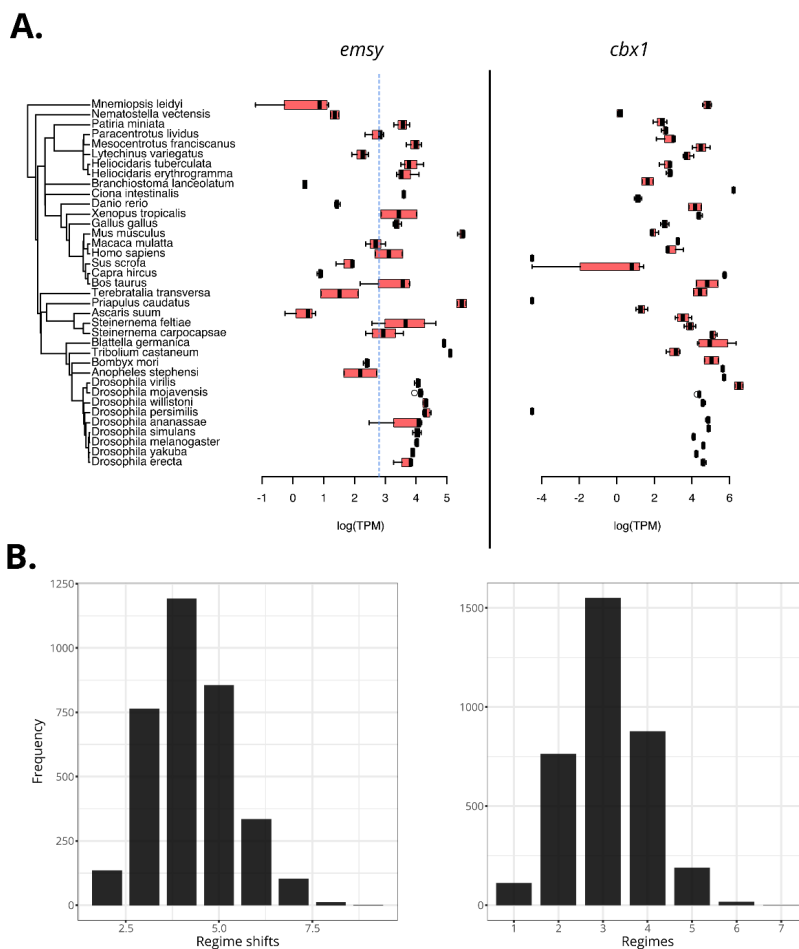
Figure 25. Gene expression across the studied species tree. Expression distributions of *emsy* and *cbx1* are presented across the species phylogeny (A). As *emsy* was found to follow an OU process the estimated Ө is also highlighted with the blue dotted line. *Cbx1* follows a Brownian motion model, therefore no Ө was estimated. Distribution regime shift numbers and the absolute number of regimes inferred by the surface algorithm (B).

3.7.2 Second round of model fittings

I have performed a second round of more complex trait evolutionary model fittings. This analysis was performed in a way that the best fitting model was expanded with a variable. For Brownian motion models the new variables introduced are multiple σ^2 parameters. For OU model a new regime is added in the form of new optima. Competing model fits were compared using likelihood ratio tests or comparing AICc scores.

My analysis returned more than 4000 orthogroups with multiple regimes (Figure 25 B). Shifts were found at varying depths of the tree and varying numbers with multiple of them being inferred as convergent shifts. A common pattern for newly added regimes was that of decreasing gene expressions in species or clades. A few example genes for such complex models are for *tektin, prdm6* or *dmrtd*. *Tektins* are cytoskeletal proteins involved in various functions, most notably they are enriched in sperm cilia and contribute to sperm motility (Roy et al., 2009). Their maternal expression was found 34 species. Two regimes were detected for this gene, one with expression and one without expression. Expression has been lost in ecdysozoan and vertebrate species. The gene, Doublesex- and mab-3-related transcription factor C2 (*dmrtd*) (Figure 26) was detected with two regimes with similar expression patterns as in the case of *tektins*. Based on these findings it is unknown if expression has been gained in Bilateria and subsequently lost in drosophilids and *P.caudatus* or if expression has been convergently lost in basal groups and the mentioned ecdysozoans. PR domain zinc finger proteins (*prdm6*) are known to be involved in transcriptional inhibition through recruitment of histone methyltransferases (Davis et al., 2006). was found in the dataset which was previously connected with male germ cell specification (Li et al., 2020). Here 3 regimes were detected. The regimes range from not expressed to low expression and finally to high expression. This gene displayed expressions within Placentalia and *B.mori.* Apart from these groups the expression is lost or weak.
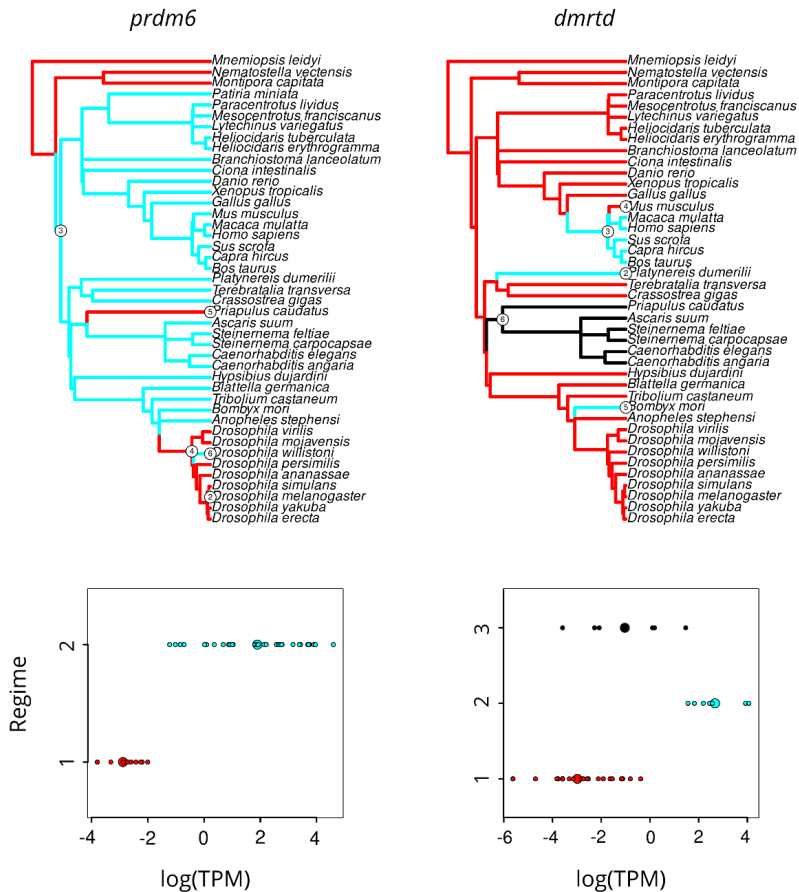
Figure 26. Examples of multi-regime model fittings. The inferred regimes are color coded both on the tree and on the adjacent plot below with corresponding colors. The adjacent plot displays the TPM distribution for each inferred regime with a bigger circle indicating the Ө for the respective regime.

I have detected AICc score improvement when considering the more complex heterogenous Brownian models in 4 genes from the degraded subset. For the rest of the genes 349 model fits improved upon modifying the base Brownian motion models. Generally, 2 rate models were the most common. The σ^2 could be either increased or decreased in one of the lineages when 2 rate models were the best fitting ones. An example for the former is the gene *mafA* which is known to partake in activating gene

expression upon insulin signaling (Martin et al., 2002). The default Brownian motion model for this gene was improved by a two rate model in which an increase is noticeable in the σ^2 within vertebrate lineages. Conversely a decrease of σ^2 can be observed in the *rab1* gene, intracellular membrane trafficking regulator (Mukhopadhyay et al., 2014). Hexapod species display lower σ^2 compared to the rest of the species. The *src42* gene which has a role in cell motility (M. Takahashi et al., 2005) had a similar pattern. Spiralian species show a decrease in σ^2 compared to the rest of the tree.

Figure 27. Examples of heterogeneous model fittings. Two rate Brownian motion models are exemplified with branch length transformations scaled by the respective σ^2 rate. These branches are also colored red.
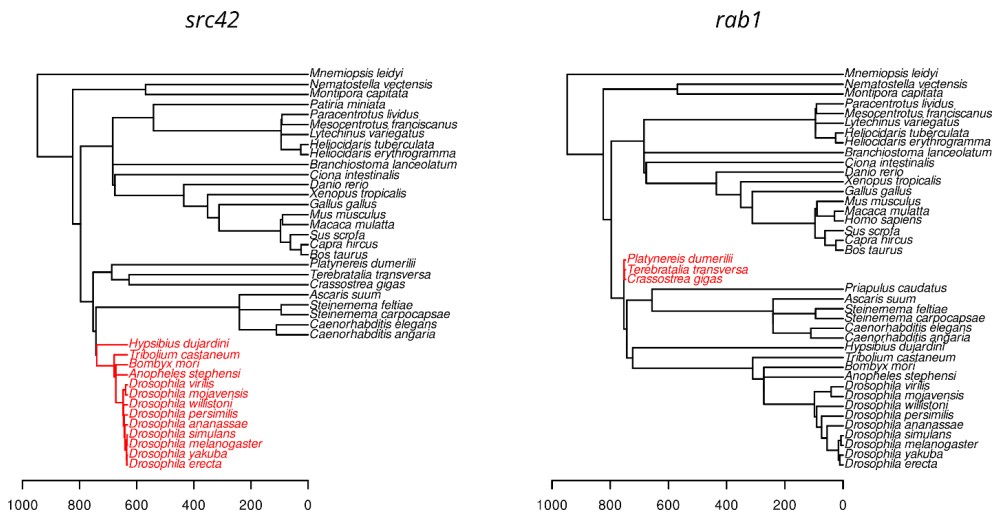
4.Discussion

4.1 The maternal transcriptome and the MZT of non-traditional model species

In the three selected species I have successfully identified a catalogue of maternal genes which could be further investigated. These species hold a great potential for inquiring evolutionary questions due to their phylogenetic positions and evolutionary rates (Carlson, 2016; Gabriel et al., 2007; Wennberg et al., 2008). An extensive insight into maternal genes was gained through *D. melanogaster* (Atallah & Lott, 2018; Kobayashi et al., 1996; Nüsslein-Volhard & Wieschaus, 1980; Thomsen et al., 2010) and *C.elegans* (Kemphues et al., 1988; Stoeckius et al., 2014; L. Wang et al., 2002). *H.exemplaris* and *P.caudatus* both are valuable sources for understanding maternal gene evolution within Ecdysozoa. For the sister group of Ecdysozoa, the Lophotrochoza, information about maternal genes is restricted to only a few lineages (Chou et al., 2016; S. E. Lee & Jacobs, 1999; M. M. Liu et al., 2014; Nakamura et al., 2017). For this clade the *T.transversa* provides a novel source of information regarding maternal genes. In the following section I will discuss what novel information have these species provided.

