THE UNIVERSITY OF CHICAGO

EVOLUTION OF EMBRYONIC AXIS DETERMINANTS VIA ALTERNATIVE TRANSCRIPTION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN INTEGRATIVE BIOLOGY

BY

YOSEOP YOON

CHICAGO, ILLINOIS
JUNE 2019

To Shinhae,

For everything.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Urs Schmidt-Ott, for guiding me through 6 years of my Ph.D. journey. I was lucky to be able to pursue a project that he and I both loved, and learn tremendously from his expertise, rigor, and enthusiasm. He was always there for a discussion, encouraged me to develop my own thinking, and taught me to become not only a better scientist, but a better person. I genuinely feel sad leaving his lab. I also would like to thank my thesis committee members, Edwin Ferguson, John Reinitz, and Robert Ho for their valuable support for my research and the manuscript, as well as job searching.

My work would not have been possible without my talented collaborators, particularly Jeff Klomp, who enabled me to explore anterior determinants of different fly species. I feel grateful to Ines Martin-Martin, Frank Criscione, Eric Calvo, and Jose Ribeiro for welcoming me to their lab for experiments and providing me with samples and data. I also thank Channa Aluvihare and Robert Harell for sending me mosquitoes many times, and Elena Davydova, who kindly taught me cloninging and other important techniques for my research.

I also would like to thank my current and previous lab members, especially Chun Wai Kwan, for teaching me everything when I knew nothing. I thank Suzanne Stratton, Elise Lemp, Daniel Sanchez, and Michael Ludwig for providing help and friendship. A special thanks to my Darwinian colleagues and friends, who made my life in Chicago so memorable.

Finally, I would like to thank my family, who have always supported me from the bottom of their hearts, and my lovely Windy and Robin for tons of emotional support. I dedicate this thesis to my wife, Shinhae, who encouraged me to chase my dream despite 5 years of long-distance relationship. I would not have been able to start nor finish this journey without her.

ABSTRACT

In flies (Diptera), lineage-specific new genes (orphan genes), such as *bicoid* in the fruit fly *Drosophila melanogaster* and *panish* in the midge *Chironomus riparius*, have been found to establish head-to-tail polarity of the embryo. However, the embryonic axis determinants of most fly species have not been identified. In this study, I show that embryonic axis determinants of distantly related fly species have frequently evolved from old genes through the use of alternative transcription start or termination sites. In moth flies (Psychodidae), including the drain fly *Clogmia albipunctata* and the sand fly *Lutzomyia longipalpis*, a maternal transcript isoform of *odd-paired (Zic)* generated by alternative transcription initiation is localized in the anterior pole of freshly laid eggs, and is necessary and sufficient for anterior development of the Clogmia embryo. In the Southern house mosquito *Culex quinquefasciatus* (Culicidae), a maternal transcript isoform of a previously uncharacterized gene, named *cucoid,* is generated by alternative transcription termination (alternative polyadenylation), localized in the anterior pole of the egg, and necessary for establishing embryo polarity. This function of *cucoid* is also conserved in the yellow fever mosquito *Aedes aegypti* (Culicidae), suggesting a conserved role of this gene in axis specification in culicine mosquitoes. In the malaria mosquito *Anopheles gambiae* (Culicidae), a maternal transcript isoform of *pangolin* (*Tcf*) generated by alternative polyadenylation functions as the axis determinant. I also found localized maternal *pangolin* transcripts in the anterior eggs of *Nephrotoma suturalis*, a basal-branching crane fly (Tipulidae), suggesting that *pangolin* functioned as ancestral axis determinant in flies. This also suggests that *panish*, the *pangolin*-related orphan axis determinant gene of Chironomus, inherited its function from *pangolin*. Finally, I provide preliminary evidence for a diverged axis specification mechanism in the black soldier fly *Hermetia illucens* (Stratiomyidae), which is closely related to

species with *bicoid*. I found that, unlike in other flies, numerous maternal mRNAs of conserved genes are localized in the anterior pole of the Hermetia egg. Together, I conclude that alternative transcription played an important role in the evolution of new developmental gene functions and gene regulatory networks in fly embryos. Given that alternative transcription is a widespread phenomenon and can underlie the evolution of entirely new gene functions, it may play a much more important role in animal evolution than previously thought.

1. Introduction

1.1. Alternative transcription

Alternative transcription is a widespread phenomenon

Cis-regulatory DNA has long been recognized as an important target in evolution (Carroll, Grenier, and Weatherbee 2005; Peter and Davidson 2011; Wittkopp and Kalay 2011). However, the contribution of alternative transcription (AT) to the evolution of gene regulatory networks remains poorly understood. AT allows a single gene to generate multiple transcript isoforms with distinct 5' and 3' ends through the use of alternative promoters (alternative transcription initiation, ATI) and polyadenylation signals (alternative transcription termination, ATT) (de Klerk and t Hoen 2015), and involves transcription of multiple pre-mRNAs from a single gene. Transcript isoforms generated by ATI are characterized by alternative first exons (**Figure 1.1.1.**) (Davuluri et al. 2008). The use of alternative first exons changes the 5'UTR sequence but it can also lead to modification in the N-terminal region of the open reading frame. Conversely, ATT generates isoforms with 3'UTR sequences of various length, or alternative last exons if coupled with alternative splicing (AS) (Tian and Manley 2017) (**Figure 1.1.1.**).

Advances in high-throughput sequencing technology allowed genome-wide analysis of AT events using methods that can specifically detect 5' or 3' ends of mRNAs. For example, ATI can be detected by Cap Analysis of Gene Expression (CAGE). In this method, a linker is attached to the 5' end of full-length cDNA, which allows the restriction endonuclease to cleave the initial 20 nucleotides of cDNA. A second linker is then attached to the 3' cleavage site and the fragment is amplified for sequencing, revealing the initial 20 bp of mRNA (Shiraki et al. 2003). Sequencing results are then mapped to the genome, which can detect alternative

transcription start sites of a gene. Several advanced CAGE methods have been developed, such as deepCAGE (Valen et al. 2009), which allows sequencing of the initial 27bp of mRNA and simplifies steps for high-throughput sequencing, and HeliScopeCAGE (Kanamori-Katayama et al. 2011), which is compatible with HeliScope single molecule sequencing. For detecting ATT events, a number of methods are currently available which can be broadly classified into two categories (Chen et al. 2017). One involves Oligo(dT) priming-based methods, which target the poly(A) tail of mRNAs using an oligo(dT) primer to generate cDNA. Methods in this category include PAS-seq (Shepard et al. 2011) and 3'-seq (Lianoglou et al. 2013). The other category uses RNA manipulation-based methods that avoid the problem of internal priming caused by oligo(dT) priming. One example of this method is 3P-seq, where biotinylated primers are splint-ligated to the end of the mRNA poly(A) tail for purification prior to the cDNA synthesis (Jan et al. 2011).

Based on methods described above, genome-wide analyses of AT have been conducted, showing that that both ATI and ATT are a widespread phenomenon. For example, there are on average four alternative transcription start sites per gene in humans (Consortium et al. 2014) and at least 70% of mammalian genes are subject to alternative polyadenylation (Derti et al. 2012). Half of the genes of the fruit fly *Drosophila melanogaster* contain more than one polyadenylation signal (Smibert et al. 2012), and about 30% of protein-coding genes in the nematode *Caenorhabditis elegans* have alternative 3'ends (Jan et al. 2011). Frequent switching of transcription start sites and polyadenylation sites has been observed throughout mouse development (Zhang et al. 2017) and between mouse neuronal cell types (Jereb et al. 2018). Finally, the combined contribution of ATI and ATT to isoform diversity exceeds that of AS (Reyes and Huber 2018; Pal et al. 2011; Shabalina et al. 2010).

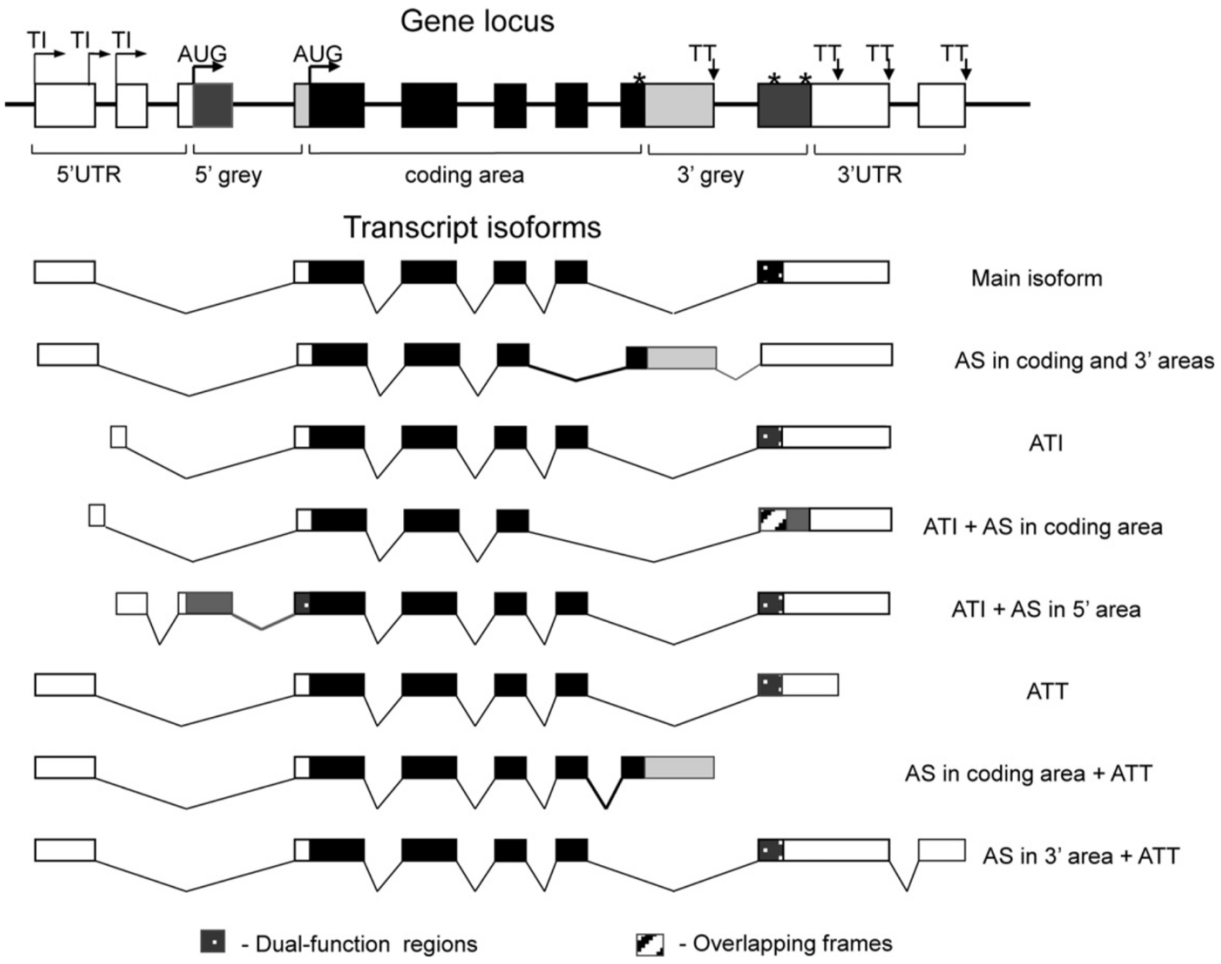


Figure 1.1.1. Types of transcript isoforms generated by alternative transcription**.** TI, transcription initiation site; AUG, translation initiation site; TT, transcription termination site; *, translation termination site; ATI, alternative transcription initiation; AS, alternative splicing; ATT, alternative transcription termination. Black box, canonical protein-coding regions; grey box, non-canonical protein-coding regions; white box, UTRs. Figure from (Shabalina et al. 2014).

Functional consequences of alternative transcription

ATI and ATT can be tightly regulated in different cells or tissues to achieve isoform-specific expression pattern (Davuluri et al. 2008). For example, a number of case studies have shown that ATI can lead to cell/tissue-specific (Pozner et al. 2007; Feng et al. 2016), developmental stage-specific (Davis and Schultz 2000), and sex-biased (Perera and Kim 2016)

expression of transcript isoforms. In addition, a global survey of transcription start sites confirmed that most human genes are expressed in a cell-type-specific manner (Consortium et al. 2014). Similarly, ATT has been shown to generate transcript isoforms with cell or tissue-specific expression (Tian and Manley 2017). For example, transcripts with different 3'UTR lengths or alternative last exons localize differently in neuronal compartments of the mammalian brain (An et al. 2008; Taliaferro et al. 2016; Ciolli Mattioli et al. 2018). The choice of a different 3'UTR length can also result in tissue-specific expression (Smibert et al. 2012; Tushev et al. 2018) and protein localization (Berkovits and Mayr 2015).

ATI and ATT can also affect translational efficiency. For example, longer 5'UTR sequences in isoforms were generally associated with lower translational efficiency in mouse, which can be explained by the existence of upstream open reading frame (uORF) or secondary RNA structure that inhibits translation (Wang et al. 2016). The choice between alternative 3'UTR sequences can affect translational efficiency as well, likely due to regulators that interact with regulatory elements in the 3'UTR (Lau et al. 2010; Pinto et al. 2011).

ATI and ATT are frequently coupled with AS. For example, in human and mouse, a positive correlation was observed between the number of alternative promoters and higher AS rates (Xin, Hu, and Kong 2008), and more than 80% of AS showed interdependency with the alternative choice of transcription start sites in human breast cancer cells (Anvar et al. 2018). A survey of mouse and human protein-coding genes also revealed strong coupling of AS with both ATI and ATT, showing that AS in 5'UTRs and 3'UTRs almost always involves ATI and ATT events, respectively (Shabalina et al. 2010). Moreover, it was found that alteration of the splicing factor hnRNP H affects ATT, providing functional evidence for coupled AS and ATT (Martinez et al. 2016).

Misregulation of AT has been associated with human diseases including cancer (Pal, Gupta, and Davuluri 2012). For example, ATI of anaplastic lymphoma kinase (ALK) generates truncated protein isoforms that stimulates several oncogene signaling pathways (Wiesner et al. 2015). Shortening of 3'UTRs of oncogene transcripts can also lead to overexpression of proteins, partly due to the loss of repressive elements in the UTR sequences (Mayr and Bartel 2009). A number of ATI and ATT events were found to be involved in cellular mechanisms similar to the epithelial-mesenchymal transition, which plays a critical role in human cancer (Shapiro et al. 2011).

Functional studies in model organisms have shown that AT can generate dominant negative and alternatively localized protein isoforms implied in different developmental functions (Davuluri et al. 2008). For example, ATI of a Wnt signaling effector gene *Tcf7l2* in mice generates an N-terminal truncated isoform, dnTcf7l2, that lacks the ß-catenin-binding domain and thus functions as a dominant negative antagonist of the canonical Wnt signaling pathway (Vacik, Stubbs, and Lemke 2011). During eye development in mice, ATI generates multiple mRNA and protein isoforms of a transcription factor MITF, which show tissue-specific expression. Some of these isoforms were functionally more important than others, as revealed by genomic deletion experiments (Bharti et al. 2008). In mice, *Runx1,* which plays a role in regulating hematopoiesis, contains two non-redundant alternative promoters (Pozner et al. 2007). In sea-urchins, E-protein ortholog (SpE-protein) generates two isoforms with N-terminal difference via ATI (Schrankel et al. 2016). The authors reported that the shorter isoform can negatively-regulate the canonical isoform in larval immunocytes.

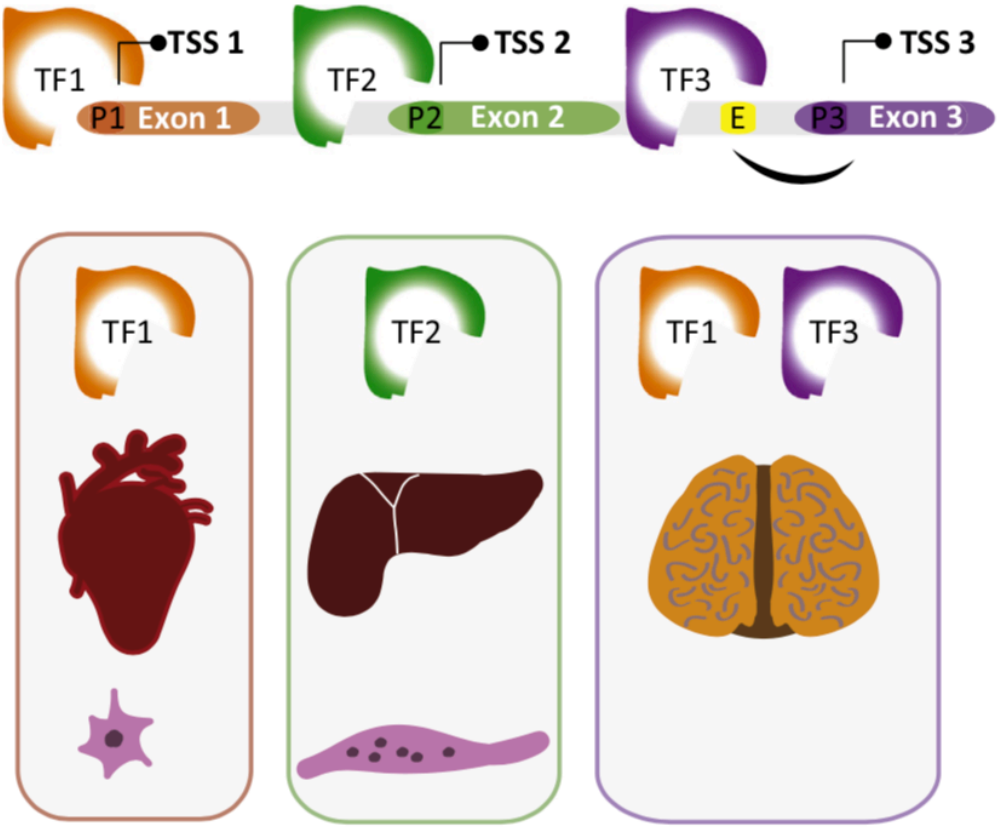


Figure 1.1.2. Illustrated examples of cell- or tissue- specific expression of isoforms via alternative transcription initiation (ATI). P1, P2, P3 represent alterative promoters; TF, transcription factors; E, enhancers; TSS, transcription start site. Figure from (de Klerk and t Hoen 2015).

Role of alternative transcription in evolution

A few studies have explored the role of AT in evolution. For example, it was noted that while a large fraction of alternative promoter sequences are conserved between human and mice, those with cell- or tissue-restricted expression have frequently changed during mammalian evolution (Baek et al. 2007; Consortium et al. 2014). Also, alternative transcription initiation and alternative transcription termination was found to have distinct effects on the evolution of gene expression. Data from mammals suggest that alternative 5'UTRs of genes mainly altered gene transcription whereas alternative 3'UTRs provided post-transcriptional regulatory options for

tissue-specific and conditional expression (Shabalina et al. 2014). Another likely substrate for evolutionary diversification are protein terminal ends generated by AT in conjunction with alternative splicing. These facultative terminal ends commonly contain intrinsically disordered regions that could facilitate the evolution of new protein-protein interactions (Buljan et al. 2013; Shabalina et al. 2014). Finally, a recent case study found that a light fur color variant of beach mice evolved repeatedly via selection for an alternative *agouti* isoform with increase translation efficiency (Linnen et al. 2013; Mallarino et al. 2017).

AT could also play an important role in driving the evolution of entirely new gene functions by altering the subcellular or cellular expression pattern of a gene. Testing this hypothesis *in vivo* requires phylogenetic comparisons to assess whether the specific function of an isoform is a new or inherited. However, AT has been studied predominantly in mice and human cells, where isoform specific loss- and gain-of-function experiments are generally difficult, laborious, and can be hard to interpret due to the complex anatomy and/or small cell sizes of these experimental systems. As a result, the role of AT in evolution has remained poorly understood.

1.2. A new experimental system for studying the evolution of gene function

Primary axis specification in embryos of flies (Diptera)

Previous research has revealed a surprising degree of evolutionary plasticity in the axis specification mechanisms of fly embryos (Klomp et al. 2015). I have taken advantage of this plasticity to study the question of how new gene functions evolve in embryogenesis. The insect order of Diptera (flies) spans about 245 million years of evolution (**Figure 1.2.1.**) (Grimaldi and Engel 2005; Wiegmann et al. 2011). The monophyletic origin of this large and diverse order is well established, which is also the case for the nested subgroups Brachycera, which spans ca. 200 million years of evolution, and Cyclorrhapha, which spans ca. 150 million years of evolution and includes the genetic model organism Drosophila. The non-brachyceran flies form a paraphyletic assemblage of multiple independent branches, such as the Culicomorpha (e.g., mosquitoes and chironomids), Psychodomorpha (e.g., moth flies), or Tipulumoprha (e.g. crane flies), and are collectively referred to as "Nematocera."

The specification of the primary axis (head-to-tail) in embryos of flies (Diptera) rests on lineage-specific maternal mRNAs that are localized at the anterior egg pole ("anterior determinants"). The existence of such localized determinants in insects were initially proposed on the basis of ligation and transplantation experiments (Sander 1976). On the basis of phenotypes obtained in chironomid midges after local ultraviolet light (UV-light) irradiation or RNase application, it was further concluded that these anterior determinants contain RNA (Yajima 1960, 1964; Kalthoff and Sander 1968; Kalthoff 1971; Kandler-Singer and Kalthoff 1976). Finally, the first anterior determinant gene (*bicoid*) was identified in a genetic screen for maternal effect mutations in the fruit fly *Drosophila melanogaster* (Fronhöfer et al 1986;

Frigerio et al. 1986; Berleth et al 1988).

Subsequent evolutionary studies suggested that *bicoid* evolved more than 150 million years ago in the lineage of higher flies (Cyclorrhapha) via gene duplication of *Hox3* (also known as *zerknüllt* or *zen*), which functions in the specification of extraembryonic tissue (Stauber et al. 1999, 2000, 2002; Rafiqi et al 2008; Lemke et al 2010; Liu et al 2018). When it was found that *bicoid* had been lost independently in Tephritid fruit flies and Tsetse flies, and that another lineage-specific anterior determinant gene (*panish*), structurally unrelated to *bicoid*, was discovered in common midges (Chironomidae), it was concluded that anterior determinants have frequently changed in dipteran evolution, despite the general conservation of downstream segmentation genes (Klomp et al. 2015).

The poor conservation of anterior determinants in flies provides an excellent model for studying the evolution of gene function. Fly embryos offer several technical advantages that facilitates the identification and functional characterization of new anterior determinants in an evolutionary context. For example, it is possible to identify candidate anterior determinants by comparing anterior and posterior transcriptomes from bisected, freshly-laid eggs (Klomp et al. 2015). This method overcomes the limitation of candidate-gene approaches and can be applied to eggs of many "non-model" fly species. In addition, anterior determinants function during the syncytial phase of fly embryogenesis, when thousands of nuclei migrate to the periphery of the egg to form the blastoderm. At this stage, the cellularization process is still incomplete and their activity can be directly suppressed or altered by injecting early embryos with appropriate reagents. The large size of eggs compared to other cell types also provides a unique opportunity for gain-of-function experiments via injection of mRNA at the posterior pole. Finally, the phenotypes of successful loss-of-function or gain-of-function experiments with anterior

determinants are predictable, as they are expected to result in mirror image duplications of the tail-end (double abdomen) or cephalon (double head). In the next sections, I will describe the mechanisms of action of *bicoid* and *panish* in more detail.

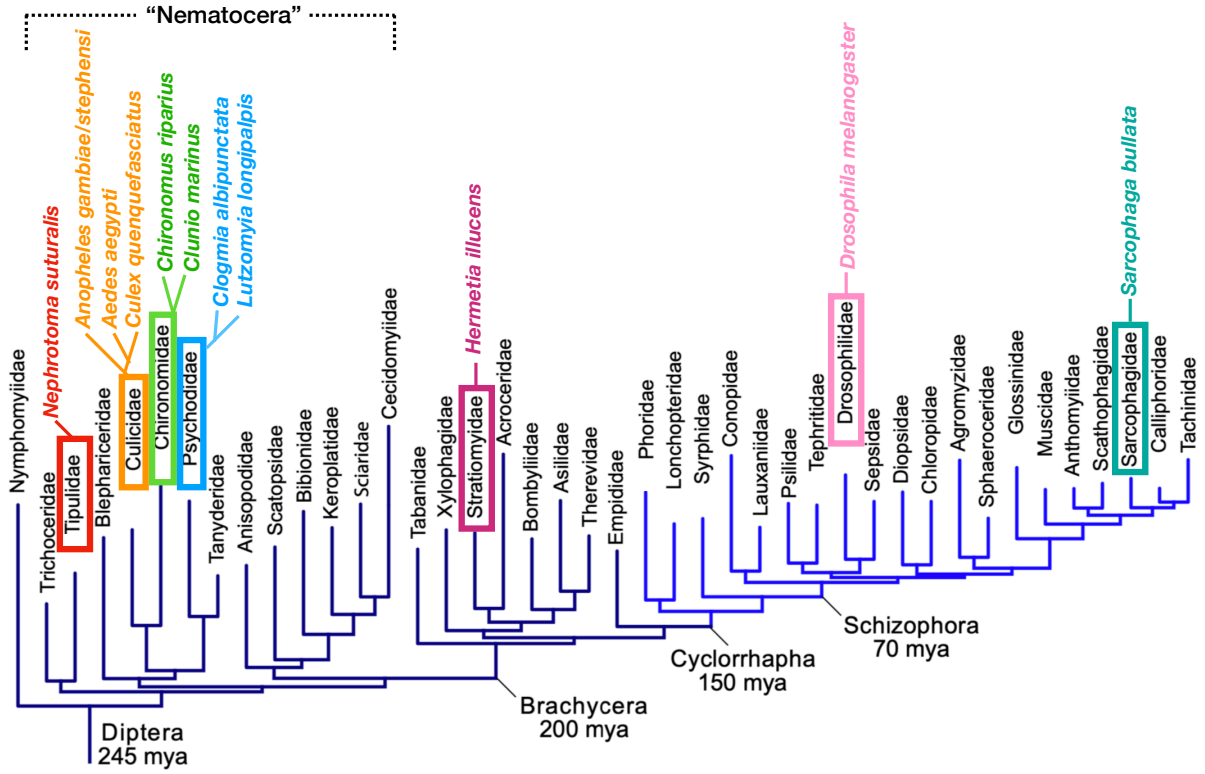


Figure 1.2.1. Phylogeny of Diptera. Species referred to in this study are indicated with colored texts. Tree adapted from (Klomp et al. 2015).

Function of *bicoid* in *Drosophila melanogaster* and other cyclorrhaphan flies

Maternal mRNA of *bicoid* is localized in the anterior pole of the freshly-laid egg, and Bicoid protein forms a gradient in the early embryo (**Figure 1.2.2.**) (Berleth et al. 1988; Driever and Nusslein-Volhard 1988b). Bicoid activates its target genes in a concentration-dependent manner (Driever and Nusslein-Volhard 1988a; Struhl, Struhl, and Macdonald 1989), with synergistic input from maternal *hunchback* (Simpson-Brose, Treisman, and Desplan 1994),

which is translationally repressed in the posterior embryo by Nanos via a Nanos-response-element in its 3'UTR (Wharton and Struhl 1991). It also directly interferes with the translation of ubiquitous maternal *caudal* transcript via 3'UTR and thereby restricts Caudal activity to the posterior embryo (Chan and Struhl 1997; Rivera-Pomar et al. 1996). Bicoid-deficient embryos fail to develop head and thorax and instead form a second set of terminal structures (Frohnhöfer and Nüsslein-Volhard 1986), or symmetrical double abdomen when the maternal activity gradient of *hunchback* is disrupted simultaneously (Driever 1993). Posterior injection of *bicoid* mRNA with heterologous UTR sequences can result in double heads, suggesting that *bicoid* is both necessary and sufficient for inducing the anterior identity (Driever, Siegel, and Nusslein-Volhard 1990). However, the loss of *bicoid* function can be compensated by simultaneously increasing the maternal dosage of *hunchback* and reducing the copy number of *knirps*, a zygotic segmentation gene that antagonizes *hunchback* (Wimmer et al. 2000).

The context of *bicoid* function varies slightly in other cyclorrhaphan flies. In the hover fly *Episyrphus balteatus* (Syrphidae), maternal input of *hunchback* is absent and knockdown of *bicoid* in this species results in mirror-image double abdomens, due to the loss of the central expression domain of *Krüppel*, a zygotic segmentation gene (Lemke et al. 2010). In the scuttle fly *Megaselia abdita* (Phoridae), *bicoid* RNA interference (RNAi), which leads to degradation of *bicoid* mRNA, results in mirror-image double abdomens despite the presence of maternal *hunchback* (Stauber, Taubert, and Schmidt-Ott 2000). This is because maternal *hunchback* in Megaselia is insufficient for the expression of other segmentation genes, such as *Krüppel* (Wotton, Jimenez-Guri, and Jaeger 2015). Megaselia also lacks maternal *caudal* expression, indicating that this species does not use Bicoid as a posttranscriptional regulator of this gene (Stauber, Lemke, and Schmidt-Ott 2008).

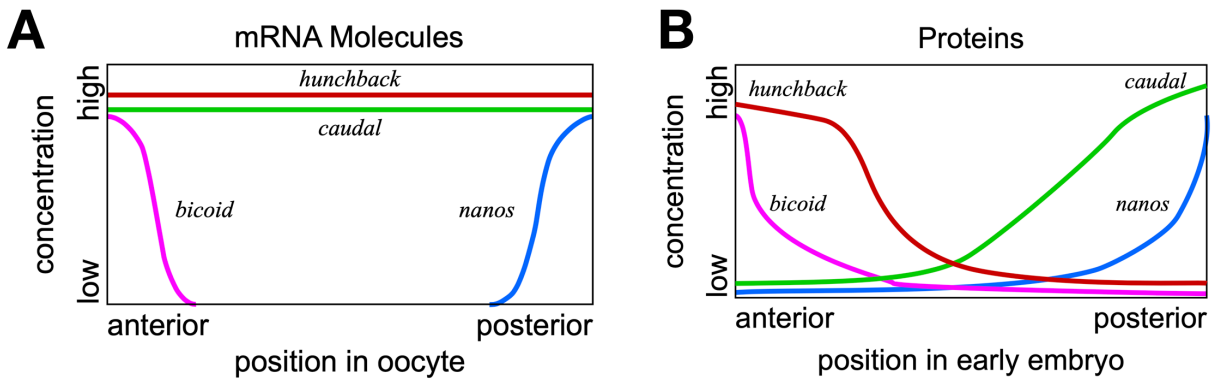


Figure 1.2.2. Examples of maternal mRNA and protein distribution in (A) Drosophila oocyte and (B) early embryo. Illustration from Wikipedia (https://en.wikipedia.org/wiki/Drosophila_embryogenesis, by Fred the Oyster, distributed under GNU Free Documentation License).

Function of *panish* in *Chironomus riparius*

panish was discovered by comparing anterior and posterior transcriptomes of freshly laid Chironomus eggs (Klomp et al. 2015). It was found that the maternal mRNA of this gene is localized at the anterior pole the egg (**Figure 1.2.3.**). *panish* encodes a novel cysteine-clamp gene and *panish* RNAi embryos develop as symmetrical double abdomens, suggesting that *panish* is required for establishing the normal head-to-tail polarity in Chironomus. However, injection of *panish* mRNA at the posterior pole is not sufficient for inducing double heads, suggesting that some posteriorly localized factor blocks its activity.

Further analyses suggested that *panish* may function permissively. The transcription of the *tailless* ortholog (*Cri-tll*) and another posterior gene, Chironomus *caudal* (*Cri-cad*), are repressed in the anterior embryo by *panish.* Unlike in Drosophila, zygotic transcripts of *Cri-tll* are initially expressed in a posterior-to-anterior gradient. *Cri-tll* RNAi results in double heads, suggesting that *Cri-tll* transcript inhibits the default anterior patterning program. This result

suggests that head development can occur in the absence of *panish* activity. Therefore, Panish may act as a transcriptional repressor, whereas Bicoid primarily functions as a transcriptional activator.

Surprisingly, *panish* is structurally unrelated to *bicoid.* This gene evolved by gene duplication and rearrangement from the *Tcf* homolog *pangolin* (*pan*) and, therefore, was named *panish* (for *pan*"ish"). Pangolin functions as the effector of the ß-catenin-dependent Wnt signaling pathway ("canonical" Wnt signaling) but Panish lacks the domain that binds to ß-catenin, and sequence similarity between Pangolin and Panish is limited to the cysteine-clamp domain (30 amino acids). The evolution of *panish* also involved gene fusion with a nucleoside kinase gene *Cri-zap3,* which is expressed maternally in the Chironomus embryo. While *Cri-zap3* shares the first two exons with *panish*, most of its ORF is not shared with *panish* due to AS, and *Cri-zap3* does not have any specific role in axis specification. However, *Cri-zap3* may have lent its promoter to *panish*, thereby allowing maternal expression of *panish*.

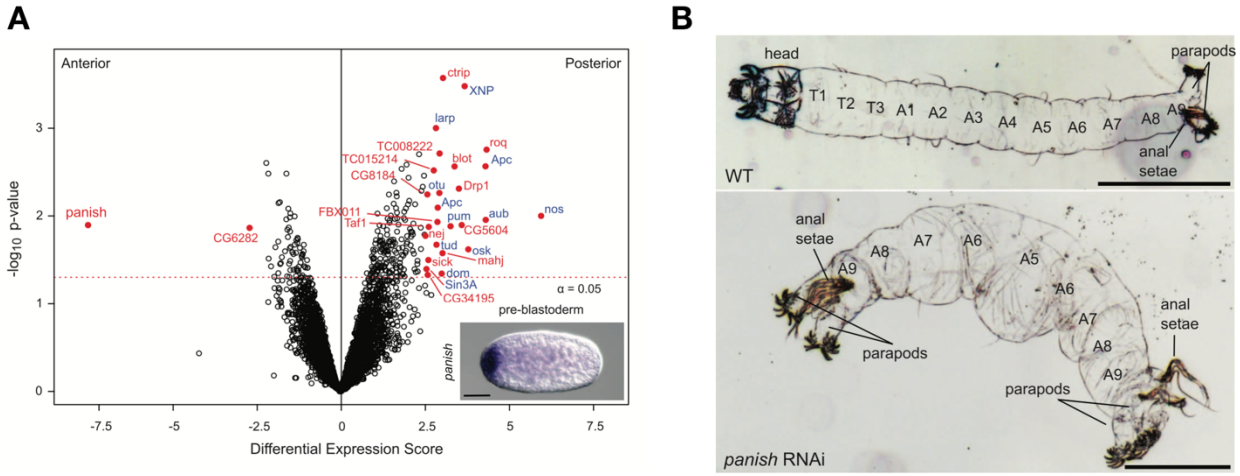


Figure 1.2.3. Expression and function of *panish* in the midge *Chironomus riparius*. **(A)** Differential expression analysis between anterior and posterior transcriptomes of 1hr-old preblastoderm embryo. Red, score > |2.5| and P < 0.05; blue, putative germ plasm components. (lower right corner) RNA in situ hybridization of *panish* in preblastoderm Chironomus embryo. Scale bar: 10µm. **(B)** First-instar larval cuticle of wild-type (top) and following *panish* RNAi (bottom). A, abdominal segment; T, thoracic segment. Scale bars, 30µm. Figures from (Klomp et al. 2015).

1.3. Research goal

The goal of this dissertation has been to identify anterior determinants in embryos of a wider range of dipteran species that lack *bicoid* and *panish* to better understand how these important developmental regulators evolved. I started the analysis with moth flies, including the drain fly *Clogmia albipunctata* (Psychodidae) and the sand fly *Lutzomyia longipalpis* (Psychodidae), subsequently extended it to three mosquito species (Culicidae: *Culex quinquefasciatus*, *Aedes aegypti, Anopheles gambiae*), crane flies (Tipulidae: *Nephrotoma suturalis*), and soldier flies (Stratiomyidae: *Hermetia* illucens) (**Figure 1.2.1.**). Many of the techniques that I used to study gene expression and function in these species had to be adapted or improved and are described in detail in the appendix of this thesis. My results show that anterior determinants of these species repeatedly evolved from conserved genes via AT and suggest that AT played an important role in the evolution of new gene functions and gene regulatory networks in fly embryos.

2. RESULTS

2.1. An alternative maternal transcript of the conserved segmentation gene *odd-paired* functions as anterior determinant in moth flies[1]

Cal-opa generates transcript isoforms with spatially and temporally distinct expression patterns during early embryogenesis

Clogmia albipunctata (Psychodidae) is a moth midge that is also commonly known as the drain fly. It is an emerging model for developmental genetic studies that can be easily reared in the laboratory, with a life cycle of about three to four weeks (Jimenez-Guri et al. 2014). Previous studies analyzed the expression of Clogmia gap genes and pair-rule genes in a comparative context to Drosophila (Rohr, Tautz, and Sander 1999; Bullock and Ish-Horowicz 2001; Garcia-Solache, Jaeger, and Akam 2010; Crombach, Garcia-Solache, and Jaeger 2014; Janssens et al. 2014). However, maternal inputs that determine the polarity of the embryo have been unknown. A classical experiment has shown that centrifugation of Clogmia eggs can result in double heads (Sander 1993), but the genetic basis of this phenotype is unknown, Also, neither *bicoid* nor *panish* were found in genome and transcriptome sequences of this species (Stauber, Prell, and Schmidt-Ott 2002; Klomp et al. 2015; Jimenez-Guri et al. 2014).

To identify the anterior determinant of Clogmia, we compared the expression levels of 5,602 annotated maternal transcripts from the anterior and posterior transcriptomes of 1hr-old bisected Clogmia embryos and ranked them according to the magnitude of their differential

[1] A manuscript describing results 1) through 3) has been submitted for publication. Differential expression analyses of transcriptomes were performed by Jeff Klomp, a previous member of our laboratory. Ines Martin-Martin from the Ribeiro lab at NIAID performed the expression level analysis of Lutzomyia pair-rule genes.

expression scores and P values (**Figure 2.1.1A**). Differential expression analyses were performed with pre-processed contigs that exclude short (<200 nucleotides), low in count (<20), and non-animal genes (e.g. bacterial, fungal, etc.), using quasi-likelihood F-tests based on negative binominal linear models. This test is useful for controlling for an error rate when the sample number is small. The false-discovery rate (FDR) was controlled by the Benjamini-Hochberg method (see methods for details). In the anterior embryo, the most enriched transcript was homologous to *odd-paired*, the Drosophila homolog of mammalian *Zic* (*zinc finger of the cerebellum*) genes. ZIC proteins are known to function as transcription factors or co-factors (Houtmeyers et al. 2013). *odd-paired* was discovered in a screen for early Drosophila segmentation genes and subsequently classified as a 'pair-rule' gene, since *odd-paired* mutants had a periodic segment defect with periodically of two segments (Jürgens et al. 1984). During the Drosophila segmentation process, *odd-paired* is expressed in a single broad domain and controls the 'frequency-doubling' of other pair-rule genes (Clark and Akam 2016).

The Clogmia genome contains a single *odd-paired* locus (Vicoso and Bachtrog 2015). Using RNA-seq data from preblastoderm and blastoderm embryos and Rapid Amplification of cDNA End (RACE), I identified maternal and zygotic *Cal-opa* transcripts with alternative first exons that I mapped onto a 54kb genomic scaffold (**Figure 2.1.1B**). The maternal transcript (*Cal-opa*^{Mat}) was detected in preblastoderm embryos (0.5hr-old) and syncytial blastoderm embryos (4hr-old). The zygotic transcript (*Cal-opa*^{Zyg}) was found in cellularized blastoderm embryos (7hr-old) and gastrulating embryos (9hr-old). Protein alignments with homologs from other flies suggest that *Cal-opa*^{Zyg} encodes the full-length Cal-Opa protein (655 amino acids), while *Cal-opa*^{Mat} encodes a truncated protein variant (635 aa), lacking the N-terminal 20 amino acids of Cal-Opa^{Zyg} **(Figure 2.1.2.)**.

To confirm the alternative *Cal-opa*^{Mat} and *Cal-opa*^{Zyg} transcripts and their non-overlapping expression patterns, I performed whole mount RNA in situ hybridization experiments with transcript-specific probes. The *Cal-opa*^{Mat} transcript was anteriorly localized in preblastoderm embryos but absent at the cellular blastoderm stage. Conversely, the *Cal-opa*^{Zyg} transcript was absent in preblastoderm embryos but expressed broadly in the trunk region of 7hr-old blastoderm embryos (**Figure 2.1.3.**), like *odd-paired* in Drosophila. These observations suggest that *Cal-opa* produces transcript isoforms with spatially and temporally distinct expression patterns.

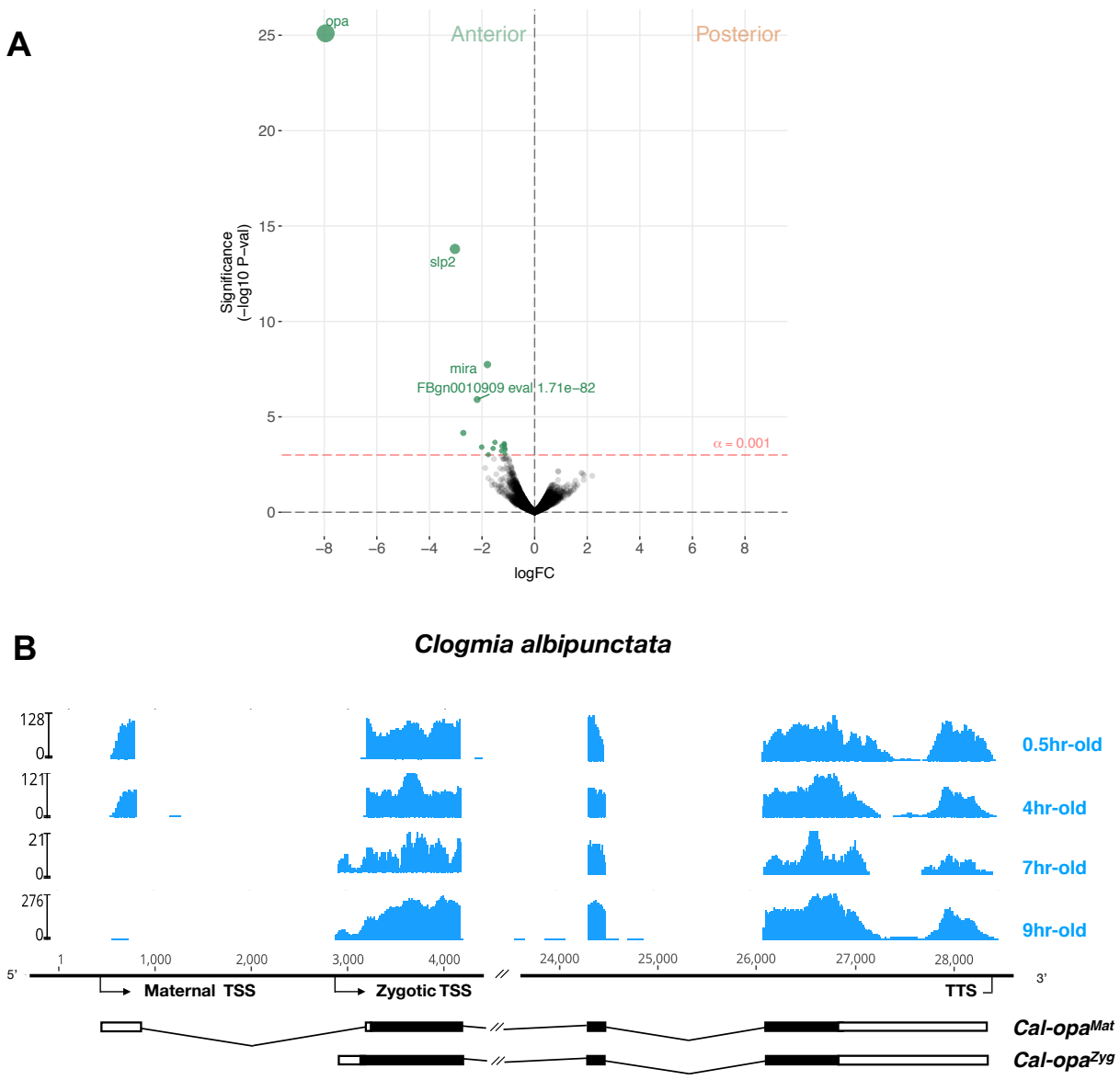


Figure 2.1.1. Maternal expression of *Cal-opa* and alternative *Cal-opa* transcripts in Clogmia embryos. **(A)** Differential expression analysis of maternal transcripts between anterior and posterior halves of 1hr-old Clogmia embryos. logFC: log fold-change. **(B)** Stage-specific RNA-seq read coverage of *Cal-opa* locus. Transcription start sites (TSS) and transcription termination sites (TTS) are indicated on the genomic scaffold (solid line with 1000 bp intervals marked) and were confirmed by RACE. Exon-intron diagram of *Cal-opa*^{Mat} and *Cal-opa*^{Zyg} transcript variants are shown with the open reading frame in black.

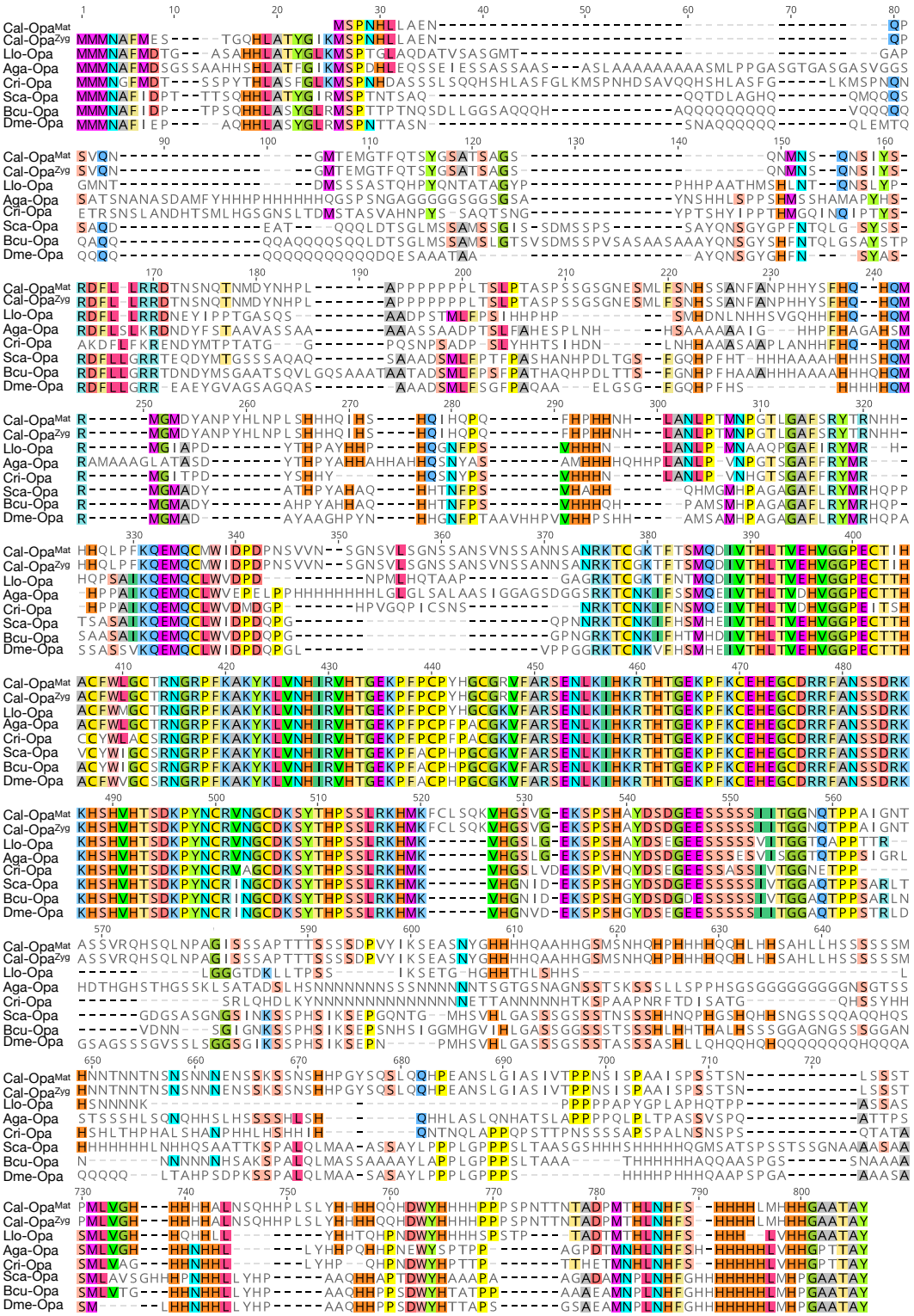


Figure 2.1.2. Protein alignment of predicted dipteran Odd-paired orthologs. Odd-paired orthologs were aligned using *ClustalW 2.1* (Larkin et al. 2007). Cal: *Clogmia albipunctata* (Psychodidae); Llo: *Lutzomyia longipalpis* (Psychodidae); Aga: *Anopheles gambiae* (Culicidae); Cri: *Chironomus riparius* (Chironomidae); Sca: *Stomoxys calcitrans* (Muscidae); Bcu: *Bactrocera cucurbitae* (Tephritidae); Dme: *Drosophila melanogaster* (Drosophilidae).

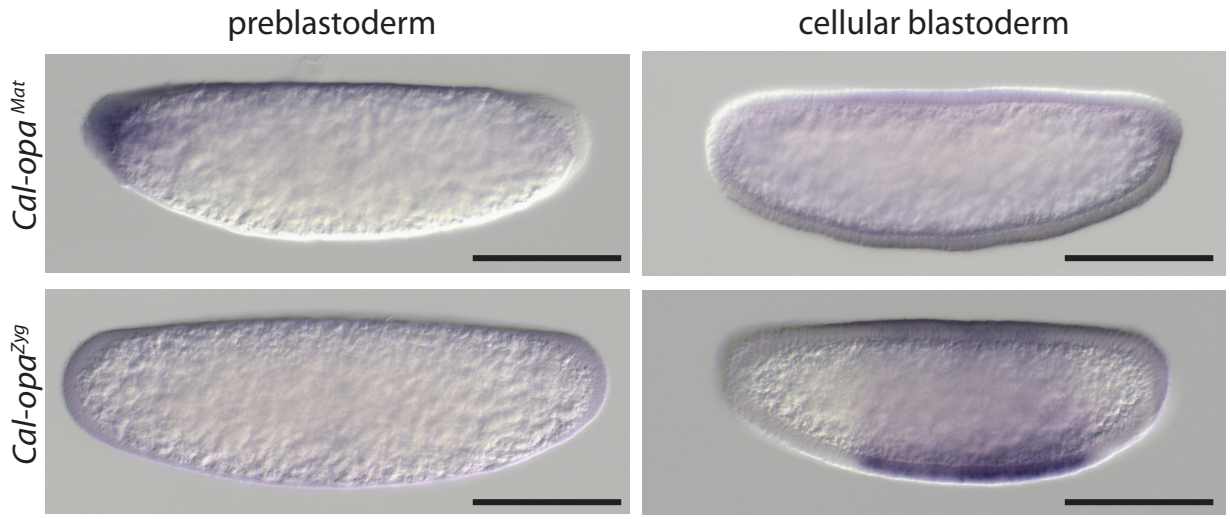


Figure 2.1.3. Expression of *Cal-opa* transcript isoforms in early Clogmia embryos. RNA in situ hybridization of *Cal-opa*^{Mat} and *Cal-opa*^{Zyg} in 1hr-old preblastoderm and 7hr-old cellular blastoderm embryos. Anterior is left and dorsal up. Scale bar: 100μm.

Cal-opa^{Mat} is required for establishing embryo polarity while *Cal-opa*^{Zyg} is required for segmentation

To determine the function of *Cal-opa*^{Mat} and *Cal-opa*^{Zyg}, I established a protocol for microinjecting early Clogmia embryos and conducted transcript-specific RNAi experiments. Injection of *Cal-opa*^{Mat} double-stranded RNA (dsRNA) led to mirror-image duplications of the tail end (double abdomen; **Figure 2.1.4. and Figure 2.1.6.**). In contrast, injection of dsRNA targeting *Cal-opa*^{Zyg} interfered with frequency-doubling of the expression domains of *Cal-slp* (the ortholog of pair-rule gene *sloppy-paired*) and caused defects in segmentation, dorsal closure, and head development but did not alter embryo polarity (**Figure 2.1.5.**). Finally, injection of dsRNA targeting both *Cal-opa*^{Mat} and *Cal-opa*^{Zyg} resulted in double abdomens with missing segments (**Figure 2.1.7.)**. These observations indicate distinct roles of *Cal-opa*^{Mat} and *Cal-opa*^{Zyg} in specifying embryo polarity and in segmentation, respectively.

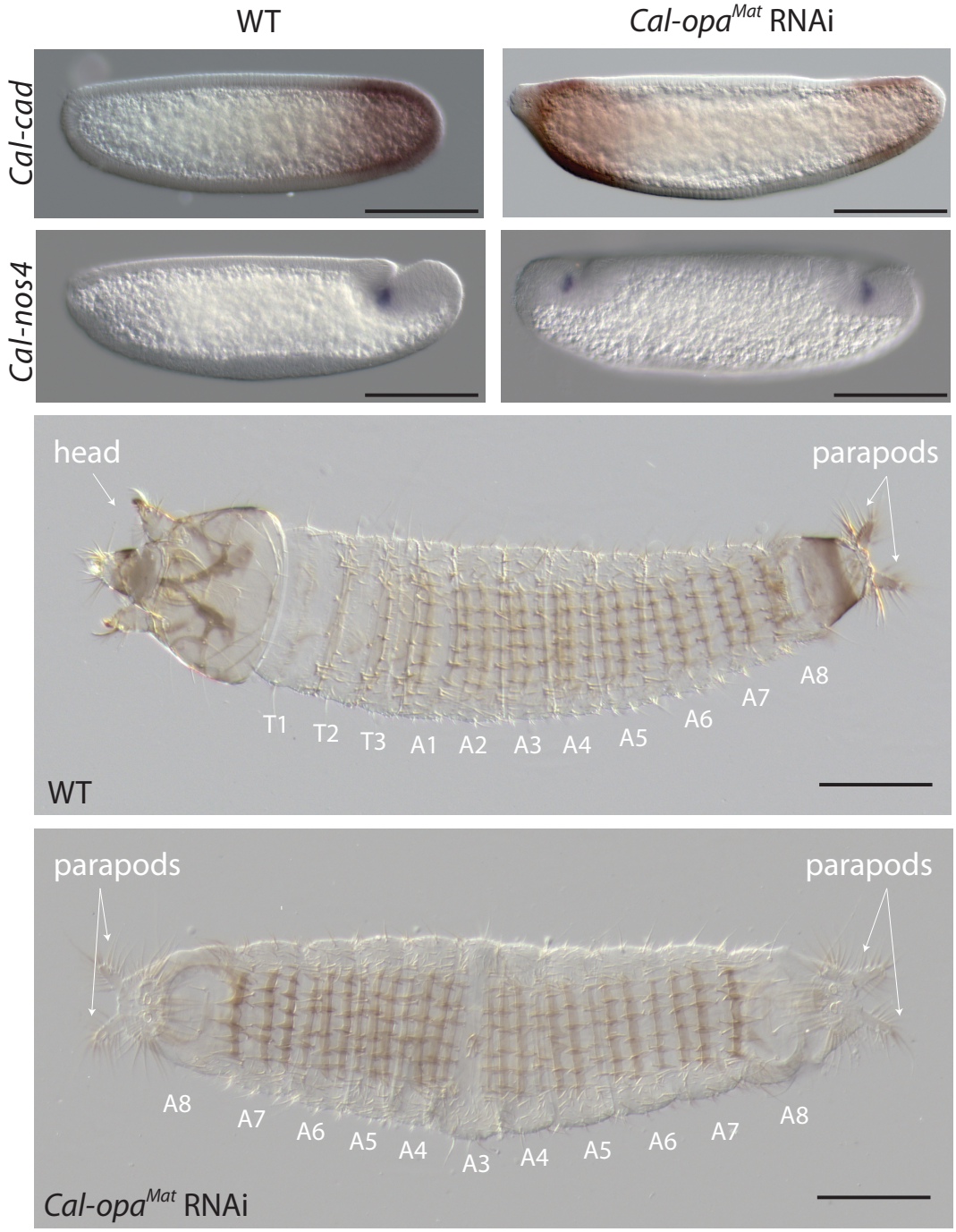


Figure 2.1.4. Function of *Cal-opa*^{Mat} transcript in early Clogmia embryo. RNA in situ hybridization of *Cal-cad* in a wild-type preblastoderm embryo and of *Cal-nos4* in a gastrulating embryo (left) and stage-matched *Cal-opa*^{Mat} RNAi embryos (right). 1st instar larval cuticle of wild type and following *Cal-opa*^{Mat} RNAi (bottom). Anterior is left and dorsal up. T: thoracic segment; A: abdominal segment. Scale bar: 100μm

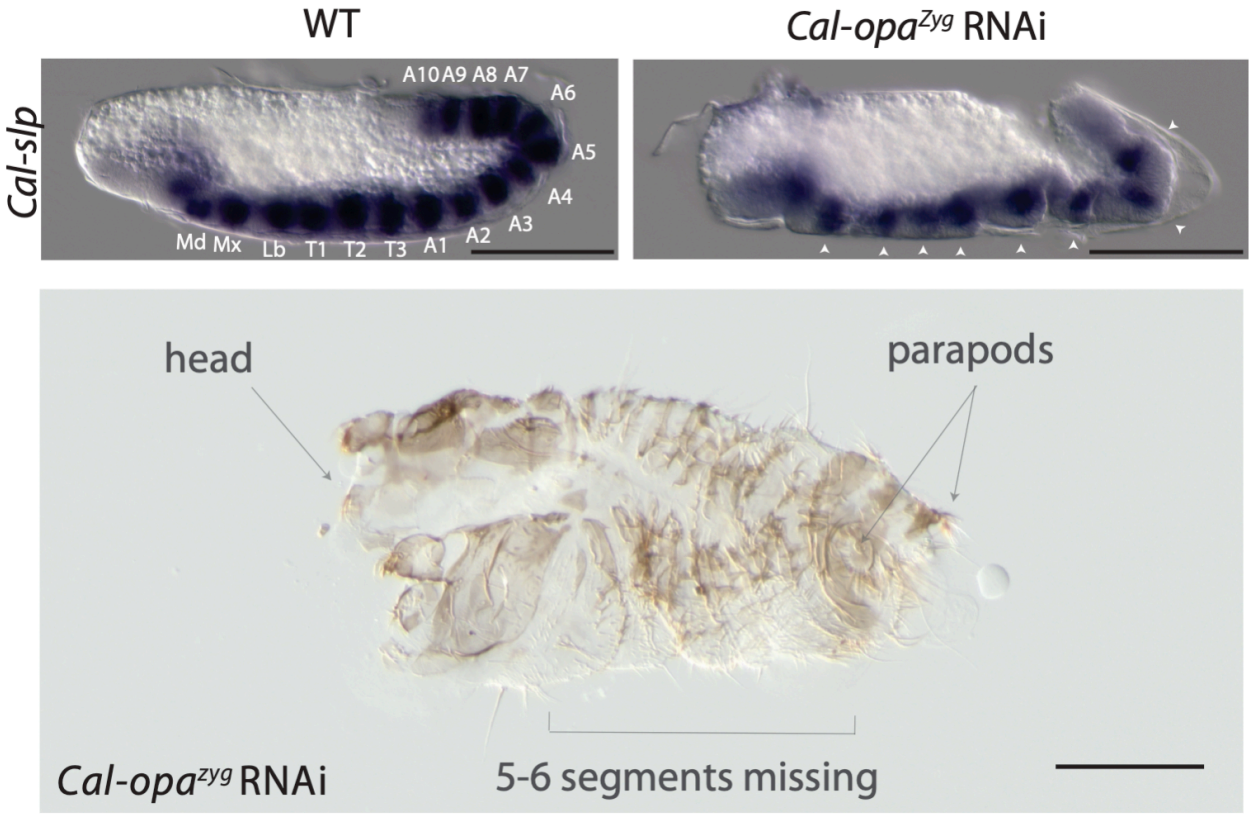


Figure 2.1.5. Function of *Cal-opa*^{Zyg} transcript in early Clogmia embryo. RNA in situ hybridization of *Cal-slp* in extending wild-type germband and stage-matched *Cal-opa*^{Zyg} RNAi embryo, and 1st instar larval cuticle phenotype following *Cal-opa*^{Zyg} RNAi. Anterior is left and dorsal up. Md: mandibular segment; Mx: maxillary segment; Lb: labial segment; T: thoracic segment; A: abdominal segment. Scale bar: 100μm.

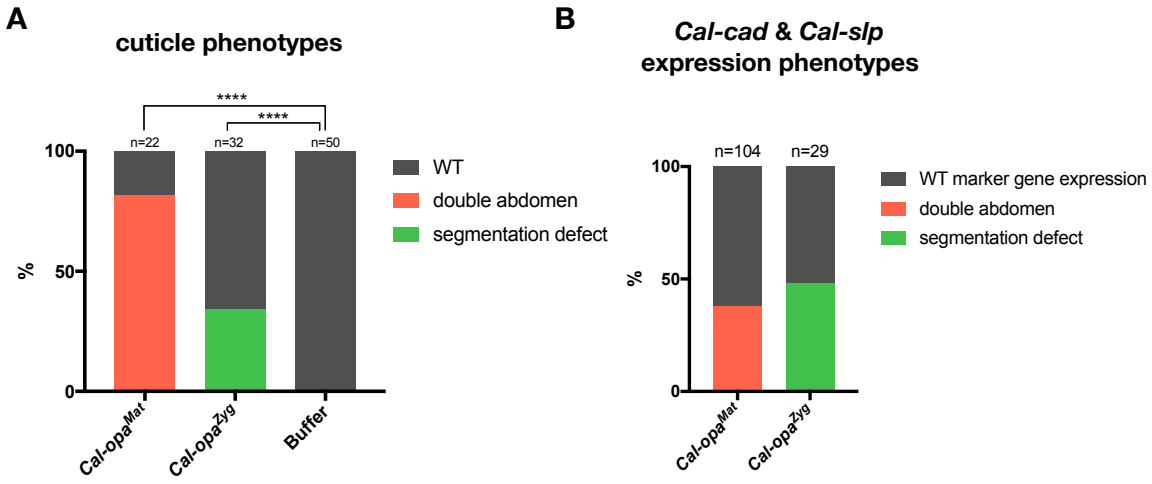


Figure 2.1.6. Frequency of strong *Cal-opa*^{Mat} and *Cal-opa*^{Zyg} RNAi phenotypes. **(A)** Cuticle phenotypes of 1st instar larvae. For representative examples see Figure 2.1.4. and Figure 2.1.5.. ****: P<0.0001, Fisher's exact test. **(B)** *Cal-cad* and *Cal-slp* expression phenotypes in embryos. For representative examples see Figure 2.1.4. and 2.1.5.. *Cal-cad* in situ hybridizations were done on *Cal-opa*^{Mat} RNAi embryos that reached cellular blastoderm stage. *Cal-slp* in situ hybridizations were done on 14- to 18hr-old *Cal-opa*^{Zyg} RNAi embryos. Embryos that had failed to reach the stage of germband extension were excluded from the analysis.

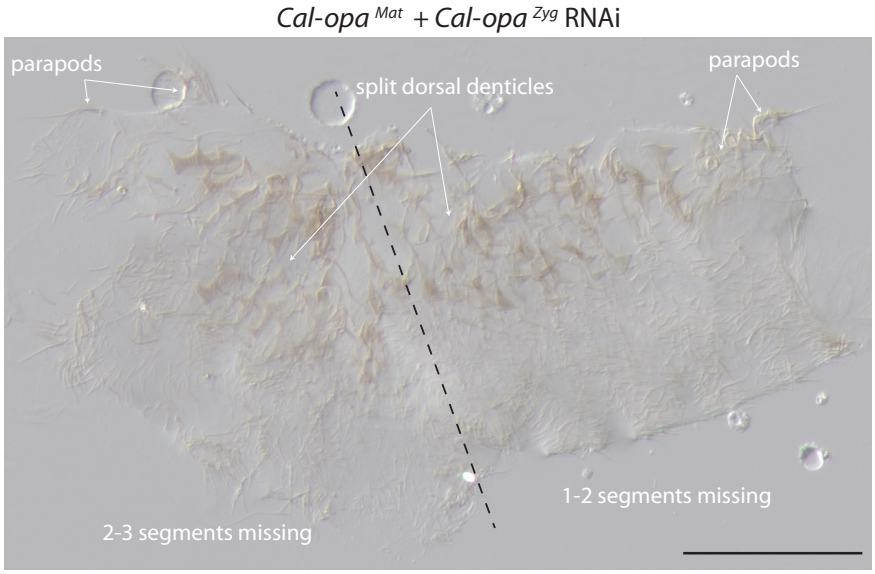


Figure 2.1.7. Cuticle phenotype of a 1st instar larva following *Cal-opa*^{Mat} and *Cal-opa*^{Zyg} double RNAi. A double abdomen with several missing segments is shown in lateral view. More severe phenotypic defects were not found in 1st instar larvae, possibly due to embryo death prior to cuticle secretion. Scale bar: 100μm.

Clogmia lacks maternal germ plasm and uses an inductive mechanism for germ cell specification

Expression profiling did not reveal any posteriorly localized maternal transcripts (**Figure 2.1.1A**, alpha = 0.001, unadj.), raising the question of whether Clogmia embryos lack maternal germ plasm. In Drosophila and other dipterans, the maternal germ plasm in the posterior embryo specifies primordial germ cells and stabilizes embryo polarity via *nanos*, which suppresses the translation of anterior determinants in the posterior embryo (Gavis and Lehmann 1992; Lemke and Schmidt-Ott 2009). The activity of *nanos* in the posterior preblastoderm is dependent on *oskar* (Lehmann 2016), which is conserved in many insects (Ewen-Campen, Schwager, and Extavour 2010) but absent from our Clogmia transcriptomes and the Clogmia genome (Vicoso and Bachtrog 2015). To test whether Clogmia lacks maternal germ plasm, I examined the expression of candidate germ cell markers, including Clogmia homologs of *nanos* (*Cal-nos1,*

Cal-nos2, Cal-nos3, and *Cal-nos4*, see **Figure 2.1.8A** for protein alignment), *vasa* (*Cal-vas*), *tudor* (*Cal-tud*), and *germ cell-less* (*Cal-gcl*). *Cal-nos1, Cal-nos3,* and *Cal-nos4* were not localized in the posterior of preblastoderm embryos but were expressed in a small set of cells at the posterior pole of cellular blastoderm and gastrulating embryos along with *Cal-vas*, *Cal-tud*, and *Cal-gcl* that were expressed more broadly (**Figure 2.1.8B-C**). These observations suggest that Clogmia lacks maternal germ plasm and that Clogmia may induce the germ cell fate zygotically. To test this hypothesis, I examined *Cal-nos* expression in *Cal-opa*^{Mat} RNAi embryos. *Cal-nos* positive cells were duplicated in double abdomens (**Figure 2.1.4.**), indicating that Clogmia uses an inductive mechanism for germ cell specification, which is repressed in the anterior embryo by *Cal-opa*^{Mat}.

Cal-opa^{Mat} localization alone is sufficient for establishing embryo polarity

I noticed that maternal transcripts of *Cal-slp* and a homolog of *miranda* (*Cal-mira*), which encodes an adaptor protein for cell fate determinants in Drosophila (Adams et al. 2000; Ikeshima-Kataoka et al. 1997), were also slightly enriched anterior (**Figure 2.1.1A**). This observation was confirmed by RNA in situ hybridizations (**Figure 2.1.9A**). Injection of *Cal-slp* dsRNA resulted in head and dorsal closure defects while *Cal-mira* dsRNA caused labrum and antennal defects, but in both cases embryo polarity was retained (**Figure 2.1.9B**).

To test whether *Cal-opa*^{Mat} requires the localized *Cal-slp* and *Cal-mira* transcripts to specify anterior development, I injected *Cal-opa*^{Mat} mRNA into the posterior pole of 1hr-old embryos. These embryos expressed a head marker, *Cal-otd* (ortholog of *ocelliless*/*orthodenticle*), on both ends of the embryo and developed a symmetrical double head, including some

duplicated thoracic elements (**Figure 2.1.10. and 2.1.11.**). These observations indicate that *Cal-opa*^{Mat} localization is sufficient for establishing embryo polarity.

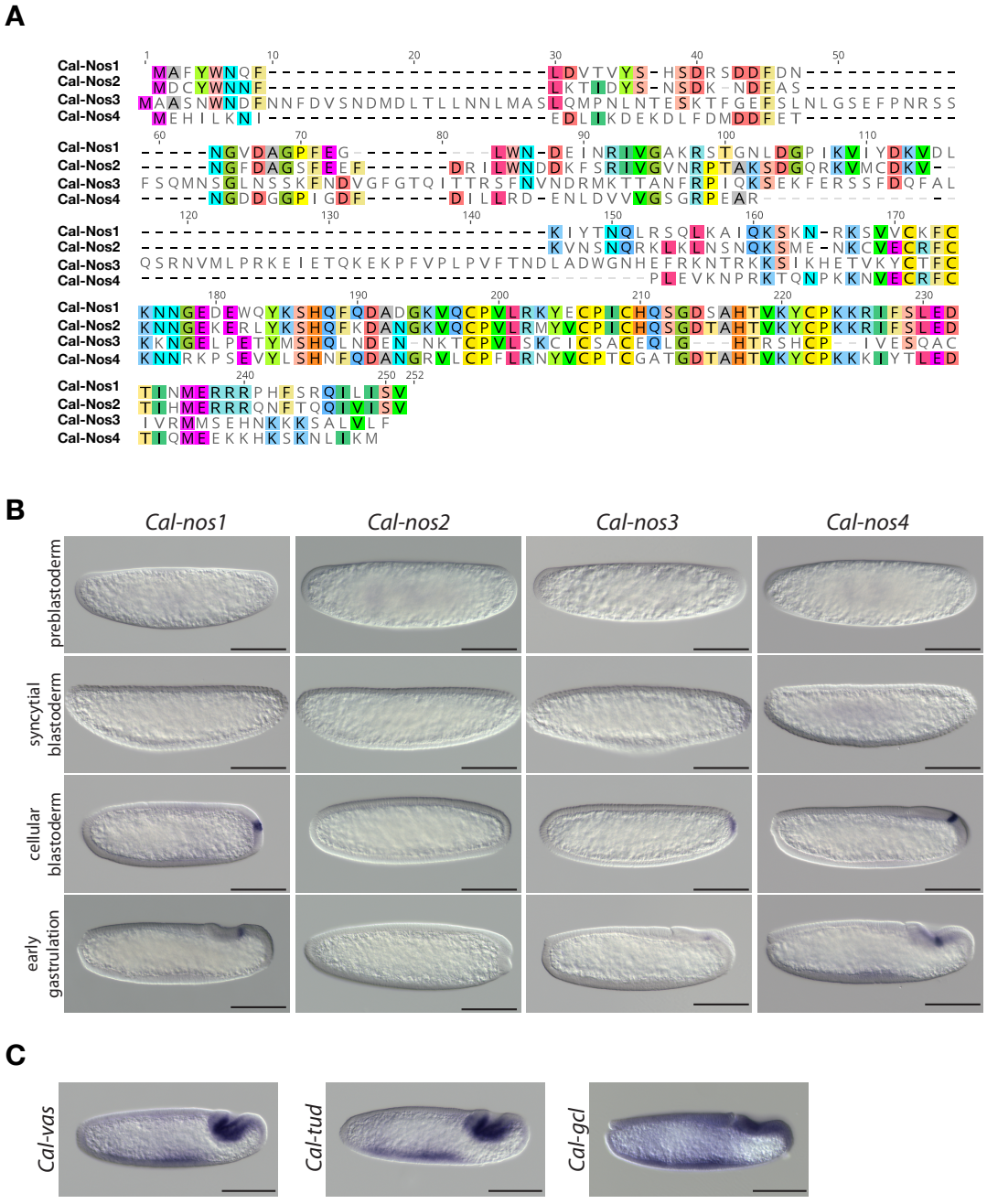


Figure 2.1.8. Clogmia homologs of *nanos*, *vasa*, *tudor*, and *germ cell-less*. **(A)** Protein alignment of Cal-Nos paralogs using *ClustalW 2.1* (Larkin et al. 2007). Conserved amino acids are highlighted. **(B)** *Cal-nos1, Cal-nos2, Cal-nos3,* and *Cal-nos4* RNA in situ hybridizations in embryos at preblastoderm, syncytial blastoderm, cellular blastoderm, and gastrulation stages. Anterior is left and dorsal up. Scale bar: 100μm. **(C)** *Cal-vas, Cal-tud,* and *Cal-gcl* RNA in situ hybridizations in embryos at gastrulation stage. Anterior is left and dorsal up. Scale bar: 100μm.

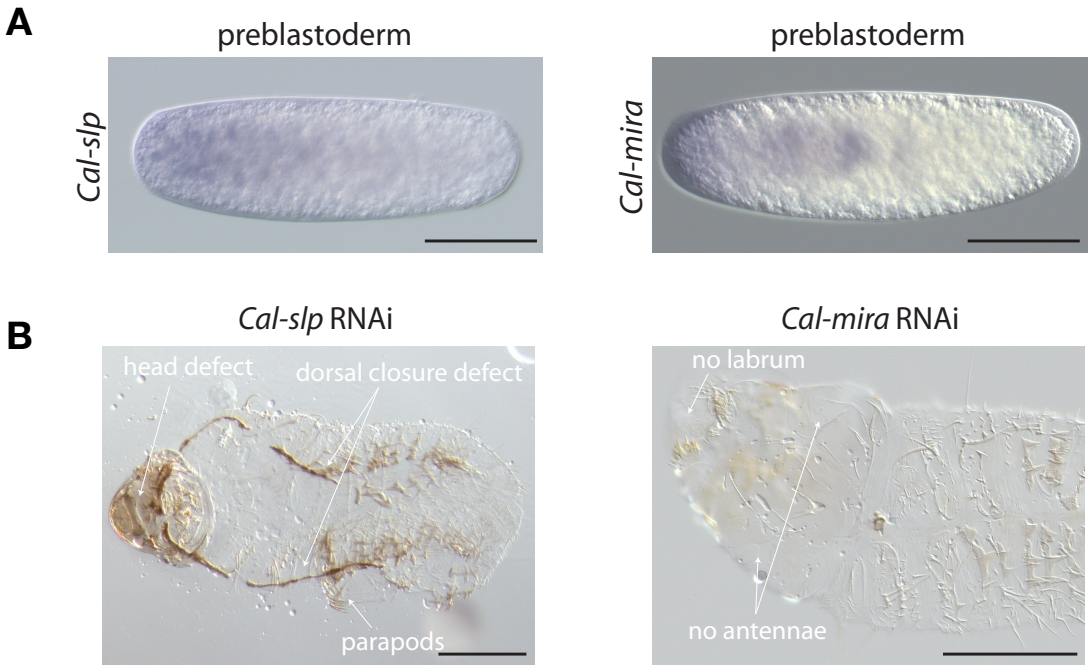


Figure 2.1.9. Maternal transcript localization and RNAi phenotypes of *Cal-slp* and *Cal-mira*. **(A)** *Cal-slp* and *Cal-mira* RNA in situ hybridization in embryos at preblastoderm stage. Anterior is left; Scale bar: 100μm. **(B)** Cuticle phenotypes of 1st instar larvae following *Cal-slp* RNAi (left) and *Cal-mira* RNAi (right) in ventral view. Anterior is left. Scale bar: 100μm.

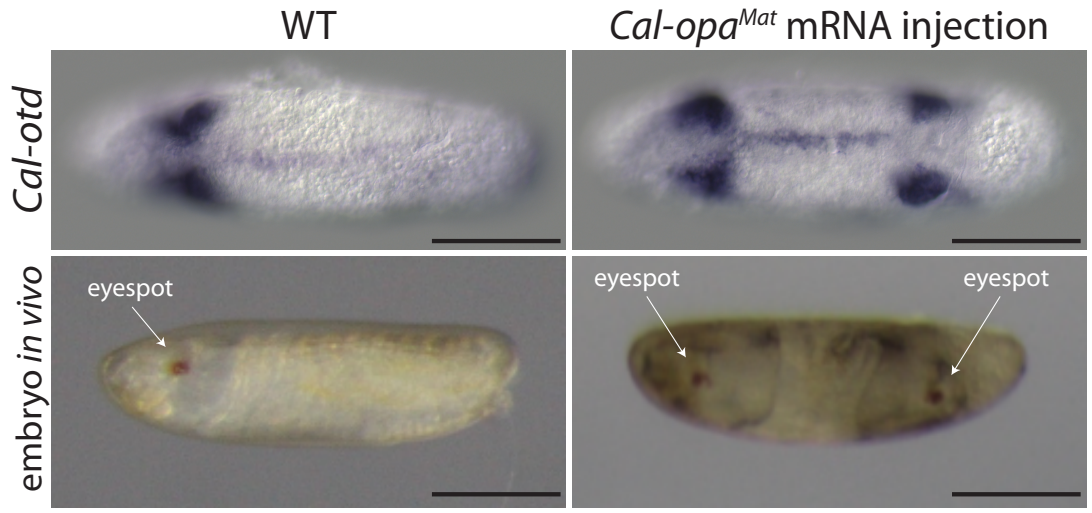


Figure 2.1.10. Ectopic head induction by *Cal-opa*^{Mat} in Clogmia. RNA in situ hybridizations of *Cal-otd* in wild-type gastrula (top left) and stage-matched embryo following posterior *Cal-opa*^{Mat} mRNA injection (top right) are shown in ventral view. A live wild-type embryo (bottom left) and a stage-matched embryo following posterior *Cal-opa*^{Mat} mRNA injection (bottom right) in lateral view. Anterior is left. Scale bar: 100μm.

Figure 2.1.11. Cuticle phenotype of a 1st instar larva following *Cal-opa*^{Mat} mRNA injection. Dorsal view. Scale bar: 100μm.