For a competent oocyte several molecular complexes are required, ranging from meiosis regulators (von Stetina & Orr-Weaver, 2011) to RNA binding proteins (Kobayashi et al., 1996; Mosquera et al., 1993; Semotok et al., 2005; Tadros et al., 2007; Zhu et al., 2018) and protein modification factors (Gu et al., 2010; Vitale et al., 2007; Y. Wu et al., 2021). All such expected housekeeping functions are present in all three species, adding further confirmation to the conservation of such functions (M. M. Liu et al., 2014). A notable difference was spotted in the use of RBPs across the three species. These differences were apparent in the RBDs and the enriched binding sites alike. In all three species transcripts for the core RNA degradation complexes were present in the maternal transcriptomes such as deadenylating CCR4-NOT complex (Denis, 1984; H.-Y. Liu et al., 1998) or factors from the decapping complex (E. van Dijk et al., 2002; Zuoren et al., 2002) and exonucleases (Muhlrad et al., 1994, 1995) . Endonucleases (E. L. van Dijk et al., 2001) were absent except from *H.exemplaris*. As endonucleases have been connected

previously with RNA interference (Lingel & Izaurralde, 2004; Martinez & Tuschl, 2004) 2004) and the presence of endonucleases was shown only in *H.exemplaris* I propose that small RNAs are the main players in targeting the maternal transcripts for degradation within this species. Additional evidence to support this is the enrichment of motifs recognized by the RISC complex (Lingel & Izaurralde, 2004; Martinez & Tuschl, 2004) and the functional enrichment of piRNA pathway. What is left unknow is the key small interfering RNAs and micro RNAs targeting the maternal transcripts. Interestingly in *P.caudatus* the RNAi does not play a pivotal role in targeting maternal transcripts for degradations. As we have evidence support RNA interference having a role in early development from ecdysozoan model systems (Dexheimer et al., 2020; Ninova et al., 2016) but not from *P.caudatus*, it would imply that the RNAi during the MZT is lost in *P.caudatus*. The reason behind this possible loss remains to be elucidated. My results also indicate that despite having the RNA degradation complexes available, *T.transversa* is not utilizing maternal transcripts encoding RBPs extensively during MZT. This is further reinforced by the lack enriched motifs recognized by RBPs recruiting the RNA degradation complexes (Laver et al., 2015; Semotok et al., 2005). In contrast, *P.caudatus* is both enriched with RBD containing transcripts and those transcripts are also more abundant, implying a functional role for them (Nadler et al., 2006). My results are insufficient in identifying what are the mechanisms recruiting the RNA degradation complexes. Many mechanisms could be utilized (see 1.1.3 Clearance of maternally provided transcripts). My results also hint towards ELAV-like proteins being of great importance for maternal RNA stability in the studied species. ELAV proteins are known to be involved in stabilizing their targets by binding to ARE elements (Soller & White, 2003, 2005). All studied species had motifs recognized by ELAV proteins. This could be attributed to a bias in the motif databases used for enrichment. A caveat for RBD motif enrichment is the limited availability for these sequences in databases (Giudice et al., 2016; Ray, Kazan, Cook, Weirauch, Najafabadi, Li, Gueroussov, Albu, Zheng, & Yang, 2013). Furthermore, divergence of motifs could hinder the transfer of functional knowledge from species to species. However, such divergences have been shown to be

bypassed by ELAV proteins (Haussmann et al., 2011). Moreover, another line of evidence supporting ELAV stabilizing functions is that a big proportion of the target transcripts displayed no decay during the MZT. A great amount of attention has been given to factors regulating the degradation of maternal transcripts (see 1.1.3 Clearance of maternally provided transcripts), less so on the factors stabilizing maternal transcripts for longer periods of times. Taken together ELAV proteins could offer a great insight into the maternal to zygotic transition.

Temporal dynamics of maternal genes in the studied species displayed similarities. Clustering revealed that the studied species most maternal genes are relevant and specific to early developmental stages. This is also supported by the lack of expression in later developmental stages for most identified clusters. Moreover in *P.caudatus* and *T. transversa* the availability of intermediate stages provided an insight into the dynamics of maternal transcript decay during MZT. Some clusters displayed an increase in expression during MZT, instead of decay. A phenomenon which could be explained by two processes: *de novo* transcription or better transcript capture during library preparation due to cytoplasmic polyadenylation (Winata et al., 2018). The data presented in this project is insufficient to accurately distinguish between the two alternatives. Nevertheless, the decline of such transiently spiking transcripts after MZT suggests that cytoplasmic polyadenylation is the favorable alternative.

From the maternal gene list, I choose several genes for validation by *in situ* hybridizations. I have chosen *nanos* because it is a well know maternal gene from all studied metazoan species (D'Agostino et al., 2006; Extavour et al., 2005; Jaruzelska et al., 2003; Mochizuki et al., 2000; Mosquera et al., 1993; Zhu et al., 2018). It's conserved maternal expression was further reinforced in *P.caudatus*. It's germ cell specifying function (Asaoka-Taguchi et al., 1999; D'Agostino et al., 2006; Extavour et al., 2005; Jaruzelska et al., 2003) is not confirmed by my analysis. The *nanos* spatial localization within *P.caudatus* oocytes is reminiscent of an intermediate step during oocyte maturation in *X.laeves* (Figure 2) (Kloc & Etkin, 1995). The signal in *P.caudatus* is a

snapshot of the early steps of Vgl1 pathway (Kloc & Etkin, 1995). Later stages of this transportation pathway were undetected. A reason behind this is the difficulty of eggshell penetration of *P.caudatus* oocytes, despite attempting different penetration approaches. Without proper permeability probes cannot diffuse into the samples. As the eggshell is formed upon maturation (Nørrevang, 1965) several events occurring during later oogenetic steps is difficult to visualize. Exploratory *in situ* hybridization experiments in *T.transversa* oocytes indicated that the late stage of the transportation pathway is present in this species. Nevertheless, what is apparent is that *nanos* RNAseq expression follows the MZT profile in both species. Following which its expression decreases to a lower level. In other model species *nanos* is not degraded, rather its expression is restricted to primordial germ cells (Asaoka-Taguchi et al., 1999; D'Agostino et al., 2006; Extavour et al., 2005; Jaruzelska et al., 2003). This restriction to a smaller subset of cells would be detected as a degradation in RNAseq datasets due to a dilution in the bulk libraries. Furthermore, *nanos* displays a spike in expression in cleavage stage *P.caudatus* embryos. This could be a result of cytoplasmic polyadenylation, although in *D.melanogaster nanos* does not undergo cytoplasmic polyadenylation (Fernando et al., 1994) so careful interpretation is warranted. *Nanos* exerts its function through binding RNAs (Asaoka-Taguchi et al., 1999), yet its binding motif was not found in the enriched motif list. The reason behind this is that despite combining multiple motif databases (Giudice et al., 2016; Ray, Kazan, Cook, Weirauch, Najafabadi, Li, Gueroussov, Albu, Zheng, & Yang, 2013) for the analysis, none of those contained the *nanos* motif.

Emsy was chosen out the list of maternal genes as it is poorly characterized (Hughes-Davies et al., 2003; Rana et al., 2011; Viré et al., 2014) and it has been associated with different carcinoma types (Hughes-Davies et al., 2003; Rodriguez et al., 2004; Viré et al., 2014). The data presented in my project indicates that *emsy* is expressed in oocytes. In *P.caudatus* oocytes e*msy* is localized in the germinal vesicle without signal from the cytoplasm. Such a nuclear retention expression pattern has been attributed to several processes. Most notably it has been attributed to a quality control process which ensures

no aberrant transcripts are translated (Fasken & Corbett, 2009; Wegener & Müller-McNicoll, 2018). This is unlikely to be the case in *P.caudatus* as all *in situ* hybridization signal originated from the GV and no cytoplasmic signal was detected. This would imply all transcripts are aberrant and undergoing degradation which is unlikely. Moreover, the RNAseq dataset displays expression of *emys* in cleavage stage embryos. Two alternatives could be more probable: stockpiling of transcripts or prevention of cytoplasmic degradation of the transcripts (Wegener & Müller-McNicoll, 2018). My results cannot clearly distinguish between the two alternatives. The temporal dynamics of *emsy* follows the MZT, indicating a possible degradation of the transcripts. Although the expression is not completely lost despite the significant reduction. A similar explanation as in the case of *nanos* could be suggested, where expression is restricted to subset of cells.

My results provided evidence that *emsy* is not the exception to this localization pattern. The third gene inspected in *P.caudatus* maternal transcriptome was the *cbx1* gene, which was chosen because of its interaction with *emsy* (Y. Huang et al., 2006). *Cbx1* is known as part of the heterochromatin and is involved in epigenetic repression through interacting with methylation signals of H3 histone lysine 9 (Lachner et al., 2001). In *P.caudatus* it exhibits the same spatial and temporal dynamics as *emsy*. The temporal co-expression of these two genes was present in *T.transversa*. Together with their known interaction (Y. Huang et al., 2006), these results hint a conserved shared role in early embryogenesis.

In the list of maternal genes, I found axonal guidance molecules as potential regulators of early development. The Slit-Robo signaling components have been detected in ovarian tissues of several vertebrate species (Dickinson & Duncan, 2010; Qin et al., 2015). No information was recovered of their expression in invertebrate species. In vertebrates, evidence suggests a crosstalk between oocytes and granulosa cells with the Slit-Robo interaction (Dickinson & Duncan, 2010; Qin et al., 2015). This role could be conserved as in *T.transversa* follicular cells play a critical role during oogenesis (Stricker & Folsom, 1997). Interestingly in *T.transversa* oocytes only *robo* expression was detected. The *slit* ligand absence could be explained by expression of it in follicle cells and by paracrine

communication between the cells. This communication channel in *P.caudatus* and *H.exemplaris* oocytes is completely absent. A possible reason behind this is that oogenesis in these two species is a cell autonomous process (Jezierska et al., 2021; Nørrevang, 1965). Without the presence of follicle cells supporting the oogenesis there is no need for paracrine communication through Robo-Slit interactions. This would be in line with the findings from *T.transversa* and vertebrate species and would imply a Metazoa wide conserved role for the Robo-Slit interaction in oogenesis. My results are insufficient to unravel the exact connection and its role in oogenesis.

The Plexin-Semaphorin pairing has also been described by previous efforts as being involved in oogenesis. In *M.musculus* it has been connected to the development of follicular development (Regev et al., 2005, 2007) and in *Drosophila melanogaster* several *semaphorin* expressions have been detected during oogenesis and their role has been connected with guiding the collective cell migration of follicular cells (Khare et al., 2000; Stedden et al., 2019). The role of this signaling in oocytes is yet to be elucidated. Here I found evidence of the signaling to be present in several of the studied species. In *T.transversa*, both partners of the interaction are present (Ayoob et al., 2006), in contrast *P.caudatus* only *semaphorin* expression was detected. *H.exemplaris* displays maternal expression for some of the interacting partners (such as *plexin-A3* and *semaphoring-5B*), but no interaction has been previously described between these two. *Plexin-A3* binds class 3 *semaphorins* as part of a complex together with *neuropillin* (T. Takahashi et al., 1999). *Semaphorin-5B* has no validated binding partner. Without a matching ligand-receptor interaction this signaling is possibly not functional during early development or oogenesis for both *P*.caudatus and *H.exemplaris.* Nevertheless, in *T.transversa* co-expression of the interacting partners suggests a functioning signaling during early development or oogenesis. The *plexin-B* spatial distribution is reminiscent of an intermediate step from the Vgl1 pathway in *X.laevis* (Figure 2). The temporal dynamics for *plexin-B* also highlighted an involvement in early developmental stages as it has two peaks during development: early development and during prospective neurogenetic

stages (Altenburger et al., 2011; Santagata et al., 2012). Its paralog, *plexin-A2* shows a more restricted temporal pattern with peaks only during early development. This could imply a strict function during cleavage stages for *plexin-A2*.

Together, the Robo-Slit and Plexin-Semaphorin cell-cell signaling partners described here add to the notion that these pathways could be important in oogenesis. Their exact function cannot be elucidated with the results presented here, nevertheless the above presented genes all have been implicated in various human cancer types, especially ovarian cancers (Dai et al., 2011; Hollis et al., 2019; Hughes-Davies et al., 2003; Rodriguez et al., 2004; Viré et al., 2014; Ye et al., 2010). Further investigation of maternal genes as these could provide valuable insights for cancer biology.

4.2 Evolution of selected maternal genes

Seeing the possible importance of the selected maternal genes I have set out to investigate if these findings could be transferred to other species. For this I have investigated in depth sequence evolution and expression evolution of these genes. *Nanos* sequence evolution has been investigated before (Curtis et al., 1995; de Keuckelaere et al., 2018; Gribouval et al., 2018). Studies regarding its evolution are restricted by low sampling numbers (Curtis et al., 1995; de Keuckelaere et al., 2018) or the focus is on vertebrate species (Gribouval et al., 2018). Here I have analyzed invertebrate sequences to get a better understanding of *nanos* evolution in invertebrates. My analysis is in line with these previous studies and confirmed the high conservation of the zf-nanos domain. It also provided indication that the dynamic duplication-loss landscape is not restricted to vertebrates only (Gribouval et al., 2018). Furthermore, the domain conservation was complemented by expression evolution analysis. My analysis provided further confirmation of evolutionary conservation of expression within for *nanos*. This is supported by most species having a maternal expression. Additionally, the best model explaining the gene expression evolution is one which models a constraint in them. The missing expressions in some species could be explained by a heterochronic shift as has been shown before (Fresques et al., 2016).