Changes in expression timing rather than protein sequences conferred the new function in *Cal-opa*

Since posterior injection of *Cal-opa*^{Mat} mRNA can induce ectopic anterior structures, I used this assay to identify essential requirements for this function. I first asked whether the timing of *Cal-opa*^{Mat} expression is critical for its function. To test this hypothesis, I conducted posterior injections of *Cal-opa*^{Mat} mRNA during the syncytial blastoderm stage (4hr) and examined *Cal-otd* expression after gastrulation. These embryos developed with normal head-to-

tail polarity (**Figure 2.1.12A)**. This result places the axis specification role of *Cal-opa* mRNA prior to the syncytial blastoderm stage and suggest that timing of *Cal-opa* activity is essential for its ability to induce head development.

I then asked whether *Cal-opa*^{Zyg} mRNA can induce head development when expressed early despite the N-terminal difference between Cal-Opa^{Mat} and Cal-Opa^{Zyg}. Posterior injection of *Cal-opa*^{Zyg} mRNA into preblastoderm embryos resulted in double heads (**Figure 2.1.12A)**. Because the translation start site of Cal-Opa^{Mat} is located downstream of the *Cal-opa*^{Zyg} translation start site, I also tested *Cal-opa*^{Zyg} mRNA in which the putative start codon for Cal-Opa^{Mat} was mutated to encode leucine (*Cal-opa*$^{Zyg\text{-}Met21Leu}$). Posterior injection of *Cal-opa*$^{Zyg\text{-}Met21Leu}$ mRNA into early Clogmia embryos also induced double heads (**Figure 2.1.12A)**. To better define the portion of Cal-Opa^{Mat} that is essential for its function as anterior determinant, I also tested more severely truncated protein variants (**Figure 2.1.12A and Figure 2.1.2.**). Cal-Opa$^{115\text{-}655}$ (Cal-Opa$^{1\text{-}655}$ = Cal-Opa^{Zyg}) and Cal-Opa$^{182\text{-}655}$ mRNAs retained the ability to induce double heads but mRNAs of Cal-Opa$^{182\text{-}445}$ and Cal-Opa$^{1\text{-}337+22}$ (a hypothetical splice variant) did not perturb embryo polarity. These results indicate that the ability of Cal-Opa to specify embryo polarity is largely insensitive to N-terminal truncation but requires the C-terminal portion of the protein.

Cal-Opa^{Mat} could have evolved into an anterior determinant independent of any protein change through acquisition of new target genes. Alternatively, changes in the protein could have resulted in a new function. To distinguish between these possibilities, I injected *odd-paired* mRNA from *Lutzomyia longipalpis* (Psychodidae), *Chironomus riparius*, and *Drosophila melanogaster* into early Clogima embryos. All of these *odd-paired* homologs induced double heads in Clogmia, provided that the endogenous kozak sequence of *Cal-opa*^{Mat} was used for

optimal translation efficiency (**Figure 2.1.12B)**. Since neither Drosophila nor Chironomus uses *odd-paired* for specifying embryo polarity and maternal *odd-paired* transcript is absent in freshly deposited eggs of chironomids (Klomp et al. 2015), mosquitoes (Akbari et al. 2013), and beetles (Choe, Stellabotte, and Brown 2017), these results suggest that Cal-Opa^{Mat} acquired its function as embryonic axis determinant independent of protein changes in the Clogmia lineage.

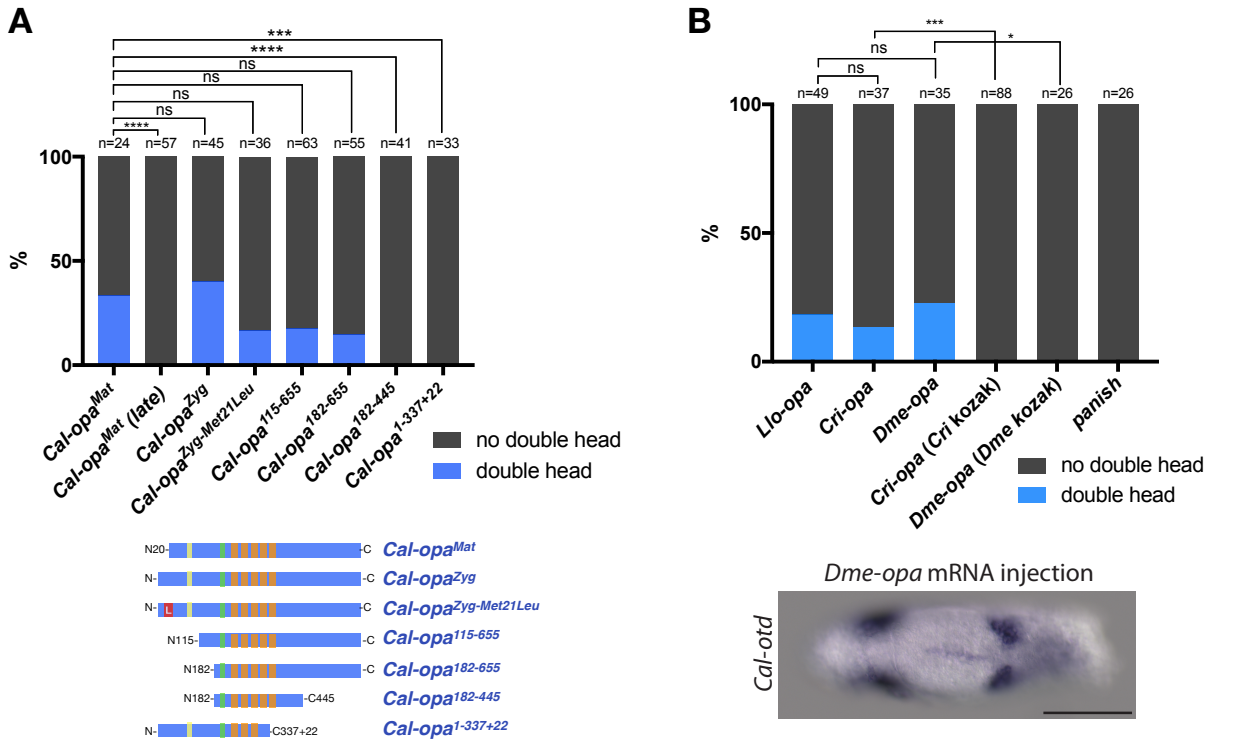


Figure 2.1.12. Ectopic head induction by mRNAs variants derived from *Cal-opa* and *Cal-opa* orthologs in Clogmia **(A)** Posterior injection of *Cal-opa* mRNA and mutated variants. Complete, symmetrical duplication of the bilateral *Cal-otd* expression domain in gastrulating embryos was counted as double head (blue, see Figure 3A). All other phenotypes, including incomplete duplications and wild type, were counted as "no double head" (black). Diagrams of predicted Cal-Opa proteins are shown with ZIC/Opa conserved motif (ZOC) in yellow, the ZIC family protein N-terminal conserved domain (ZFNC) in green, and zinc finger domains in orange. The Met21Leu mutation in Cal-Opa$^{Zyg\text{-}Met21Leu}$ is marked in red. *Cal-opa^{Mat}* (late): *Cal-opa^{Mat} was* injected during the syncytial blastoderm stage (4hr). ns: P>0.05; ***: P<0.001; ****: P<0.0001, Fisher's exact test. **(B)** Posterior injection of *odd-paired* mRNAs from Lutzomyia, Chironomus, and Drosophila. Phenotypes were counted as in (A) and double head from *Dme-opa* mRNA injection is shown as example. *Cri-opa* and *Dme-opa* mRNAs include the Kozak sequence of *Cal-opa^{Mat}* (TAAG). *Cri*-kozak and *Dme*-kozak refer to *odd-paired* sequences with donor-specific Kozak-sequences from Chironomus (AAAA) and Drosophila (GACC), respectively. ns.: P>0.05; *: P<0.05; **: P<0.01; ***: P<0.001, Fisher's exact test.

Alternative transcription and expression of *odd-paired* is conserved in the sand fly Lutzomyia

Next, I asked whether the role of *Cal-opa*^{Mat} is conserved in the sand fly Lutzomyia, another member of the Psychodidae family that is of public health concern due to its role in the transmission of visceral leishmaniasis. In Lutzomyia, an *odd-paired* homolog was also the most enriched transcript in the anterior of 1-2hr old embryo and was named *Llo-opa*^{Mat} **(Figure 2.1.13A)**. In the posterior embryo, the most enriched transcript was homologous to *oskar*, indicating that Lutzomyia eggs contain maternal germ plasm at the posterior pole, unlike Clogmia eggs. These findings suggest that sand fly embryos also determine embryo polarity via anterior localization of maternal *odd-paired* transcript, and that maternal germ plasm was lost only in the Clogmia lineage. Close examination of Lutzomyia transcriptomes from 1hr-old and 24hr-old embryos also revealed zygotic *odd-paired* transcript (*Llo-opa*^{Zyg}) (**Figure 2.1.13B**). *Llo-opa*^{Mat} and *Llo-opa*^{Zyg} share the same open reading frame but differ at their untranslated 5' and 3' ends. Since the N-terminal ends of Llo-Opa^{Mat}/Llo-Opa^{Zyg} and Cal-Opa^{Zyg} proteins are homologous (**Figure 2.1.2.**), I infer that the N-terminal truncation of Cal-Opa^{Mat} occurred after the transcript had evolved maternal expression and anterior localization. The detection of *Llo-opa*^{Zyg} transcript in 24hr-old embryos coincided with that pair-rule segmentation gene homologs (**Figure 2.1.14.**), indicating that *Llo-opa*^{Zyg} functions during segmentation.

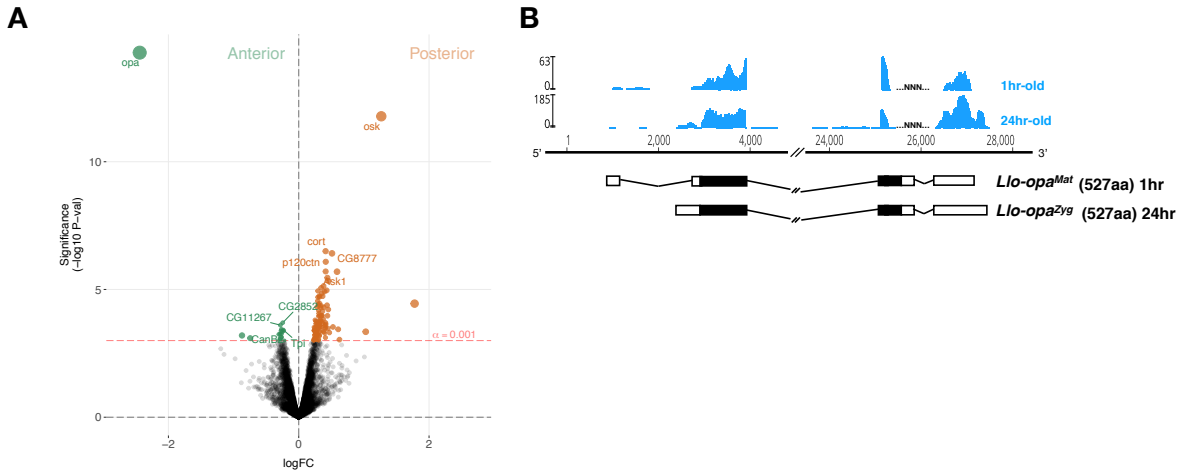


Figure 2.1.13. Expression of alternative *Llo-opa* transcripts. **(A)** Differential expression analysis of maternal transcripts between anterior and posterior halves of 1-2hr-old Lutzomyia embryos. **(B)** Stage-specific RNA-seq read coverage of *Llo-opa* genomic locus and diagrams of *Llo-opa*^{Mat} and *Llo-opa*^{Zyg} transcripts. logFC: log fold-change (see also legend to Figure 2.1.1.).

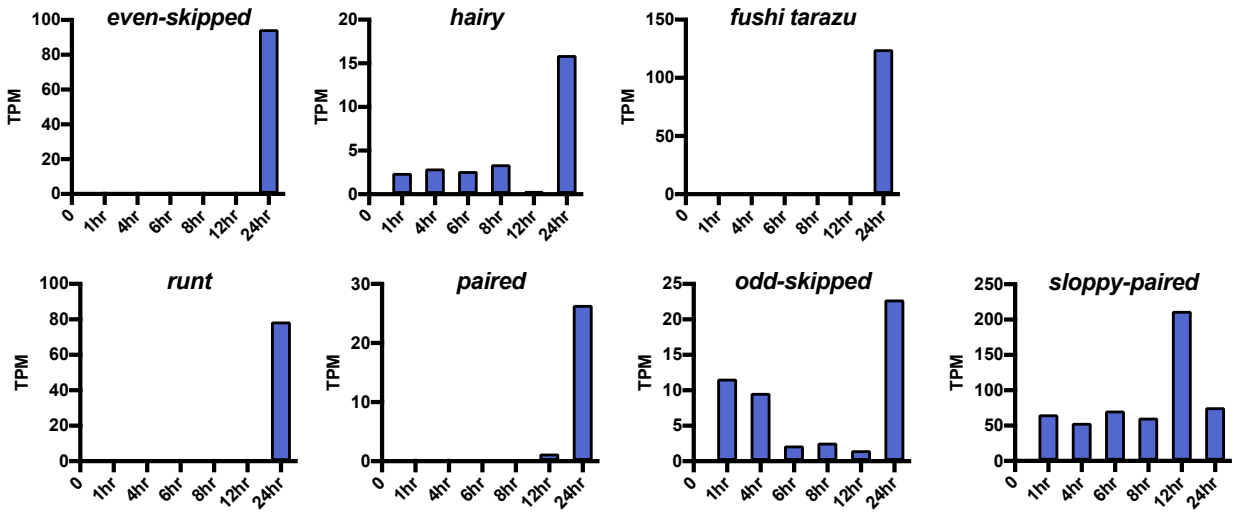


Figure 2.1.14. Expression level of select Lutzomyia pair-rule gene homologs**.** Based on RNA-seq data from 1, 4, 6, 8, 12, and 24 hr-old embryos. TPM: transcripts per million.

2.2. An alternative maternal transcript of a conserved C2H2 zinc finger gene *cucoid* functions as anterior determinant in Culicine mosquitoes[2]

Transcript homologous to *cucoid* is enriched in the anterior pole of freshly laid Culex eggs

The southern house mosquito *Culex quinquefasciatus* and the yellow fever mosquito *Aedes aegypti* represent the mosquito subfamily Culicinae and are species of enormous public health concern. Culex is a vector of West Nile virus (the leading cause of mosquito-borne disease in the continental United States) and the nematode *Wuchereria bancrofti* (the major cause of lymphatic filariasis), while Aedes transmits yellow fever, dengue, chikungunya and Zika viruses. Despite their medical importance, the embryonic development of these species remain poorly known, and are difficult to study due their eggshell properties.

Freshly deposited mosquito eggs lack maternal transcript orthologous to *odd-paired* (Akbari et al. 2013). Their genome also lacks orthologs of *bicoid* and *panish* (Klomp et al. 2015). These findings prompted me to extend my search for anterior determinants to mosquitoes. Initially, I focused on Culex, because their eggs are large and have clearly distinguishable anterior and posterior egg poles. Culex also lays eggs in a form of a raft, with all eggs aligned together with the same anterior-to-posterior orientation. This allowed me to easily bisect multiple eggs simultaneously for RNA-seq analysis.

8,239 Culex transcripts were annotated from the anterior and posterior transcriptomes of 1hr-old preblastoderm embryos and were ranked them according to the magnitude of their differential expression scores and P values (**Figure 2.2.1A**). In the posterior embryo, the most

[2] Differential expression analyses of transcriptomes were performed by Jeff Klomp, a previous member of our laboratory. Ines Martin-Martin and Frank Criscione from the Ribeiro lab at NIAID performed *cucoid* RNAi experiments.

enriched transcript was related to *nanos*, consistent with the presence of maternal germ plasm in this species (**Figure 2.2.3B**) (Juhn et al. 2008). The most enriched transcript in the anterior embryo was closely related to an uncharacterized gene of Drosophila (*CG9215*), which encodes a protein with five C2H2 zinc finger domains. However, reciprocal BLAST searches suggest that *CG9215* belongs to a poorly defined larger gene family in Drosophila that might be represented by a single gene in mosquitoes. I therefore named this Culex gene as *cucoid* (culicine mosquito gene (cu-) with a similar expression to *bicoid* (-coid)) (**Figure 2.2.2.**).

RACE experiments with cDNA from 0-7hr-old embryos revealed three alternative *cucoid* transcripts with distinct 3' ends (*cucoid*^{A}, *cucoid*^{B}, and *cucoid*^{C}) (**Figure 2.2.1B**), but only *cucoid*^{B} and *cucoid*^{C} were recovered from cDNA of 0-2hr-old preblastoderm embryos, suggesting that one or both these transcripts might be maternally localized at the anterior pole. These transcript isoforms encode protein isoforms with different C-terminal ends: *cucoid*^{A} (678aa), *cucoid*^{B} (553aa), and *cucoid*^{C} (490aa), respectively. Due to AS in conjunction with alternative transcription termination, the predicted protein of *cucoid*^{B} lacks the last C2H2 zinc finger while the predicted protein of *cucoid*^{C} lacks the last two C2H2 zinc fingers (**Figure 2.2.2**).

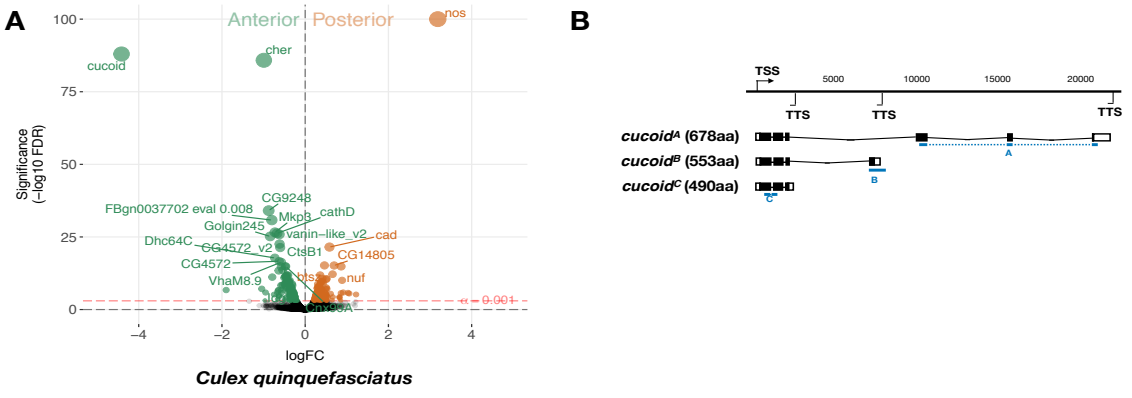


Figure 2.2.1. Maternal expression of *cucoid* and alternative *cucoid* isoforms **(A)** Differential expression analysis of maternal transcripts between anterior and posterior halves of 1hr-old Culex embryos. **(B)** Diagrams of *cucoid* transcript isoforms based on RNA-seq and RACE experiments (see also legend to Figure 2.1.1.). Blue lines represent sequence complementary to RNA probes used in Fig 2.2.3.

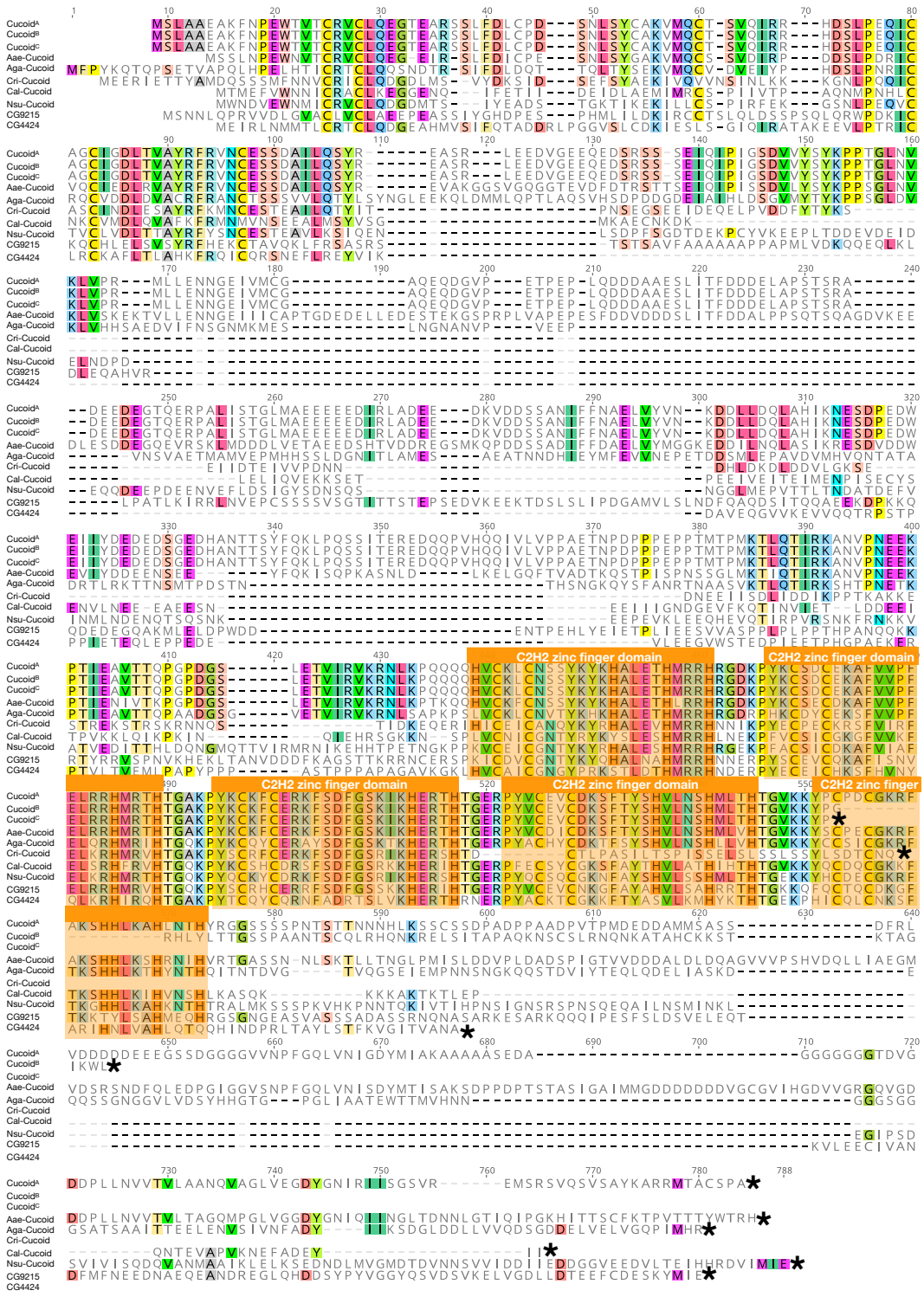


Figure 2.2.2. Protein alignment of dipteran Cucoid orthologs. Proteins were aligned using *ClustalW 2.1* (Larkin et al. 2007). Conserved zinc finger domains are indicated with colored boxes. Black asterisk(*) indicates the last amino acids of proteins. Nsu: *Nephrotoma suturalis* (Tipulidae); Cal: *Clogmia albipunctata* (Psychodidae); Aae: *Aedes agypti* (Culicide); Cqu: *Culex quinquefasciatus* (Culicidae); Aga: *Anopheles gambiae* (Culicidae); Cri: *Chironomus ripariu* (Chironomidae).

cucoid^{C} is localized at the anterior pole of freshly laid Culex eggs and *cucoid* controls embryo polarity

To test for isoform specific expression patterns, I performed RNA in situ hybridizations with specific probes for *cucoid*^{A} (probe A) or *cucoid*^{B} (probe B) and a probe against all three isoforms (probe C). *cucoid*^{A} and *cucoid*^{B} expression was detected in the fore and hind gut of extended germbands but not in 1hr-old preblastoderm embryos. In contrast, the probe against all three isoforms detected maternally localized *cucoid* transcript at the anterior pole in addition to the zygotic expression pattern (**Figure 2.2.3A**). Taken together, these results suggest that only the *cucoid*^{C} isoform is maternally localized at the anterior pole and could function as anterior determinant. To test this hypothesis, I collaborated with Ines Martin-Martin and Frank Criscione in the Ribeiro lab at the National Institute of Health in Rockville (MD) to inject *cucoid* dsRNA from the shared 5' region and examined the expression of a posterior marker (*Cqu-cad*) in gastrulating embryos. Many of these embryos expressed *Cqu-cad* in the anterior and underwent ectopic gastrulation at the anterior pole, suggesting that normal head-to-tail polarity was lost (**Figure 2.2.3C**).

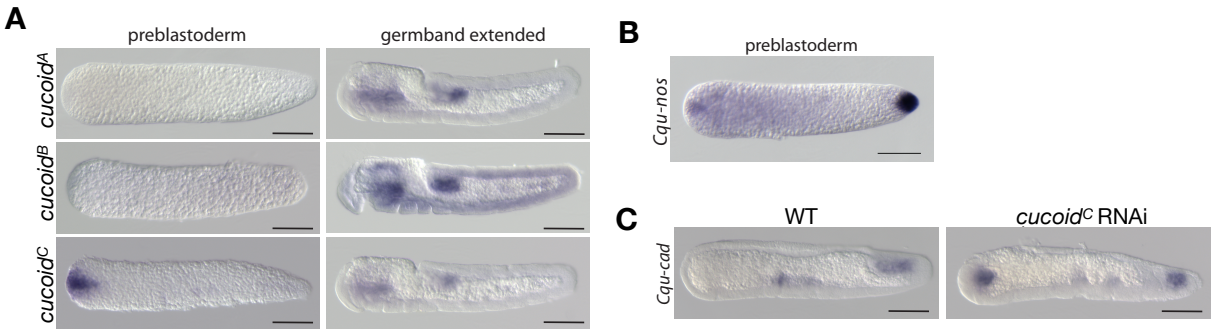


Figure 2.2.3. Expression and function of alternative *cucoid* isoforms in Culex. **(A)** RNA in situ hybridization of *cucoid*^{A}, *cucoid*^{B}, and *cucoid*^{C} transcripts in 1hr-old preblastoderm and germband extending embryos. Anterior is left and dorsal up. Scale bar: 100μm. **(B)** RNA in situ hybridization of *Cqu*-nos in 1hr-old preblastoderm embryo. Anterior is left. Scale bar: 100μm **(C)** RNA in situ hybridization of *Cqu-cad* in a wild-type gastrulating embryo (left) and in a stage-matched *cucoid* RNAi embryos (right; 16/48 with *cucoid* dsRNA versus 0/25 with lacZ control dsRNA; P<0.0005). Anterior is left and dorsal up. Scale bar: 100μm.

Expression and function of *cucoid* ortholog is conserved in the yellow fever mosquito *Aedes aegypti*

I then asked whether the function of *cucoid* as the anterior determinant was conserved in the yellow fever mosquito *Aedes aegypti*, which also transmits dengue, chikungunya and Zika viruses. 5,802 Aedes transcripts from the anterior and posterior transcriptomes of 1hr-old preblastoderm embryos were annotated and ranked according to the magnitude of their differential expression scores and P values. I found that transcript of Aedes *cucoid* (*Aae-cucoid*) to be highly enriched in the anterior embryo and was able to confirm this result by RNA in situ hybridization (**Figure 2.2.4A-B**). In the posterior embryo, no highly enriched transcripts were observed. This was unexpected given that whole mount in situ hybridizations revealed posterior localized transcript of Aedes *nanos* in 1hr-old embryos (**Figure 2.2.4B**) and that maternal transcript of Aedes *oskar* is also localized at the posterior pole (Juhn and James 2006). Low statistical power of our differential expression analysis in Aedes might explain this discrepancy. This is because it is difficult to reliably distinguish the anterior and posterior pole in Aedes eggs. Thus, I could have confounded the anterior and posterior pole in some of the bisected eggs (see Methods).

RACE experiment with cDNA from 1-6hr old embryos revealed three *Aae-cucoid* isoforms with distinct 3'ends (*Aae-cucoid* ^{A}, *Aae-cucoid* ^{B}, *Aae-cucoid* ^{C}). Unlike *Aae-cucoid* ^{A} and *Aae-cucoid* ^{B}, which encoded protein with five C2H2 zinc finger domains, *Aae-cucoid* ^{C} encoded a truncated protein with only two zinc finger domains (**Figure 2.2.2**). Injection of dsRNA targeting all three *Aae-cucoid* isoforms resulted in double abdomens (**Figure 2.2.4D**), suggesting that the mosquito homologue of *cucoid* evolved its function as anterior determinant prior to the divergence of the Culex and Aedes lineages.

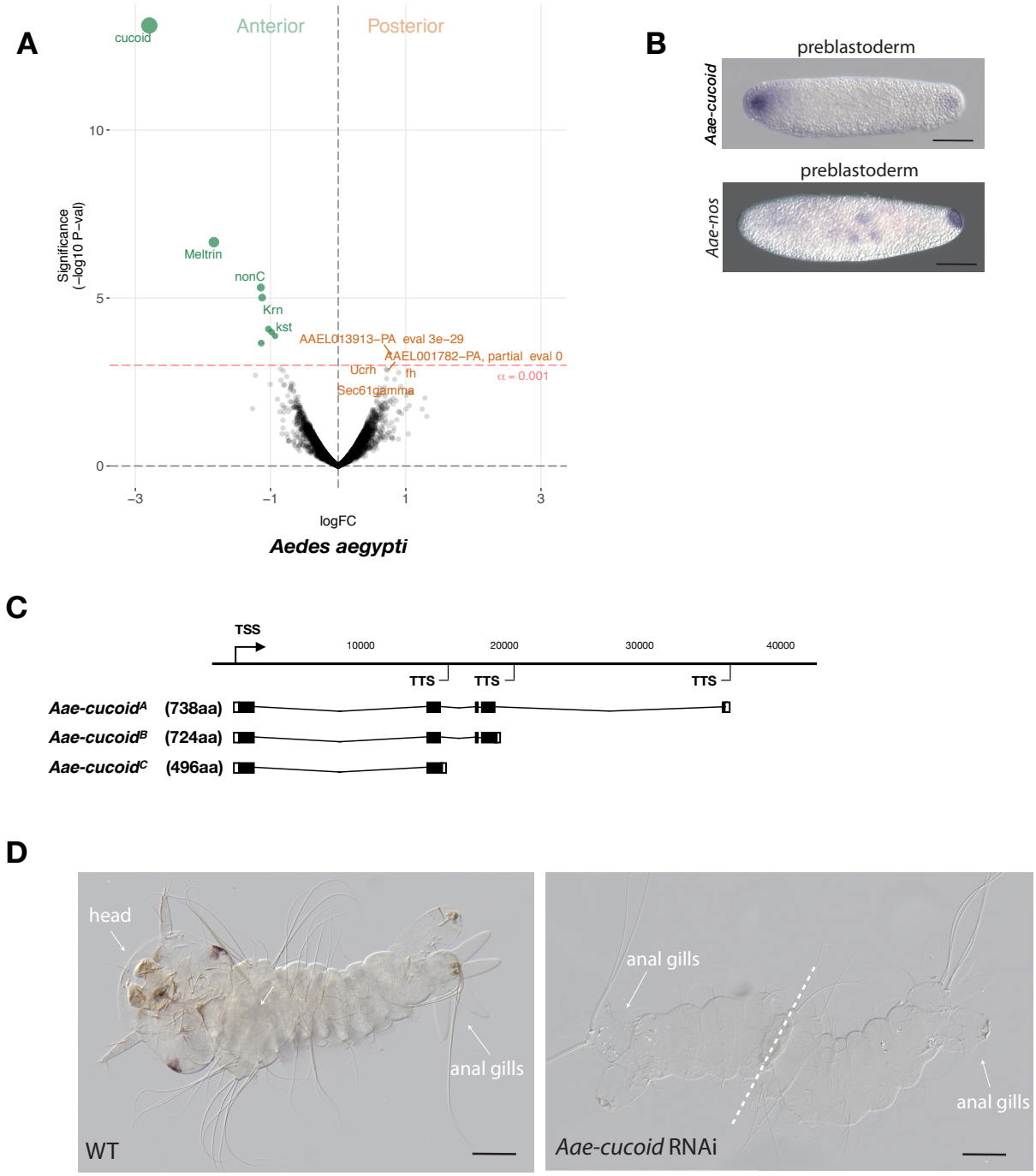


Figure 2.2.4. Expression and function of *cucoid* in Aedes. **(A)** Differential expression analysis of maternal transcripts between anterior and posterior halves of 1hr-old Aedes embryos. **(B)** RNA in situ hybridization of *Aae-cucoid* (probe targeting all three isoforms) (top) and *Aae-nos* (bottom) transcript in 1hr-old preblastoderm embryo. Anterior is left and dorsal up. Scale bar: 100μm. **(C)** Diagrams of *Aae-cucoid* transcript isoforms based on RNA-seq and RACE experiments (see also legend to Figure 2.1B). **(D)** 1st instar Aedes larval cuticle of wild type (top) and following *Aae-cucoid* RNAi (bottom; 9/26 versus 0/22 with control dsRNA; P<0.005). Scale bar: 100μm.

2.3. Anterior localization and function of a maternal *pangolin* isoform in anopheline mosquitoes and crane flies suggests that *panish* inherited its function from *pangolin*[3]

An alternative maternal transcript isoform of *pangolin* is localized at the anterior pole of a freshly laid Anopheles eggs

The *Anopheles gambiae* species complex constitutes an outgroup to the Culex-Aedes clade. It includes 8 or more sub-Saharan species that are difficult to distinguish due to widespread genealogical heterogeneity across the genome, incomplete lineage sorting and introgression (Thawornwattana, Dalquen, and Yang 2018). We interchangeably used *A. gambiae* and *A. coluzzii*, two sibling species within this species complex that are responsible for the majority of malaria transmission in Africa, to identify the anterior determinant of this important mosquito lineage. Whole mount RNA in situ hybridizations with a probe against the *Anopheles gambiae* homolog of *cucoid* did not detect any anterior localized transcript in 1hr-old embryos, suggesting that Anopheles uses a different anterior determinant than Culex and Aedes.

To identify this gene, 9,353 transcripts were annotated from anterior and posterior transcriptomes of 1hr-old preblastoderm embryos of *A. gambiae* and were ranked according to the magnitude of their differential expression scores and P values. In the posterior embryo, the most enriched transcript was homologous to *nanos.* In the anterior embryo, the most enriched transcript was homologous to *pangolin* (also known as *Tcf*) (**Figure 2.3.1A**). To test for potential alternative maternal and zygotic isoforms of *pangolin* in Anopheles, I mapped the assembled

[3] Differential expression analyses of transcriptomes were performed by Jeff Klomp, a previous member of our laboratory.

transcripts and 5' and 3' RACE products from 1-6hr-old embryos onto an available *A. gambiae* genome assembly (AgamP4**)**. I identified two alternative transcript variants with non-overlapping 3'UTRs but nearly identical open reading frames that I named *Aga-pan*^{Mat} and *Aga-pan*^{Zyg}, respectively (**Figure 2.3.1B and Figure 2.3.2.**). *Aga-pan*^{Mat} was tightly localized at the anterior pole of 1-2hr-old preblastoderm embryos and only weakly expressed in elongated germbands, whereas *Aga-pan*^{Zyg} was expressed segmentally in elongated germbands but not in embryos younger than 2hr-old preblastoderm embryos (**Figure 2.3.1C)**. Both *pangolin* isoforms were conserved in *Anopheles coluzzii* with sequence identity above 99% and the maternal isoform (*Aco-pan*^{Mat}) was localized at the anterior pole (**Figure 2.3.3.**). The stage-specific expression of both *pangolin* isoforms was also conserved outside the *Anopheles gambiae* species complex in *Anopheles stephensi* (**Figure 2.3.4.**). Alignments of dipteran Pangolin proteins suggest that, in Anopheles, the maternal variant includes an additional 7 amino acids at the C-terminal end due to alternative polyadenylation and splicing (**Figure 2.3.2.)**.

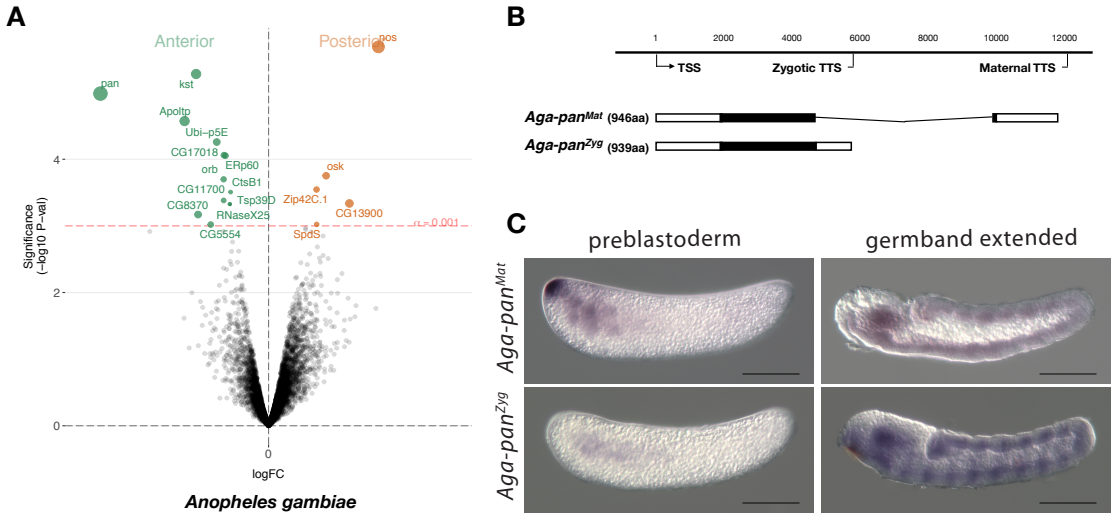


Figure 2.3.1. Expression of alternative *pangolin* isoforms in Anopheles. **(A)** Differential expression analysis of maternal transcripts between anterior and posterior halves of 1hr-old Anopheles embryos. **(B)** Diagrams of *Aga-pan*^{Mat} and *Aga-pan*^{Zyg} transcripts based on RNA-seq and RACE experiments (see also legend to Figure 2.1.1.). **(C)** RNA in situ hybridization of *Aga-pan*^{Mat} and *Aga-pan*^{Zyg} transcripts in 1hr-old preblastoderm and germband extending embryos of *A.gambiae*. Anterior is left and dorsal up. Scale bar: 100μm.

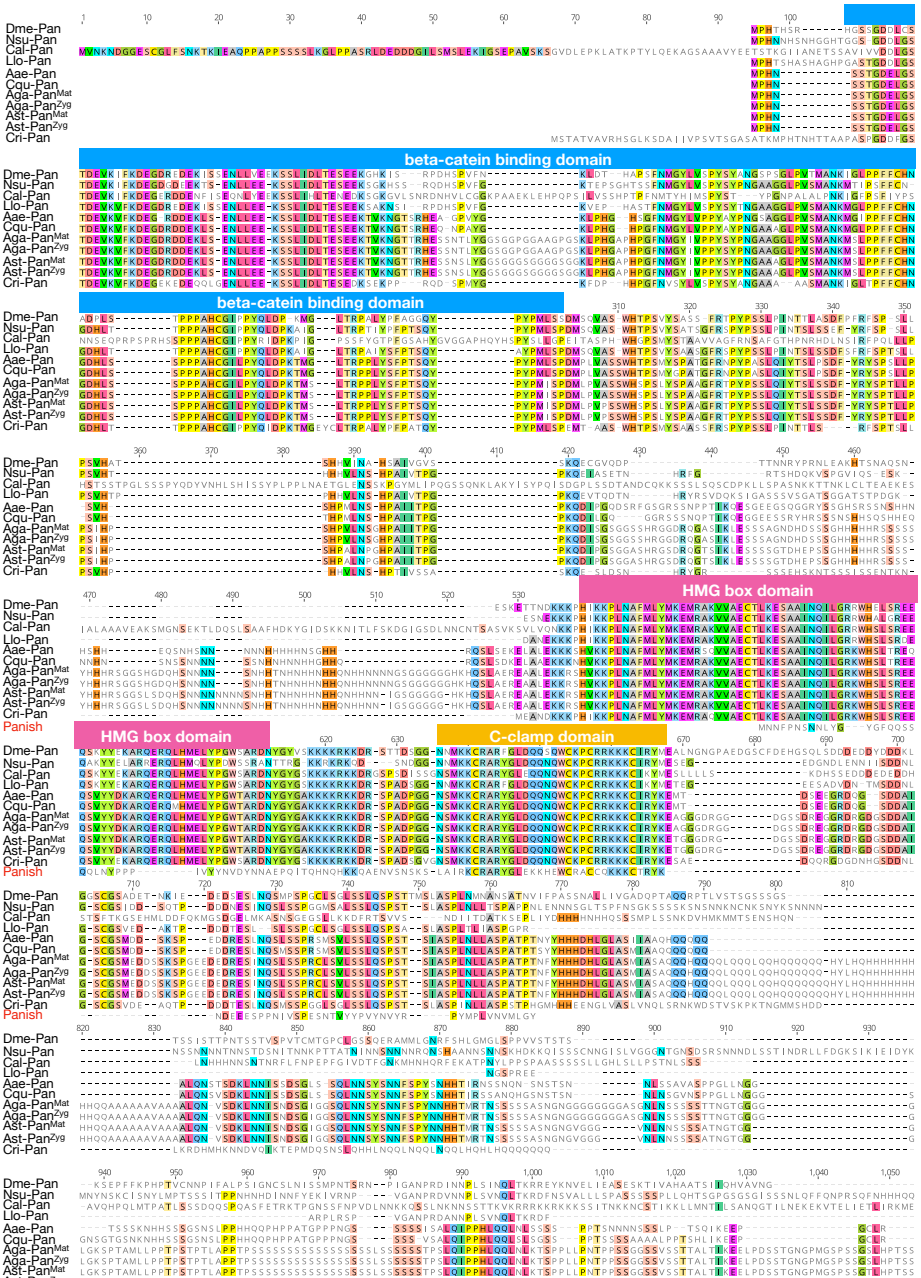


Figure 2.3.2. Protein alignment of dipteran Pangolin orthologs and Panish. Proteins were aligned using *ClustalW 2.1* (Larkin et al. 2007). Conserved domains are indicated with colored boxes. Dme: Drosophila melanogaster (Drosophilidae); Nsu: *Nephrotoma suturalis* (Tipulidae); Cal: *Clogmia albipunctata* (Psychodidae); Llo: *Lutzomyia longipalpis* (Psychodidae); Aae: *Aedes agypti* (Culicide); Cqu: *Culex quinquefasciatus* (Culicidae); Aga: *Anopheles gambiae* (Culicidae); Ast: *Anopheles stephensi* (Culicidae); Cri: *Chironomus riparius* (Chironomidae). Conserved amino acids are shown in color.

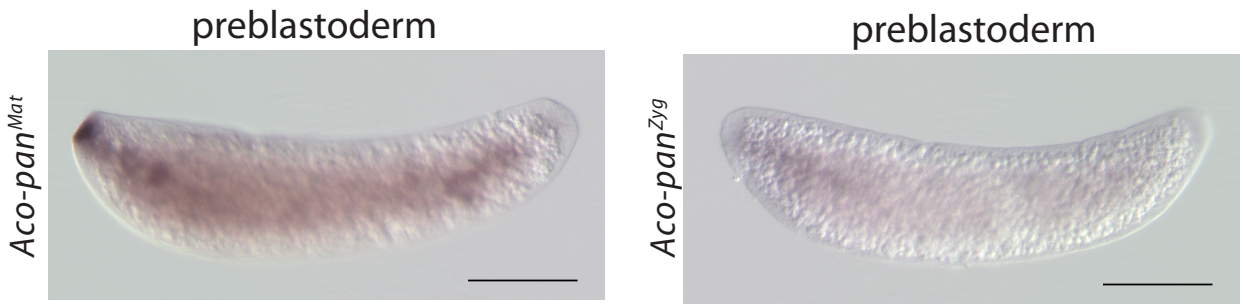


Figure 2.3.3. RNA in situ hybridization of *Aco-pan*^{Mat} and *Aco-pan*^{Zyg} transcripts in 1hr-old *A.coluzzii* preblastoderm embryos. Anterior is left and dorsal up. Scale bar: 100μm.

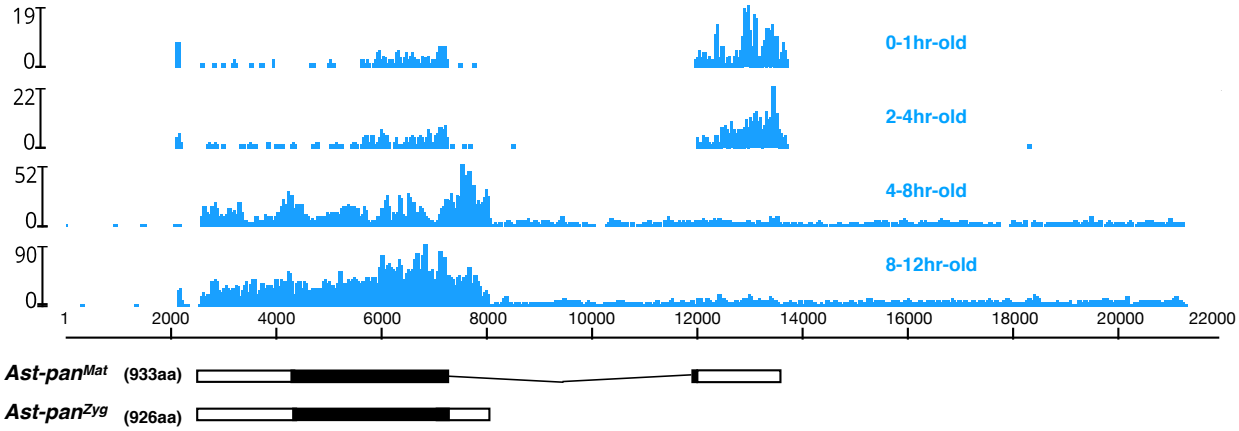


Figure 2.3.4. Stage-specific RNA-seq read coverage of *Ast-pan* genomic locus and diagrams of *Ast-pan*^{Mat} and *Ast-pan*^{Zyg} transcripts (see also legend to Figure 1A). Black box, Open reading frame; white box, untranslated regions.

Anterior localization of maternal *pangolin* transcripts is conserved in Nephrotoma eggs

Anterior-localized maternal *pangolin* (*Tc-pan*) transcript has also been observed in the eggs of a beetle (*Tribolium castaneum*) (Bucher, Farzana et al. 2005). It is therefore possible that this localization pattern evolved before the radiation of Diptera. To test whether ancestral dipterans localized maternal *pangolin* transcript at the anterior pole of the egg, I collected bisected freshly deposited eggs from the crane fly *Nephrotoma suturalis* (Tipulidae), which belongs to one of the oldest branches of dipterans (Wiegmann et al. 2011; Grimaldi and Engel

2005) and sequenced the anterior and posterior transcriptomes. 5,371 transcripts were annotated and ranked according to the magnitude of their differential expression scores and P values (**Figure 2.3.5A**). The most enriched transcript in the posterior embryo was related to *oskar*, suggesting that crane fly eggs contain maternal germ plasm at the posterior pole. The most enriched transcript in the anterior embryo was homologous to *pangolin* and therefore named *Nsu-pan*^{Mat}. The anterior localization of this transcript was confirmed by RNA in situ hybridization (**Figure 2.3.5B**). RACE experiments with cDNA from 1hr-old embryos identified multiple isoforms with slightly variable 5' ends but the same open reading frame. An alignment of the predicted Nsu-Pan protein from this open reading frame with other dipteran Pangolin homologs revealed conserved N-terminal and C-terminal ends in Nsu-Pan. These results suggest that anterior localization of maternal *pangolin* transcript was present in ancestral dipteran insects and that conserved Pangolin protein preceded the modified variant (Aga-Pan^{Mat} and Panish) as an anteriorly localized isoform.

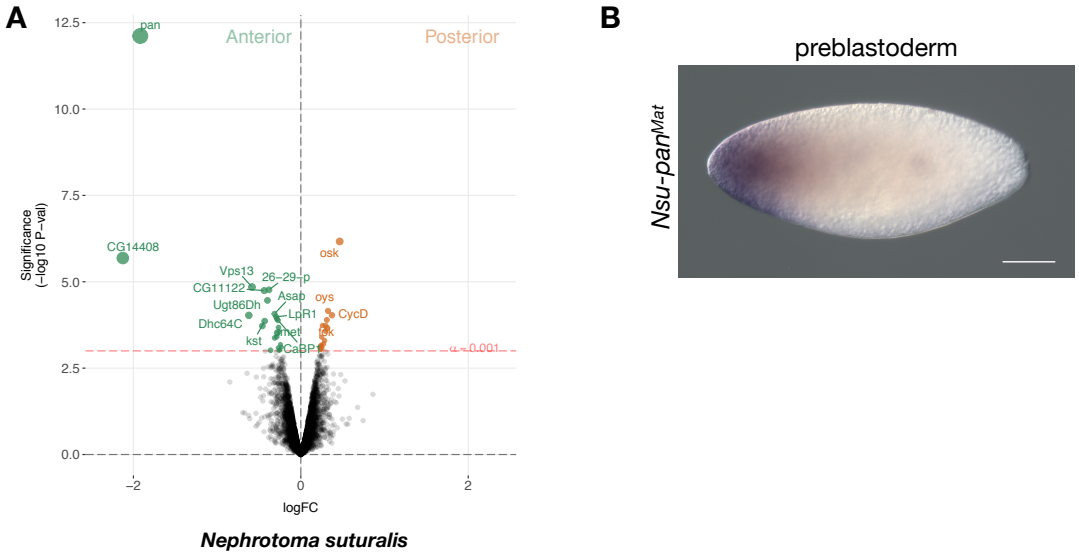


Figure 2.3.5. Expression of maternal *pangolin* transcript in Nephrotoma. **(A)** Differential expression analysis of maternal transcripts between anterior and posterior halves of 1hr-old Nephrotoma (Tipulidae) embryos. logFC: log fold-change. **(B)** RNA in situ hybridization of *Nsu-pan*^{Mat}.

Maternal *pangolin* transcript isoform is required for establishing head-to-tail polarity in Anopheles

In the beetle Tribolium, the function of maternal *pangolin* remains unknown. In previous *Tc-pan* RNAi experiments targeting both maternal and zygotic transcripts, only posterior defects were observed, which can be explained by the known role of zygotic *Tc-pan* in canonical Wnt signaling in the posterior growth zone (Bolognesi et al. 2008; Fu et al. 2012; Prühs, Beermann, and Schröder 2017; Ansari et al. 2018). However, it is possible that the reported phenotype corresponds to a hypomorphic phenotype. To investigate the role of localized maternal *pangolin* transcript in flies, I decided to study *Anopheles coluzzii,* where we can specifically target the maternal isoform. Injection of *Aco-pan*^{Mat}-specific dsRNA into several hundred 1hr-old *A. coluzzii* embryos resulted in only 37 cuticles with variable phenotypes, including anterior truncations and, in extreme cases, double abdomens (**Figure 2.3.6A-C**). I noticed perturbed segmentation boundaries in the double abdomens, suggesting that *Aga-pan*^{Mat} may also function in segmentation, as suggested by its weak zygotic expression pattern. Injection of *Aco-pan*^{Zyg}-specific dsRNA into 1hr-old *A. coluzzii* embryos resulted in severe segmentation defects that were difficult to characterize, but no double abdomens or anterior-specific truncation defects were found (**Figure 2.3.6D**).

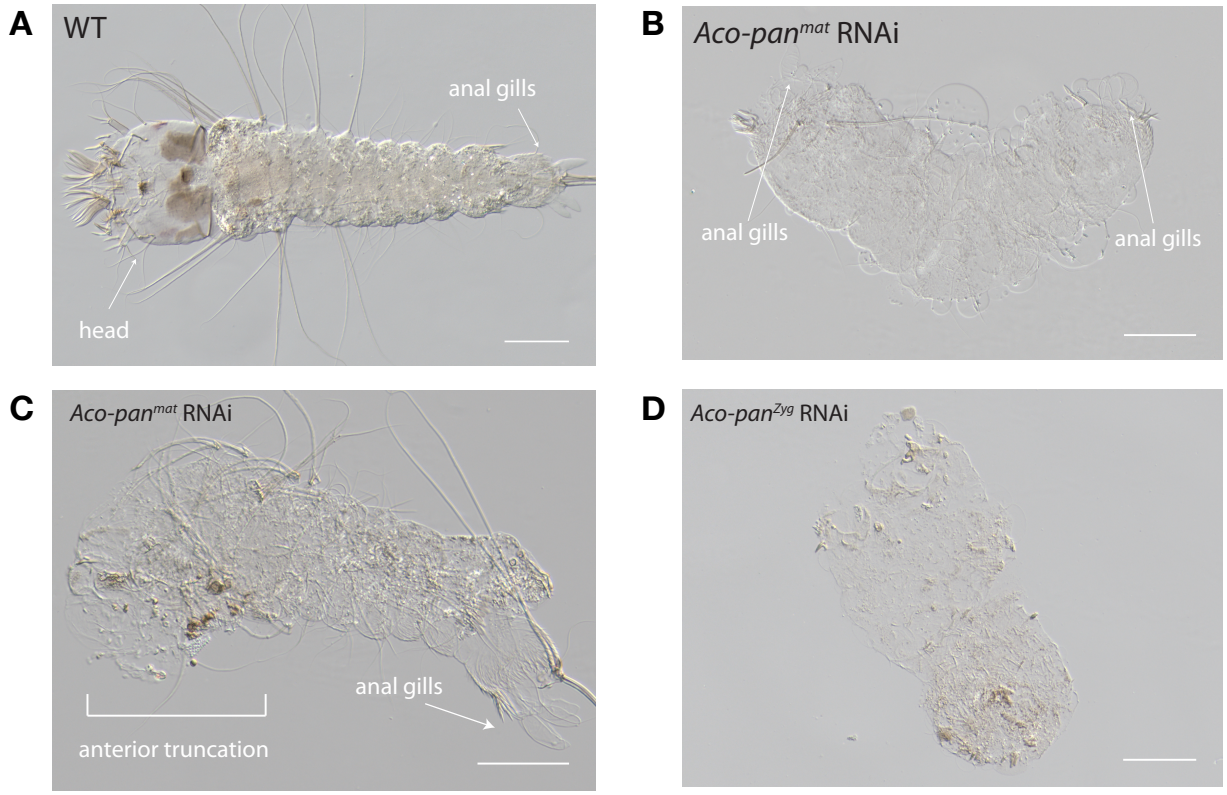


Figure 2.3.6. Function of a maternal *pangolin* isoform in Anopheles. **(A)** 1st instar *A.coluzzii* larval cuticle of wild type. **(B)** Cuticle phenotype of double abdomen (3/37) following *Aco-pan*^{Mat} RNAi. **(C)** Intermediate, anterior truncation phenotype (8/37) following *Aco-pan*^{Mat} RNAi. Scale bar: 100μm. **(D)** Cuticle phenotype of a *A.coluzzii* 1st instar larva following *Aco-pan*^{Zyg} RNAi (11/23). Phenotype difference between *Aco-pan*^{Mat} RNAi and *Aco-pan*^{Zyg} is significant; $p<0.05$, Fisher's exact test.

pangolin cannot substitute the role of *panish* in Chironomus embryos

The findings in Anopheles and Nephrotoma suggest that localized maternal *pangolin* transcript functioned as an anterior determinant in ancestral dipterans. In Chironomus, the ortholog of *pangolin* (*Cri-pan*) is not expressed maternally and its diverged paralog *panish* functions as an anterior determinant (Klomp et al. 2015). Given that *panish* evolved from *pangolin* via gene duplication, *panish* probably inherited its role from *pangolin*. Therefore, it is possible that *Cri-pan* and *panish* are still functionally equivalent when expressed at the anterior

pole of preblastoderm Chironomus embryos. Alternatively, *panish* may have co-evolved with its targets. In this case, *Cri-pan* should no longer be able to fulfill the function of *panish*. To distinguish between these possibilities, I examined the ability of *panish* and *Cri-pan* mRNAs to rescue the RNAi phenotype of *panish*. It has been shown previously that dsRNA of the *panish* 3'UTR can induce the double abdomen phenotype with a penetrance of nearly 100%, and that this phenotype can be rescued in roughly half of the embryos by injecting *panish* mRNA with heterologous UTRs at the anterior pole, shortly after the injection of dsRNA (Klomp et al. 2015). I used this assay to compare the functions of *panish* mRNA (positive control), frame shifted *panish* mRNA (negative control), *Cri-pan* mRNA, and a modified *Cri-pan* mRNA designed to better resemble *panish* mRNA (*Cri-pan*$^{trunc.}$ mRNA)**.** *Cri-pan*$^{trunc.}$ mRNA encodes a N-terminal truncated Cri-Pan variant lacking the ß-Catenin binding and HMG box domains with two mutations in the cysteine-clamp domain to mimic conserved changes of the Panish cysteine-clamps (**Figure 2.3.7A)**. Only *panish* mRNA rescued *panish* RNAi embryos (**Figure 2.3.7B)**, suggesting that *panish* co-evolved with its targets and functionally diverged after its origin via gene duplication from *pangolin.*

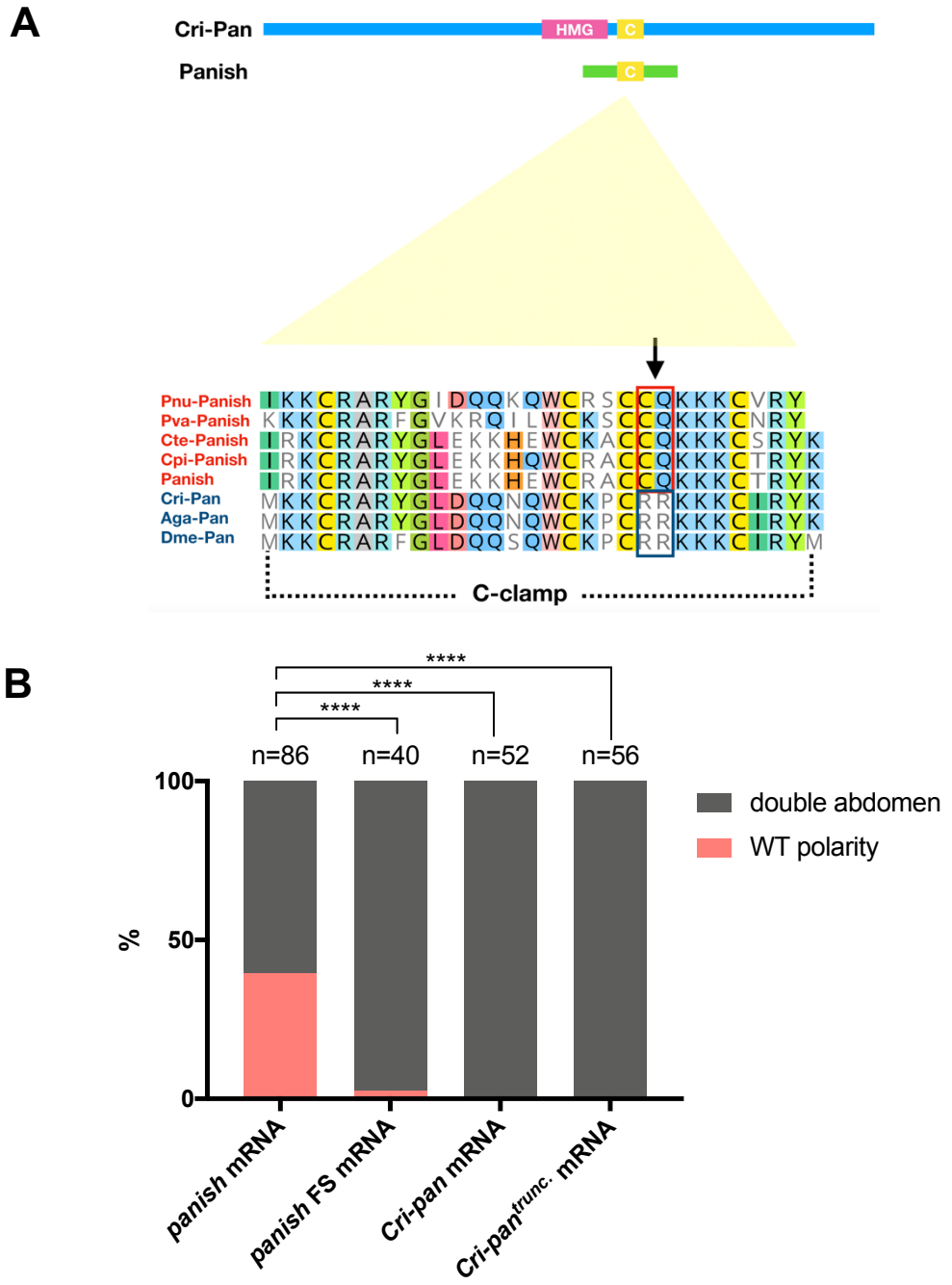


Figure 2.3.7. *panish* RNAi rescue experiments with *Cri-pan* mRNA. **(A)** Diagram of Cri-Pan and Panish (HMG: HMG-box domain; C: Cysteine-clamp domain) and protein alignment of conserved C-clamp domains between Panish and Pan orthologs. Panish-specific sequence change (RR->CQ) is indicated as an arrow. Pnu: *Polypedilum nubifer*; Pva: *Polypedilum vanderplanki*; Cte: *Chironomus tentans*; Cpi: *Chironomus piger*; Cri: *Chironomus riparius;* Aga: *Anopheles gambiae;* Dme: *Drosophila melanogaster.* **(B)** *panish* RNAi rescue experiments using *panish* mRNA, panish FS (frameshift mutation) mRNA, *Cri-pan* mRNA, and *Cri-pan*^{trunc} mRNA. *Cri-pan*^{trunc} mRNA encodes Cri-Pan protein that lacks N-terminal sequences before C-clamp domain (N-terminal 365 aa deleted, including HMG-box domain) and with Panish-specific sequence change (RR->CQ) in C-clamp domain, as in (A). ****: P<0.0001

Alternative *pangolin* isoforms and their expressions in *Clunio marinus*

Using publicly available genomes, I surveyed the occurrence of *panish* in Chironomidae midges. Orthologs of *panish* in *Chironomus piger* and *C. tentans* have been reported previously (Klomp, Athy et al. 2015). Because Panish is short (131aa) and contains introns in the diagnostic C-clamp domain sequence, it was not possible to identify *panish* orthologs using BLAST. Therefore, I utilized the fact that the *panish* locus overlaps with *Cri-zap3,* and directly examined the *Cri-zap3* locus to search for sequences homologous to C-clamp residues. Using this approach, I found *panish* orthologs in two other species within the subfamily Chironominae, including *Polypedilum vanderplanki* and *P. nubifer* (**Figure 2.3.7A and Figure 2.3.8**). However, *panish* orthologs were not found outside Chironominae, while orthologs of *pangolin* and *zap3* were identified in all species (positive control) (**Figure 2.3.8.**). These results suggest that *panish* evolved in the Chironominae subfamily.

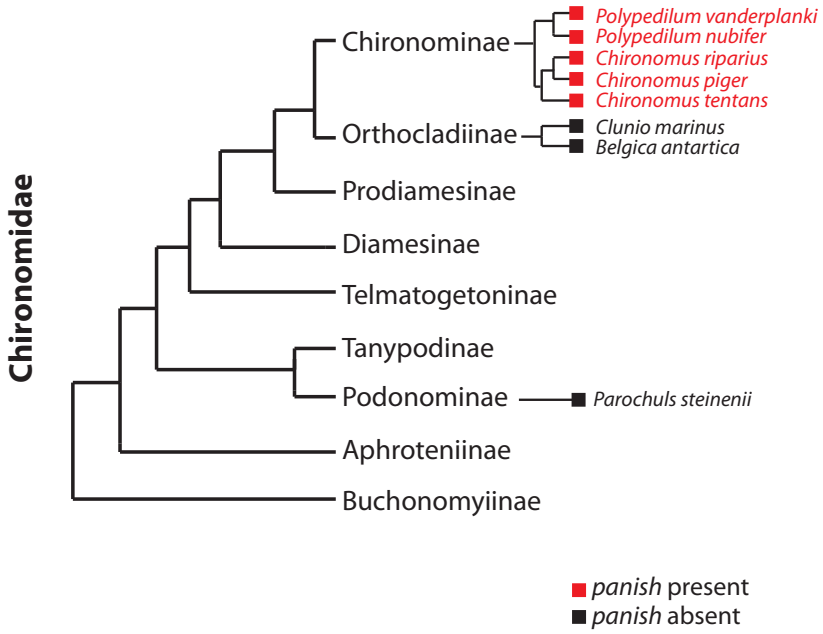


Figure 2.3.8. Occurrence of *panish* in Chironomidae genomes**.** *panish* occurrence in recently published genome sequences of *Polypedilum vanderplanki* (Gusev et al. 2014), *P. nubifer* (Gusev et al. 2014), *Clunio marinus* (Kaiser et al. 2016), *Belgica antarctica* (Kelley et al. 2014), and *Parochuls steinenii* (Kim et al. 2017). The phylogeny is based on Cranston *et al.* (Cranston, Hardy, and Morse 2011).

To test whether chironomid species that represent outgroups to Chironominae use *pangolin* as anterior determinant, I collected samples of *Clunio marinus*, an orthocladiinae midge that lacks *panish* (**Figure 2.3.8.)** to examine the expression of Clunio *pangolin* (*Cma-pan*). For this experiment, fixed embryos were provided by Dusica Brisevac from the Tobias Kaiser Lab at the Max-Planck-Institute for Evolutionary Biology in Plön, Germany. Interestingly, according to the annotation provided by the Kaiser lab, *Cma-pan* had four transcript isoforms generated by alternative transcription initiation and alternative splicing (**Figure 2.3.9A**). Among them, *Cma-pan^{D}* encodes a protein that lacks a significant portion of N-terminal region compared to other isoforms and, thus, is more similar to Panish than other predicted Cma-Pan proteins (**Figure 2.3.9B**). To test whether *Cma-pan^{D}* is expressed maternally, I conducted 5'RACE experiments with cDNA from 0-2hr Clunio embryos. I found that *Cma-pan^{D}* is strongly expressed at this developmental stage. This observation raises the question of whether the *Cma-pan^{D}* transcript is anteriorly localized in preblastoderm embryo. To study the expression of *Cma-pan* in *Clunio marinus*, Dusica Brisevac in the Kaiser lab provided fixed embryos. However, RNA in situ hybridization using probes that recognize *Cma-pan^{D}, Cma-pan^{A+B+D},* and *Cma-pan^{A+B+C}* did not reveal localization of any *Cma-pan* isoforms (**Figure 2.3.10.**). One possibility is that *Cma-pan^{D}* is ubiquitously expressed in the preblastoderm embryo with no relevant function in primary axis specification of this species. Alternatively, its activity could be concentrated in the anterior by some other mechanism, such as translational inhibition in the posterior embryo. Another possibility is that mRNAs in the fixed embryo samples were degraded during the sample delivery from Germany.

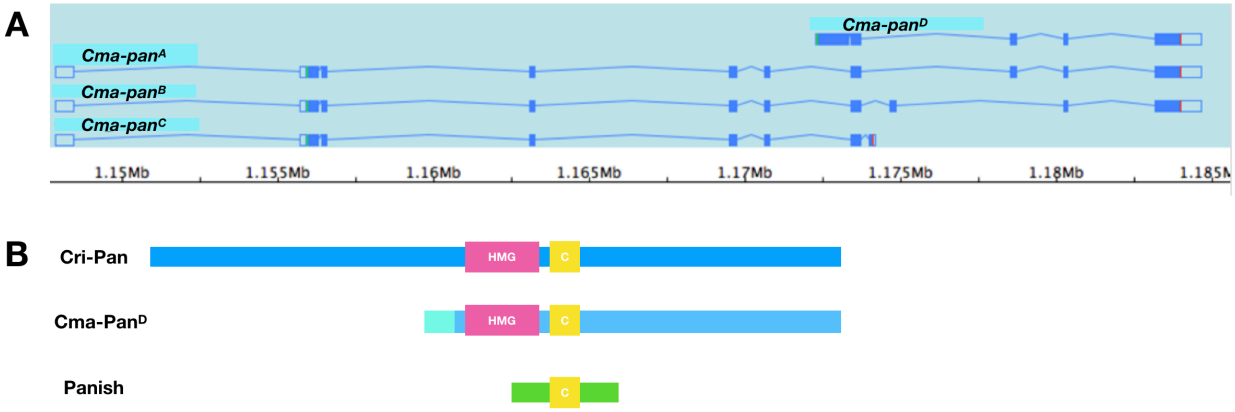


Figure 2.3.9. Alternative *Cma-pan* transcript isoforms. **(A)** Transcript annotation of *Cma-pan* locus (provided by Tobias Kaiser at Max Plank Institute of Evolutionary Biology). **(B)** Comparison of Predicted *Cma-pan*^{D} protein isoform with Cri-Pan and Panish. HMG: HMG box domain; C: Cysteine-clamp domain.

Figure 2.3.10. RNA in situ hybridization of *Cma-pan* transcript isoforms in 0-2hr-old Clunio preblastoderm embryos. Anterior is left and dorsal up.

2.4. Asymmetrical localization of maternal mRNAs in the black soldier fly *Hermetia illucens*[4]

Establishment of *Hermetia illucens* as a developmental genetics model

Basal-branching brachyceran flies are most promising species for filling the "gap" between the primary axis specification mechanisms found in cyclorrhaphan flies and nematoceran flies (see **figure 1.2.1.** for tree). Especially, these flies could provide evidence for testing alternative hypotheses on *bicoid* evolution, since they might retain the mechanism that preceded the *bicoid*-dependent axis specification mechanism in evolution. One prevalent hypothesis is that Bicoid replaced the ancestral function of Orthodenticle, a homeodomain protein with similar DNA binding affinity and a role in head development of many animals (Datta et al. 2018; Lynch et al. 2006; Schröder 2003; Wimmer et al. 2000) (see DISCUSSION for details). An alternative hypothesis is that *bicoid* inherited its function from its progenitor gene *zerknüllt*. Therefore, asking whether these flies use an *orthodenticle*-like or *zerknüllt*-like anterior determinant gene could potentially provide important information for understanding the evolutionary origin of *bicoid.* Among the basal-branching Brachycera, many species have a long generation time (e.g., one year) and/or cannot easily be reared in the laboratory. This is also the case for the dance flies (Empididae), which are among the closest relatives of cyclorrhaphan flies. However, with the help of Danny Chen, an undergraduate student in our lab, I found that the black soldier fly *Hermetia illucens* (Stratiomyidae), another representative of lower Brachycera, is suitable for developmental genetic studies in embryos, can be easily reared, and has a reasonably short generation time (54 days at 28C).

[4] Differential expression analysis of transcriptomes was performed by Jeff Klomp, a previous member of our laboratory.

I established embryo fixation, eggshell removal, *in situ* hybridization, and microinjection protocols for early Hermetia embryos. Hermetia embryos have a tough eggshell but it can be removed manually using a sharp tungsten needle under the microscope. Test *in situ* hybridization experiments using RNA probes for Hermetia *orthodenticle* (*Hil-otd)* and Hermetia *caudal* (*Hil-cad*) were successful (**Figure 2.4.2A-B**). Interestingly, both *Hil-otd* and *Hil-cad* were expressed extensively along the AP axis from the anterior and posterior, respectively, in early cellularizing embryos (**Figure 2.4.2A-B).** Microinjection of *Hil-cad* resulted in posterior truncation of the 1st instar larva, which is an expected phenotype, suggesting that embryonic RNAi works efficiently in Hermetia (**Figure 2.4.3.**).

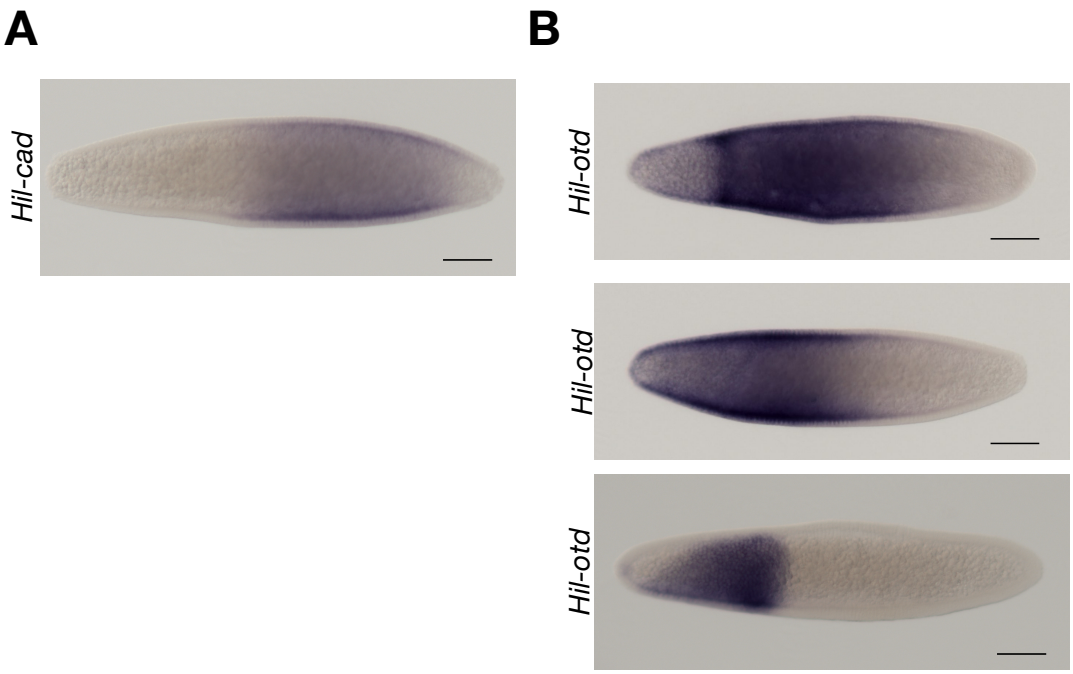


Figure 2.4.1. Establishment of RNA in situ hybridization protocol in *Hermetia illucens* embryo. **(A)** *Hil-cad* expression in early cellular blastoderm embryos **(B)** *Hil-otd* expression in syncytial (top), early cellular blastoderm (middle), and late cellular blastoderm embryos. Anterior is left; Scale bar: 100μm.

Figure 2.4.2. Establishment of RNAi protocol in *Hermetia illucens*. WT cuticle (top) and a cuticle with posterior defect from *Hil-cad* RNAi (bottom) are shown. Anterior is left; dorsal view; Scale bar: 100μm.

Asymmetrical localization of mRNAs in Hermetia egg

Our former lab member Chun Wai Kwan bisected 1hr-old Hermetia embryos and sequenced the anterior and posterior transcriptomes. 7,883 transcripts could be annotated and were ranked according to the magnitude of their differential expression scores and P values. In the posterior embryo, no transcripts were enriched above the significance level, raising the question of whether Hermetia lacks maternal germ plasm, like Clogmia (**Figure 2.4.3. See Figure 2.1.1. for Clogmia**). While transcripts of *oskar* and *nanos* (*Hil-nos*) homologs were

present in the maternal transcriptomes of early Hermetia embryos, RNA in situ hybridization of *Hil-nos* did not show posterior expression in preblastoderm and syncytial blastoderm embryos (in collaboration with Henry Scheffer, a summer student in our lab) (**Figure 2.4.4.**). Surprisingly, many maternal mRNAs were enriched in the anterior embryo (**Figure 2.4.3.**). To test whether this anterior bias in transcript localization was an artifact, I performed RNA in situ hybridization experiments to confirm the localization of transcripts that appeared to be highly enriched in the anterior embryo, including Hermetia homologs of *short stop* (*Hil-shot*), *staufen* (*Hil-stau*)*, TBP-related factor 1* (*Hil-trf1*)*, daughters against dpp* (*Hil-dad*)*,* and *Hil-CG9432.* Transcripts from all of these genes were anteriorly localized in freshly-laid Hermetia eggs, confirming the result from differential expression analysis (**Figure 2.4.5.**). It is therefore likely that transcripts of other genes that are shown to be differentially expressed are also enriched in the anterior pole of the Hermetia egg.

Some of the anteriorly enriched transcripts in Hermetia eggs encode orthologs of Drosophila proteins that function in axis specification, including Shot and Stau. Shot is a cytoskeletal linker molecule that associates with the fusome and is required for microtubule organization during oogenesis (Roper and Brown 2004; Nashchekin, Fernandes, and St Johnston 2016). Importantly, the fusome is required for establishing the body axis within the oocyte by interacting with centrosomes (Riechmann and Ephrussi 2001; Nashchekin, Fernandes, and St Johnston 2016). Therefore, the anterior localization of *Hil-shot* transcript may indicate a polarized microtubule network in the preblastoderm Hermetia embryo and could underlie the massive anterior bias of maternal transcripts of many genes. Stau is a double-stranded RNA binding protein and functions in mRNA localization. Especially, Stau is required during Drosophila oogenesis for the microtubule-dependent localization of *bicoid* and *oskar* mRNAs to

the anterior and posterior poles, respectively. Therefore, also *Hil-stau* may contribute to the anterior transport or localization of transcripts via microtubule network established in Hermetia preblastoderm embryos (Ferrandon et al. 1994; St Johnston, Beuchle, and Nusslein-Volhard 1991).