The peculiar expression pattern of *emsy* and previous smaller scale phylogenetic analysis prompted an in-depth phylogenetic analysis of this gene. I have found selective differences across two major lineages. Negative selection is more restrictive in the ENT domain compared to the whole alignment, in line with previous descriptions of this domain (Hughes-Davies et al., 2003). The negative selection varied across branches too. Although the ENT domain showed less signal for differences across branches. The reason behind the relaxation within invertebrate lineages is unknown. Without clear functions assigned to *emsy* (Hughes-Davies et al., 2003) it is difficult to hypothesize any selective forces driving this selective pattern. Additionally, the intense duplications occurring in the gene tree are compensated by loss events and a single copy is generally maintained across most lineages. In the case of *emsy* I would advise caution when extrapolating information between the two major lineages shown in my project. Unfortunately, the sequence divergence of invertebrate lineages is of unknown significance. Nevertheless, expression of this gene indicates selective forces maintaining its expression in oocytes.

The third investigated gene, *cbx1* was chosen for further analysis due to its known interaction and co-expression in oocytes with *emsy* (Y. Huang et al., 2006; Hughes-Davies et al., 2003). Furthermore, previously it was shown that *cbx1* is involved in early developmental stages, possibly as a heterochromatic protein (X.-Y. Liu et al., 2017; Ruddock-D'Cruz et al., 2008). My analyses showed that *cbx1* displayed a similar evolutionary pattern as *emsy*. It also displayed a relaxation of selection in the invertebrate lineages. Additionally, sites under differential selection indicated that the relaxation is not affecting the functional domains of the protein. Therefore, the relaxation in this gene is likely to be negligible for its functional conservation. I have also provided an overview of its expression in oocytes across a wide range of species. Previous results indicated that it is expressed in chordate oocytes (X.-Y. Liu et al., 2017; Ruddock-D'Cruz et al., 2008). My results expand this notion with invertebrates. Given the temporal and spatial co-expression, similar phylogenetic patterns and known interaction between *emsy* and *cbx1* could indicate that these two genes are involved in early development. Given that both

proteins are chromatin associated (Hughes-Davies et al., 2003; Ruddock-D'Cruz et al., 2008) it is possible that they fulfill the similar roles in oocytes.

Gene expression evolution analyses reflected the sequence conservation found in *emsy*. The notion that *emsy* is expressed across a wide array of species and also a constraint on its expression was detected supports a conserved expression pattern across the metazoan tree. However, *cbx1* despite being expressed across a wide range of species, I found that an unconstrained model explained best its gene expression evolution. The expression values (Figure 25) showed higher variability due to several species losing expression for this gene. The rest of the tree displayed expression values indicative of constraint. Interestingly the data also suggests that the two genes might function separately as in some species expression is lost for *cbx1* yet *emsy* expression is strongly expressed (e.g. *H.sapiens*). The above outline contradiction indicates the need for further experiments in order to elucidate the possible relationship of these two genes and their roles in early development.

4.3 Evolution of maternal genes

Analysis of the maternal gene expression evolution revealed several patterns across the metazoan tree. Species with more recent divergences displayed more similarities as previously described (Felsenstein, 1985). This also had implications downstream, during model fitting steps. The less variation of recently diverged clades could be inferred by the model fittings as a signal of selection. To overcome this a parametric bootstrapping framework was utilized, similar to previous approaches (Boettiger et al., 2012). Furthermore, a caution is warranted when interpreting heterogeneous and multi-regime model fittings as these models could misidentify the signal from recently diverged clades. Another caveat of the dataset is stemming from the unbalanced sampling. The spiralian lineage possesses an immense species diversity (Guo, 2009), yet it is underrepresented in my dataset. In parallel drosophiliid species are extensively represented by several species. These sampling restrictions will be freed up with new data being produced.

A technical pitfall of this comparative framework is that of orthology inference. A systematic error regarding these issues has been uncovered recently (Natsidis et al., 2021). The error stems from sampled species having varied evolutionary rates. This issue has yet to be resolved, in the meantime in the lack of better alternatives the current methodologies of orthology inference must be used. Despite limitations arising from data sampling this comparative framework still provided valuable information regarding the evolution of maternal gene expression.

Maternal gene expression is expected to have high divergences (1.2 Gene expression evolution). This project with new species and the use of phylogenetic comparative methods adds new molecular insights into the debate. Orthology inference results does indeed hint that maternal genes are diverse as they are grouped into marginally smaller orthogroups compared to the reference group. This could be a result of maternal gene sequences being more divergent. With higher sequence divergence homology relationships are more difficult to identify. (Natsidis et al., 2021) and therefore orthology relationships are more difficult to elucidate. My analysis also suggests a stark difference between the degraded subset of genes and the non-degraded ones. The orthology inference suggests that degraded subset of genes is more difficult to cluster into orthogroups. Furthermore, the tip distance differences also highlight that non-degraded maternal gene orthologs are more conserved as they cover more distantly related species. Maternal orthologs between diverged lineages are still identifiable despite the higher amount of duplication events inferred (Ohno, 1970) suggesting that sequence conservation in at least one of the paralogs is maintained. Presence of purifying selection acting on duplicate genes has been evidenced by previous studies both on sequence and expression level (Gillard et al., 2021; Kondrashov et al., 2002; Lynch & Conery, 2000). Therefore, maternal genes could display signs of purifying selection despite the pervasive duplication events.

The pattern of sequence conversation could be explained by the maternal transcriptome containing an abundance of housekeeping genes which would have less sequence

divergences, therefore grouped better into orthogroups (Lv et al., 2015). Genes not fulfilling a housekeeping role, rather contributing to the unique early developmental steps, would have higher sequence divergences due to weaker selection and therefore being more difficult to assign to appropriate orthogroups.

The housekeeping role within the maternal transcriptome is also reflected by the GO enrichment analysis, where many terms shared across species would be characteristic of housekeeping genes. In contrast for the degraded subset of genes it was more difficult to annotate with functional terms, let alone to find overlapping enriched terms. As the annotation is alignment dependent, the former could be traced back to higher sequence divergences within the degraded subset of genes. Because of fewer annotations, the enrichment will have difficulties identifying enriched terms. Furthermore, the reconstructed phylogenies based on the gene expressions and presence/absence datasets also support this argument. At its core the dataset analyzed in that approach is the same orthology inference output. The difference is in the approach utilized. Here, the non-degraded maternal genes show higher occupancy compared to the degraded ones and are more reflective of the species phylogeny.

A third approach of modelling conferred further insight into the dynamics of expression evolution. Expression presence or absence of a given maternal gene suggested that losing or gaining expression in both maternal gene categories is equally likely. Degraded genes display a more dynamic landscape of expression gain or loss, suggesting less constraint in maintaining expressions. Moreover, if omitting the transformation of expression values into a binary dataset further insight can be gained. The gene expression evolution of maternal genes is overwhelmingly populated by orthogroups with expressions evolving according to an OU process. Conversely the degraded subset of these genes is evolving according to a random walk process. Put differently maternal genes which do not undergo clearance have gene expression values restricted to certain values. This could either be a selective force maintaining expression of genes within that orthogroup or alternatively a selective force keeping certain genes at low expression values. To contrast this the

degraded subset of genes lose this selective signal and rather show signs of lack for a restrictive force. This could be in line with previous theoretical work (Kalinka & Tomancak, 2012; Mousseau & Fox, 1998) (see 1.1.5 Maternal effects through maternal genes). In this framework genes under maternal effect would show evolutionary variability as a Brownian motion. Therefore, the degraded subset of maternal genes does display the theoretical divergence. This could be attributable to a force maintaining such divergence or a consequence of generational uncoupling of the phenotypes (Mousseau & Fox, 1998). With a constant shuffling of maternal genes new niches could be occupied. In this scenario the non-degraded maternal genes function as housekeeping genes and are required to fulfill basic cellular requirements. This is reflected by their constant gene expression patterns throughout the MZT. Therefore, the housekeeping genes are under constraint as these processes are shared across all species.

It must be acknowledged that this could also possibly be an artefact emerging from model misspecifications arising due to smaller tree size as these models are sensitive to tree sizes (Boettiger et al., 2012; Cooper et al., 2016). Indeed, in a portion of degraded genes the full maternal orthogroups were found to follow an OU process. A caution is warranted then when interpreting the degraded maternal gene expression evolution. Nevertheless, maternal gene expressions display high degree of conservation, a notion supporting an early conservation model (Garstang, 1922; Riedl, 1978.

The implications of these findings could be relevant for basic biology and medicine. My project hopefully added nuances to the debate of conservation of maternal genes. Furthermore, I hope that the provided list of conserved maternal genes is a great source of further explorations.

5.Summary

In conclusion maternal genes show peculiar evolutionary patterns. The maternal-to-zygotic transition is a complex phenomenon which as a process is shared across many lineages. The specifics of the degradation mechanisms of maternal transcripts are divergent and species specific.

Genes which do not undergo clearance during MZT show high levels of conservation. This subgroup due to its conservation has potential to be exploited for further studies related to human diseases. The subset of maternal genes undergoing clearance in turn showed high divergences. Due to these divergences this subgroup could be held accountable for species specific adaptions to different reproductive modes and therefore an excellent candidate for evolutionary studies.

6.Bibliography

Aanes, H., Winata, C. L., Lin, C. H., Chen, J. P., Srinivasan, K. G., Lee, S. G. P., Lim, A. Y. M., Hajan, H. S., Collas, P., Bourque, G., Gong, Z., Korzh, V., Aleström, P., & Mathavan, S. (2011). Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Research*, *21*(8), 1328–1338. https://doi.org/10.1101/gr.116012.110

Altenburger, A., Martinez, P., & Wanninger, A. (2011). Homeobox gene expression in Brachiopoda: The role of Not and Cdx in bodyplan patterning, neurogenesis, and germ layer specification. *Gene Expression Patterns*, *11*(7), 427–436. https://doi.org/10.1016/j.gep.2011.07.001

Anger, M., Stein, P., & Schultz, R. M. (2005). CDC6 requirement for spindle formation during maturation of mouse oocytes. *Biology of Reproduction*, *72*(1), 188–194.

Antonio Marco. (2017). *Oocytes* (Malgorzata Klock, Ed.; 1st ed.). Springer, Cham.

Arhat, A., Meredith, P., Rosemary, G. B., Peter, R. G., & Clifford, J. T. (2004). Bmp4 and Morphological Variation of Beaks in Darwin's Finches. *Science*, *305*(5689), 1462–1465. https://doi.org/10.1126/science.1098095

Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., & Kobayashi, S. (1999). Maternal Pumilio acts together with Nanos in germline development in Drosophila embryos. *Nature Cell Biology* 1, 431–437

Atallah, J., & Lott, S. E. (2018). Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo. *PLoS Genetics*, *14*(12). https://doi.org/10.1371/journal.pgen.1007838

Ayoob, J. C., Terman, J. R., & Kolodkin, A. L. (2006). Drosophila Plexin B is a Sema-2a receptor required for axon guidance. *Development*, *133*(11), 2125–2135. https://doi.org/10.1242/dev.02380

Barckmann, B., Pierson, S., Dufourt, J., Papin, C., Armenise, C., Port, F., Grentzinger, T., Chambeyron, S., Baronian, G., Desvignes, J. P., Curk, T., & Simonelig, M. (2015). Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs Involved in Germ Cell Development in the Early Embryo. *Cell Reports*, *12*(7), 1205–1216. https://doi.org/10.1016/j.celrep.2015.07.030

Baroux, C., Autran, D., Gillmor, C. S., Grimanelli, D., & Grossniklaus, U. (2008). The Maternal to Zygotic Transition in Animals and Plants. *Cold Spring Harbor Symposia on Quantitative Biology*, *73*, 89–100. https://doi.org/doi: 10.1101/sqb.2008.73.053

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. da, Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Castro, L. G., … Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(1), 480–489. https://doi.org/10.1093/nar/gkaa1100

Bazzini, A. A., Viso, F., Moreno-Mateos, M. A., Johnstone, T. G., Vejnar, C. E., Qin, Y., Yao, J., Khokha, M. K., & Giraldez, A. J. (2016). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *The EMBO Journal*, *35*(19), 2087–2103. https://doi.org/10.15252/embj.201694699

Beckel-Mitchener, A. C., Miera, A., Keller, R., & Perrone-Bizzozero, N. I. (2002). Poly(A) Tail Length-dependent Stabilization of GAP-43 mRNA by the RNA-binding Protein HuD. *Journal of Biological Chemistry*, *277*(31), 27996–28002. https://doi.org/10.1074/jbc.M201982200

Bertolani, R. (2001). Evolution of the Reproductive Mechanisms in Tardigrades — A Review. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, *240*(3), 247–252. https://doi.org/10.1078/0044-5231-00032

Boettiger, C., Coop, G., & Ralph, P. (2012). Is your phylogeny informative? measuring the power of comparative methods. *Evolution*, *66*(7), 2240–2251. https://doi.org/10.1111/j.1558-5646.2011.01574.x

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grützner, F., Bergmann, S., Nielsen, R., Pääbo, S., & Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343–348. https://doi.org/10.1038/nature10532

Bushati, N., Stark, A., Brennecke, J., & Cohen, S. M. (2008). Temporal Reciprocity of miRNAs and Their Targets during the Maternal-to-Zygotic Transition in Drosophila. *Current Biology*, *18*(7), 501–506. https://doi.org/10.1016/j.cub.2008.02.081

Calarco, P. G. (1995). Polarization of Mitochondria in the Unfertilized Mouse Oocyte. In *Developmental Genetics* 16(1), 36-43 https://doi.org/doi: 10.1002/dvg.1020160108

Carlson, S. J. (2016). The Evolution of Brachiopoda. *Annual Review of Earth and Planetary Sciences*, *44*(1), 409–438. https://doi.org/10.1146/annurev-earth-060115-012348

Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic Analysis: Models and Estimation Procedures. *Evolution*, *21*(3), 550–570. https://doi.org/10.2307/2406616

Cecile, M., Nicole, M. T., Raimond, G. R. B., Wieger, H., Roland, A. R., Piet, G., Stephen, C., & Andrew, A. M. (2007). Structural insights into the Slit-Robo complex. *Proceedings of the National Academy of Sciences*, *104*(38), 14923–14928. https://doi.org/10.1073/pnas.0705310104

Chou, H.-C., Pruitt, M. M., Bastin, B. R., & Schneider, S. Q. (2016). A transcriptional blueprint for a spiral-cleaving embryo. *BMC Genomics*, *17*(1), 552. https://doi.org/10.1186/s12864-016-2860-6

Christopher A. Davis, Michael, H., Arnold, A. M., Lillian, B. S., Oliver, G. M., James, A. R., Geoffrey, C., Stephen, H., Gary, K. O., & Eric, N. O. (2006). PRISM/PRDM6, a Transcriptional Repressor That Promotes the Proliferative Gene Program in Smooth Muscle Cells. *Molecular and Cellular Biology*, *26*(7), 2626–2636. https://doi.org/10.1128/MCB.26.7.2626-2636.2006

Church, S. H., Munro, C., Dunn, C. W., & Extavour, C. G. (2022). The evolution of ovary-specific gene expression in Hawaiian Drosophilidae. *BioRxiv*. https://doi.org/10.1101/2021.11.30.470652

Cohen, B. L. (2013). Rerooting the rDNA gene tree reveals phoronids to be "brachiopods without shells"; dangers of wide taxon samples in metazoan phylogenetics (Phoronida; Brachiopoda). *Zoological Journal of the Linnaean Society*, *167*(1), 82–92. https://doi.org/10.1111/j.1096-3642.2012.00869.x

Collart, C., Owens, N. D. L., Bhaw-Rosun, L., Cooper, B., de Domenico, E., Patrushev, I., Sesay, A. K., Smith, J. N., Smith, J. C., & Gilchrist, M. J. (2014). High-resolution analysis of gene activity during the *Xenopus* mid-blastula transition. *Development*, *141*(9), 1927–1939. https://doi.org/10.1242/dev.102012

Conklin, E. G. (1905a). Mosaic development in ascidian eggs. *Journal of Experimental Zoology*, *2*(2), 145–223. https://doi.org/10.1002/jez.1400020202

Conklin, E. G. (1905b). Organ-forming substances in the eggs of ascidians. *The Biological Bulletin*, *8*(4), 205–230. https://doi.org/10.2307/1535879

Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., & Hughes, T. R. (2011). RBPDB: a database of RNA-binding specificities. *Nucleic Acids Research*, *39*(1), 301–308. https://doi.org/10.1093/nar/gkq1069

Cooper, N., Thomas, G. H., Venditti, C., Meade, A., & Freckleton, R. P. (2016). A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society. Linnean Society of London*, *118*(1), 64–77. https://doi.org/10.1111/bij.12701

Cruickshank, T., & Wade, M. J. (2008). Microevolutionary support for a developmental hourglass: Gene expression patterns shape sequence variation and divergence in Drosophila. *Evolution and Development*, *10*(5), 583–590. https://doi.org/10.1111/j.1525-142X.2008.00273.x

Curtis, D., Apfeld, J., & Lehmann, R. (1995). nanos is an evolutionarily conserved organizer of anterior-posterior polarity. *Development*, *121*(6), 1899–1910. https://doi.org/10.1242/dev.121.6.1899

D'Agostino, I., Merritt, C., Chen, P. L., Seydoux, G., & Subramaniam, K. (2006). Translational repression restricts expression of the *C. elegans* Nanos homolog NOS-2 to the embryonic germline. *Developmental Biology*, *292*(1), 244–252. https://doi.org/10.1016/j.ydbio.2005.11.046

Dai, C. F., Jiang, Y. Z., Li, Y., Wang, K., Liu, P. S., Patankar, M. S., & Zheng, J. (2011). Expression and roles of Slit/Robo in human ovarian cancer. *Histochemistry and Cell Biology*, *135*(5), 475–485. https://doi.org/10.1007/s00418-011-0806-2

Dalcq, A. M. (1955). Processes of synthesis during early development of rodents' eggs and embryos. *Studies on Fertility 1955*(7), 113–122

Dapper, A. L., & Wade, M. J. (2020). Relaxed Selection and the Rapid Evolution of Reproductive Genes. *Trends in Genetics*, *36*(9), 640–649. https://doi.org/10.1016/j.tig.2020.06.014

de Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Genetics*, *49*(6), 941–945. https://doi.org/10.1038/ng.3858

de Keuckelaere, E., Hulpiau, P., Saeys, Y., Berx, G., & van Roy, F. (2018). Nanos genes and their role in development and beyond. *Cellular and Molecular Life Sciences*, *75*(11), 1929–1946. https://doi.org/10.1007/s00018-018-2766-3

de Renzis, S., Elemento, O., Tavazoie, S., & Wieschaus, E. F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biology*, *5*(5), 1036–1051. https://doi.org/10.1371/journal.pbio.0050117

Demuth, J. P., & Wade, M. J. (2007). Maternal expression increases the rate of bicoid evolution by relaxing selective constraint. *Genetica*, *129*(1), 37–43. https://doi.org/10.1007/s10709-006-0031-4

Deng, M., Chen, B. B., Liu, Z., Cai, Y., Wan, Y., Zhang, G., Fan, Y., Zhang, Y., & Wang, F. (2020). YTHDF2 Regulates Maternal Transcriptome Degradation and

Embryo Development in Goat. *Frontiers in Cell and Developmental Biology*, *8*. https://doi.org/10.3389/fcell.2020.580367

Denis, C. L. (1984). Identification of new genes involved in the regulation of yeast alcohol dehydrogenase II. *Genetics*, *108*(4), 833–844. https://doi.org/10.1093/genetics/108.4.833

Detivaud, L., Pascreau, G., Karaïskou, A., Osborne, H. B., & Kubiak, J. Z. (2003). Regulation of EDEN-dependent deadenylation of Aurora A/Eg2-derived mRNA via phosphorylation and dephosphorylation in *Xenopus laevis* egg extracts. *Journal of Cell Science*, *116*(13), 2697–2705. https://doi.org/doi: 10.1242/jcs.00477

Dexheimer, P. J., Wang, J., & Cochella, L. (2020). Two MicroRNAs Are Sufficient for Embryonic Patterning in *C. elegans*. *Current Biology*, *30*(24), 5058-5065.e5. https://doi.org/10.1016/j.cub.2020.09.066

Dickinson, R. E., & Duncan, W. C. (2010). The SLIT–ROBO pathway: a regulator of cell function with implications for the reproductive system. *Reproduction*, *139*(4), 697–704. https://doi.org/10.1530/REP-10-0017

Domazet-Lošo, T., & Tautz, D. (2010). A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*, *468*(7325), 815–818. https://doi.org/10.1038/nature09632

Driever, W., Siegel, V., & Nusslein-Volhard, C. (1990). Autonomous determination of anterior structures in the early *Drosophila* embryo by the bicoid morphogen. *Development*, *109*(4), 811–820. https://doi.org/10.1242/dev.109.4.811

Drost, H.-G., Gabel, A., Grosse, I., & Quint, M. (2015). Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Molecular Biology and Evolution*, *32*(5), 1221–1231. https://doi.org/10.1093/molbev/msv012

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development*, *1994*(Supplement), 135–142.

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., Sørensen, M. v., Haddock, S. H. D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M., Wheeler, W. C., Martindale, M. Q., & Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, *452*(7188), 745–749. https://doi.org/10.1038/nature06614

Dunn, C. W., Zapata, F., Munro, C., Siebert, S., & Hejnol, A. (2018). Pairwise comparisons across species are problematic when analyzing functional genomic

data. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(3), 409–417. https://doi.org/10.1073/pnas.1707515115

Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: support for time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, *4*(7), 688–691. https://doi.org/10.1111/2041-210X.12051

Edgar, B. A., & Datar, S. A. (1996). Zygotic degradation of two maternal Cdc25 mRNAs terminates Drosophila's early cell cycle program. *Genes & Development*, *10*(15), 1966–1977. http://genesdev.cshlp.org/content/10/15/1966.abstract

Edgar, B. A., & O'Farrell, P. H. (1990). The three postblastoderm cell cycles of Drosophila embryogenesis are regulated in G2 by string. *Cell*, *62*(3), 469–480. https://doi.org/10.1016/0092-8674(90)90012-4

Edwards, R. G., & Beard, H. K. (1997). Oocyte polarity and cell determination in early mammalian embryos. In *Molecular Human Reproduction 3*(10),863-905. https://doi.org/10.1093/molehr/3.10.863

Eichhorn, S. W., Subtelny, A. O., Kronja, I., Kwasnieski, J. C., Orr-Weaver, T. L., & Bartel, D. P. (2016). mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. *ELife*. 5:e16955. https://doi.org/10.7554/eLife.16955.001

Emms, D. M., & Kelly, S. (2021). SHOOT: phylogenetic gene search and ortholog inference. *BioRxiv*, 2021.09.01.458564. https://doi.org/10.1101/2021.09.01.458564

Extavour, C. G., Pang, K., Matus, D. Q., & Martindale, M. Q. (2005). vasa and nanos expression patterns in a sea anemone and the evolution of bilaterian germ cell specification mechanisms. *Evolution & Development*, *7*(3), 201–215. https://doi.org/10.1111/j.1525-142X.2005.05023.x

Fasken, M. B., & Corbett, A. H. (2009). Mechanisms of nuclear mRNA quality control. *RNA Biology*, *6*(3), 237–241. https://doi.org/10.4161/rna.6.3.8330

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist, 125*(1) 1-15. https://doi.org/10.1086/284325

Fernando, J. S., Marshal, E. L., Christopher, W., Peter, G. J., & Sidney, S. (1994). Coordinate Initiation of *Drosophila* Development by Regulated Polyadenylation of Maternal Messenger RNAs. *Science*, *266*(5193), 1996–1999. https://doi.org/10.1126/science.7801127

Flanagan, J. F., Blus, B. J., Kim, D., Clines, K. L., Rastinejad, F., & Khorasanizadeh, S. (2007). Molecular Implications of Evolutionary Differences in CHD Double

Chromodomains. *Journal of Molecular Biology*, *369*(2), 334–342. https://doi.org/10.1016/j.jmb.2007.03.024

Fluck, R. A., Miller, A. L., & Jaffe, L. F. (1991). Slow calcium waves accompany cytokinesis in medaka fish eggs. *The Journal of Cell Biology*, *115*(5), 1259–1265. https://doi.org/10.1083/jcb.115.5.1259

Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. *The American Naturalist*, *160*(6), 712–726. https://doi.org/10.1086/343873

Fresques, T., Swartz, S. Z., Juliano, C., Morino, Y., Kikuchi, M., Akasaka, K., Wada, H., Yajima, M., & Wessel, G. M. (2016). The diversity of nanos expression in echinoderm embryos supports different mechanisms in germ cell specification. *Evolution & Development*, *18*(4), 267–278. https://doi.org/10.1111/ede.12197