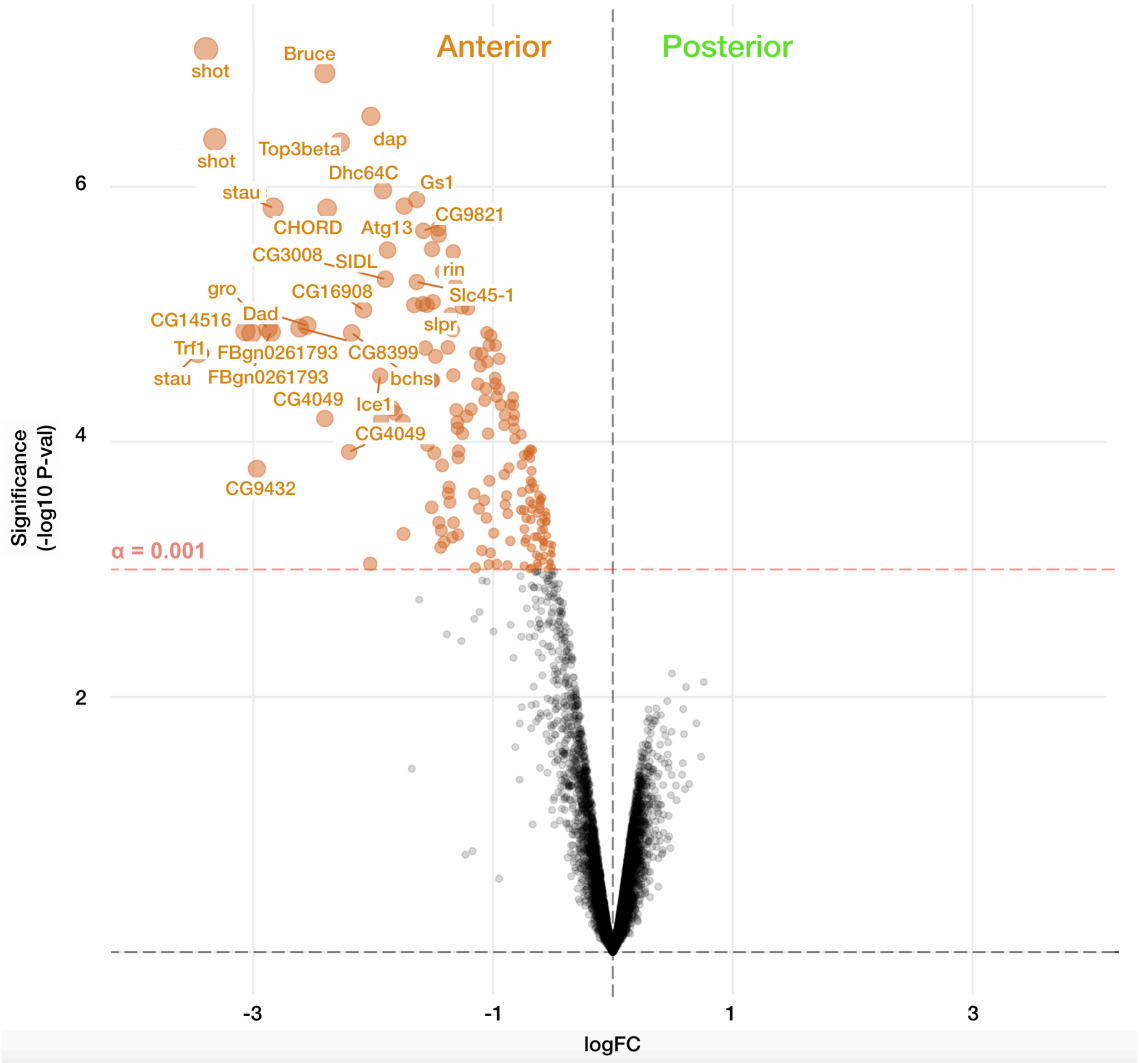


Figure 2.4.3 Differential expression analysis of maternal transcripts between anterior and posterior halves of 1hr-old Hermetia embryos. logFC: log fold-change.

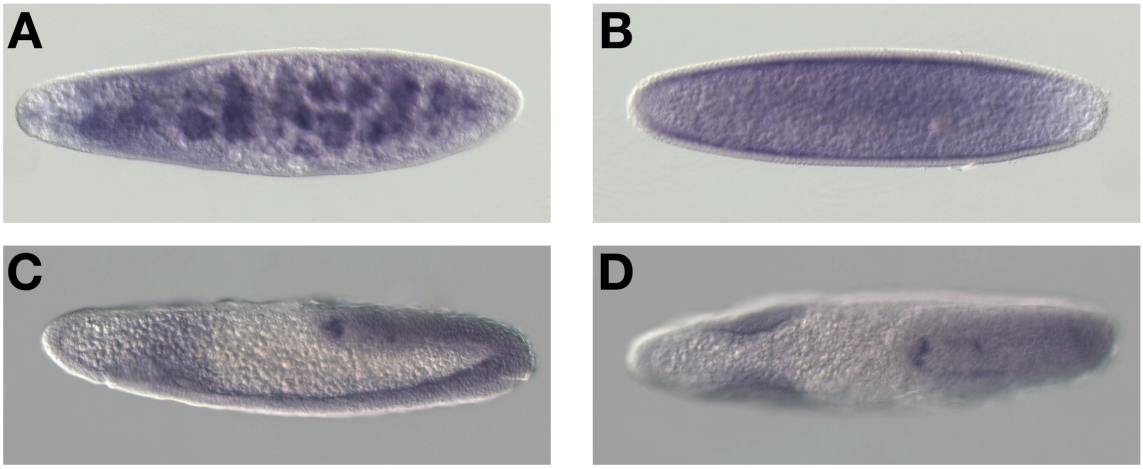


Figure 2.4.4. RNA in situ hybridization of *Hil-nos* in **(A)** preblastoderm embryo **(B)** syncytial blastoderm embryo **(C)** germband extending embryo, lateral view. **(D)** Germband extending embryo, dorsal view. Anterior is left.

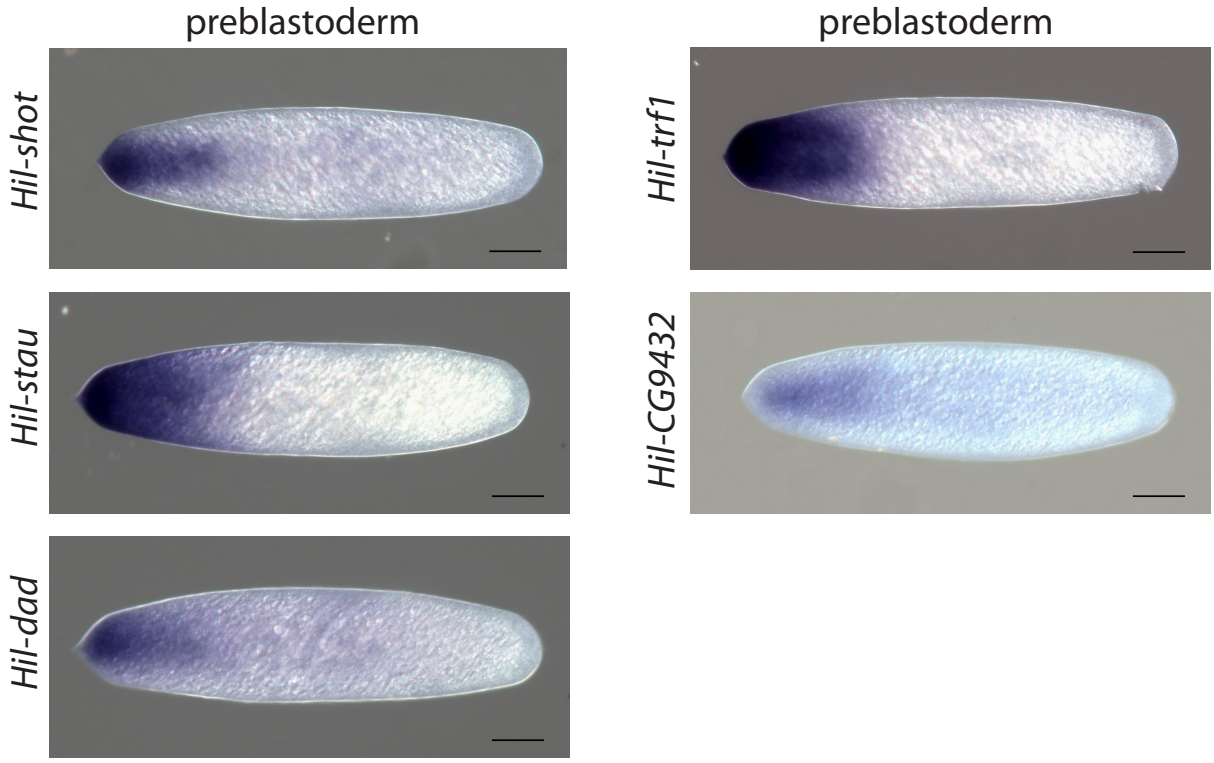


Figure 2.4.5. RNA in situ hybridization of anteriorly localized maternal mRNAs in Hermetia embryo. Anterior is left; Scale bar: 100μm.

BMP signaling is not required for establishing head-to-tail polarity in Hermetia embryo

Our lab previously found an anterior-to-posterior gradient of BMP activity in early Hermetia embryos using antibody staining against phosphorylated Mad protein (pMad), which functions as transcriptional effector of BMP signaling (performed by Chun Wai Kwan, a former member of our laboratory; data not shown). I tested whether the BMP signaling pathway has been co-opted for AP axis formation in Hermetia by knocking down *Hil-dpp,* the ortholog of *decapentaplegic, a* key component of this pathway. *Hil-dpp* RNAi embryos exhibited head and segmentation defects and a twisted abdomen phenotype but retained normal head-to-tail polarity **(Figure 2.4.6B**). Therefore, BMP signaling does not seem to be involved in establishing head-to-tail polarity of Hermetia embryos.

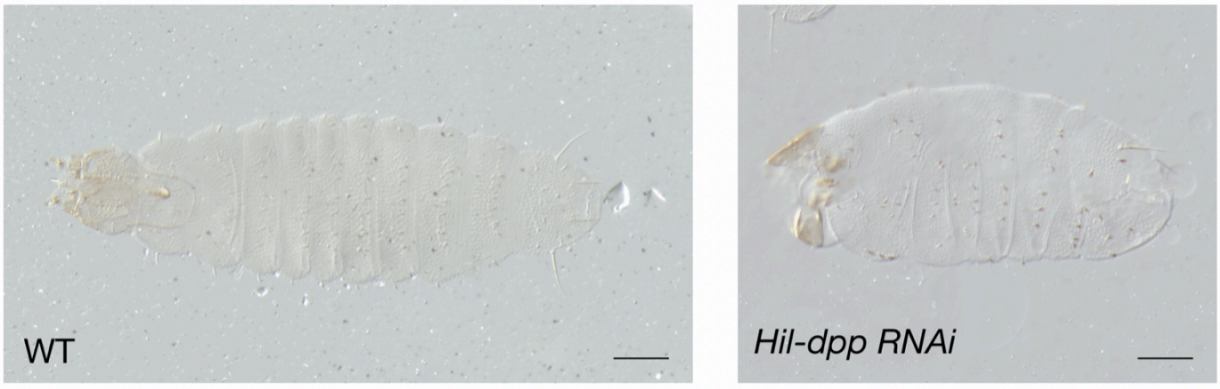


Figure 2.4.6. Function of BMP signaling components in Hermetia. 1st instar Hermetia larval cuticle of wild type (left) and following *Hil-dpp* RNAi. Anterior is left. Dorsal view. Scale bar: 100μm.

2.5. Preliminary data on asymmetrical localization of mRNAs in the egg of the grey flesh fly *Sarcophaga bullata*

The grey flesh fly *Sarcophaga bullata* (Sarcophagidae) is a readily available cyclorrhaphan fly that has lost *bicoid* (J. Klomp, unpublished data). Therefore, I included this species in the search for new anterior determinants, even though follow-up analyses will be complicated by the fact that this species is viviparous. I sequenced the anterior and posterior transcriptomes of 1hr-old preblastoderm Sarcophaga embryos and ranked transcripts according to the magnitude of their differential expression scores and false discovery rate (FDR) values. Among 626 transcripts that were found to be differentially expressed (FDR <0.001; fold change >2), the most anteriorly enriched transcript was homologous to *translocation protein 1* (*trp1*), which has been implicated in the translocation of proteins to the ER membrane (Willer et al. 2003) (**Figure 2.5.1.**). In the posterior embryo, *exuperantia* (*exu*) appeared to be significantly enriched. Exu is an RNA-binding pseudonuclease and is required for the localization of *bicoid* and *oskar* mRNAs during Drosophila oogenesis by associating with microtubules (Schupbach and Wieschaus 1986; Berleth et al. 1988; Cha, Koppetsch, and Theurkauf 2001; Wilhelm et al. 2000; Lazzaretti et al. 2016).

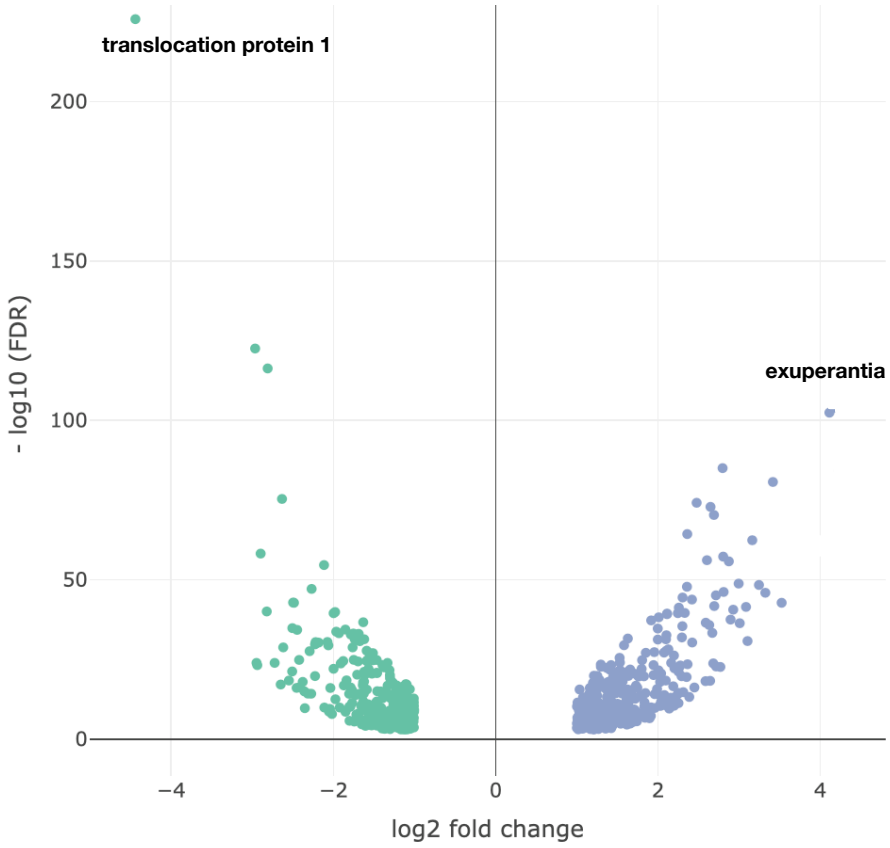


Figure 2.5.1. Differential expression analysis of maternal transcripts between anterior and posterior halves of 1hr-old Sarcophaga embryos. Cutoff; FDR <0.001; fold change >2.

3. DISCUSSION

3.1. Alternative transcription as a primary mechanism for the evolution of embryonic axis determinants

Our study shows that anterior determinants of fly embryos repeatedly evolved via AT from conserved genes (**Figure 4.1.1.**). Unlike AS, AT primarily modifies the 5' or 3' ends of a transcript. In case of *Cal-opa,* the use of alternative transcription start sites generates two transcript isoforms that differ in their 5' ends (ATI). In contrast, *cucoid* and *Aga-pan* use alternative transcription termination sites in conjunction with AS to generate transcript isoforms that differ in their 3'ends (ATT). Therefore, all these anterior determinants contain either 5' or 3' UTR sequence that is not shared with the corresponding zygotic isoforms.

In order to function as an anterior determinant in fly embryos, the transcript must be transferred to the anterior of the egg or otherwise be restricted in its activity (e.g., via translational repression of a ubiquitous transcript in the posterior). Signals for subcellular transcript localization or translational repression are typically found in UTRs (Holt and Bullock 2009). Therefore, it is possible that AT facilitates the evolution of anterior determinants by providing the substrate for isoform-specific localization of signals that do not interfere with other gene functions. For example, it has been shown that alternative last exons of transcript isoforms confer isoform-specific localization in neurons (Taliaferro et al. 2016; Ciolli Mattioli et al. 2018). During the evolution of new anterior determinants, AT could have exposed a preexisting cryptic localization signal in its unique UTR of the maternal isoforms. Alternatively, localization signals may have evolved gradually by enhancing spurious transcript localization. Additional

experiments will be needed to test whether the unique UTR sequences of anterior determinants are essential for their localization at the anterior egg pole.

In addition to changes in UTR sequences, AT also can result in the truncation or elongation of the open reading frame. For example, the anterior determinant of Clogmia (Cal-Opa^{Mat}) lacks the N-terminal 20 amino acids (**Figure 2.1.2.**), and the anterior determinant of Anopheles (Aga-pan^{Mat}) encodes protein that includes additional 7 amino acids at the C-terminal end (**Figure 2.3.2.**). However, the truncation in the maternal Opa protein is not conserved in Lutzomyia, in which *Llo-opa^{Mat}* and *Llo-opa^{Zyg}* encode the same protein, and Odd-paired homologs from these and other species can function as anterior determinant in Clogmia (**Figure 2.1.12.**). Also, the elongation of Aga-Pan^{Mat} protein is not conserved in Nephrotoma, in which the localized *Nsu-pan^{Mat}* transcript encodes a Pangolin protein with a conserved C-terminal end (**Figure 2.3.2.**). Therefore, modifications in the open reading frame of these genes may reflect secondary changes.

It is unclear whether the C-terminal truncation in the anterior determinant of Culex (Cucoid^{C}) is functionally relevant. Due to alternative transcription termination, Cucoid^{C} lacks the last two C2H2 zinc fingers compared to the canonical isoform, which contains five. RACE experiments in Aedes, which also uses a Cucoid homolog as anterior determinant, revealed a C-terminal truncated isoform that lacks three C2H2 zinc fingers (Aae-Cucoid^{C}) but whether this isoform is anteriorly localized remains to be tested. If Aae-Cucoid^{C} is the anterior determinant in Aedes, it is possible that protein modifications in Aedes and Culex Cucoid are functionally important. However, if the full-length isoform (Aae-Cucoid^{A}) is the anterior determinant, modification in Cucoid^{C} may reflect a secondary change. Gain-of-function experiments using

mRNAs encoding different Cucoid isoforms in Aedes and Culex could help to distinguish between these possibilities.

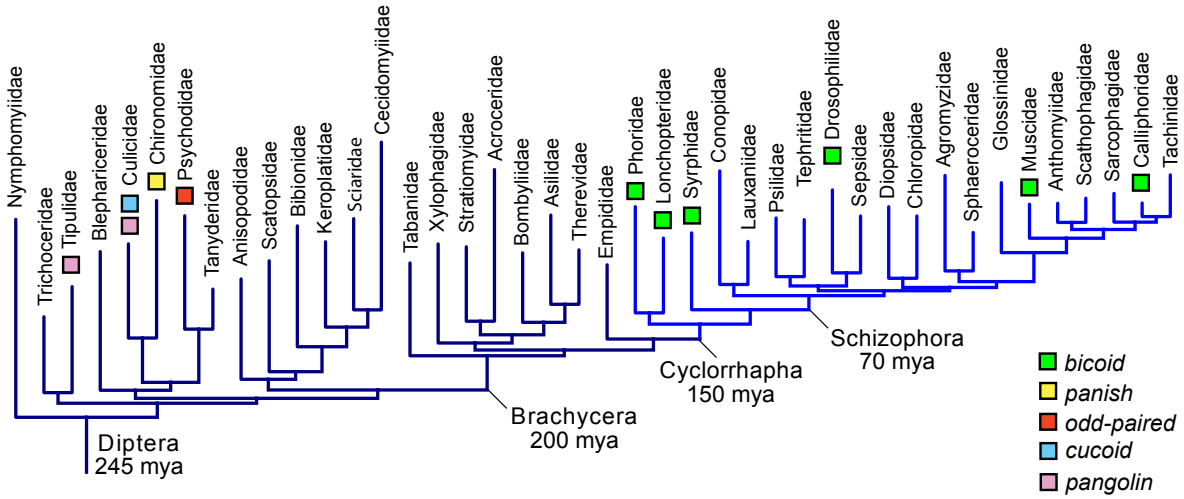


Figure 3.1.1. Anterior determinants in Diptera. The phylogenetic tree of dipteran families is based on published data (Wiegmann et al. 2011). Mya, million years ago.

3.2. Ancestral mechanism for primary axis specification in dipteran insects

It has been proposed that Orthodenticle, or some other homeodomain protein with lysine at position 50 (K50), like Orthodenticle, functioned as ancestral anterior patterning gene in Diptera. One reason for this hypothesis has been that Bicoid also contains a K50 within the homeodomain, which confers DNA binding affinity similar to Orthodenticle, whereas its progenitor gene Zerknüllt and other closely related Hox proteins contain glutamine (Q) at this position (Wilson et al. 1996; Burglin and Affolter 2016). Based on these observations, several studies hypothesized that Bicoid replaced the ancestral role of Orthodenticle as an anterior determinant (Datta et al. 2018; Lynch et al. 2006; Schröder 2003; Wimmer et al. 2000), which I will discuss in more detail in the next section. Another reason is that Orthodenticle is a highly conserved gene that functions in head patterning in animals, and that a role of maternal *orthodenticle* mRNA in anterior pattering has been reported from several long germ insects. For example, in the wasp *Nasonia vitripennis* (Hymenoptera), maternal transcripts of an *orthodenticle* ortholog (*Nv-otd1*) are anteriorly localized in the egg and are required for anterior repression of *caudal* and anterior patterning by regulating head and thoracic gap genes (Lynch et al. 2006), while another anteriorly localized transcript orthologous to *giant* (*Nv-gt*) provides a permissive anterior patterning role (Brent et al. 2007). In the honeybee (*Apis mellifera*), another hymenopteran insect, two *orthodenticle* orthologs are maternally expressed and have been shown to have an anterior patterning role, while their transcripts are not localized at the anterior egg pole (Wilson and Dearden 2011). Finally, anteriorly localized *orthodenticle* mRNA was also found in lepidopteran insects, such as the silk moth *Bombyx mori* (Nakao 2012) and a butterfly *Pararge aegeria* (Carter, Gibbs, and Breuker 2015).

However, my findings from lower dipterans, together with previous findings from Chironomus (Klomp et al. 2015), strongly suggest that a maternal *pangolin* isoform with conserved open reading frame functioned as an anterior determinant in ancestral dipterans (**Figure 3.1.1.)**. The role of *pangolin* as an anterior determinant could be much older than the dipteran lineage. Anteriorly localized maternal *pangolin* transcript has also been observed in the eggs of a beetle (*Tribolium castaneum*, Coleoptera), even though its function remains unresolved (Ansari et al. 2018; Prühs, Beermann, and Schröder 2017; Bolognesi et al. 2008; Bucher et al. 2005). Instead, a homolog of *germ cell-less* (*Tc-gcl*) is required for establishing embryo polarity during oogenesis (Ansari et al. 2018). One of the functions of *Tc-gcl* is to enable the transport of maternal *axin* mRNA from nurse cells to the oocyte, where *axin* mRNA is localized at the anterior pole. Axin protein functions as a negative regulator of ß-catenin dependent Wnt signaling (aka canonical Wnt signaling) and is required for anterior patterning in Tribolium but does not establish embryo polarity per se (Fu et al. 2012). While current models of primary axis specification in the Tribolium embryo do not attribute any specific role to anterior-localized maternal *pangolin* transcript, it could be an evolutionary relic. Phylogenetic comparisons of maternal *pangolin* expression and function across a wide range of beetles and other insects might be necessary to test this hypothesis.

3.3. Evolution of unique anterior determinant genes

Unlike the anterior determinants identified in this study, the previously described anterior determinants of Drosophila and Chironomus are encoded by newly evolved genes, *bicoid* and *panish.* These genes seem to be dispensable outside the context of axis specification (Driever 1993; Klomp et al. 2015), suggesting that they evolved specifically for this function. They could have acquired their function de novo via protein evolution or via inheritance from the progenitor gene. Our findings suggest that the role of *pangolin* in axis specification was already present in ancestral dipterans (**Figure 3.1.1.)**. Accordingly, *panish,* which evolved from *pangolin* via gene duplication in the Chironominae lineage, should have inherited its function from *pangolin*. Another important finding is that, once Panish sequence diverged from Pangolin, it was followed by neutral evolution of developmental gene regulatory networks (developmental systems drift) such that Pangolin can longer function in anterior patterning in Chironomus embryo **(Figure 2.3.7.** and **Figure 7.3.2.**). Future examinations of *pangolin* isoforms and their expression in the eggs of chironomids that lack *panish* orthologs (species representing basal chironomid lineages) could reveal intermediate steps in this process, such as a localized truncated *pangolin* isoform.

Similarly, *bicoid* could have acquired its function de novo via protein evolution or via inheritance from its progenitor gene, *zerknüllt*. As discussed in the section above, several previous studies have hypothesized that Bicoid replaced Orthodenticle, a conserved homeodomain protein with similar DNA-binding affinity that functions in animal head development (Wimmer et al. 2000; Schröder 2003; Lynch et al. 2006; Datta et al. 2018). Ancestrally reconstructed homeodomains confirmed that a single amino acid change in the homeodomain of Bicoid (Q50K), which is shared by Bicoid and Orthodenticle, caused a dramatic shift of Bicoid's DNA-binding affinity *in vitro* and target recognition *in vivo* (Liu et al.

2018). However, my study does not support the underlying assumption that *orthodenticle* functioned as an anterior determinant in ancestral dipterans. Furthermore, we did not find localized *orthodenticle* mRNA or mRNA that encodes a K50 homeodomain protein in any of the lower dipterans or basal-brachyceran flies.

An alternative hypothesis on *bicoid* evolution is that, in analogy with our findings, *zerknüllt* functioned as an anterior determinant in ancestral Cyclorrhaphan flies before *bicoid* evolved and that *bicoid* inherited its function. I found maternal (*Cal-zen*^{Mat}) and zygotic *zerknüllt* (*Cal-zen*^{Zyg}) isoforms with distinct 5'UTRs in preblastoderm and blastoderm Clogmia embryos (**Figure 3.1.2.**) but the maternal transcript is distributed ubiquitously in early embryos. Maternal *zerknüllt* expression is common in lower Diptera but was probably lost in the stem lineage of Cyclorrhapha (Stauber, Prell, and Schmidt-Ott 2002). It is possible that *zerknüllt* evolved a localization signal in this stem lineage of Cyclorrhapha, which spans a time window of about 20 million years (Grimaldi and Engel 2005). This scenario could explain why, in *Megaselia abdita* (Phoridae), a cyclorrhaphan fly with maternal *bicoid* expression and zygotic *zerknüllt* expression, the maternal *bicoid* transcript is localized at the anterior pole of the embryo while the zygotic *zerknüllt* transcript is localized on the apical side of blastoderm cells (Stauber, Jäckle, and Schmidt-Ott 1999), presumably through the same microtubule-dependent machinery (Bullock and Ish-Horowicz 2001). Furthermore, previous findings suggest that Clogmia *hunchback* (*Cal-hb*), the ortholog of a gap gene that primarily patterns anterior segments, contains a *zerknüllt*-dependent enhancer (Lemke et al. 2008) (unpublished data from our laboratory). This observation suggests that Zerknüllt may control the expression of zygotic *hunchback* in some flies. If *bicoid* inherited its function from *zerknüllt*, the Q50K mutation in the homeodomain of Bicoid must have been a secondary, potentially maladaptive change. In this case, it may have

been fixed in the cyclorrhaphan stem lineage via a compensatory or balancing mechanism and would have been the cause of substantial developmental systems drift.

In general, my study provides an example in which regulatory changes in cellular and subcellular gene expression (maternal expression and anterior localization in my case) drives the evolution of new gene functions, whereas protein evolution drives subsequent divergence of the newly established developmental mechanism (developmental systems drift). As the case of *panish* suggests, highly diverged genes with a specific and unique function can be younger than the function they represent. However, developmental systems drift can easily obscure this history, unless intermediate steps are preserved in some extant species. We may thus overestimate the role of protein divergence in the evolution of new genes with essential functions.

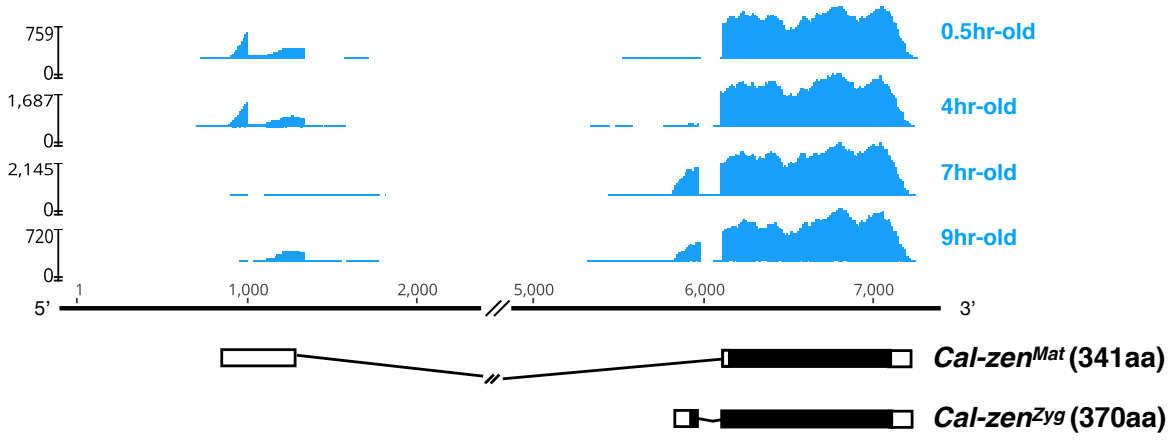


Figure 3.1.2 Maternal and zygotic transcript variants of Clogmia *zerknüllt* (*Cal-zen*). Stage-specific RNA-seq read coverage of *Cal-zen* genomic locus above schematic diagrams of *Cal-zen*^{Mat} and *Cal-zen*^{Zyg} transcript variants with open reading frame (black) and recovered UTR sequences.

3.4. Why did anterior determinants change frequently in dipteran evolution?

The evolutionary lability of anterior determinants in lower flies raises the question of why they change frequently during evolution. An observation from classical UV-irradiation experiments using embryos from chironomid midges (Yajima 1964; Kalthoff and Sander 1968) and Drosophila (Bownes and Kalthoff 1974) was that double abdomens are frequently induced in midges but not in Drosophila, which is much more resistant to such treatment. These observations may reflect differences in the mechanisms of action of the anterior determinants between these species. Lower dipterans may use a relatively simple genetic mechanism for axis specification compared to that of cyclorrhaphan flies, where *bicoid* binds to many target genes (Chen et al. 2012; Ochoa-Espinosa et al. 2005) and additional maternal transcripts, such as *hunchback* and *nanos,* play an additional role in axis specification. In Chironomus, *hunchback* and *nanos* may not be involved in axis specification and it is likely that Panish has only few targets, such as *caudal* and *tailless* (Klomp et al. 2015). In Clogmia, no posteriorly localized maternal transcripts were found and double heads were frequently obtained from posterior *Cal-opa* mRNA injection, while only few double heads could be induced by posterior *bicoid* mRNA injection in the Drosophila embryo (Driever, Siegel, and Nusslein-Volhard 1990). Therefore, in lower dipterans, anterior determinants could have fewer target genes than in cyclorrhaphan flies and this may explain why anterior determinants in lower dipterans change more frequently.

Mechanistically, in lower dipterans, it is possible that axis specification is driven by a similar core mechanism and that only subcomponents are being replaced as anteriorly localized determinant. For example, moth fly Odd-paired could function as a component of a larger transcription factor complex, possibly together with Pangolin. It has been shown that in neuronal precursors of nematodes, orthologs of Pangolin and Odd-paired form a complex that mediates

atypical transcriptional activation of specific targets (Murgan et al. 2015). In this respect, Cal-Opa^{Mat}/Llo-Opa^{Mat}, or Cucoid/Aae-Cucoid might function by forming a complex with the respective Pangolin orthologs to regulate gene expression, given that Clogmia and Aedes *pangolin* transcripts are detected ubiquitously in preblastoderm embryos.

Another possibility is that anterior determinants are co-opted from genes that share a similar feature, such as controlling the chromatin state of target genes. A recent study utilizing ATAC-seq and biochemical assays suggests that Bicoid functions by changing the chromatin state of its targets in a concentration-dependent manner (Hannon, Blythe, and Wieschaus 2017). Similarly, Drosophila *opa* seems to regulate its target pair-rule genes by modifying their chromatin state during segmentation (Shelby Blythe, personal communication). If controlling chromatin states of early developmental genes is the key-feature of anterior determinants, it is possible that *opa* homologs in moth flies have been co-opted for such a function in axis specification by promoting open chromatin states for anterior-patterning genes in a concentration-dependent manner.

4. FUTURE DIRECTIONS

Significance of AT in the evolution of new gene functions

My results suggest that AT may play an important role in evolution than previously thought. It will be interesting to conduct genome-wide analyses of AT events in species beyond the model organisms, using high-throughput sequencing methods such as CAGE and 3'-seq. These analyses will identify species-specific AT events and may allow us to study the significance of AT in the origin of new gene function. For example, this comparative approach could reveal a frequent pattern in which AT underlies the origin of new gene functions before any gene duplication events have taken place in derived species. The results will help us to reassess the significance of gene duplication and protein evolution in the functional evolution of the genome.

Molecular and genetic mechanism of newly identified anterior determinants

Investigating the mechanism of action of anterior determinants identified in this study could inform the question of why they change frequently in dipteran evolution. For example, protein tagging of anterior determinants using CRISPR/Cas9 will allows us to explore their target binding sites or co-factors via ChIP-seq or co-immunoprecipitation. These experiments may reveal similar targets or binding partners shared among different anterior determinants. Also, it will be interesting to test whether canonical Wnt signaling underlies axis specification in lower dipterans, given that a Wnt signaling gene (*pangolin*), its variant (*panish*), and its potential co-factor (*odd-paired*) are functioning as anterior determinant in different species. In Chironomus, an antagonist of Wnt signaling (*Cri-APC*) is enriched in the posterior pole of preblastoderm embryo (**Figure 6.3.5A**) and knockdown of Wnt signaling genes such as *axin*

(*Cri-axin*) and *Cri-APC* have produced severe defects including ectopic eyespots, which may reflect a disrupted axis formation (preliminary data, results not shown).

Axis specification mechanism in Hermetia embryos

My preliminary data in Hermetia will be useful to study how Hermetia embryos establish head-to-tail polarity and how *bicoid* evolved. The massive anterior bias of maternal mRNAs in these embryos suggest that Hermetia embryos might use a different mechanism for axis specification and maternal mRNA localization than midges and mosquitoes. For example, anterior localization of many mRNAs could be a byproduct of *Hil-shot* localization, which may polarize the microtubule network in the preblastoderm embryo, given that it does so in Drosophila oocytes (Nashchekin, Fernandes, and St Johnston 2016). In this case, the anterior determinant of Hermetia may not have to be maternally localized at the anterior pole; it is possible that its ubiquitous early zygotic transcript is transported to the anterior embryo to exert its function. Also, since Hermetia is a basal-brachyceran fly that lacks *bicoid,* it will be interesting to ask whether Hermetia *zerknüllt* (*Hil-zen*) or *Hil-otd* has a role in axis specification or regulates the expression of *hunchback* ortholog (*Hil-hb*) in early embryo.

5. MATERIALS AND METHODS

Cloning procedures and mRNA/dsRNA synthesis

Coding sequences from Clogmia, Lutzomyia, Chironomus, Anopheles, Nephrotoma, Culex, Aedes, Hermetia were amplified from embryonic cDNA with primers constructed from RNA-seq data. The coding sequence of *odd-paired* was amplified from cDNA (FI01113) that was obtained from the BDGP Gold collection of the Drosophila Genomics Resource Center. Amplified cDNA was cloned into the expression vector pSP35T (Amaya, Musci, and Kirschner 1991), using In-Fusion® HD Cloning Kit (Clontech), and PstI- or EcoRI-linearized vector was used for mRNA synthesis using mMESSAGE mMACHINE^{TM} SP6. *panish, Cri-pan,* and *Cri-pan^{trunc}* mRNAs were synthesized from PCR template (containing T7 polymerase binding site) using mMESSAGE mMACHINE^{TM} T7ultra. Mutations in the open reading frame (*Cal-opa$^{Zyg\text{-}Met21Leu}$*, *panish FS, Cri-pan^{trunc}*) and in the kosak sequences of *Dme-opa* and *Cri-opa* were generated using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent). Double-stranded RNA (dsRNA) was generated from PCR-amplified templates using embryonic cDNA and primers containing T7 polymerase binding sites as described (Klomp et al. 2015). Forward and reverse primer sequences for generating templates for mRNA synthesis were:

Cal-opa^{Mat}:

5’-TAAGATGAGTCCGAATCACTTACTGGCC

5’-TTAATAGGCCGTCGCTGCACC

Cal-opa^{zyg}:

5’-CAACATGATGATGAACGCTTTTATGGAA

5’-TTAATAGGCCGTCGCTGCACC

Cal-opa$^{115\text{-}655}$:

5’-ATGCTCTTCTCAAATCACTCTTCAGC

5’-TTAATAGGCCGTCGCTGCACC

Cal-opa$^{182\text{-}655}$:

5’-ATGAACCCGGGAACCTTGGG

5’-TTAATAGGCCGTCGCTGCACC

Cal-opa$^{182\text{-}445}$:

5’-ATGAACCCGGGAACCTTGGG

5’-TTACGCGGGATTCAGCTGACTATG

Cal-opa$^{1\text{-}337+22}$:

5’-CAACATGATGATGAACGCTTTTATGGAA

5’-TCACAAAATTTCACTGAATTCCGTCAAAATATCACTAGA

Llo-opa:

5’-AAAGATGATGATGAATGCATTTATGGACACAG

5’-TCAGTACGCCGTGGCGGCG

Dme-opa$^{Dme\ kozak}$:

5’-GACCATGATGATGAACGCCTTCA

5’-GTCAATACGCCGTCGCTGCGCCGGG

Cri-opa$^{Cri\ kozak}$:

5’-AAAAATGATGATGAATGGTTTTATGGACACA

5’-TCAATAAGCTGTCGTTGGACCGTGAT

Cri-pan$^{trunc.}$:

5’- AAAAATGTATCCAGATTGGAGCTCGC

5’- TTACGTCACACTAATAGCATTTCCATCATCCC

Forward and reverse primer sequences for dsRNA were (gene specific sequence underlined):

Cal-opa^{Mat}:

5'-CAGAGATGCATAATACGACTCACTATAGGGAGAAAACAATTGTGAAGTGCGACA

5'-

CAGAGATGCATAATACGACTCACTATAGGGAGACAAATTTCCAAACGATGACAGA

Cal-opa^{zyg}:

5'- CAGAGATGCATAATACGACTCACTATAGGGAGAACTACCGCCGCGAACACACG

5'-CAGAGATGCATAATACGACTCACTATAGGGAGAGTCCAGTCGATTCCATAAAAGC

Cal-slp:

5'- CAGAGATGCATAATACGACTCACTATAGGGAGATCGATCAGCTCCCTTTTGCC

5'-CAGAGATGCATAATACGACTCACTATAGGGAGATGAGATCGTTCCCGTTGGAC

Cal-mira:

5'- CAGAGATGCATAATACGACTCACTATAGGGAGAACAGCAAAAAGGAAGCGAAA

5'-CAGAGATGCATAATACGACTCACTATAGGGAGAGGGATTCAATTTGCCTTTGA

Aga-pan^{Mat}:

5'- CAGAGATGCATAATACGACTCACTATAGGGAGACACACAGGGCACAATAATCG

5'- CAGAGATGCATAATACGACTCACTATAGGGAGAGACTGCATGTCCGTCGTCTA

Aga-pan^{Zyg}:

5'- CAGAGATGCATAATACGACTCACTATAGGGAGAACATCACACACCCCACACAC

5'- CAGAGATGCATAATACGACTCACTATAGGGAGATTGGTCCGTTCGTGATTGTA

cucoid:

5' - CAGAGATGCATAATACGACTCACTATAGGGAGACGAGGATGTTGCTGGAGAAT

5'- CAGAGATGCATAATACGACTCACTATAGGGAGAACTCCCGAAATCGGAAAACT

Aae-cucoid:

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACCGGAATCGAATCTGAGCTA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATGATCTGGATGTTGCCGTAG

Cri-pan:

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAGCGAGAAGGAAGATGAGCAG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAGCGAGCTCCAATCTGGATAC

Hil-cad:

5’-

CAGAGATGCATAATACGACTCACTATAGGGAGAAAAAGTCTAACCAATATTTTCCTT

TCC

5’- CAGAGATGCATAATACGACTCACTATAGGGAGACATTGTGAGAGCACCGATTG

Hil-dpp:

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACCAAAACGGCCTAAAATTGA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAAACACACCTTGGGCACTTTC

Microinjection of embryos (See APPENDIX 6.4. for a step-by-step protocol)

Chironomus embryo injection was done as previously described (Klomp et al. 2015). Clogmia eggs were dissected from ovaries and activated under water. Eggs of Aedes, Culex, and Anopheles were collected in a dark chamber on a moist filter paper for about 30 minutes (Fisher Scientific, Cat. No 09-795C). Hermetia eggs were collected in a piece of cardboard. These eggs were transferred to another filter paper cut into 4cm x 2cm pieces and aligned perpendicularly to the edge of a cover glass (Fisher Scientific, Cat. No 12-648-5C) with the prospective injection side pointing towards the glass edge. We noticed that injecting eggs near the anterior or posterior

pole was critical for survival of the procedure. During the alignment procedure, water was applied to the filter paper as needed to prevent eggs from desiccation. After aligning the eggs, the cover glass was removed and excess water on the filter paper was absorbed using filter paper. A second cover glass with a layer of double-sided tape (Scotch 3M) was slightly pressed against the aligned eggs to transfer the eggs to the double-sided tape. The embryos were then immediately covered with halocarbon oil to prevent desiccation. For cuticle preparations, the embryos were injected under halocarbon oil 27 (Sigma, MKBZ7202V). The oil was washed off under a gentle stream of water immediately after injection. In the case of Clogmia, Aedes, Culex, and Hermetia, the cover glass was transferred to a moist chamber (petri dish with wet kimwipe paper) and kept at 28˚C, and water was added every day to prevent desiccation. In the case of Anopheles, the eggs were allowed to develop under water. Removal of the halocarbon oil was critical to ensure embryo survival until late developmental stages and hatching. For eggs to be fixed within a day following injection, we used a 1:1 mixture of halocarbon oil 27 and halocarbon oil 700 (Sigma, MKCB5817) and left the injected eggs immersed in the oil until fixation. Embryos were injected with quartz needles using a Narishige IM-300 microinjector. Quartz capillaries (Sutter Instruments Q100-70-10) were pulled with a Sutter instrument P-2000 laser-based micropipette puller. Our settings for the needle puller were: Heat 645, Fil 4, Vel 40, Del 125, Pul 130. Needles were back-filled and the tip was broken open at the time of injection by slightly touching the first egg.

Embryo fixation (See APPENDIX 6.4.for a step-by-step protocol)

Clogmia and Hermetia embryos were dechorionated using a 10% dilution of commercial bleach (8.25% sodium hypochlorite) for 3 minutes. For Nephrotoma embryos, a 25% dilution was used for 3 minutes until the chorion became slightly transparent. Embryos of Aedes, Culex, and Anopheles were dechorionated as described (Juhn and James 2012). Dechorionated embryos were fixed in a 50mL falcon tube, using 20mL of boiling salt/detergent-solution (100µL 10% triton-X, 500µL 28% NaCl, up to 20mL of water). After 10 seconds, water was applied to the tube to cool down the embryos. If needed, the embryos were devitellinized in a 1:1 mixture of n-heptane and methanol by gentle shaking. Embryos with vitelline membrane attached were further devitellinized using sharp tungsten needles in an agar plate covered with methanol. Devitellized embryos were stored in 100% methanol at -20°C.

RNA in situ hybridization

RNA in situ hybridizations were conducted as described (Klomp et al., 2015), using digoxigenin (DIG)-labeled probes and Fab fragments from anti-DIG antibodies conjugated with alkaline phosphatase (AP) (Roche, IN, USA). Probes were prepared from PCR templates, using sequence-specific forward primers and reverse primers with T7 promoter sequence (see above for *Cal-opa*^{Mat}*, Cal-opa*^{Zyg}*, Cal-cad, Cal-slp, Cal-mira, Aga-pan*^{Mat}*, Aga-pan*^{Zyg}*,* and *Aae-cucoid*) (gene specific sequence underlined).

Cal-nos1:

5’-AGCACTTTTCCCCCAAGAGT

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAGGCATTCATATTTCCTCAGCA

Cal-nos2:

5’-AATTATTCTGTTCCAAAGTTGAGATT

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACCCCAGACTGGTGACAAAT

Cal-nos3:

5’-TGAGTTAAATAGAGTGAAAACAGCAAA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATACCGTCTCGTGCTTAATCG

Cal-nos4:

5’-GGCAAAATTTTCCAAGTGAA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACGTGTCCTCAAGCGTGTAGAT

Cal-vas:

5’-CTGAGGCGAACTTGTGTGAA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAATTGGCAATGTCCAGTCCTC

Cal-tud

5’-ATTCTGCAAGTCGTCGAGGT

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACCTGTACCAGCCATTGTCCT

Cal-gcl

5’-GCAGAACCCCTTGGACATTA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAGTAACGCCCACAATTCGTCT

cucoid^{A}:

5’-CGAGGATGTTGCTGGAGAAT

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAACTCCCGAAATCGGAAAACT

cucoid^{B}:

5’-ACGATGAGGAGGAGGGTTCT

5’- CAGAGATGCATAATACGACTCACTATAGGGAGACGCACTTCACCGTGTGTAAC

cucoid^{C}:

5’-GGGGCGACATCTATATCTCACT

5’-

CAGAGATGCATAATACGACTCACTATAGGGAGAACAGTGAGAAAAATTCCCAACTT

TAGT

Cqu-cad:

5’-CACGTGTTCCATCAGTCCAG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAATGAGGCTTAACGAGGATGG

Cqu-nos:

5’-AAGTGCCGTGAATTTTGTCC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAGCGAAACCAATTCGACAGTT

Nsu-pan:

5’-TCGCGGCAAGATCATAGTCC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACCTGCAGGGTTTACACCACT

Aae-nos:

5’- CAAACGTGAAGCGGAAGATT

5’- CAGAGATGCATAATACGACTCACTATAGGGAGAATTACGTCCGGAAGTGTTCG

Hil-nos:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATCAGGCGTAAGCAGTTCAAA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATGATGATTGGCTTCTGTGGA

Hil-otd:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATCGCAATCACATTCAACACA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATGTGGATATTCGAGGCCATT

Hil-cad:

5’-

CCAAGCCTTCATTAACCCTCACTAAAGGGAGAAAAAGTCTAACCAATATTTTCCTTT

CC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACATTGTGAGAGCACCGATTG

Hil-shot:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGACCTGGTGACCCACAACTCTT

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAAACAAGTTGGACCAGGCATC

Hil-stau:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATGCTCTTGCCATGAAACTTG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATTCAGTGCCTGCTGAATTTG

Hil-trf1:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAAGTCCACCGCAAATACCAAG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACTGCATCATGGAAGGGATCT

Hil-dad:

5’-TTCATTAACCCTCACTAAAGGGAGAGCGAATGGAGATTGATGGAT

5’-CATAATACGACTCACTATAGGGAGAAGTTAGGCCTCGAAGTGCAA

Hil-CG9432

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATATACCGAAACCGCAGGAAC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAGGATTTGCAGCTGAACCATT

Cma-pan^{ABC}:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAAAAATTCCACTTGGCTGTGC

5’- CAGAGATGCATAATACGACTCACTATAGGGAGATGCTGCACTCATTTCTGGAC

Cma-pan^{ABD}:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATGCTCTGAGCGGTCTGTCTA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACTTTGCAGGTGGCACTATGA

Cma-pan^{D}:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGGAGCCGTGTTTAAATGAGC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAACTTCGCCTTCAGTGTTGCT

Cal-bowl:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGCCGGTTTTACCCTCGAGAA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACGATCCCACCTCCAAAACGA

Cal-pan:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAACCTCCTCTACACCGGGACT

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAACTTTTTCTTTCGCCTGCAA

Cal-arm:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGTCAATGTGGTGACCTGTGC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAGCTCCATTGACAGCCTCTTC

Cal-slp:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATCGATCAGCTCCCTTTTGCC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATGAGATCGTTCCCGTTGGAC

Cal-blch:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATGCCTCCCTGGGCGGAAAAC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAACTTCGCCAGCTTCAGTCGCA

Cal-gro:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATCGACGCCTTCCAACAAGAA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATTTCGTTCCATCGGCACTGA

Cal-stau:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAAACGGGAATGCTGAACAAAC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACCTACCATCAGCACCACCTT

Cal-unknown (ORF is unknown so two probes were made in both directions)*:*

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAAACCAGATGTGCGCCAAGAA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACTCGATTTCAGGGCGTGGAT

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAAACCAGATGTGCGCCAAGAA

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGACTCGATTTCAGGGCGTGGAT

Cri-puf:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAATTTGCCACGTCGTAGTTCC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATGATTCTGGTTTTGCTGCTG

Cri-APC:

5’-#1972

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATGGTGGTCGAATGCTGTCAA

Cri-cic:

5’-#1486

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATGGATATCTGCAGAATTCGCCC

Cri-dor:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGCTTTCGGACGTTTTGTTTG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATCTTGCATCGAGATTGACCA

Cri-opa:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAAGCACTCACATCTTGCATCG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAGCCTATCACAACCCTCGTGT

Cri-otd:

5’-ATTAACCCTCACTAAAGGGAGACAAATAGCTCGAATGCAACG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACACTCATTGCACCACCAGAG

Aga-CG9125:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGTATTCCTTCGCCAACCGTA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAACCACCACTACCACCACCAT

Aga-cad:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGTCGGGTGTTGCTCTTGTT

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACCTTCCCCAATTACACGGCT

Aae-nej:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGCTAGTGCAAGCCATCTTCC

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATAGGCTATTCCGCTGCTGTT

Aae-ds:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATTTCGGATGAAATCCAGGAG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGACAAACATGGGAGTGTTGTCG

Aae-spen:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGGTCGAAGAGTTGGGATGAA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATTTCCGATCCTTCGACATTC

Nsu-cad:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGATCGTGCAGATCAAAAACGAG

5’-CAGAGATGCATAATACGACTCACTATAGGGAGAAGTTCCAAGCGTTGATGGTC

Cfu-pan:

5’-CCAAGCCTTCATTAACCCTCACTAAAGGGAGAGGGAGGCCATCAAAGGTGTA

5’-CAGAGATGCATAATACGACTCACTATAGGGAGATGCTGCAGGATGTATGGGTG

Rapid Amplification of cDNA Ends (RACE)

Total RNA was phenol/chloroform extracted from Clogmia (1hr-old and 9hr-old embryos), Anopheles (1-6hr-old embryos), Culex (0-7hr old embryos), and Nephrotoma (1-29hr-old embryos) fixed in TRIzol™ Reagent (Invitrogen) and precipitated with isopropanol. 5’/3’ RACE was performed using SMARTer® RACE 5'/3' Kit (Clontech) with the custom-made primers (including at the 5’ end 15 nucleotides of pRACE vector sequence). Gene specific sequences are underlined.

Cal-opa 5’RACE primer: 5’-

GATTACGCCAAGCTTCTGGGTGACGCCGTGGGCAAGGACGTCA *Cal-opa* 3’RACE primer: 5’-GATTACGCCAAGCTTCGCGTCGATCGTCACGCCCCCAAATTCG

Aga-pan 5’RACE primer: 5’-

GATTACGCCAAGCTTCGAATCTCCGGCCGCGGAATTGAGACTT

Aga-pan 3’RACE primer: 5’-

GATTACGCCAAGCTTAGCTTCACGCGACCAGCAAAACCAACGG

cucoid 5’RACE primer:

5’- GATTACGCCAAGCTTCGTGACGGCTTCGATGGTTGGTTTTTCC

cucoid 3’RACE primer:

5’- GATTACGCCAAGCTTCGCACGTGTTGAACAGTCACATGTTGAC

Nsu-pan 5’RACE primer:

5’-GATTACGCCAAGCTTTCTGGTCGTGCGACGTTCTTCCAAATCG

Nsu-pan 3’RACE primer:

5’-GATTACGCCAAGCTTTCCCGTTGGTGCAAATCCACGAGATGTG

Aae-cucoid 5’RACE primer A:

5’-GATTACGCCAAGCTTGTAGGCCACCCGAAGATCCTCGATGCAC

Aae-cucoid 5’RACE primer B:

5’-GATTACGCCAAGCTTGGTGATGTGACTTGGCGAAGCGTTTGCC

Aae-cucoid 3’RACE primer A:

5’-GATTACGCCAAGCTTGTGCATCGAGGATCTTCGGGTGGCCTAC

Aae-cucoid 3’RACE primer B:

5’- GATTACGCCAAGCTTGGCAAACGCTTCGCCAAGTCACATCACC

Cma-pan 5’RACE primer:

5’- GATTACGCCAAGCTTGGAAGAAAACGGGCACCTGAAGTCTCGT

Cuticle preparations

Cuticles were prepped four to five days after injection. Eggshells were removed with tungsten needles and the embryos were transferred to a glass block dish with a drop of 1:4 glycerol/acetic acid. Following incubation in 1:4 glycerol/acetic acid overnight at room temperature, the cuticles were transferred onto a glass slide, oriented, mounted in 1:1 Hoyer’s medium/lactic acid (Stern and Sucena 2000), covered with a cover glass, and dried overnight at 65˚C.

RNA-seq sample preparation and sequencing

Bisection of anterior and posterior embryo halves, RNA extraction, and sequencing were conducted as described (Klomp et al. 2015). In the case of Clogmia, anterior or posterior embryo halves from three 1hr-old embryos were pooled and RNA-seq data were obtained from two replicates. In the case of Lutzomyia, embryo halves from ten 1-2hr-old embryos were pooled and four replicates were generated. In the case of Anopheles (G-3 strain), embryo halves from five 1hr-old embryos were pooled and three replicates were generated. In the case of Culex, embryo halves from seven 1hr-old embryos were pooled and three replicates were generated. In the case of Aedes (Liverpool "black eye" strain), embryo halves from five 1hr-old embryos were pooled and four replicates were generated. In the case of Nephrotoma, embryo halves from nine 1hr-old embryos were pooled and three replicates were generated. In the case of Hermetia, embryo halves from two 1hr-old embryos were pooled three replicates were generated. In the case of Sarcophaga, embryo halves from four 1hr-old embryos were pooled and two replicates were generated.

Stage-specific Clogmia transcriptomes were generated from the offspring of a single mother and total RNA from five embryos was used for each stage. In the case of Lutzomyia, about 100 staged embryos were pooled for RNA extraction, and two independent RNA extractions from each time point were combined and submitted for sequencing.

Prior to library construction, RNA integrity, purity, and concentration were assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip (Agilent Technologies, USA). Purification of messenger RNA (mRNA) was performed using the oligo-dT beads provided in the Illumina TruSEQ mRNA RNA-SEQ kit (Illumina, USA). Complementary DNA (cDNA) libraries for Illumina sequencing were constructed using the Illumina TruSEQ mRNA RNA-SEQ

kit (Illumina, USA), using the manufacturer-specified protocol. Briefly, the mRNA was chemically fragmented and primed with random oligos for first strand cDNA synthesis. Second strand cDNA synthesis was then carried out with dUTPs to preserve strand orientation information. The double-stranded cDNA was then purified, end repaired, and "a-tailed" for adaptor ligation. Following ligation, the samples were selected a final library size (adapters included) of 400-550 bp using sequential AMPure XP bead isolation (Beckman Coulter, USA). The libraries were sequenced in an Illumina HiSeq 4000 DNA sequencer, utilizing a pair end sequencing flow cell with a HiSeq Reagent Kit v4 (Illumina, USA).

RNA-seq data preprocessing

The TrimGalore (Krueger 2012) wrapper for Cutadapt (Martin 2011) and FastQC (Andrews 2010) was used to remove adapters and low quality sequences from raw fastq files. Overlapping reads were combined with Flash (Magoc and Salzberg 2011) prior to assembly.

Transcriptome assembly and annotation

Trinity 2.4.0 (Grabherr et al. 2011) on the Indiana University Karst high-performance computing cluster was used for assembling contiguous sequences (contigs) from the paired end (PE) sequence data of Clogmia, Lutzomyia, Anopheles, and Nephrotoma. ABySS 2.0 (Jackman et al. 2017) was used for assembling contigs from Culex and Aedes data. Only contigs of 200 nucleotides or greater were retained. BLAST+ tools (Camacho et al. 2009) were used to annotate contigs by conducting best-reciprocal-blast first against the *Drosophila melanogaster* transcriptome (BDGP6) peptide sequences (blastx/tblastn) and then the coding sequence (tblastx) with a maximum threshold evalue of 1e-10. Biomart and AnnotationDbi packages were used for

gene ids and names. The longest open reading frames (ORFs) of unannotated transcripts were compared to the RefSeq invertebrate protein database (downloaded 4-1-2017) using blastp (max evalue 1e-10) followed by a similar comparison to remove transcripts with ORFs matching RefSeq plant, protozoan, archaea, bacteria, fungi, plasmid, or viral sequences (downloaded 6-1-2017). Remaining transcripts were designated by the top BLAST hit in *D. melanogaster*.

Alignment and differential expression analysis

Cleaned paired-end read data was aligned and analyzed using R base (Ihaka and Gentleman 1996) and Bioconductor (Gentleman et al. 2004) software packages. Sequence alignment was conducted with the seed-and-vote aligner, Subread, as implemented in the Rsubread package (Liao, Smyth, and Shi 2013) with up to 5 multi-mapping locations, 6 mismatches, and 20 subreads/seeds per read. Sequence file manipulation, including sorting and indexing of ".bam" files, was done using Rsamtools (Morgan et al. 2013).

To avoid potential biases in transcript localization unrelated to anterior-posterior axis formation, transcripts annotated with mitochondrial, ribosomal, or ambiguous status (e.g., predicted, hypothetical, or uncharacterized) were filtered out prior to the differential expression comparisons. Transcripts with 20 or fewer counts in any of the A-P pairs were also excluded from the analysis prior to library normalization. Lower scoring, potentially related transcripts matching a given gene from the *D. melanogaster* transcriptome were retained for initial differential expression comparisons but removed for clarity of presentation in subsequent analyses and volcano plots. Trimmed mean of M-values (TMM) (Robinson and Oshlack 2010) was used for normalization and EdgeR (Robinson, McCarthy, and Smyth 2010) was used to perform quasi-likelihood F-tests between A-P samples, corrected for multiple testing using FDR

(Benjamini-Hochberg). Following filtering based on annotation and detection of >20 counts per paired samples, we used the following number of transcripts for differential expression comparisons: 5,602 for Clogmia; 5,392 for Lutzomyia; 8,239 for Culex; 5,802 for Aedes; 9,353 for Anopheles; 5,371 for Nephrotoma; 7,883 for Hermetia

Mapping RNA-seq reads to genomic loci

RNA-seq reads from stage-specific transcriptomes were mapped to genomic scaffolds containing a gene of interest using TopHat RNA-seq aligner (Trapnell, Pachter, and Salzberg 2009). Publicly available *Anopheles stephensi* transcriptomes used in this paper were: SRR515316, SRR515341, SRR514863, and SRR515304.

Data availability

Sequencing data from this project (described under Results, sections 1-3) was deposited at the National Center for Biotechnology Information under Bioproject ID PRJNA454000. Sequencing data from Hermetia and Sarcophaga are available in the Schmidt-Ott laboratory.

6. APPENDIX

6.1. *Clogmia albipunctata*

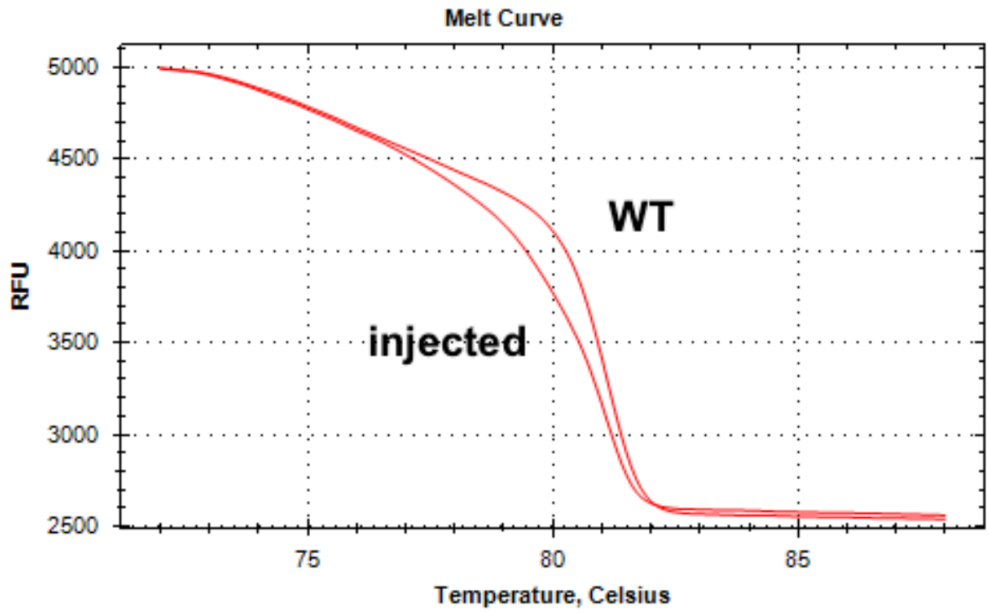
6.1.1. Establishment of CRISPR/Cas9 genome-editing in Clogmia

To obtain proof-of-concept that CRISPR/Cas9 genome editing is possible in Clogmia, I synthesized guide RNA (gRNA) targeting the beginning of *Cal-opa*^{Mat} open reading frame, which effects both Cal-Opa^{Mat} and Cal-Opa^{Zyg} (see below for gRNA sequence and **APPENDIX 6.4.** for step-by-step instructions). I injected the gRNA (1ug/ul) together with Cas9 protein (1ug/ul – PNA Bio CP0120) into early syncytial blastoderm Clogmia embryos and prepped genomic DNA from each hatched larva. Using the genomic DNA and primers below, I conducted high resolution melting (HRM) analysis to check whether mutations had occurred in the targeted *Cal-opa* sequence. Among 64 prepped larvae, 17 revealed a shift in the melt curve compared to WT (**Fig 6.1.1A**), and Sanger sequencing using the same primer revealed indel mutations near the target site (**Fig 6.1.1B).** Therefore, I conclude that CRISPR/Cas9 genome editing works in Clogmia embryos.

- *Cri-tll* gRNA forward primer (gene specific sequence underlined):
    - GAAATTAATACGACTCACTATAGGATGAGTCCGAATCACTTACGTTTAAG AGCTATGCTGGAA
- gRNA reverse primer:
    - AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTT ATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC
- *Cal-opa* HRM F:
    - AACGCTTTTATGGAATCGACTGGAC

- *Cal-opa* HRM R:
    - GTCTGAAAAGTCCCCATCTCTGTCA

A



B

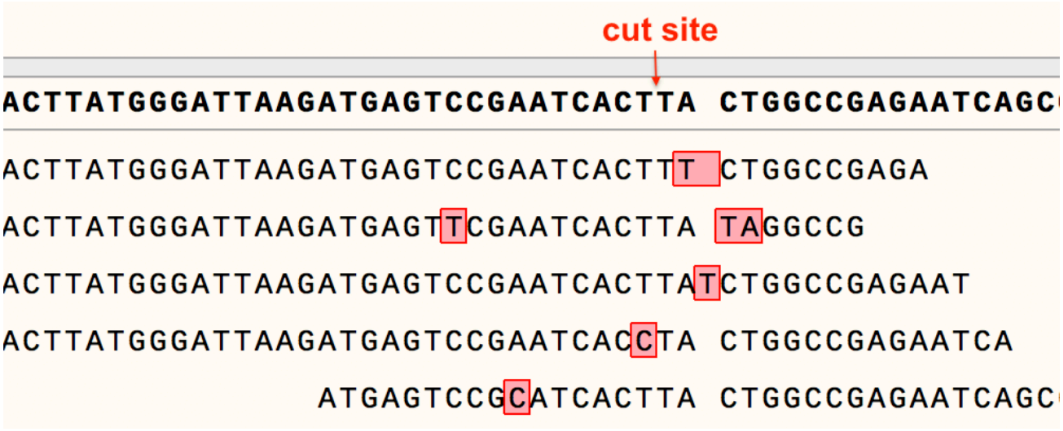


Figure 6.1.1. Mutations induced by CRISPR/Cas9 genome editing in Clogmia. (A) High resolution melting analysis of *Cal-opa* target sequence. (B) Sanger sequencing of *Cal-opa* target site from injected individuals. Red square indicates mismatches induced by CRISPR/Cas9 genome editing.

6.1.2. Gene Expression Studies in Clogmia embryos

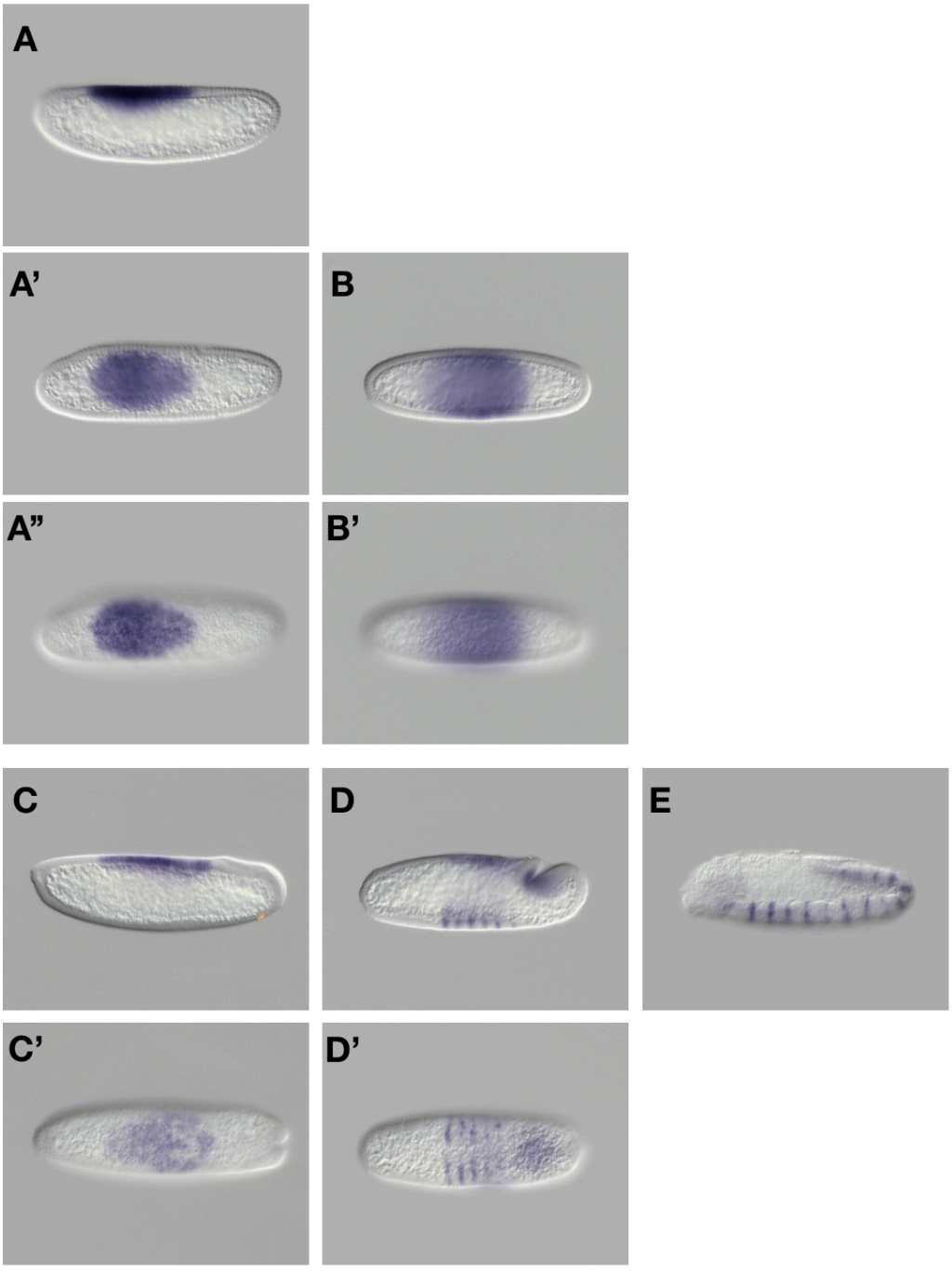


Figure 6.1.2. Expression of Clogmia *engrailed* (*Cal-en*). Early dorsal expression of *engrailed* has not been reported in other flies. Anterior is left. (A,B) Cellular blastoderm embryo, lateral view. (A',A'',B') Cellular blastoderm embryo, dorsal view. (C,D) Early gastrulating embryo, lateral view. (C',D') Early gastrulating embryo, dorsal view. (E) Germband extending embryo, lateral view.

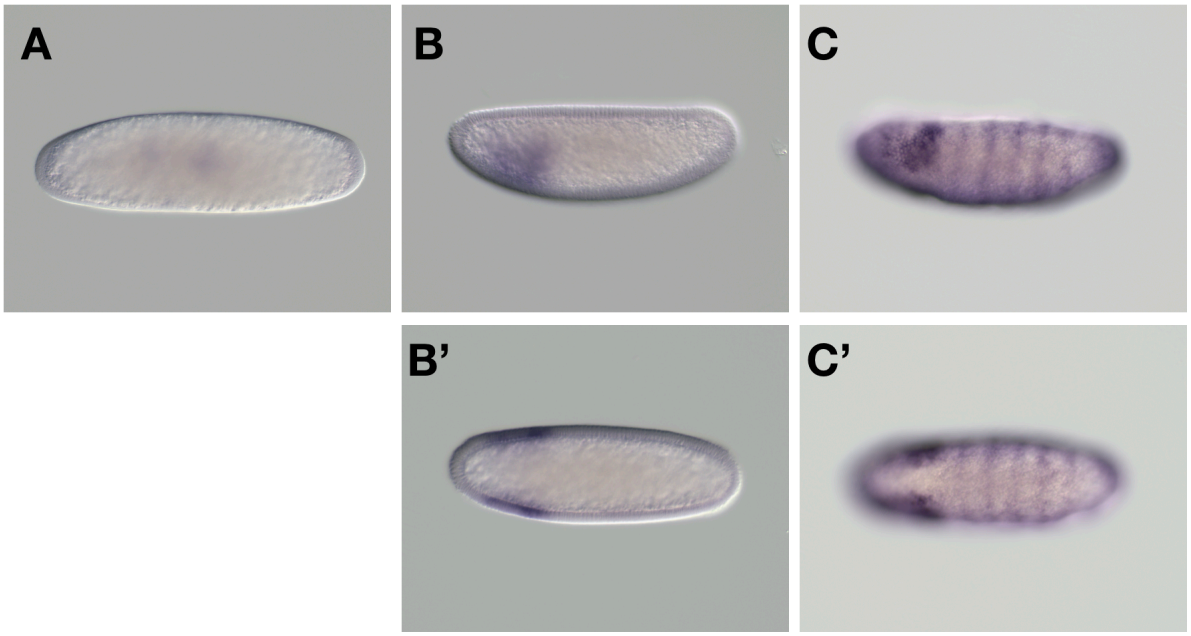


Figure 6.1.3. Expression of Clogmia *brother of odd with entrails limited* (*Cal-bowl*). Anterior is left. (A) Preblastoderm embryo (B) Cellular blastoderm embryo, lateral view (B') Cellular blastoderm embryo, dorsal view. (C) Early gastrulating embryo, lateral view. (C') Early gastrulating embryo, dorsal view.

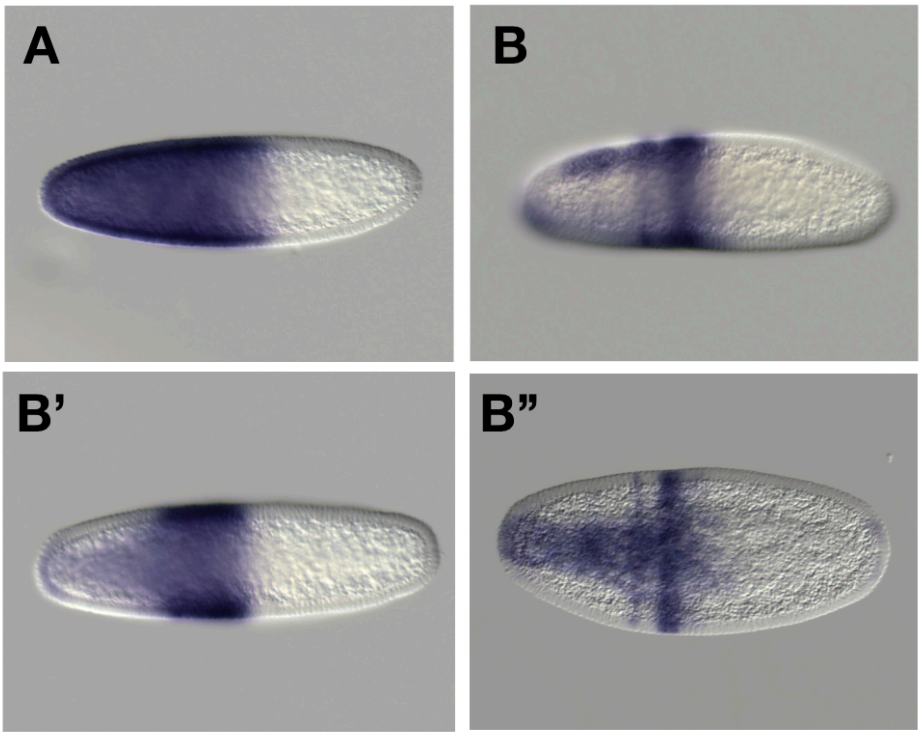


Figure 6.1.4. Expression of Clogmia *hunchback* (*Cal-hb*). (A) Syncytial blastoderm embryo (B) Cellular blastoderm embryo, lateral view. (B') Cellular blastoderm embryo, ventral view. (B'') Cellular blastoderm embryo, dorsal view

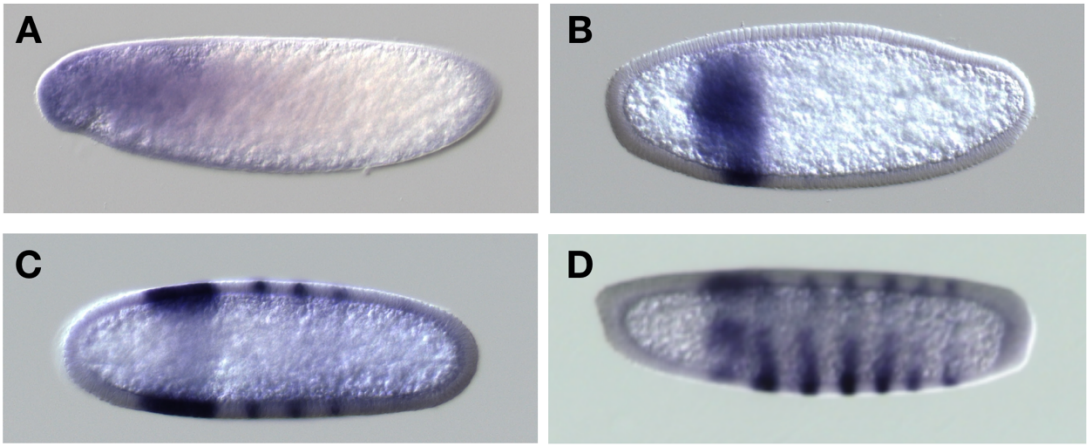


Figure 6.1.5. Expression of Clogmia *sloppy-paired* (*Cal-slp*). (A) Preblastoderm embryo (B) Early cellular blastoderm embryo. (C) Cellular blastoderm embryo (D) Late cellular blastoderm embryo.

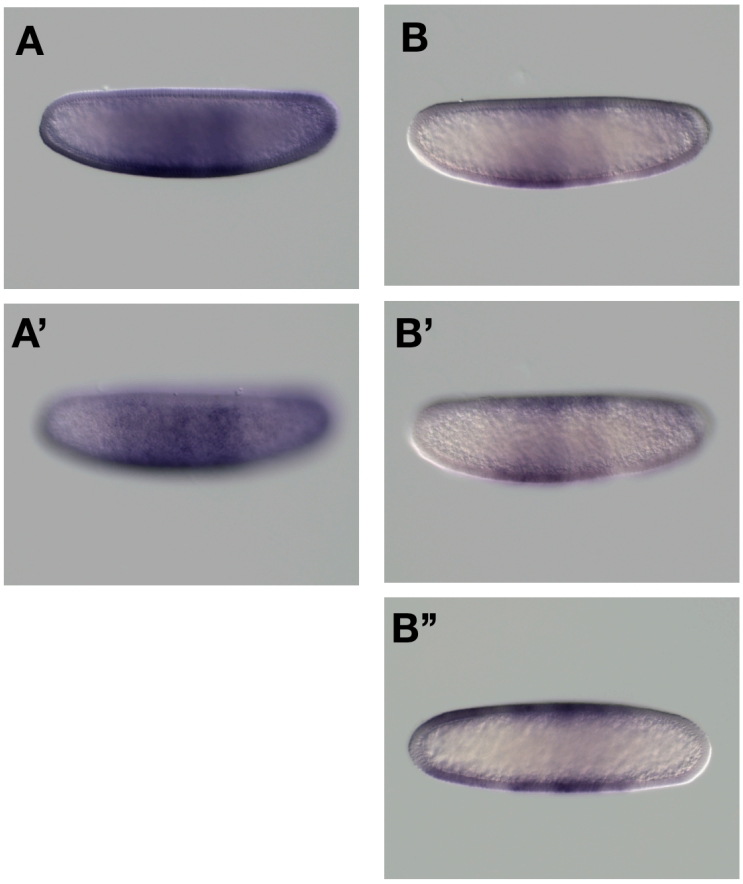


Figure 6.1.6. Expression of Clogmia *groucho* (*Cal-gro*). (A-B') Cellular blastoderm embryo, lateral view. (B'') Cellular blastoderm embryo (same embryo with B), dorsal view. *Cal-gro* is expressed similarly with *Cal-opa*^{Zyg} at this stage, in contrast to *Dme-gro* which is expressed ubiquitously.