Frohnhöfer, H. G., & Nüsslein-Volhard, C. (1986). Organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid. *Nature*, *324*(6093), 120–125. https://doi.org/10.1038/324120a0

Fukushima, K., & Pollock, D. D. (2020). Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18090-8

Gabriel, W. N., McNuff, R., Patel, S. K., Gregory, T. R., Jeck, W. R., Jones, C. D., & Goldstein, B. (2007). The tardigrade *Hypsibius dujardini*, a new model for studying the evolution of development. *Developmental Biology*, *312*(2), 545–559. https://doi.org/10.1016/j.ydbio.2007.09.055

Garland Jr., T., Dickerman, A. W., Janis, C. M., & Jones, J. A. (1993). Phylogenetic Analysis of Covariance by Computer Simulation. *Systematic Biology*, *42*(3), 265–292. https://doi.org/10.1093/sysbio/42.3.265

Garstang, W. (1922). The theory of recapitulation: a critical re-statement of the biogenetic law. *Zoological Journal of the Linnean Society*, *35*(232), 81–101. https://doi.org/10.1111/j.1096-3642.1922.tb00464.x

Ghanbarian, A. T., & Hurst, L. D. (2015). Neighboring Genes Show Correlated Evolution in Gene Expression. *Molecular Biology and Evolution*, *32*(7), 1748–1766. https://doi.org/10.1093/molbev/msv053

Gilbert, D. G. (2019). Genes of the pig, Sus scrofa, reconstructed with EvidentialGene. *PeerJ*, *7*, e6374–e6374. https://doi.org/10.7717/peerj.6374

Gildor, T., Cary, G. A., Lalzar, M., Hinman, V. F., & Ben-Tabou de-Leon, S. (2019). Developmental transcriptomes of the sea star, *Patiria miniata*, illuminate how gene expression changes with evolutionary distance. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-52577-9

Gillard, G. B., Grønvold, L., Røsæg, L. L., Holen, M. M., Monsen, Ø., Koop, B. F., Rondeau, E. B., Gundappa, M. K., Mendoza, J., Macqueen, D. J., Rohlfs, R. v, Sandve, S. R., & Hvidsten, T. R. (2021). Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biology*, *22*(1), 103. https://doi.org/10.1186/s13059-021-02323-0

Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Dongen, S. van, Inoue, K., Enright, A. J., & Schier, A. F. (2006). Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal mRNAs. *Science 312*(5770), 75-9. https://doi.org/10.1126/science.1122689

Giudice, G., Sánchez-Cabo, F., Torroja, C., & Lara-Pezzi, E. (2016). ATtRACT—a database of RNA-binding proteins and associated motifs. *Database*, *2016*, baw035. https://doi.org/10.1093/database/baw035

Goldstein, B., Hird, S. N., & White, J. G. (1993). Cell polarity in early *C. elegans* development. *Development*, *119*(Supplement), 279–287. https://doi.org/10.1242/dev.119.Supplement.279

Graindorge, A., le Tonquèze, O., Thuret, R., Pollet, N., Osborne, H. B., & Audic, Y. (2008). Identification of CUG-BP1/EDEN-BP target mRNAs in *Xenopus tropicalis*. *Nucleic Acids Research*, *36*(6), 1861–1870. https://doi.org/10.1093/nar/gkn031

Gribouval, L., Sourdaine, P., Lareyre, J.-J., Bellaiche, J., le Gac, F., Mazan, S., Guiardiere, C., Auvray, P., & Gautier, A. (2018). The nanos1 gene was duplicated in early Vertebrates and the two paralogs show different gonadal expression profiles in a shark. *Scientific Reports*, *8*(1), 6942. https://doi.org/10.1038/s41598-018-24643-1

Gu, L., Wang, Q., & Sun, Q.-Y. (2010). Histone modifications during mammalian oocyte maturation: Dynamics, regulation and functions. *Cell Cycle*, *9*(10), 1942–1950. https://doi.org/10.4161/cc.9.10.11599

Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A., & Kingsley, D. M. (2014). A molecular basis for classic blond hair color in Europeans. *Nature Genetics*, *46*(7), 748–752. https://doi.org/10.1038/ng.2991

Guo, X. (2009). Use and exchange of genetic resources in molluscan aquaculture. *Reviews in Aquaculture*, *1*(3–4), 251–259. https://doi.org/10.1111/j.1753-5131.2009.01014.x

Guven-Ozkan, T., Nishi, Y., Robertson, S. M., & Lin, R. (2008). Global Transcriptional Repression in *C. elegans* Germline Precursors by Regulated Sequestration of TAF-4. *Cell*, *135*(1), 149–160. https://doi.org/10.1016/j.cell.2008.07.040

Haeckel, E. (1866). Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von C. Darwin reformirte Descendenz-Theorie. https://doi.org/10.5962/bhl.title.3953

Harrison, P. W., Wright, A. E., & Mank, J. E. (2012). The evolution of gene expression and the transcriptome–phenotype relationship. *Seminars in Cell & Developmental Biology*, *23*(2), 222–229. https://doi.org/10.1016/j.semcdb.2011.12.004

Hart, T., Komori, H. K., LaMere, S., Podshivalova, K., & Salomon, D. R. (2013). Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics*, *14*(1), 778. https://doi.org/10.1186/1471-2164-14-778

Harvey, E. B. (1936). Parthenogenetic Merogony or Cleavage, without Nuclei in *Arbacia punctulata. Biological Bulletin 71*(1), 101-121. https://doi.org/10.2307/1537411

Haussmann, I. U., Li, M., & Soller, M. (2011). ELAV-Mediated 3′-End Processing of ewg Transcripts Is Evolutionarily Conserved Despite Sequence Degeneration of the ELAV-Binding Site. *Genetics*, *189*(1), 97–107. https://doi.org/10.1534/genetics.111.131383

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W. E. G., Seaver, E., Wheeler, W. C., Martindale, M. Q., Giribet, G., & Dunn, C. W. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1677), 4261–4270. https://doi.org/10.1098/rspb.2009.0896

Hodgins-Davis, A., & Townsend, J. P. (2009). Evolving gene expression: from G to E to G×E. *Trends in Ecology & Evolution*, *24*(12), 649–658. https://doi.org/10.1016/j.tree.2009.06.011

Hogvall, M., Vellutini, B. C., Martín-Durán, J. M., Hejnol, A., Budd, G. E., & Janssen, R. (2019). Embryonic expression of priapulid Wnt genes. *Development Genes and Evolution*, *229*(4), 125–135. https://doi.org/10.1007/s00427-019-00636-6

Hollis, R. L., Churchman, M., Michie, C. O., Rye, T., Knight, L., McCavigan, A., Perren, T., Williams, A. R. W., McCluggage, W. G., Kaplan, R. S., Jayson, G. C., Oza, A., Paul, D. H., Herrington, C. S., Kennedy, R., & Gourley, C. (2019). High EMSY expression defines a BRCA-like subgroup of high-grade serous ovarian carcinoma

with prolonged survival and hypersensitivity to platinum. *Cancer*, *125*(16), 2772–2781. https://doi.org/10.1002/cncr.32079

Hölzer, M., & Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience*, *8*(5), giz039. https://doi.org/10.1093/gigascience/giz039

Howard, R. J., Giacomelli, M., Lozano-Fernandez, J., Edgecombe, G. D., Fleming, J. F., Kristensen, R. M., Ma, X., Olesen, J., Sørensen, M. v, Thomsen, P. F., Wills, M. A., Donoghue, P. C. J., & Pisani, D. (2022). The Ediacaran origin of Ecdysozoa: integrating fossil and phylogenomic data. *Journal of the Geological Society*, jgs2021-107. https://doi.org/10.1144/jgs2021-107

Hu, H., Uesaka, M., Guo, S., Shimai, K., Lu, T.-M., Li, F., Fujimoto, S., Ishikawa, M., Liu, S., Sasagawa, Y., Zhang, G., Kuratani, S., Yu, J.-K., Kusakabe, T. G., Khaitovich, P., Irie, N., & Consortium, the E. (2017). Constrained vertebrate evolution by pleiotropic genes. *Nature Ecology & Evolution*, *1*(11), 1722–1730. https://doi.org/10.1038/s41559-017-0318-0

Huang, D., Vannier, J., & Chen, J. Y. (2004). Recent *Priapulidae* and their Early Cambrian ancestors: Comparisons and evolutionary significance. *Geobios*, *37*(2), 217–228. https://doi.org/10.1016/j.geobios.2003.04.004

Huang, Y., Myers, M. P., & Xu, R.-M. (2006). Crystal Structure of the HP1-EMSY Complex Reveals an Unusual Mode of HP1 Binding. *Structure*, *14*(4), 703–712. https://doi.org/10.1016/j.str.2006.01.007

Hughes-Davies, L., Huntsman, D., Ruas, M., Fuks, F., Bye, J., Chin, S.-F., Milner, J., Brown, L. A., Hsu, F., Gilks, B., Nielsen, T., Schulzer, M., Chia, S., Ragaz, J., Cahn, A., Linger, L., Ozdag, H., Cattaneo, E., Jordanova, E. S., … Kouzarides, T. (2003). EMSY Links the BRCA2 Pathway to Sporadic Breast and Ovarian Cancer. *Cell*, *115*(5), 523–535. https://doi.org/10.1016/S0092-8674(03)00930-9

Hulse, A. M., & Cai, J. J. (2013). Genetic variants contribute to gene expression variability in humans. *Genetics*, *193*(1), 95–108. https://doi.org/10.1534/genetics.112.146779

Irie, N., & Kuratani, S. (2011). Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Communications 2*(248). https://doi.org/10.1038/ncomms1248

Ivanova, I., Much, C., di Giacomo, M., Azzi, C., Morgan, M., Moreira, P. N., Monahan, J., Carrieri, C., Enright, A. J., & O'Carroll, D. (2017). The RNA m6A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal

Transcriptome and Oocyte Competence. *Molecular Cell*, *67*(6), 1059-1067.e4. https://doi.org/10.1016/j.molcel.2017.08.003

Jacobs, S. A., Taverna, S. D., Zhang, Y., Briggs, S. D., Li, J., Eissenberg, J. C., Allis, C. D., & Khorasanizadeh, S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *The EMBO Journal*, *20*(18), 5232–5241.

Jaruzelska, J., Kotecki, M., Kusz, K., Spik, A., Firpo, M., & Reijo Pera, R. A. (2003). Conservation of a Pumilio-Nanos complex from *Drosophila* germ plasm to human germ cells. *Development Genes and Evolution*, *213*(3), 120–126. https://doi.org/10.1007/s00427-003-0303-2

Jezierska, M., Miernik, A., Sojka, J., Student, S., Śliwińska, M. A., Gross, V., & Poprawa, I. (2021). Oogenesis in the tardigrade *Hypsibius exemplaris* Gąsiorek, Stec, Morek & Michalczyk, 2018 (Eutardigrada, Hypsibiidae). *Micron*, *150*. https://doi.org/10.1016/j.micron.2021.103126

Junqueira Alves, C., Yotoko, K., Zou, H., & Friedel, R. H. (2019). Origin and evolution of plexins, semaphorins, and Met receptor tyrosine kinases. *Scientific Reports*, *9*(1), 1970. https://doi.org/10.1038/s41598-019-38512-y

Kærn, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics 6*(6), 451-464. https://doi.org/10.1038/nrg1615

Kalinka, A. T., & Tomancak, P. (2012). The evolution of early animal embryos: conservation or divergence? *Trends in Ecology & Evolution*, *27*(7), 385–393. https://doi.org/10.1016/j.tree.2012.03.007

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kelly, S. M., Leung, S. W., Apponi, L. H., Bramley, A. M., Tran, E. J., Chekanova, J. A., Wente, S. R., & Corbett, A. H. (2010). Recognition of Polyadenosine RNA by the Zinc Finger Domain of Nuclear Poly(A) RNA-binding Protein 2 (Nab2) Is Required for Correct mRNA 3`-End Formation. *Journal of Biological Chemistry*, *285*(34), 26022–26032. https://doi.org/10.1074/jbc.M110.141127

Kemphues, K. J., Kusch, M., & Wolf, N. (1988). Maternal-effect lethal mutations on linkage group II of *Caenorhabditis elegans*. *Genetics*, *120*(4), 977–986. https://doi.org/10.1093/genetics/120.4.977

Khare, N., Fascetti, N., DaRocha, S., Chiquet-Ehrismann, R., & Baumgartner, S. (2000). Expression patterns of two new members of the Semaphorin family in *Drosophila*

suggest early functions during embryogenesis. *Mechanisms of Development*, *91*(1), 393–397. https://doi.org/10.1016/S0925-4773(99)00297-X

Kidd, T., Brose, K., Mitchell, K. J., Fetter, R. D., Tessier-Lavigne, M., Goodman, C. S., & Tear, G. (1998). Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. *Cell*, *92*(2), 205–215.