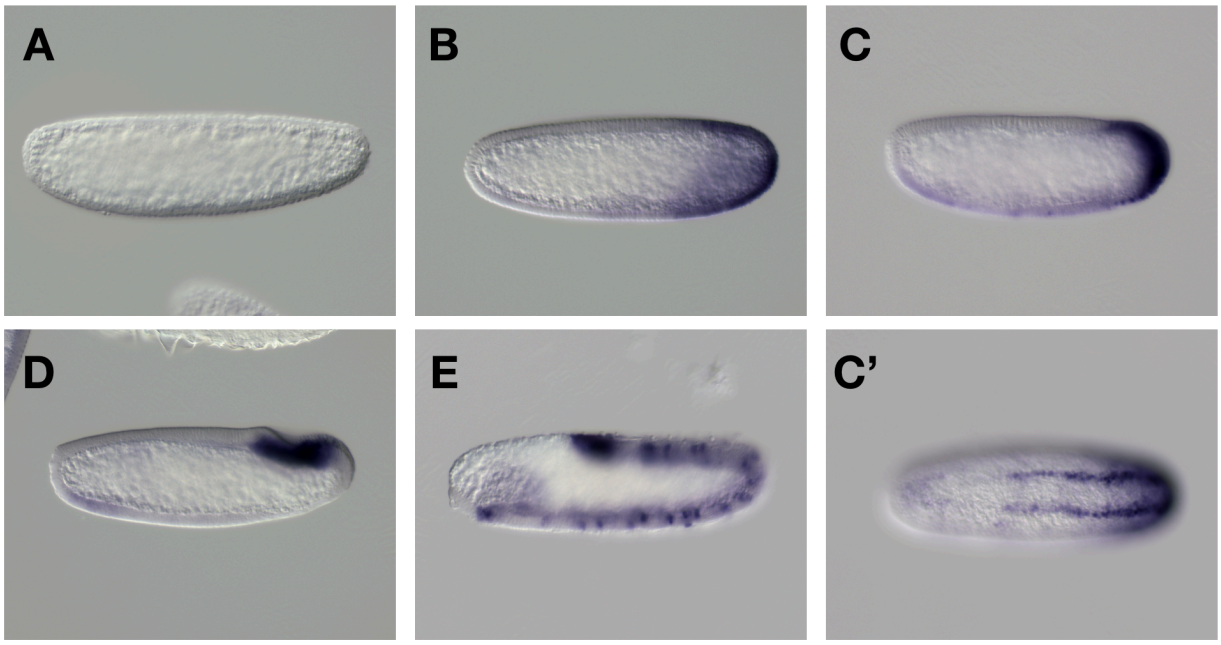


Figure 6.1.7. Expression of an orphan Clogmia transcript (*Cal-unknown*). Differential expression analysis of Clogmia embryo (Figure 2.1.1) suggested that this transcript is enriched in posterior of preblastoderm embryo, but I did not find such pattern. No homologs or conserved domains were found from NCBI protein database (A) Preblastoderm embryo. (B-C) Cellular blastoderm embryo, lateral view. (C’) Cellular blastoderm embryo, ventral view. (D) Early gastrulating embryo, lateral view. (E) Germband extended embryo, lateral view. Anterior is left.

6.1.3. Description 1st instar larval cuticle of Clogmia

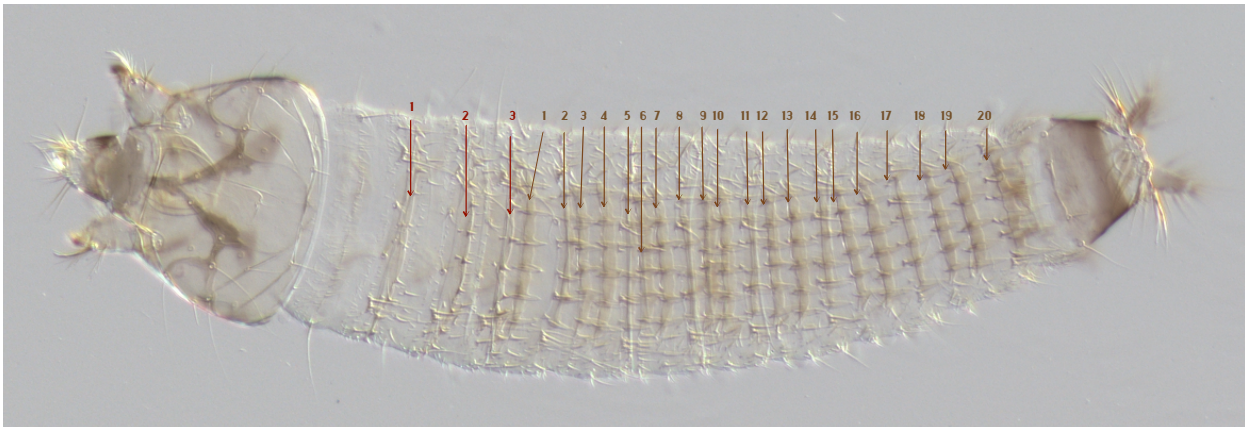


Figure 6.1.8. 1st instar Clogmia larval cuticle in dorsal view. 1-3 in red: thoracic dorsal denticles. 1-20 in brown: abdominal dorsal denticles.

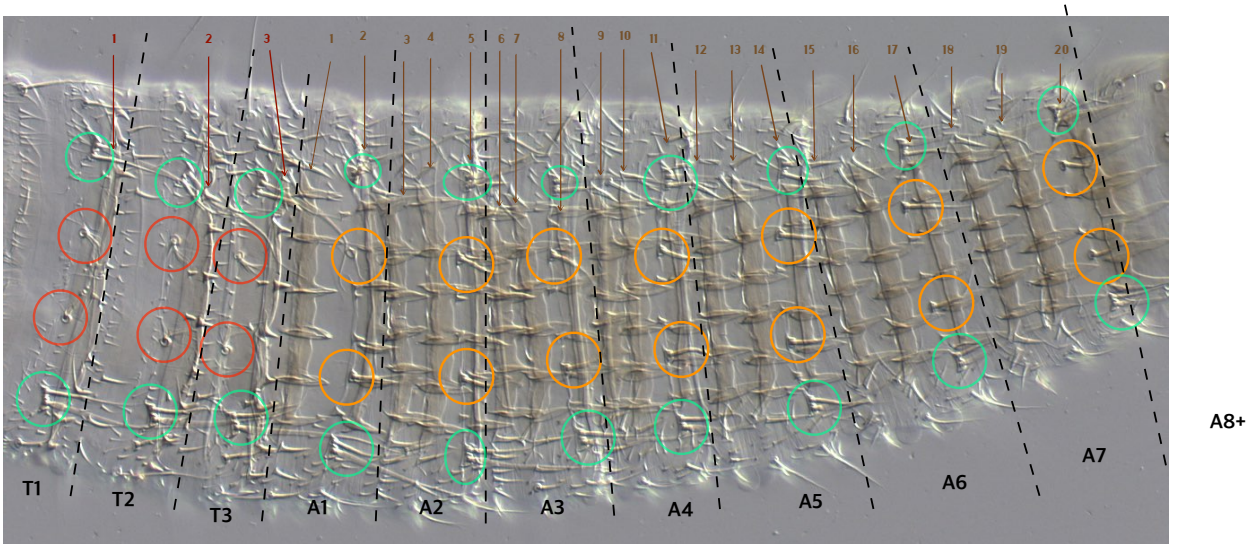


Figure 6.1.9. Dorsal landmark structures in 1st instar larval cuticle of Clogmia. T1-T3 contains one dorsal denticle each. A1 contains two dorsal denticles. A2-A7 contain three dorsal denticles each. Red circle: thoracic segment specific dorsal hair (brush tip-like shape). Orange circle: abdominal segment specific dorsal hair (single-point tip shape). Green circle: dorsolateral hairs in each segment.

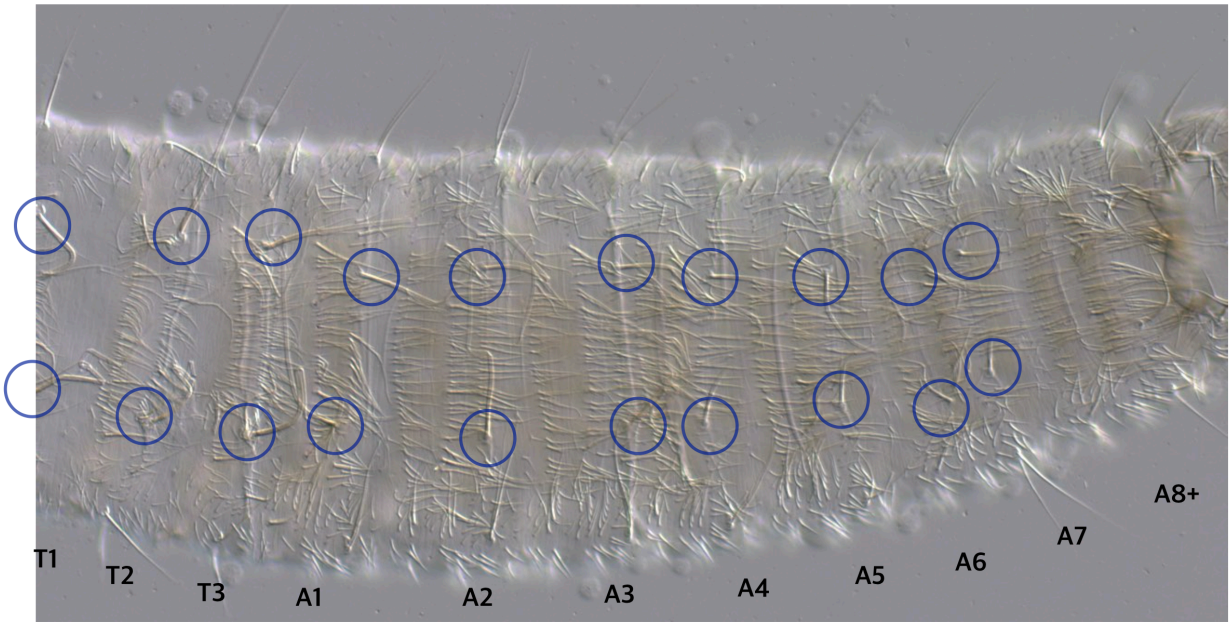


Figure 6.1.10. Ventral landmark structures in 1st instar larval cuticle of Clogmia. Each segment contains two very long ventral hairs (marked in blue circles).

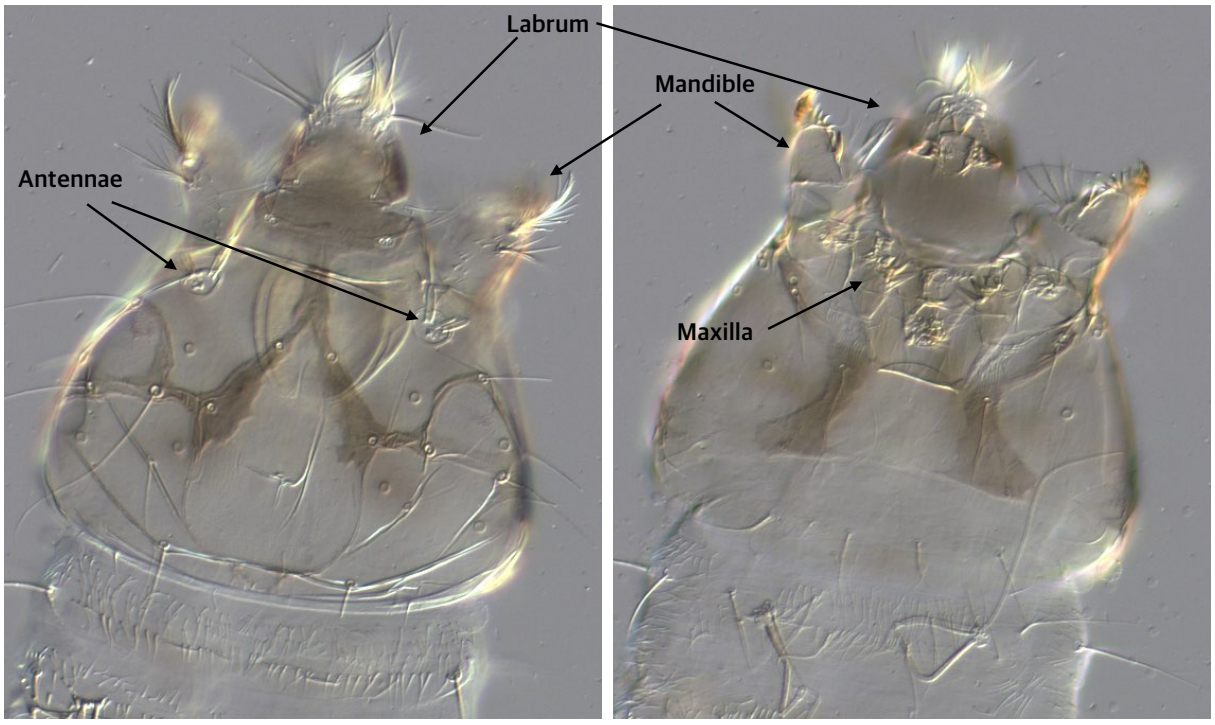


Figure 6.1.11. Head of 1st instar larval cuticle of Clogmia. Left: dorsal view. Right: ventral view.

6.2. *Chironomus riparius*

6.2.1. Preliminary CRISPR/Cas9 experiments in Chironomus

To obtain a proof-of-concept that CRISPR/Cas9 genome editing is possible in Chironomus, I synthesized gRNA targeting *Cri-tll* (see below for gRNA sequence and **APPENDIX 6.4.** for step-by-step instructions) and injected the gRNA (1ug/ul) together with Cas9 protein (1ug/ul – PNA Bio CP0120) into 175 Chironomus embryos of 4-8 pole cell stage. The gRNA was designed to cut the *Cri-tll* gene sequence encoding 45th and 46th residue of Cri-Tll. Most injected embryos (158) died but all surviving embryos (17) developed as double heads, suggesting that mutations had disrupted the *Cri-tll* open reading frame with high efficiency. As a control, I injected 166 embryos with *Cri-tll* gRNA without Cas9 protein. In this batch, 82 embryos were indistinguishable from WT and 84 embryos died early, suggesting that the double heads were induced by CRISPR/Cas9-mediated genome editing.

-*Cri-tll* gRNA forward primer (gene specific sequence underlined):

GAAATTAATACGACTCACTATAG<u>GTATGCCGTGATCATAGCTC</u>GTTTAAGAGCTATG

CTGGAA

-gRNA reverse primer:

AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAAA

CTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC

6.2.2. Troubleshooting RNAi effect induced by overexpression of *panish* mRNAs in Chironomus

While testing the effect of ectopic posterior *panish* mRNA expression, I found that over-expression of *panish* mRNA can induce double abdomens with high efficiency. This counter-intuitive phenomenon occurred only when capped *panish* mRNA had been synthesized with an SP6 protomer from the linearized psp35T plasmid containing the *panish* open reading frame (Clone #1026). *panish* mRNA generated in this way induced double abdomens depending on the concentration of the injected mRNA and depending on the developmental stage at which the mRNA was injected. Specifically, I found that when the concentration of *panish* mRNA exceeds ~4μg/μl, most injected embryos develop as double abdomens (**Figure 6.2.1A**). Furthermore, the sensitive phase for efficient double abdomen formation in response to *panish* overexpression ended with the beginning of the two-pole-cell stage (**Figure 6.2.1B**), suggesting that this phenomenon is restricted to very early stages of embryogenesis (before ~2hr). I also found that the site of mRNA injection along the AP axis affected the double abdomen rate. While almost all posteriorly injected embryos produced double abdomens (48/49), only about half of the anteriorly injected embryos developed as double abdomen (24/47). This observation suggests that anteriorly injected mRNA can partly rescue wild-type polarity.

Also, since injected *panish* mRNA has different UTR sequences compared to the endogenous mRNA, it is possible that this phenomenon is mediated by UTR of *panish* transcript. To further test this idea, I performed RNA *in situ* hybridization with *panish* mRNA (6.3μg/μl) posteriorly injected embryos that had been fixed 3 hours after the injection. For this experiment, I used a probe that recognizes both endogenous *panish* mRNA and the injected

panish mRNA. In posteriorly injected embryos, I detected *panish* mRNA only in the posterior half, and endogenous anterior *panish* expression was absent (**Figure 6.2.1C**). This result suggests that endogenous *panish* mRNA was specifically degraded, most likely via its UTR sequences.

The above findings suggest that *panish* might initiate an auto-regulatory negative feedback loop which is concentration and developmental stage dependent. Another hypothesis is that this phenomenon occurs due to the generation of small amount of dsRNA during mRNA synthesis, leading to an RNAi effect when the reagent is injected above a certain concentration threshold. For example, dsRNA could be generated with low efficiency if the *panish* open reading frame contain a sequence similar to SP6 promoter in 3'->5' direction. Alternatively, dsRNA might be generated from non-linearized circular plasmids. To avoid these issues, I generated *panish* mRNA using a PCR-amplified template containing T7 promoter sequences. *panish* mRNA synthesized this way did not induce double abdomen when injected itself and effectively rescued *panish* RNAi embryos (see **Figure 2.2.7B**), suggesting that the double abdomen phenomenon was a technical artifact rather than a biological effect. Specific degradation of endogenous *panish* mRNA observed by RNA in situ hybridization might also be a technical artifact, because injected mRNA is so much higher in concentration and therefore could not be fully degraded or takes much longer time to get degraded (**Figure 2.2.1C**). It is still uncertain whether the dsRNA was generated from putative SP6 promoter in the *panish* open reading frame, or from the circular plasmid. Therefore, caution should be taken when planning *panish* mRNA overexpression experiments, as unintended RNAi effects can occur.

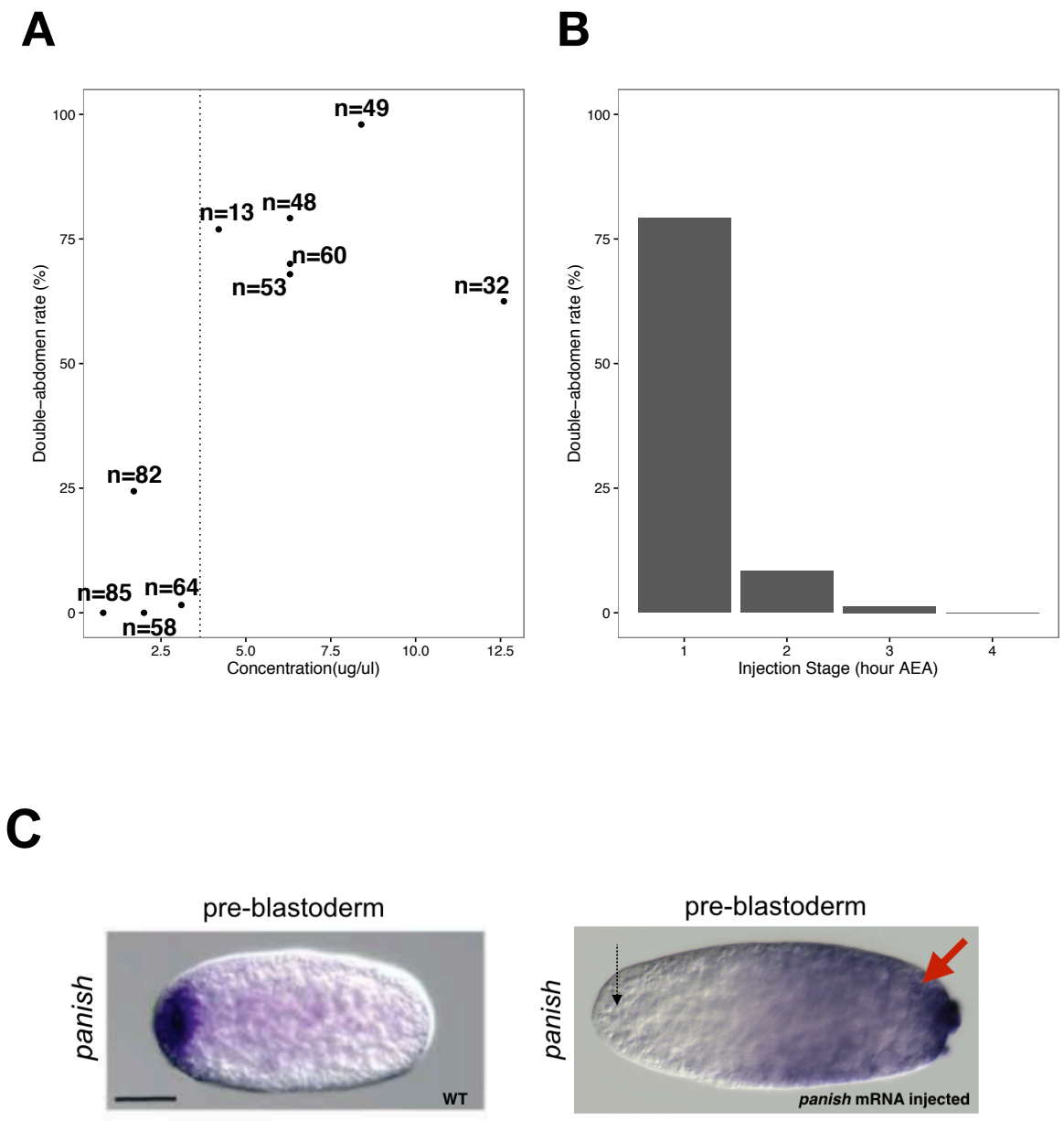


Figure 6.2.1. Overexpression of *panish* mRNA induces double abdomen (when the mRNA is synthesized from a SP6 promoter in a linearized plasmid containing *panish* open reading frame) **(A)** Occurrence of double abdomens depending on the concentration of injected *panish* mRNA. Each dot represents individual experiments. The concentration threshold for the induction of double abdomens is indicated by a dotted line. All embryos were injected at 1hr after egg activation (AEA). **(B)** Occurrence of double abdomens depending on the developmental stage. *panish* mRNA was injected at a concentration of 6µg/µl. **(C)** RNA in situ hybridization of *panish* in a wild-type embryo (left) and a representative example (22/29) of an embryo that had been injected with *panish* mRNA at a concentration of 6µg/µl at 1hr AEA. Anterior is left. Red arrow: injected position. Note the absence of endogenous *panish* expression (black dotted arrow).

6.2.3. *Cri-opa* mRNA cannot induce an ectopic head in Chironomus

I tested whether *Cri-opa* mRNA can induce head development when injected into the posterior pole of 2-pole-cell stage Chironomus embryos. Injection of full-length *Cri-opa* mRNA resulted in failed germband retraction (25/79) or did not perturb development (54/79). Double heads were not observed. The N-terminal region of Cri-Opa contains three copies of a short sequence motif with partial homology to the N-terminal region of Cal-Opa^{Mat} (**Figure 6.2.2.**). I synthesized three *Cri-opa* mRNA variants that encode N-terminal truncated Cri-Opa proteins, resembling Cal-Opa^{Mat} (*Cri-opa* N-del 1 mRNA, *Cri-opa* N-del 2 mRNA, *Cri-opa* N-del 3 mRNA). Posterior injection of these mRNAs resulted in WT larvae (n= 70, 79, 80, respectively). Taken together, my results suggest that *Cri-opa* mRNA cannot induce head development in Chironomus when expressed posteriorly in 2-pole-cell stage embryos.

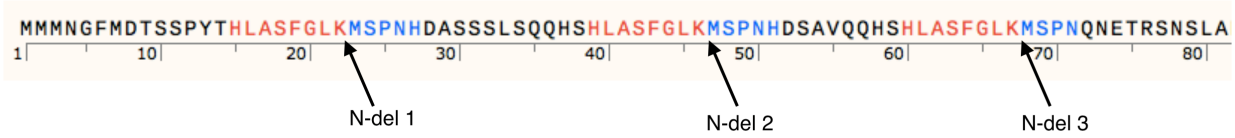


Figure 6.2.2. N-terminal sequence of Cri-Opa. Blue sequences are homologous to N-terminal sequences of Cal-Opa^{Mat}. Arrows indicate the start codons encoded by *Cri-opa* N-del 1 mRNA, *Cri-opa* N-del 2 mRNA, *Cri-opa* N-del 3 mRNA, respectively. See Figure 2.1.2. for full protein sequence.

6.2.4. Expression of *Cri-otd* in *Cri-tll* RNAi embryos

Figure 6.2.3. RNA in situ hybridization of Cri-*otd* in *Cri-tll* RNAi embryos. Anterior is left. Dorsal view. Note that the ectopic posterior expression domain of *Cri-otd* in the *Cri-tll* RNAi embryo is smaller than the endogenous *Cri-otd* domain. This suggests that posteriorly expressed genes, such as Chironomus *caudal* (*Cri-cad*), might be repressing ectopic *Cri-otd* expression.

6.2.5. Gene Expression Studies in Chironomus embryos

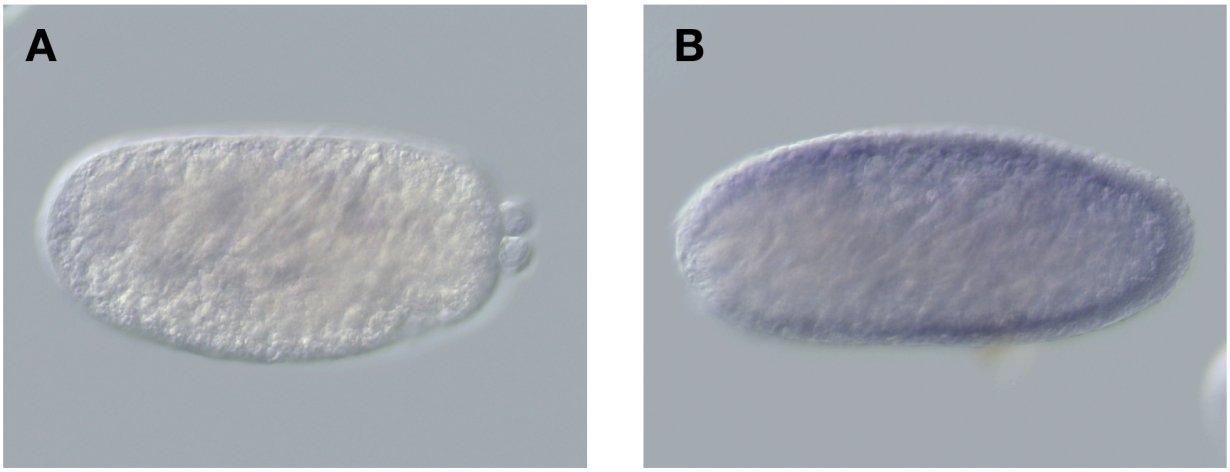


Figure 6.2.4. RNA in situ hybridization of Chironomus *puffyeye* (*Cri-puf*). Differential expression analysis from (Klomp et al. 2015) suggested that *Cri-puf* is posteriorly enriched, but I did not find such pattern in preblastoderm embryo. (A) preblastoderm embryo at two-pole cell stage. (B) Syncytial blastoderm embryo. Anterior is left.

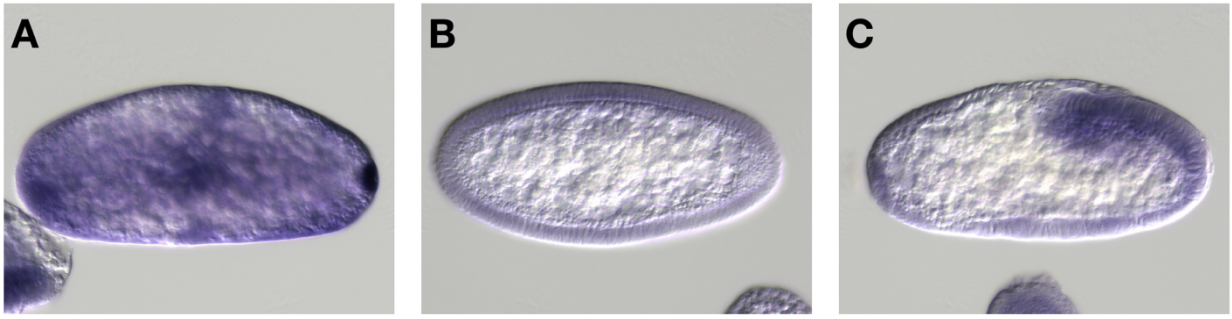


Figure 6.2.5. RNA in situ hybridization of Chironomus *adenomatous polyposis coli* (*Cri-APC*). Differential expression analysis from (Klomp et al. 2015) suggested that *Cri-APC* is posteriorly enriched, and I found that *Cri-APC* transcript is localized at the posterior pole of preblastoderm embryo. (A) Preblastoderm embryo. (B) Cellular blastoderm embryo. (C) Gastrulating embryo. Anterior is left.

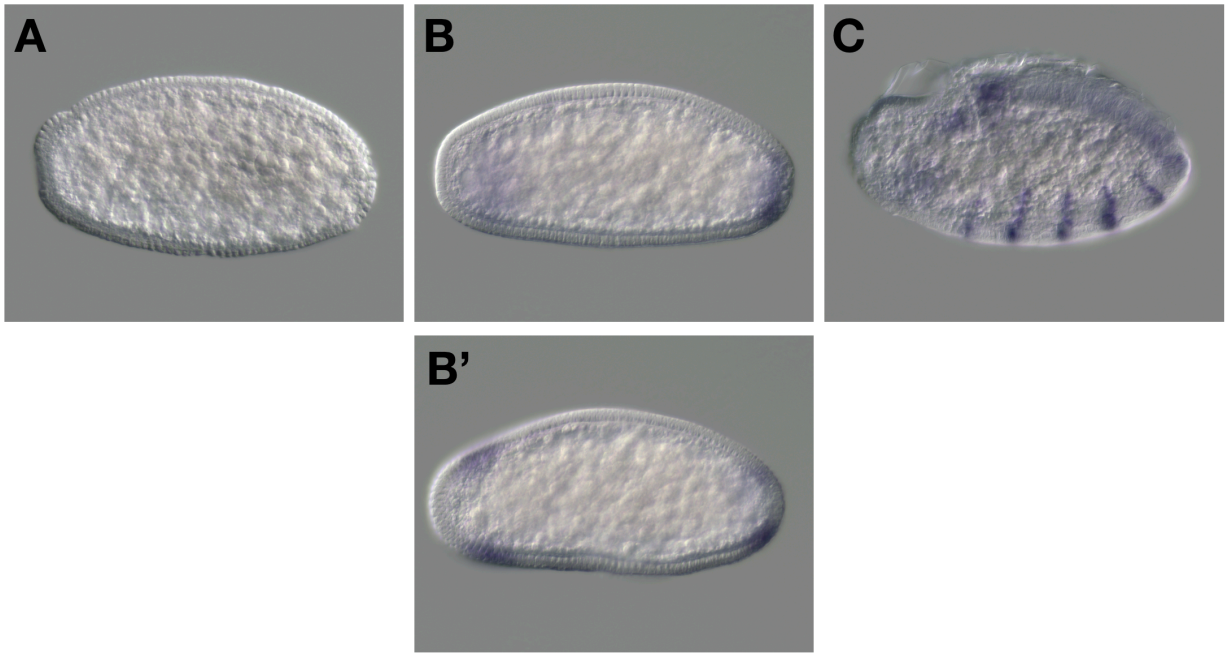


Figure 6.2.6. RNA in situ hybridization of Chironomus *wingless* (*Cri-wg*). (A) Syncytial blastoderm embryo. Lateral view. (B) Cellular blastoderm embryo. Lateral view. (B') Cellular blastoderm embryo. Dorsal view. (C) Germband extended embryo. Lateral view. Anterior is left.

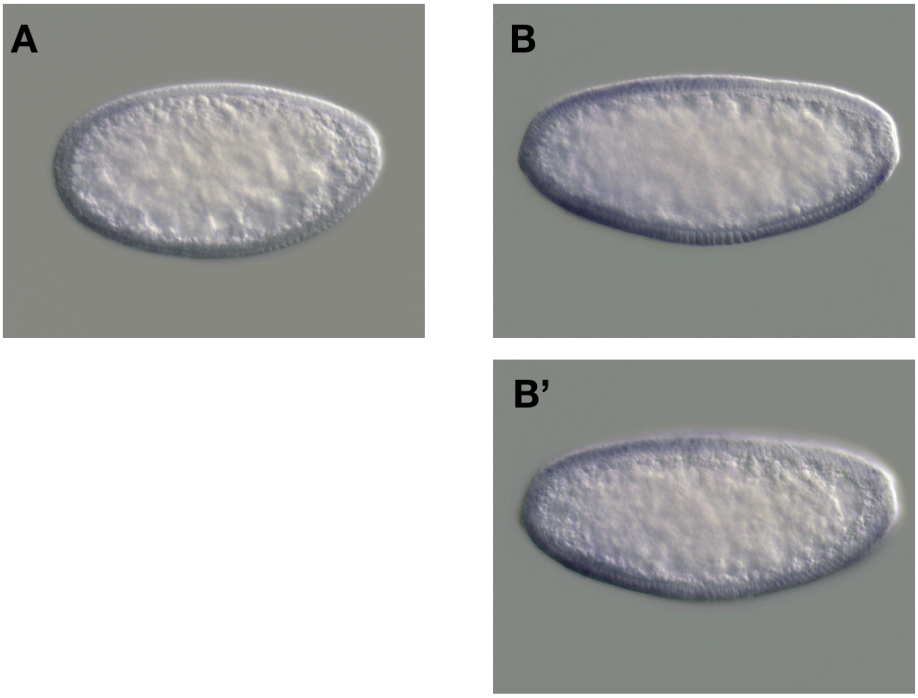


Figure 6.2.7. RNA in situ hybridization of Chironomus *capicua* (*Cri-cic*). Note that *Cri-cic* appears to be slightly enriched in anterior region of blastoderm embryos. (A) Syncytial blastoderm embryo. Lateral view. (B-B') Cellular blastoderm embryo. Lateral views. Anterior is left.

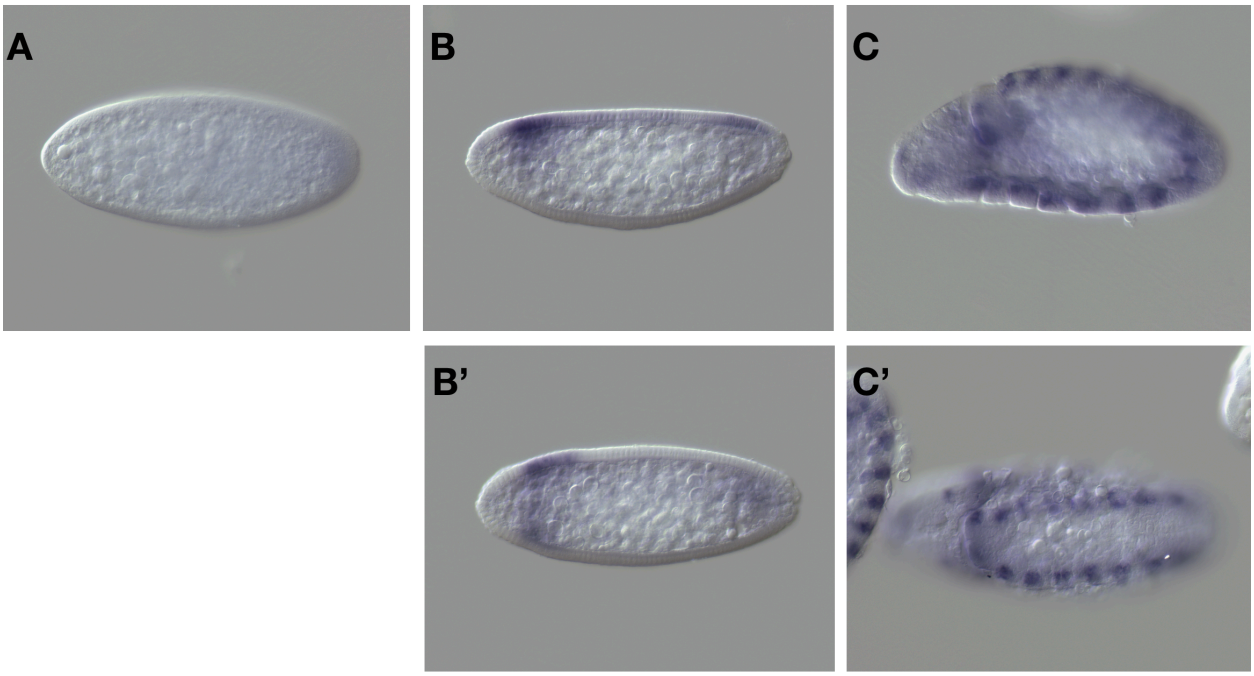


Figure 6.2.8. RNA in situ hybridization of Chironomus *dorsocross* (*Cri-dor*). (A) Preblastoderm embryo. (B) Cellular blastoderm embryo, lateral view. (B') Cellular blastoderm embryo. Dorsal view. (C) Germband extended embryo, lateral view. (C') Germband extended embryo, dorsal view. Anterior is left.

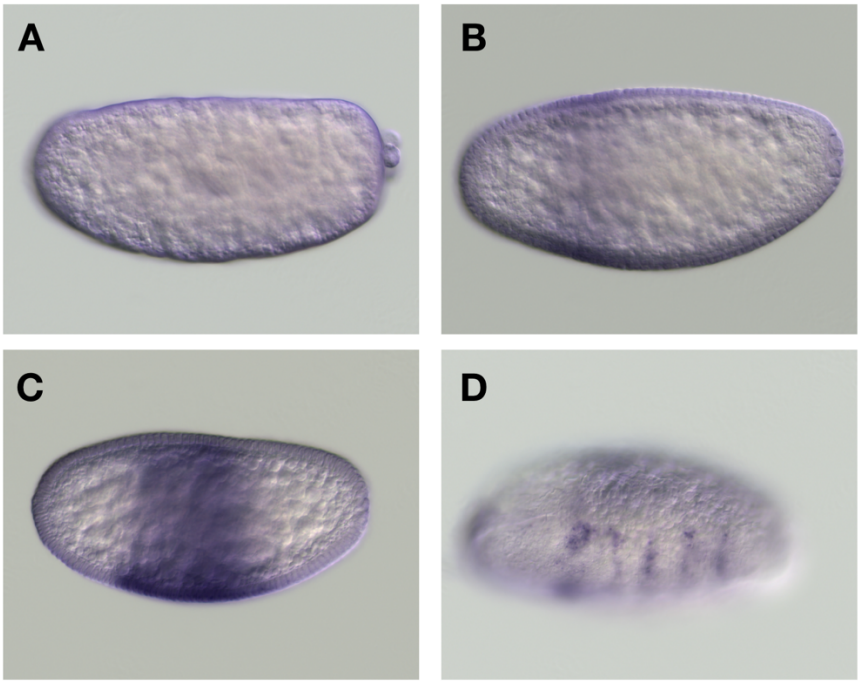


Figure 6.2.9. RNA in situ hybridization of Chironomus *odd-paired* (*Cri-opa*). (A) Preblastoderm embryo. (B) Syncytial blastoderm embryo, lateral view. (C) Cellular blastoderm embryo. Lateral view. (D) Germband extended embryo, ventrolateral view. Anterior is left.