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

Kirkpatrick, M., & Lande, R. (1989). The evolution of maternal characters. *Evolution*, *43*(3), 485–503. https://doi.org/10.1111/j.1558-5646.1989.tb04247.x

Kloc, M., & Etkin, L. D. (1995). Two distinct pathways for the localization of RNAs at the vegetal cortex in *Xenopus* oocytes. *Development*, *121*(2), 287–297. https://doi.org/10.1242/dev.121.2.287

Kloc, M., Jedrzejowska, I., Tworzydlo, W., & Bilinski, S. M. (2014). Balbiani body, nuage and sponge bodies - The germ plasm pathway players. *Arthropod Structure and Development 43*(4) 341–348. https://doi.org/10.1016/j.asd.2013.12.003

Kobayashi, S., Yamada, M., Asaoka, M., & Kitamura, T. (1996). Essential role of the posterior morphogen nanos for germline development in Drosophila. *Nature*, *380*(6576), 708–711. https://doi.org/10.1038/380708a0

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., & Koonin, E. v. (2002). Selection in the evolution of gene duplications. *Genome Biology*, *3*(2), 1–9.

Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Molecular Biology and Evolution*, *22*(5), 1208–1222. https://doi.org/10.1093/molbev/msi105

Krauchunas, A. R., Horner, V. L., & Wolfner, M. F. (2012). Protein phosphorylation changes reveal new candidates in the regulation of egg activation and early embryogenesis in *D. melanogaster*. *Developmental Biology*, *370*(1), 125–134. https://doi.org/10.1016/j.ydbio.2012.07.024

Kryuchkova-Mostacci, N., & Robinson-Rechavi, M. (2016). Tissue-Specificity of Gene Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs. *PLoS Computational Biology*, *12*(12). https://doi.org/10.1371/journal.pcbi.1005274

Kumar, L., & Futschik, M. E. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformation*, *2*(1), 5. https://doi.org/10.6026/97320630002005

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, *34*(7), 1812–1819. https://doi.org/10.1093/molbev/msx116

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., & Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, *410*(6824), 116–120. https://doi.org/10.1038/35065132

Laver, J. D., Li, X., Ray, D., Cook, K. B., Hahn, N. A., Nabeel-Shah, S., Kekis, M., Luo, H., Marsolais, A. J., Fung, K. Y. Y., Hughes, T. R., Westwood, J. T., Sidhu, S. S., Morris, Q., Lipshitz, H. D., & Smibert, C. A. (2015). Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0659-4

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P., & Krause, H. M. (2007). Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. *Cell*, *131*(1), 174–187. https://doi.org/10.1016/j.cell.2007.08.003

Lee, M. T., Bonneau, A. R., & Giraldez, A. J. (2014). Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annual Review of Cell and Developmental Biology*, *30*(1), 581–613. https://doi.org/10.1146/annurev-cellbio-100913-013027

Lee, S. E., & Jacobs, D. K. (1999). Expression of Distal-less in molluscan eggs, embryos, and larvae. *Evolution & Development*, *1*(3), 172–179. https://doi.org/10.1046/j.1525-142x.1999.99016.x

Lemaître, J.-M., Bocquet, S., & Méchali, M. (2002). Competence to replicate in the unfertilized egg is conferred by Cdc6 during meiotic maturation. *Nature*, *419*(6908), 718–722. https://doi.org/10.1038/nature01046.

Levin, M., Anavy, L., Cole, A. G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D. H., Feder, M., Fernandez-Valverde, S. L., Nakanishi, N., Simmons, D., Simakov, O., Larsson, T., Liu, S. Y., Jerafi-Vider, A., Yaniv, K., Ryan, J. F., … Yanai, I. (2016). The mid-developmental transition and the evolution of animal body plans. *Nature*, *531*(7596), 637–641. https://doi.org/10.1038/nature16994

Levin, M., Hashimshony, T., Wagner, F., & Yanai, I. (2012). Developmental Milestones Punctuate Gene Expression in the *Caenorhabditis Embryo*. *Developmental Cell*, *22*(5), 1101–1108. https://doi.org/10.1016/j.devcel.2012.04.004

Li, T., Zhang, H., Wang, X., Yin, D., Chen, N., Kang, L., Zhao, X., & Ma, Y. (2020). Cloning, Molecular Characterization and Expression Patterns of DMRTC2 Implicated in Germ Cell Development of Male Tibetan Sheep. *International Journal of Molecular Sciences*, *21*(7). https://doi.org/10.3390/ijms21072448

Lingel, A., & Izaurralde, E. (2004). RNAi: Finding the elusive endonuclease. *RNA*, *10*(11), 1675–1679. https://doi.org/10.1261/rna.7175704

Liu, H.-Y., Badarinarayana, V., Audino, D. C., Rappsilber, J., Mann, M., & Denis, C. L. (1998). The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. *The EMBO Journal*, *17*(4), 1096–1106. https://doi.org/10.1093/emboj/17.4.1096

Liu, M. M., Davey, J. W., Jackson, D. J., Blaxter, M. L., & Davison, A. (2014). A conserved set of maternal genes? Insights from a molluscan transcriptome. *The International Journal of Developmental Biology*, *58*(6–8), 501–511. https://doi.org/10.1387/ijdb.140121ad

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., & Pan, T. (2015). N6 -methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, *518*(7540), 560–564. https://doi.org/10.1038/nature14234

Liu, X.-Y., Zhang, X.-B., Li, M.-H., Zheng, S.-Q., Liu, Z.-L., Cheng, Y.-Y., & Wang, D.-S. (2017). Genome-wide identification, evolution of chromobox family genes and their expression in Nile tilapia. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *203*, 25–34. https://doi.org/10.1016/j.cbpb.2016.09.001

Lomberk, G., Wallrath, L., & Urrutia, R. (2006). The Heterochromatin Protein 1 family. *Genome Biology*, *7*(7), 228. https://doi.org/10.1186/gb-2006-7-7-228

Long, J. A. (1964). The embryology of three species representing three superfamilies of articulate Brachiopoda. University of Washington ProQuest Dissertations Publishing

Lund, E., Liu, M., Hartley, R. S., Sheets, M. D., & Dahlberg, J. E. (2009). Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos. *RNA*, *15*(12), 2351–2363. https://doi.org/10.1261/rna.1882009

Lv, W., Zheng, J., Luan, M., Shi, M., Zhu, H., Zhang, M., Lv, H., Shang, Z., Duan, L., Zhang, R., & Jiang, Y. (2015). Comparing the evolutionary conservation between human essential genes, human orthologs of mouse essential genes and human housekeeping genes. *Briefings in Bioinformatics*, *16*(6), 922–931. https://doi.org/10.1093/bib/bbv025

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, *290*(5494), 1151–1155.

Majewski, J., & Pastinen, T. (2011). The study of eQTL variations by RNA-seq: From SNPs to phenotypes. In *Trends in Genetics 27*(2), 72–79. https://doi.org/10.1016/j.tig.2010.10.006

Mallatt, J., Craig, C. W., & Yoder, M. J. (2012). Nearly complete rRNA genes from 371 Animalia: Updated structure-based alignment and detailed phylogenetic analysis. *Molecular Phylogenetics and Evolution*, *64*(3), 603–617. https://doi.org/10.5061/dryad.1v62kr3q

Martin, O., Jonathan, R., G, M. L., & Arun, S. (2002). Identification of β-cell-specific insulin gene transcription factor RIPE3b1 as mammalian MafA. *Proceedings of the National Academy of Sciences*, *99*(10), 6737–6742. https://doi.org/10.1073/pnas.102168499

Martín-Durán, J. M., & Hejnol, A. (2015). The study of *Priapulus caudatus* reveals conserved molecular patterning underlying different gut morphogenesis in the Ecdysozoa. *BMC Biology*, *13*(29). https://doi.org/10.1186/s12915-015-0139-z

Martinez, J., & Tuschl, T. (2004). RISC is a 5′ phosphomonoester-producing RNA endonuclease. *Genes & Development*, *18*(9), 975–980. https://doi.org/10.1101/gad.1187904

Matsumoto, K., Meric, F., & Wolffe, A. P. (1996). Translational repression dependent on the interaction of the Xenopus Y- box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition. *Journal of Biological Chemistry*, *271*(37), 22706–22712. https://doi.org/10.1074/jbc.271.37.22706

Mishima, Y., & Tomari, Y. (2016a). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Molecular Cell*, *61*(6), 874–885. https://doi.org/10.1016/j.molcel.2016.02.027

Mishima, Y., & Tomari, Y. (2016b). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Molecular Cell*, *61*(6), 874–885. https://doi.org/10.1016/j.molcel.2016.02.027

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E., Tosatto, S., Paladin, L., Raj, S., Richardson, L., Finn, R., & Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*. https://doi.org/10.1093/nar/gkaa913

Møbjerg, N., Halberg, K. A., Jørgensen, A., Persson, D., Bjørn, M., Ramløv, H., & Kristensen, R. M. (2011). Survival in extreme environments - on the current knowledge of adaptations in tardigrades. *Acta physiologica 202*(3) 409–420. https://doi.org/10.1111/j.1748-1716.2011.02252.x

Mochizuki, K., Sano, H., Kobayashi, S., Nishimiya-Fujisawa, C., & Fujisawa, T. (2000). Expression and evolutionary conservation of nanos-related genes in *Hydra*. *Development Genes and Evolution*, *210*(12), 591–602. https://doi.org/10.1007/s004270000105

Mosquera, L., Forristall, C., Zhou, Y., & King, M. L. (1993). A mRNA localized to the vegetal cortex of *Xenopus* oocytes encodes a protein with a nanos-like zinc finger domain. *Development*, *117*(1), 377–386. https://doi.org/10.1242/dev.117.1.377

Mousseau, T., & Fox, C. (1998). The adaptive significance of maternal effects. *Trends in Ecology & Evolution*, *13*, 403–407. https://doi.org/10.1016/S0169-5347(98)01472-4

Muhlrad, D., Decker, C. J., & Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. *Genes & Development*, *8*(7), 855–866. https://doi.org/10.1101/gad.8.7.855

Muhlrad, D., Decker, C. J., & Parker, R. (1995). Turnover mechanisms of the stable yeast PGK1 mRNA. *Molecular and Cellular Biology*, *15*(4), 2145–2156. https://doi.org/10.1128/MCB.15.4.2145

Mukhopadhyay, A., Quiroz, J. A., & Wolkoff, A. W. (2014). Rab1a regulates sorting of early endocytic vesicles. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, *306*(5), G412–G424. https://doi.org/10.1152/ajpgi.00118.2013

Munro, C., Zapata, F., Howison, M., Siebert, S., & Dunn, C. W. (2022). Evolution of gene expression across species and specialized zooids in Siphonophora. *Molecular Biology and Evolution*, *39*(2). https://doi.org/10.1101/2021.07.30.454354

Muto, A., Kume, S., Inoue, T., Okano, H., & Mikoshiba, K. (1996). Calcium waves along the cleavage furrows in cleavage-stage *Xenopus* embryos and its inhibition by heparin. *The Journal of Cell Biology*, *135*(1), 181–190. https://doi.org/10.1083/jcb.135.1.181

Nachtomy, O., Shavit, A., & Yakhini, Z. (2007). Gene expression and the concept of the phenotype. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, *38*(1), 238–254. https://doi.org/10.1016/j.shpsc.2006.12.014

Nadler, J. J., Zou, F., Huang, H., Moy, S. S., Lauder, J., Crawley, J. N., Threadgill, D. W., Wright, F. A., & Magnuson, T. R. (2006). Large-Scale Gene Expression Differences Across Brain Regions and Inbred Strains Correlate With a Behavioral Phenotype. *Genetics*, *174*(3), 1229–1236. https://doi.org/10.1534/genetics.106.061481

Nakamura, T., Shiomi, I., & Shimizu, T. (2017). Embryonic expression of festina lente (fel), a novel maternal gene, in the oligochaete annelid *Tubifex tubifex*. *Gene Expression Patterns*, *25–26*, 29–35. https://doi.org/10.1016/j.gep.2017.05.001

Natsidis, P., Kapli, P., Schiffer, P. H., & Telford, M. J. (2021). Systematic errors in orthology inference and their effects on evolutionary analyses. *IScience*, *24*(2), 102110. https://doi.org/10.1016/j.isci.2021.102110

Ninova, M., Ronshaugen, M., & Griffiths-Jones, S. (2016). MicroRNA evolution, expression, and function during short germband development in *Tribolium castaneum*. *Genome Research*, *26*(1), 85–96. https://doi.org/10.1101/gr.193367.115

Nørrevang, Arne. (1965). Oogenesis in Priapulus Caudatus Lamarck : an electron microscopical study correlated with light microscopical and histochemical findings.

Nüsslein-Volhard, C. (1977). Genetic analysis of pattern-formation in the embryo ofDrosophila melanogaster. *Wilhelm Roux's Archives of Developmental Biology*, *183*(3), 249–268. https://doi.org/10.1007/BF00867325

Nüsslein-Volhard, C., Lohs-Schardin, M., Sander, K., & Cremer, C. (1980). A dorso-ventral shift of embryonic primordia in a new maternal-effect mutant of Drosophila. *Nature*, *283*(5746), 474–476. https://doi.org/10.1038/283474a0

Nüsslein-Volhard, C., & Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature*, *287*(5785), 795–801. https://doi.org/10.1038/287795a0

Ohnishi, Y., Totoki, Y., Toyoda, A., Watanabe, T., Yamamoto, Y., Tokunaga, K., Sakaki, Y., Sasaki, H., & Hohjoh, H. (2010). Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. *Nucleic Acids Research*, *38*(15), 5141–5151. https://doi.org/10.1093/nar/gkq229

Ohno, S. (1970). *Evolution by gene duplication*. Springer Science & Business Media.

O'Meara, B. C., Ané, C., Sanderson, M. J., & Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, *60*(5), 922–933. https://doi.org/10.1111/j.0014-3820.2006.tb01171.x

Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, *26*(4), 331–348. https://doi.org/10.1111/j.1463-6409.1997.tb00423.x

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. https://doi.org/10.1038/44766

Pan, H., O'Brien, M. J., Wigglesworth, K., Eppig, J. J., & Schultz, R. M. (2005). Transcript profiling during mouse oocyte development and the effect of gonadotropin priming and development in vitro. *Developmental Biology*, *286*(2), 493–506. https://doi.org/10.1016/j.ydbio.2005.08.023

Paris, J., Osborne, H. B., Couturier, A., le Guellec, R., & Philippe, M. (1988). Changes in the polyadenylation of specific stable RNA during the early development of *Xenopus laevis*. *Gene*, *72*(1), 169–176. https://doi.org/10.1016/0378-1119(88)90139-4

Persson, D., Halberg, K. A., Jørgensen, A., Ricci, C., Møbjerg, N., & Kristensen, R. M. (2011). Extreme stress tolerance in tardigrades: Surviving space conditions in low earth orbit. *Journal of Zoological Systematics and Evolutionary Research*, *49*(Supplement 1), 90–97. https://doi.org/10.1111/j.1439-0469.2010.00605.x

Picelli, S., Björklund, Å. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research*, *24*(12), 2033–2040. http://genome.cshlp.org/content/24/12/2033.abstract

Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R., & Coller, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, *160*(6), 1111–1124. https://doi.org/10.1016/j.cell.2015.02.029

Qin, N., Fan, X. C., Zhang, Y. Y., Xu, X. X., Tyasi, T. L., Jing, Y., Mu, F., Wei, M. L., & Xu, R. F. (2015). New insights into implication of the SLIT/ROBO pathway in the prehierarchical follicle development of hen ovary. *Poultry Science*, *94*(9), 2235–2246. https://doi.org/10.3382/ps/pev185

Rabani, M., Pieper, L., Chew, G. L., & Schier, A. F. (2017). A Massively Parallel Reporter Assay of 3′ UTR Sequences Identifies In Vivo Rules for mRNA Degradation. *Molecular Cell*, *68*(6), 1083-1094.e5. https://doi.org/10.1016/j.molcel.2017.11.014

Radford, H. E., Meijer, H. A., & de Moor, C. H. (2008). Translational control by cytoplasmic polyadenylation in Xenopus oocytes. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms 1779*(4), 217–229. https://doi.org/10.1016/j.bbagrm.2008.02.002

Ramos, S. B. v, Stumpo, D. J., Kennington, E. A., Phillips, R. S., Bock, C. B., Ribeiro-Neto, F., & Blackshear, P. J. (2004). The CCCH tandem zinc-finger protein Zfp36l2 is crucial for female fertility and early embryonic development. *Development, 131*(19):4883-93. https://doi.org/10.1242/dev.01336

Rana, A. A., Roper, S. J., Palmer, E. A., & Smith, J. C. (2011). Loss of *Xenopus tropicalis* EMSY causes impairment of gastrulation and upregulation of p53. *New Biotechnology*, *28*(4), 334–341. /https://doi.org/10.1016/j.nbt.2010.10.010

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., & Yang, A. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, *499*(7457), 172–177. https://doi.org/10.1038/nature12311

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecenas, D., … Hughes, T. R. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, *499*(7457), 172–177. https://doi.org/10.1038/nature12311

Redelings, B. D., & Holder, M. T. (2017). A supertree pipeline for summarizing phylogenetic and taxonomic information for millions of species. *PeerJ*, *2017*(3). https://doi.org/10.7717/peerj.3058

Regev, A., Goldman, S., & Shalev, E. (2005). Expression of plexin-B1 in the mouse ovary and its possible role in follicular development. *Fertility and Sterility*, *84*, 1210–1219. https://doi.org/10.1016/j.fertnstert.2005.05.011

Regev, A., Goldman, S., & Shalev, E. (2007). Semaphorin-4D (Sema-4D), the Plexin-B1 ligand, is involved in mouse ovary follicular development. *Reproductive Biology and Endocrinology*, *5*(1), 12. https://doi.org/10.1186/1477-7827-5-12

Riedl, R. (1978). Order in living organisms: a systems analysis of evolution. John Wiley & Sons.

Rodriguez, C., Hughes-Davies, L., Vallès, H., Orsetti, B., Cuny, M., Ursule, L., Kouzarides, T., & Theillet, C. (2004). Amplification of the BRCA2 Pathway Gene EMSY in Sporadic Breast Cancer Is Related to Negative Outcome. *Clinical Cancer Research*, *10*(17), 5785–5791. https://doi.org/10.1158/1078-0432.CCR-03-0410

Rohlfs, R. v, & Nielsen, R. (2015). Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. *Systematic Biology*, *64*(5), 695–708. https://doi.org/10.1093/sysbio/syv042

Rothberg, J. M., Hartley, D. A., Walther, Z., & Artavanis-Tsakonas, S. (1988). slit: an EGF-homologous locus of *D. melanogaster* involved in the development of the embryonic central nervous system. *Cell*, *55*(6), 1047–1059. https://doi.org/10.1016/0092-8674(88)90249-8

Roy, A., Lin, Y.-N., Agno, J. E., DeMayo, F. J., & Matzuk, M. M. (2009). Tektin 3 is required for progressive sperm motility in mice. *Molecular Reproduction and Development*, *76*(5), 453–459. https://doi.org/10.1002/mrd.20957

Ruddock-D'Cruz, N. T., Prashadkumar, S., Wilson, K. J., Heffernan, C., Cooney, M. A., French, A. J., Jans, D. A., Verma, P. J., & Holland, M. K. (2008). Dynamic changes in localization of chromobox (CBX) family members during the maternal to embryonic transition. *Molecular Reproduction and Development*, *75*(3), 477–488. https://doi.org/10.1002/mrd.20752

Rudwick, M. J. S. (1970). Living and fossil brachiopods. Hutchinson

Santagata, S., Resh, C., Hejnol, A., Martindale, M. Q., & Passamaneck, Y. J. (2012). Development of the larval anterior neurogenic domains of *Terebratalia transversa* (Brachiopoda) provides insights into the diversification of larval apical organs and the spiralian nervous system. *EvoDevo*, *3*(1), 3. https://doi.org/10.1186/2041-9139-3-3

Schüpbach, T., & Wieschaus, E. (1986). Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. *Roux's Archives of Developmental Biology*, *195*, 302–317. https://doi.org/10.1007/BF00376063

Schuster, R. O., Nelson, D. R., Grigarick, A. A., & Christenberry, D. (1980). Systematic Criteria of the Eutardigrada. *Transactions of the American Microscopical Society 99*(3), 284-303. https://doi.org/10.2307/3226004

Semotok, J. L., Cooperstock, R. L., Pinder, B. D., Vari, H. K., Lipshitz, H. D., & Smibert, C. A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early *Drosophila* embryo. *Current Biology*, *15*(4), 284–294. https://doi.org/10.1016/j.cub.2005.01.048

Shen-Orr, S. S., Pilpel, Y., & Hunter, C. P. (2010). Composition and regulation of maternal and zygotic transcriptomes reflects species-specific reproductive mode. *Genome Biology*, *11*(6). https://doi.org/10.1186/gb-2010-11-6-r58

Shirley, T. C., & Storch, V. (1999). *Halicryptus higginsi* n.sp. (Priapulida): A Giant New Species from Barrow. In *Biology* (Vol. 118, Issue 4).

Signor, S. A., & Nuzhdin, S. v. (2018). The Evolution of Gene Expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. /https://doi.org/10.1016/j.tig.2018.03.007

Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F., Yue, J. X., Chen, Y. C., Freeman, R. M., Sasaki, A., Hikosaka-Katayama, T., Sato, A., Fujie, M., Baughman, K. W., Levine, J., … Gerhart, J. (2015). Hemichordate genomes and deuterostome origins. *Nature*, *527*(7579), 459–465. https://doi.org/10.1038/nature16150

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

Skovsted, C. B., Brock, G. A., Topper, T. P., Paterson, J. R., & Holmer, L. E. (2011). Scleritome construction, biofacies, biostratigraphy and systematics of the tommotiid *Eccentrotheca helenia* sp. nov. from the Early Cambrian of South Australia. *Palaeontology*, *54*(2), 253–286. https://doi.org/10.1111/j.1475-4983.2010.01031.x

Soller, M., & White, K. (2003). ELAV inhibits 3′-end processing to promote neural splicing of ewg pre-mRNA. *Genes & Development*, *17*(20), 2526–2538. https://doi.org/10.1101/gad.1106703

Soller, M., & White, K. (2005). ELAV multimerizes on conserved AU4-6 motifs important for ewg splicing regulation. *Molecular and Cellular Biology*, *25*(17), 7580–7591. https://doi.org/10.1128/MCB.25.17.7580-7591.2005

Sperling, E. A., Pisani, D., & Peterson, K. J. (2011). Molecular paleobiological insights into the origin of the Brachiopoda. *Evolution and Development*, *13*(3), 290–303. https://doi.org/10.1111/j.1525-142X.2011.00480.x

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., & Cohen, S. M. (2005). Animal microRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell*, *123*(6), 1133–1146. https://doi.org/10.1016/j.cell.2005.11.023

Stedden, C. G., Menegas, W., Zajac, A. L., Williams, A. M., Cheng, S., Özkan, E., & Horne-Badovinac, S. (2019). Planar-Polarized Semaphorin-5c and Plexin A Promote the Collective Migration of Epithelial Cells in Drosophila. *Current Biology*, *29*(6), 908-920.e6. /https://doi.org/10.1016/j.cub.2019.01.049

Stein, P., Savy, V., Williams, A. M., & Williams, C. J. (2022). Modulators of calcium signalling at fertilization. *Open Biology*, *10*(7), 200118. https://doi.org/10.1098/rsob.200118

Stoeckius, M., Grün, D., Kirchner, M., Ayoub, S., Torti, F., Piano, F., Herzog, M., Selbach, M., & Rajewsky, N. (2014). Global characterization of the

oocyte-to-embryo transition in *Caenorhabditis elegans* uncovers a novel m RNA clearance mechanism . *The EMBO Journal*, *33*(16), 1751–1766. https://doi.org/10.15252/embj.201488769

Stricker, S. A., & Folsom, M. W. (1997). Oocyte Maturation in the Brachiopod *Terebratalia transversa*: Role of Follicle Cell-Oocyte Attachments During Ovulation and Germinal Vesicle Breakdown. *The Biological Bulletin*, *193*(3), 324–340. https://doi.org/10.2307/1542935

Stroband, H. W. J., te Krounie, G., & van Gestel, W. (1992). Differential susceptibility of early steps in carp (Cyrinus carpio) development to α-amanitin. *Roux's Archives of Developmental Biology*, *202*(1), 61–65. https://doi.org/10.1007/BF00364597

Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H., & Bartel, D. P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, *508*(1), 66–71. https://doi.org/10.1038/nature13007

Surget-Groba, Y., & Montoya-Burgos, J. I. (2010). Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Research*, *20*(10), 1432–1440. http://genome.cshlp.org/content/20/10/1432.abstract

Tadros, W., Goldman, A. L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T. R., Westwood, J. T., Smibert, C. A., & Lipshitz, H. D. (2007). SMAUG Is a Major Regulator of Maternal mRNA Destabilization in *Drosophila* and Its Translation Is Activated by the PAN GU Kinase. *Developmental Cell*, *12*(1), 143–155. https://doi.org/10.1016/j.devcel.2006.10.005

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: A play in two acts. *Development*, *136*(18), 3033–3042. https://doi.org/10.1242/dev.033183

Takahashi, M., Takahashi, F., Ui-Tei, K., Kojima, T., & Saigo, K. (2005). Requirements of genetic interactions between Src42A, armadillo and shotgun, a gene encoding E-cadherin, for normal development in *Drosophila*. *Development*, *132*, 2547–2559. https://doi.org/10.1242/dev.01850

Takahashi, T., Fournier, A., Nakamura, F., Wang, L.-H., Murakami, Y., Kalb, R. G., Fujisawa, H., & Strittmatter, S. M. (1999). Plexin-Neuropilin-1 Complexes Form Functional Semaphorin-3A Receptors. *Cell*, *99*(1), 59–69. https://doi.org/10.1016/S0092-8674(00)80062-8

Thomsen, S., Anders, S., Janga, S. C., Huber, W., & Alonso, C. R. (2010). Genome-wide analysis of mRNA decay patterns during early *Drosophila* development. *Genome Biology*, *11*(9). https://doi.org/10.1186/gb-2010-11-9-r93

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Physical Review*, *36*(5), 823–841. https://doi.org/10.1103/PhysRev.36.823

van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E., & Séraphin, B. (2002). Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *The EMBO Journal*, *21*(24), 6915–6924. https://doi.org/10.1093/emboj/cdf678

van Dijk, E. L., Sussenbach, J. S., & Holthuizen, P. E. (2001). Kinetics and regulation of site-specific endonucleolytic cleavage of human IGF-II mRNAs. *Nucleic Acids Research*, *29*(17), 3477–3486. https://doi.org/10.1093/nar/29.17.3477

Vannier, J., Calandra, I., Gaillard, C., & Zylińska, A. (2010). Priapulid worms: Pioneer horizontal burrowers at the Precambrian-Cambrian boundary. *Geology*, *38*(8), 711–714. https://doi.org/10.1130/G30829.1

Vastenhouw, N. L., Cao, W. X., & Lipshitz, H. D. (2019). The maternal-to-zygotic transition revisited. *Development*, *146*(11), dev161471. https://doi.org/10.1242/dev.161471

Viré, E., Curtis, C., Davalos, V., Git, A., Robson, S., Villanueva, A., Vidal, A., Barbieri, I., Aparicio, S., Esteller, M., Caldas, C., & Kouzarides, T. (2014). The Breast Cancer Oncogene EMSY Represses Transcription of Antimetastatic microRNA miR-31. *Molecular Cell*, *53*(5), 806–818. https://doi.org/10.1016/j.molcel.2014.01.029

Vitale, A. M., Calvert, M. E. K., Mallavarapu, M., Yurttas, P., Perlin, J., Herr, J., & Coonrod, S. (2007). Proteomic profiling of murine oocyte maturation. *Molecular Reproduction and Development*, *74*(5), 608–616. https://doi.org/10.1002/mrd.20648

von Baer, K. E. (1828). Über Entwickelungsgeschichte der Thiere: Beobachtung und Reflexion. Bei den Gebrüdern Bornträger. https://doi.org/10.5962/bhl.title.6303

von Stetina, J. R., & Orr-Weaver, T. L. (2011). Developmental Control of Oocyte Maturation and Egg Activation in Metazoan Models. *Cold Spring Harbor Perspectives in Biology*, *3*(10). http://cshperspectives.cshlp.org/content/3/10/a005553.abstract

Wagner, G. P., & Altenberg, L. (1996). Perspective: complex adaptations and the evolution of evolvability. *International Journal of Organic Evolution*, *50*(3), 967–976.

Wagner, G. P., Kin, K., & Lynch, V. J. (2013). A model based criterion for gene expression calls using RNA-seq data. *Theory in Biosciences*, *132*(3), 159–164. https://doi.org/10.1007/s12064-013-0178-3

Wang, J., Garrey, J., & Davis, R. E. (2014). Transcription in pronuclei and one- to four-cell embryos drives early development in a nematode. *Current Biology*, *24*(2), 124–133. https://doi.org/10.1016/j.cub.2013.11.045

Wang, L., Eckmann, C. R., Kadyk, L. C., Wickens, M., & Kimble, J. (2002). A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*. *Nature*, *419*(6904), 312–316. https://doi.org/10.1038/nature01039

Wang, Q. T., Piotrowska, K., Ciemerych, M. A., Milenkovic, L., Scott, M. P., Davis, R. W., & Zernicka-Goetz, M. (2004). A Genome-Wide Study of Gene Activity Reveals Developmental Signaling Pathways in the Preimplantation Mouse Embryo. *Developmental Cell*, *6*(1), 133–144. https://doi.org/10.1016/S1534-5807(03)00404-0

Webster, B., Copley, R., Jenner, R., Mackenzie-Dodds, J., Bourlat, S., Rota-Stabelli, O., Littlewood, D. T. J., & Telford, M. (2006). Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral Ecdysozoan. *Evolution & Development*, *8*, 502–510. https://doi.org/10.1111/j.1525-142X.2006.00123.x

Wegener, M., & Müller-McNicoll, M. (2018). Nuclear retention of mRNAs – quality control, gene regulation and human disease. *Seminars in Cell & Developmental Biology*, *79*, 131–142. https://doi.org/10.1016/j.semcdb.2017.11.001

Wei, Z., Angerer, R. C., & Angerer, L. M. (2006). A database of mRNA expression patterns for the sea urchin embryo. *Developmental Biology*, *300*(1), 476–484. https://doi.org/10.1016/j.ydbio.2006.08.034

Wennberg, S. A., Janssen, R., & Budd, G. E. (2008). Early embryonic development of the priapulid worm *Priapulus caudatus*. *Evolution & Development*, *10*(3), 326–338. https://doi.org/10.1111/j.1525-142X.2008.00241.x

Wennberg, S. A., Janssen, R., & Budd, G. E. (2009). Hatching and earliest larval stages of the priapulid worm Priapulus caudatus. *Invertebrate Biology*, *128*(2), 157–171. https://doi.org/10.1111/j.1744-7410.2008.00162.x

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., & Scheffler, K. (2015). RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. *Molecular Biology and Evolution*, *32*(3), 820–832. https://doi.org/10.1093/molbev/msu400

Wilk, R., Hu, J., Blotsky, D., & Krause, H. M. (2016). Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. *Genes & Development*, *30*(5), 594–609. https://doi.org/10.1101/gad.276931.115

Williams, A., Carlson, S. J., Brunton, C. H. C., Holmer, L. E., & Popov, L. (1996). A supra-ordinal classification of the Brachiopoda. *Philosophical Transactions of the*

Royal Society of London. Series B: Biological Sciences, *351*(1344), 1171–1193. https://doi.org/10.1098/rstb.1996.0101

Willmore, K. E., Young, N. M., & Richtsmeier, J. T. (2007). Phenotypic variability: Its components, measurement and underlying developmental processes. *Evolutionary Biology*, *34*(3–4), 99–120. https://doi.org/10.1007/s11692-007-9008-1

Winata, C. L., Łapiński, M., Pryszcz, L., Vaz, C., bin Ismail, M. H., Nama, S., Hajan, H. S., Lee, S. G. P., Korzh, V., Sampath, P., Tanavde, V., & Mathavan, S. (2018). Cytoplasmic polyadenylation-mediated translational control of maternal mRNAs directs maternal-to-zygotic transition. *Development*, *145*(1), dev159566. https://doi.org/10.1242/dev.159566

Winberg, M. L., Noordermeer, J. N., Tamagnone, L., Comoglio, P. M., Spriggs, M. K., Tessier-Lavigne, M., & Goodman, C. S. (1998). Plexin A Is a Neuronal Semaphorin Receptor that Controls Axon Guidance. *Cell*, *95*(7), 903–916. https://doi.org/10.1016/S0092-8674(00)81715-8

Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y. H., Abdellaoui, A., Batista, S., Butler, C., Chen, G., Chen, T. H., D'Ambrosio, D., Gallins, P., Ha, M. J., Hottenga, J. J., … Boomsma, D. I. (2014). Heritability and genomics of gene expression in peripheral blood. *Nature Genetics*, *46*(5), 430–437. https://doi.org/10.1038/ng.2951

Wu, L., Ferger, K. E., & Lambert, J. D. (2019). Gene Expression Does Not Support the Developmental Hourglass Model in Three Animals with Spiralian Development. *Molecular Biology and Evolution*, *36*(7), 1373–1383. https://doi.org/10.1093/molbev/msz065

Wu, Y., Li, M., & Yang, M. (2021). Post-Translational Modifications in Oocyte Maturation and Embryo Development. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.645318

Xu, Y.-Q., Li, X., Zhong, Y., & Zheng, Y.-F. (2021). Evolution and diversity of axon guidance Robo receptor family genes. *Journal of Systematics and Evolution*, *59*(1), 169–182. https://doi.org/10.1111/jse.12587

Yang, M., May, W. S., & Ito, T. (1999). JAZ Requires the Double-stranded RNA-binding Zinc Finger Motifs for Nuclear Localization. *Journal of Biological Chemistry*, *274*(39), 27399–27406. https://doi.org/10.1074/jbc.274.39.27399

Ye, S., Hao, X., Zhou, T., Wu, M., Wei, J., Wang, Y., Zhou, L., Jiang, X., Ji, L., Chen, Y., You, L., Zhang, Y., Xu, G., Zhou, J., Ma, D., & Wang, S. (2010). Plexin-B1

silencing inhibits ovarian cancer cell migration and invasion. *BMC Cancer*, *10*(1), 611. https://doi.org/10.1186/1471-2407-10-611

Yu, Q., Li, X.-T., Zhao, X., Liu, X.-L., Ikeo, K., Gojobori, T., & Liu, Q.-X. (2014). Coevolution of Axon Guidance Molecule Slit and Its Receptor Robo. *PLOS ONE*, *9*(5), e94970-. https://doi.org/10.1371/journal.pone.0094970

Zhao, B. S., Wang, X., Beadell, A. v., Lu, Z., Shi, H., Kuuspalu, A., Ho, R. K., & He, C. (2017). M6 A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature*, *542*(7642), 475–478. https://doi.org/10.1038/nature21355

Zhao, P., Shi, C., Wang, L., & Sun, M. (2022). The parental contributions to early plant embryogenesis and the concept of maternal-to-zygotic transition in plants. *Current Opinion in Plant Biology*, *65*, 102144.https://doi.org/10.1016/j.pbi.2021.102144

Zhu, W., Wang, T., Zhao, C., Wang, D., Zhang, X., Zhang, H., Chi, M., Yin, S., & Jia, Y. (2018). Evolutionary conservation and divergence of Vasa, Dazl and Nanos1 during embryogenesis and gametogenesis in dark sleeper (*Odontobutis potamophila*). *Gene*, *672*, 21–33. https://doi.org/10.1016/j.gene.2018.06.016

Zuoren, W., Xinfu, J., Anne, C.-S., & Megerditch, K. (2002). The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proceedings of the National Academy of Sciences*, *99*(20), 12663–12668. https://doi.org/10.1073/pnas.192445599

uib.no