6.2.6. Description of cuticular head structures of the 1st instar Chironomus larva

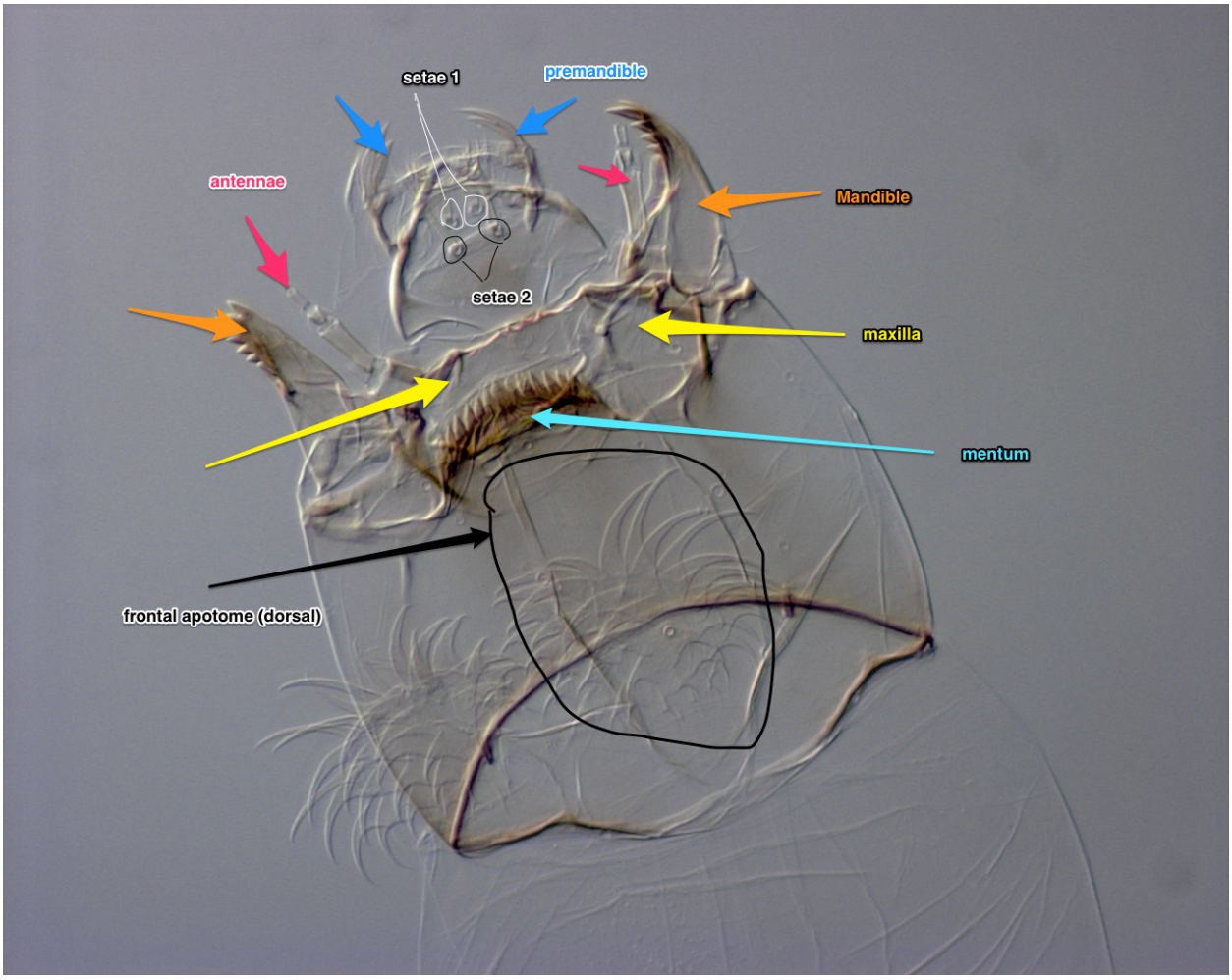


Figure 6.2.10. Description of head structures of Chironomus 1st instar larva. Ventral view. Black circle indicates frontal apotome at dorsal side.

6.3. Mosquitoes and crane flies

6.3.1. Gene expression studies in mosquito embryos

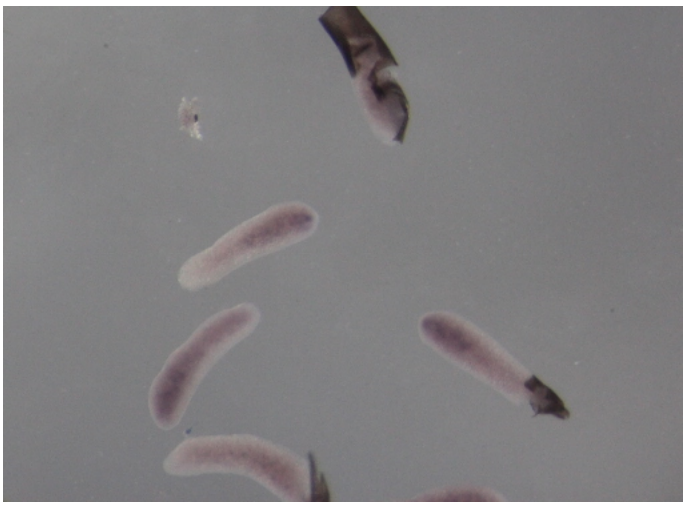


Figure 6.3.1. RNA in situ hybridization of *Anopheles gambiae cucoid* (*Aga-cucoid*).

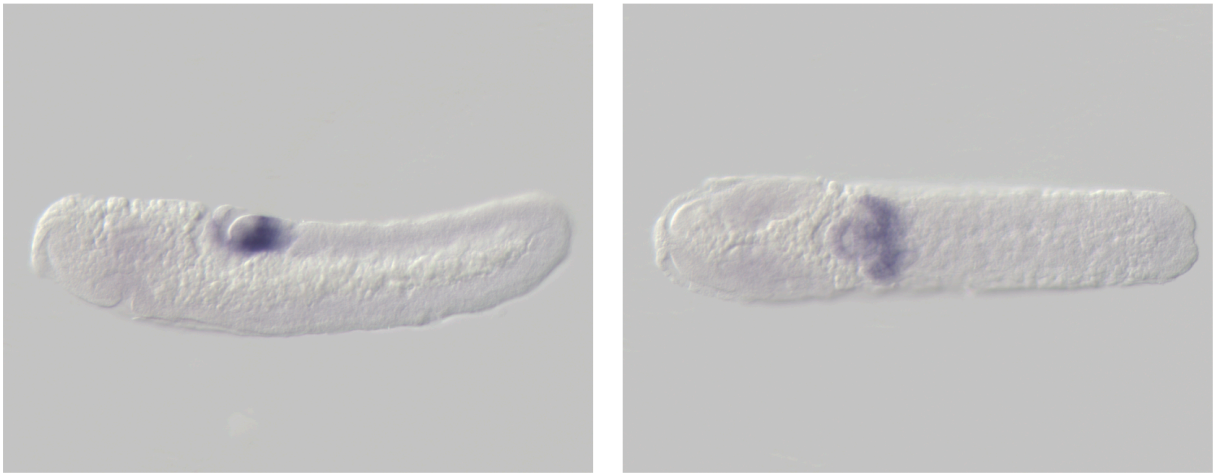


Figure 6.3.2. RNA in situ hybridization of *Anopheles gambiae caudal* (*Aga-cad*) in extended germband embryos. (left) Lateral view; (right) dorsal view. Anterior is left.

6.3.2. Expression of *Cqu-nos* in Culex embryo

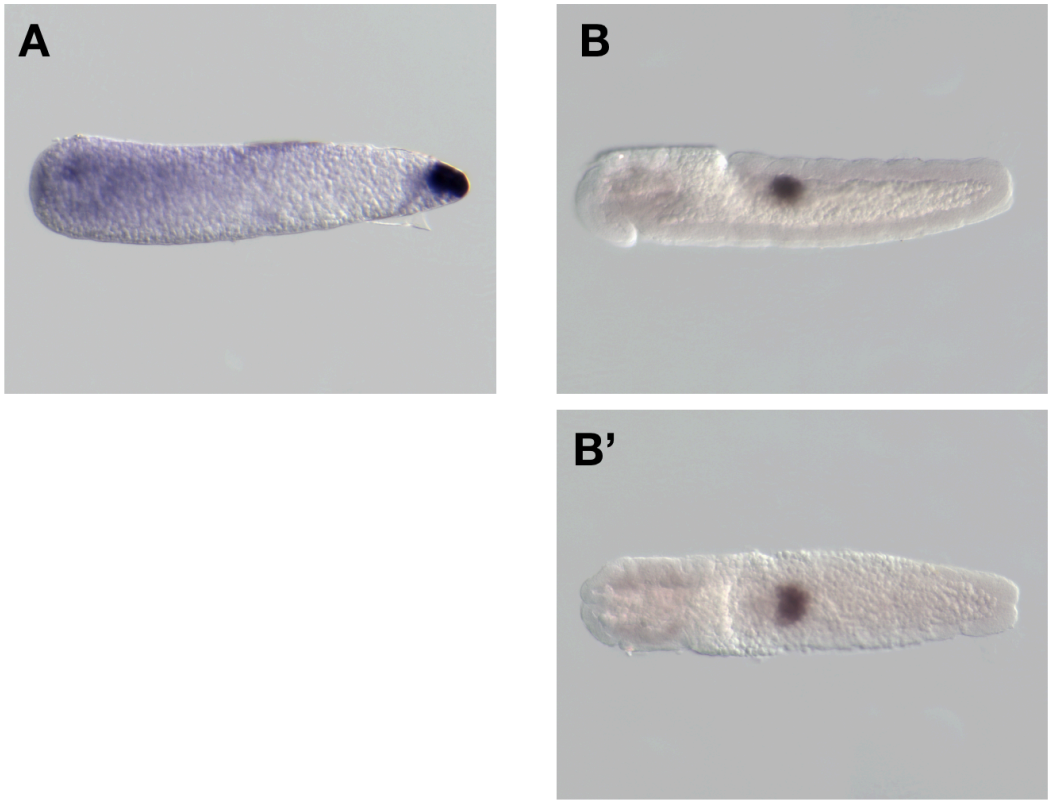


Figure 6.3.3. RNA in situ hybridization of Culex *nanos* (*Cqu-nos*). (A) Preblastoderm embryo. (B) Germband extended embryo. Lateral view. (B') Extended germband embryo. Dorsal view. Anterior is left

6.3.3. Expression of *Nsu-cad* in Nephrotoma embryo

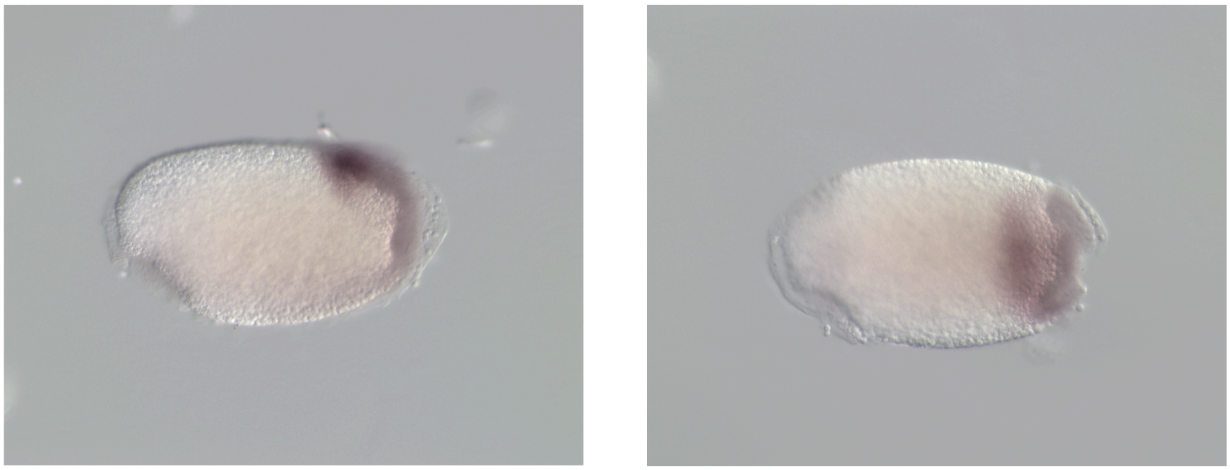


Figure 6.3.4. RNA in situ hybridization of Nephrotoma *caudal* (*Nsu-cad*) in a gastrulating embryo in lateral (left) and dorsal view (right). Anterior is left.

6.4. Detailed protocols

6.4.1. Microinjection of *Chironomus riparius* embryos[5]

A. Needle pulling

- Use borosilicate glass capillary (World Precision Instrument: #1B100F-4) and Sutter Instrument Inc. Model P-87 Flaming/Browning Micropipette puller (located in CLSC 9th floor). First time users should be trained by lab members before use.
- Note that P-87 is a filament-based puller; lifespan of the filament is limited, and it can break (twice during my Ph.D). This will affect the shape of the needle and you should reprogram the parameters accordingly. For this, it is crucial to keep good needles somewhere safe so that you can refer to their shape when reprogramming. For reprogramming, refer to P-87 puller manual and instruction notes from IGTRCN (see labserver -> Yoseop Yoon -> Methods).

1. Use Brososilicate glass capillary tubes with a filament inside (World Precision Instrument: #1B100F-4)
2. Pull needles with program 9 (Heat: 615; Pull: 87; Velocity: 50; Time: 120)

B. Needle beveling

- For Chironomus, preparing perfectly beveled needles is one of the most important steps for microinjection. Unless the tip is laser-sharp, most injected embryos will die, and you

[5] This protocol is adapted and modified from the previous protocols written by Derek Athy, Ab. Matteen Rafiqi and Natasha Bloch.

will waste a lot of time. Also, be sure to prepare more-than-enough beveled needles, prior to injection, since needles can break or clog during injections.

- Beveling can be divided into two steps. First, use Narishige EG-44 capillary grinder in our lab to grind the tip of the needle. Second, immediately pushout water inside the needle under oil using Narishige IM-300 microinjector. This will prevent dissolved water substances from clogging the tip of the needle during evaporation.

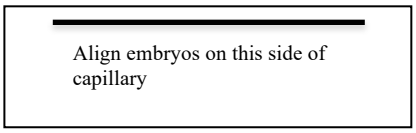
1. Use the following settings in Narishige EG-44 capillary grinder:
    - Speed: 8.0 / Degree of grinding: 25
2. Fill the syringe with Millipore water and set it up in a way that it will drop water on the diamond stone with a constant rate.
3. Before turning the grinder on, it is important to focus the scope on the stone first.
4. Put a needle on the holder and make sure the stone is very wet before starting to lower it.
5. Start lowering the needle and as soon as it touches the stone, give 1/8 of a turn to the small wheel. You will see a steady flow of water into the needle. If the water flow into the needle is too fast, it means that the tip opening is too large. In this case, discard the needle.
6. After 5-10 seconds, immediately place the needle into a needle holder connected to Narishige IM-300 microinjector. Place the tip of the needle under a drop of halocarbon 27-oil (Sigma H8773) in the microscope glass slide. If the tip looks dull, discard the needle. Next, quickly press Clear (CLR) button several times to push out water in the needle. If successful, you will see bubbles constantly coming out from the needle tip. If not, discard the needle.

7. While CLR function is still on (important), raise the level of the needle so that tip escapes the drop of oil. If you stop CLR before raising the needle, oil will flow into the needle. Store the needle for future use.

C. Alignment of embryos

1. Collect the egg packages in a glass vial (that has not been in contact with any chemicals other than bleach) with water. Select them under the scope according to the desired stage for injection. For knocking down maternal transcripts, you should inject embryos before two-pole-cell stage (roughly between 1hr 30min – 2hrs after egg laying).
2. Under the dissecting microscope, remove as much water from the dish as possible using a plastic pipette, and then add 1ml of 3% bleach (concentrated CLOROX - Sodium Hypochlorite 8.25%).
3. Shake gently until the jelly around egg packages begins to dissolve and some embryos have come loose from the egg package. You should constantly watch embryos through stereomicroscope during this process. Once 10-20 embryos start to break free from the egg package, add water to fill the dish about ¾ full to further dilute the bleach. When adding water, use a squirt bottle to generate pressure, which will further dissemble the egg package. Do not wait until embryos are completely dissembled – it's too late.
4. Wash 6x with water (I spin the embryos in a clockwise fashion so that the embryos congregate at the bottom of the dish making it easier to remove H2O during each wash. Using a plastic transfer pipette (is easier and faster for removing water from dish).
5. Remove most of the water in the dish, leaving enough water to keep the embryos wet to prevent drying out while aligning the embryos.

6. Clean a glass slide with 70% EtOH and put it under the scope (use a dissecting microscope with a substage mirror. Place a 0.25mm ∅ capillary (Hilgenberg) in the middle of the slide, parallel to the long side of the slide as shown below.



7. Apply a drop of water using a brush to the left side of the capillary to attach the capillary to the slide. You should be able to see H2O flowing through the capillary to the right end. It is better to add water in smaller increments if necessary than adding too much water to the slide. You can also take a moist brush and glide it along the capillary.
8. With a clean brush take embryos from the glass vial and drop them close to the capillary. Dry the excess water in the brush against a Kimwipe and align the embryos against the capillary. Repeat this step until the entire capillary is covered[6]. There should be enough water on the slide so that the embryos are wet. However, too much water will make embryos move along the capillary. You want to be able to guide the embryos along the capillary using the water tension; avoid touching the embryos with a dry brush.

9. Once the embryos are aligned, allow the embryos to dry for ~1 minute. Time for this step is not always the same as it depends on how much water was transferred with the embryos.

[6] This step has to be done fast enough so that the first embryos to be aligned do not dry too much.

10. Cover the line of embryos with ~20-35 µl of 1:1 halocarbon oil (one part 27-oil, Sigma H8773; 1 part 700-oil Sigma H8898). Volume varies due to viscosity and ability to pipette oil out smoothly; just need to make sure all embryos are covered completely by oil. You should avoid excess oil for covering embryos to be used for fixation and/or cuticle preparations.

D. Embryo injection

1. Turn on the Narishige IM-300 microinjector and open the valve of the nitrogen tank. You can check the pressure at the digital display of the microinjector. The pressure should be around 60-90 psi.
2. Backfill an injection needle with 1.6µl of reagent using Eppendorf microloader tips (E5242956003). Spin down the reagent at >12,000 rpm at 4C for at least 10 minutes prior to backfilling, as highly concentrated reagent such as dsRNA may clog injection needle. Always keep reagents on ice.
3. Locate the needle tip near embryos under the microscope. You should focus on embryos first, and then move around the needle to bring the tip to the right location.
4. Press the foot pedal to remove air from the tip of the needle. The default mode is a constant pressure mode ("→" on the display). Press the pedal several times until the air bubble comes out constantly from the needle. If it does not work, raise the level of the needle tip to move it outside the oil and then move it back into the oil, and repeat pressing the pedal. If you still can't get the reagent out, discard the needle and load a new needle.

5. Wait until all the air bubbles are gone. When the reagents eventually come out, you will see a constant increase in the size of a single bubble. This is your reagents. Do not inject embryo before you remove all the air bubbles.
6. Switch the injection mode to Vent (“V” on the display). In this mode, pressing the foot pedal will generate in a single burst of pressure. Adjust the settings of the micromanipulator to produce an appropriately sized bubble upon injection action. The best way for setting the injection volume is to adjust the injection time. Start with 0.05 sec and gradually increase the time if the injection bubble is too small.
7. Inject embryos. Do not use the joystick to move the needle. Instead move the stage to inject. When the embryo approaches the needle by slowly moving the stage, you will see the membrane invaginate and snap back slightly once the needle has penetrated into the embryo. Upon pressing the foot pedal, you should see a small pulse/burst into the embryo. If you feel too much tension when penetrating the membrane, discard the needle and use a sharper needle. Again, do not move on with a bad needle. You will waste a lot of time.
8. After you are done injecting the entire slide, place it on top of a wet KimWipe in a Petri dish. Make sure the KimWipe is wet throughout the embryo development.
9. Allow the embryos to develop in the 28C incubator until they have reached desired stage. Add more water to the dish over time if necessary, making sure that the water level does not cover the slide.

E. Post injection processing (cuticle preparation)

1. Wait for 3-4 days until embryos produce cuticle. Many injected embryos will not hatch on their own so the vitelline membrane may have to be removed manually. This can be done using a sharp tungsten needle, taking caution not to sever/damage the cuticle.
2. Using a tungsten needle, transfer the cuticles from the slide into a dish containing 1:4 glycerol:acetic acid (100% acetic acid works also). Separate the cuticles from any oil globules as best as possible
3. Transfer all cuticles into glass dishes with 1:4 glycerol:acetic acid and incubate at room temperature (RT) for overnight.
4. Clean slides with 70% EtOH and a KimWipe to remove dust/debris. Slides need to be as clean as possible for best cuticle preparations.
5. Transfer the cuticles onto a clean slide using a clean 1-2 hair brush. I place a ~2-4µl drop of 1:1 Hoyers Medium/Lactic Acid onto the slide to prevent the cuticles from sticking to the slide and drying out. (I normally place ~10-15 cuticles on one slide but this is personal preference. Also, make sure you transferred all the cuticles from the glass dish; sometimes it is very difficult to locate the cuticles.
6. Carefully spread cuticles evenly and in desired orientation using the brush.
7. Put an additional 8-10µl drop of 1:1 Hoyer's Medium/Lactic Acid onto slide and put a clean coverslip very slowly over the cuticles to minimize bubbles being trapped by coverslip.
8. Place slides in a 65^{o}C oven overnight.

6.4.2 Microinjection of *Clogmia albipunctata* embryos[7]

A. Needle pulling

- Use quartz glass capillary tubes without filament inside (Sutter instrument: Q1007010) and Sutter Instrument Inc. Model P-2000 Laser Micropipette puller (located in the Bezanilla lab at GCIS). First time users should be trained by the Bezanilla lab members before use.
- While P-2000 puller produces more stable result than P-87 puller over long periods of time, be sure to keep good needles somewhere safe so that you can refer to their shape when you need reprogramming (accidents can happen). For reprogramming, refer to P-2000 puller manual and instruction notes from IGTRCN (see labserver -> Yoseop Yoon -> Methods).

1. Use quartz glass capillary tubes without filament inside (Sutter instrument: Q1007010). Using model P-2000 Laser Micropipette puller, pull needles with program 14 (Heat: 645; Fil: 4; Vel: 40; Del: 125; Pul: 130).
2. Below image shows an example of expected needle tip shape.

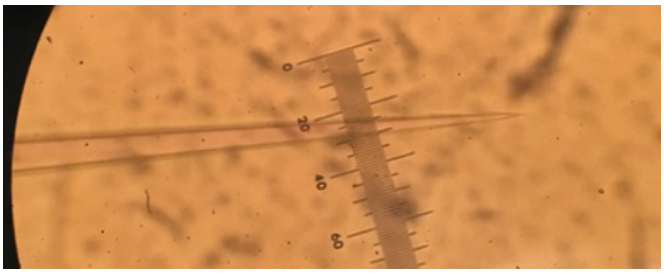


Figure 6.4.1. An expected quartz needle tip shape for Clogmia microinjection.

[7] This protocol can be used to inject embryos of mosquitoes (Aedes, Culex, and Anopheles) and Hermetia.

B. Embryo alignment

1. Prepare filter papers (Fisher Scientific, Cat. No 09-795C), one in 4cm x 2cm pieces, and another in smaller pieces. Put a cover slide (Fisher Scientific, Cat. No 12-648-5C) above these papers, as shown below. Apply a small amount of water to the filter paper.

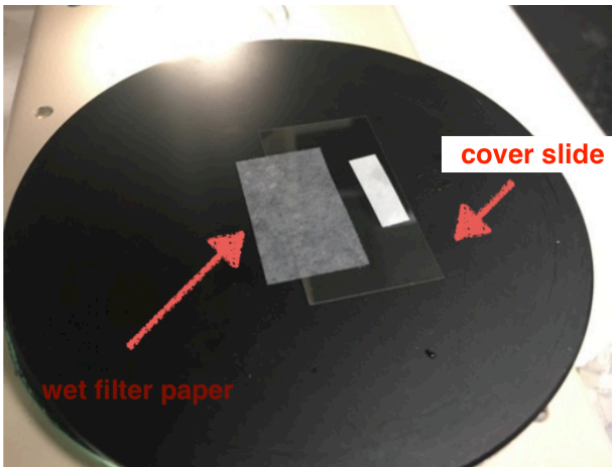


Figure 6.4.2. Preparing filter papers and a cover slide for aligning Clogmia embryos.

2. Prepare another cover slide, which will be used later for embryo transfer after the alignment. Attach a piece of double-sided tape to the cover slide, and then attach a piece of Tegaderm tape (MOORE MEDICAL LLC: 90323) on the double-sided tape, as shown below. Tegaderm tape has a better adhesiveness under water and oil than a double-sided tape alone.

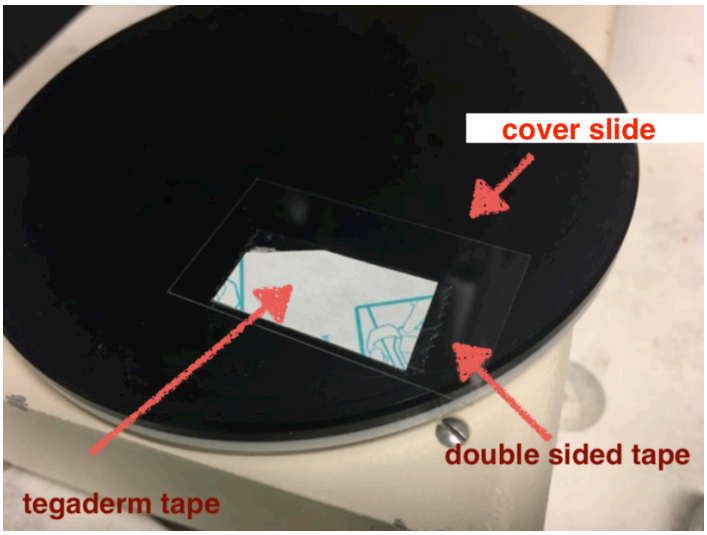


Figure 6.4.3. Preparing an injection slide for Clogmia embryos.

3. Transfer eggs to a wet filter paper (the 4cm x 2cm piece), and using brush, align them perpendicularly with the prospective injection side pointing towards the cover slide edge, as shown below. Injecting eggs near the anterior or posterior pole is critical for survival of the procedure. During the alignment procedure, water should be frequently applied to the filter paper as needed to protect eggs from desiccation. However, too much water can make embryos float and will disrupt the alignment process.

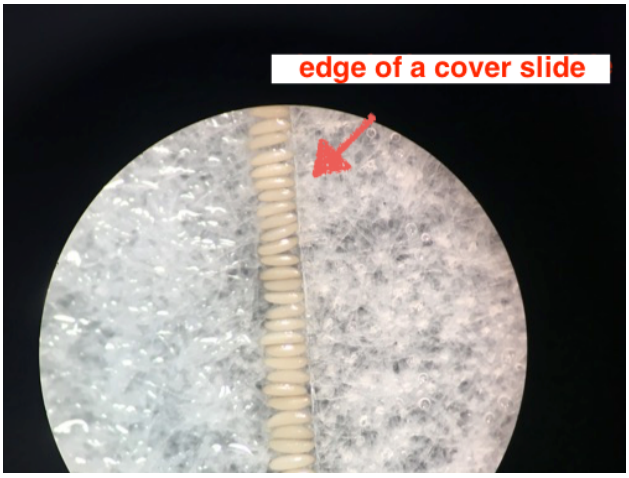


Figure 6.4.4. Alignment of Clogmia embryos.

4. After aligning the eggs, carefully remove the cover slide and excess water on the filter paper, using another piece of filter paper. Slightly press the second cover slide with Tegaderm tape against the aligned eggs to transfer the eggs to the Tegaderm tape.
5. Immediately cover the embryo with halocarbon oil to prevent desiccation. You should use different type of oils depending on the post-injection processing method. For cuticle preparations, use halocarbon oil 27 (Sigma, MKBZ7202V). This type of oil is easier to wash off after injection. For eggs to be fixed within a day following injection, use 1:1 mixture of halocarbon oil 27 and halocarbon oil 700 (Sigma, MKCB5817) and leave the injected eggs immersed in the oil until fixation. While embryos do not hatch under oil, the survival rate under oil is higher than without oil, up to one day post-injection. 1:1 mixture of halocarbon oil 27 and halocarbon oil 700 oil does not spread over time and can help to protect embryos from desiccation.

C. Embryo injection

1. Turn on the Narishige IM-300 microinjector and open the valve of nitrogen tank. You can check the pressure on the digital display of the microinjector. The pressure should be around 60-90 psi.
2. Backfill an injection needle (quartz needle) with 1.6ul of reagent (spin down the reagent with >12,000 rpm at 4C for at least 10 minutes prior to backfilling, as highly concentrated reagent such as dsRNA may clog injection needle. Always keep reagents on ice).
3. Locate the needle tip near eggs under the microscope (you should focus on embryos first, and then move around needle to bring the tip to the right location).

4. Press the foot pedal under the default, constant pressure mode (indicated on the digital display as "→").
5. Very slightly touch the first embryo with the needle to break open the tip of the needle. Each time you touch the embryo, press the food pedal. Repeat until you see bubbles coming out.
6. Inject eggs using the constant pressure mode. You cannot inject Clogmia or mosquito eggs using the Vent mode, as there will be a backflow from egg cytoplasm, clogging the needle. When injecting, you should quickly remove the needle after the penetration. The needle should penetrate the chorion smoothly. However, the tip will gradually break and become dull. This makes it necessary to change needle every 10-20 injections, or even more frequently. Also, you will have to change the needle occasionally due to clogging. When changing the needle, you should switch the injection mode to Vent (as indicated as "V" on display) first, to avoid any injuries from a needle shoot-out.
7. After you are done injecting all embryos on the slide, place the slide on top of a wet KimWipe in a Petri dish. Make sure the KimWipe is wet throughout the embryo development.
8. Allow the embryos to develop in the 28C incubator until they have reached desired stage. Add more water to the dish over time if necessary, making sure that the water level does not cover the slide.

D. Post injection processing

For cuticle preparation, see **APPENDIX 6.4.1.E**

6.4.3. Fixation of dipteran embryos using heat[8]

A. Dechorionation

1. Collect egg packages in a glass dish in a water.
2. Dechorionate embryos in diluted commercial bleach (concentrated CLOROX - Sodium Hypochlorite 8.25%):
    a. Chironomus: follow steps in **APPENDIX 6.4.1.C**
    b. Clogmia: in 10% bleach (commercial bleach) 10% for 90 sec
    c. Anopheles and Culex: in 25% bleach for 75 sec
    d. Aedes: in 25% bleach for 35 sec
    e. Hermetia: in 10% bleach for 180 sec
    f. Nephrotoma: in 25% bleach for 180 sec (until the egg shell becomes slightly transparent)
3. Wash thoroughly with water (~ 6 times) and transfer embryos to a 50 ml falcon tube using pipette.
4. Remove the water

B. Fixation (Heat)

5. Freshly prepare 20 ml of Fixing solution I with the following composition in another 50ml falcon tube:

    19.3ml of MilliQ water

[8] Modified from Chun Wai Kwan and Matteen Rafiqi's protocol, which was based on WIESCHAUS, E. and C. NÜSSLEIN-VOLHARD. 1986. Looking at embryos. In Drosophila, a practical approach (ed. D.B. Roberts), pp. 199-226. IRL Press, Oxford.}

500μl of NaCl (28% w/v, 28g in 100mL)

200μl of Triton X 100 (5%)

6. Briefly heat the solution in a microwave (28 sec) to boil (it turns turbid and has bubbles on surface). Do not fully tighten the lid to prevent the tube from exploding in the microwave.
7. Immediately pour the heated Fixing solution I into the falcon tube with embryos. Flip the tube up and down and rotate (with lid closed) 4~5 times so that all embryos get heat-fixed.
8. After flipping, add 30 ml of Mili Q water to the tube too cool it down. Let floating embryos settle to the bottom of the tube.
9. Transfer heat-fixed embryos to a fresh 1.5ml tube and remove solution.
10. Add 1ml of Mili Q water and quickly remove water. (Embryos can explode if they stay too long in hypotonic solution) This step is critical for devitellinization rate.
11. Add 600µl of Heptane and then add 600µl of MeOH and immediately shake intensely (as powerful and fast as possbile) for 20 ~ 40 seconds. (You need to be very fast at this step for effective devitellinization.)
12. Let embryos settle and remove heptane and wash embryos 3 times with methanol.

D. Formaldehyde fixation [This is a postfix and improves quality of storage and in situs]

13. Add Fixing solution II (865µl of 3:1 PBS:MeOH + 135µl of 37% formaldehyde) and incubate in rotator for 25 minutes.
14. Remove fixative.
15. Wash embryos 3 times with methanol
16. Store embryos in MeOH at -20˚C.

E. Manual devitellinization

In many cases, fixed embryos will still have their vitelline membrane attached. These embryos can be devitellized manually using tungaten needles, under methaol in a 3% agar plate.

6.4.4. sgRNA synthesis for CRISPR/Cas9 genome editing[9]

Reagents:

CloneAmp HiFi PCR premix (Takara 639298)

HiScribe™ T7 High Yield RNA Synthesis Kit (NEB E2040S)

mini Quick Spin RNA Columns (Roche, 11814427001)

Primers:

>sgRNA_F

GAAATTAATACGACTCACTATAGNNNNNNNNNNNNNNNNNNNGTTTAAGAGCTAT

GCTGGAA

>sgRNA_R

AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAAA

CTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC

Oligo design:

1. Design primers to sequence the region you plan to target with your gRNA(s). Often there are SNPs in the region relative to the reference genome. PCR and sequence the region.
2. Use your new sequence to identify potential gRNA target sites (*i.e.* 20-mers followed by a PAM [NGG]).
    a. You can do this by eye, or run your sequence through the FlyCRISPR Optimal Target Finder tool at http://flycrispr.molbio.wisc.edu/tools (Your oligo should

[9]Modified from Darli Massardo's protocol, which was based on (Bassett and Liu 2014).

start with >= 1 G in order to get good transcription with T7. If you cannot find any good oligos with a G, simply replace the first base in your oligo with a G. This should not interfere with efficient cutting because Cas9 can tolerate mismatches in the 5' end of the gRNA sequence.)

b. BLAST each potential oligo to the appropriate reference genome to check specificity. Use low stringency to ensure you pick up all potential off-targets. Pick 2 – 3 oligos with no off-targets. Remember that the PAM is absolutely required for Cas9 activity.

3. Plug your oligo sequence (MINUS THE PAM) into sgRNA_F in place of the string of Ns. Order this as a normal oligo from IDT. sgRNA_R only needs to be ordered once, as it will always work with any sgRNA_F primer you construct.

PCR amplification:

1. Dilute primers stocks from IDT to 100 µM. Make 10 µM working aliquots.
2. Set up PCRs on ice:
    1. 25 µL CloneAmp HF 2X Master Mix
    2. 2.5 µL sgRNA_F (for gene of interest)
    3. 2.5 µL sgRNA_R (universal)
    4. 20 µL H2O
    5. Vortex to mix thoroughly.
3. Run on thermocycler:
    1. 98°C 30 sec
    2. 98°C 10 sec

3. 60°C 30 sec
4. 72°C 15 sec
5. GOTO 2, rep 34
6. 72°C 1 min
7. HOLD 12°C

4. Check 5µL on 2% gel, 100V for ~25min, using 100bp ladder. There should be a fat band at ~120 bp.

PCR purification:

1. Add 4.5µl 3M NaOAc (1/10volume) (pH5.0) to each gRNA
2. Vortex, spin door
3. Add 50µl 1:1 phenol:chloroform
4. Shake, vortex for 30sec
5. Spin 10.000xg for 5min
6. Move supernatant to a new tube
7. Add 100µl of 100% ethanol, mix thoroughly
8. Chill -80C for 30min
9. Spin max speed at 4C for 20min
10. Remove supernatant
11. Add 300ul 70% ethanol
12. Spin max speed for 10min at 4C
13. Remove supernatant
14. Dry on bench

15. Resuspend in 30µl H2O or Tris

Transcription and amplification:

HiScribe™ T7 High Yield RNA Synthesis Kit (NEB E2040S)

1. Set up reaction on ice:
    - 2µl Template (0.1-0.5µg of PCR template)
    - 10µl NTP buffer mix
    - 2µl T7 enzyme mix
    - up to 20µl H20
    - Total volume 20µl
2. Incubate at 37C over night (~10~16 hours, since the template is short)
3. Expand the synthesis to 50µl by adding 30µl of H20
4. Add 2µl turbo DNase, mix and incubate for 15min at 37C

Phenol-chloroform extraction and ethanol precipitation:

1. Adjust the reaction volume to 180 µl by adding nuclease-free water. Add 20 µl of 3 M sodium acetate, pH 5.2 or 20µl of 5 M ammonium acetate and mix thoroughly.
2. Extract with an equal volume of 1:1 phenol:chloroform mixture, followed by two extractions with chloroform. Collect the aqueous phase and transfer to a new tube.
3. Precipitate the RNA by adding 2 volumes of ethanol. Incubate at –20°C for at least 30 minutes and collect the pellet by centrifugation.
4. Remove the supernatant and rinse the pellet with 500µl of ice cold 70% ethanol.
5. Resuspend the RNA in 30µl 0.1 mM EDTA. Store the RNA at –20°C or below.

Removal of free nucleotides:

Using mini Quick Spin RNA Columns (Roche, 11814427001)

A. Prepare column

1. Resuspend the sephadex matrix in the column buffer by vigorously inverting the column several times.
2. Remove the cap: First, the top cap. and then the bottom tip.
3. Remove excess buffer and place column in 1.5ML centrifuge tube.
4. Centrifuge at 1,000g (not RPM) for 1 min at RT.
5. Discard the collection tube with the eluted buffer and use the column immediately.

B. Purify the sample

6. Place the prepared column in a new 1.5mL tube.
7. Very slowly and carefully apply the sample (25µl-75µl) to the center of the column bed.
8. Centrifuge the tube at 1,000g for 4min at RT
9. The eluate contains purified sample

C. Measure the concentration on a nanodrop and dilute to 1µg/µl. You can send to RNA Bioanalyzer to get the real concentration.

D. Add 1µg/µl Cas9 protein to the 1µg/µl gRNAs and incubate for 15min at 37C before injection.

8. REFERENCES

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A. Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kennison, K. A. Ketchum, B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L. Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacleb, M. Palazzolo, G. S. Pittman, S. Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. Saunders, F. Scheeler, H. Shen, B. C. Shue, I. Siden-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z. Y. Wang, D. A. Wassarman, G. M. Weinstock, J. Weissenbach, S. M. Williams, WoodageT, K. C. Worley, D. Wu, S. Yang, Q. A. Yao, J. Ye, R. F. Yeh, J. S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X. H. Zheng, F. N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H. O. Smith, R. A. Gibbs, E. W. Myers, G. M. Rubin, and J. C. Venter. 2000. 'The genome sequence of *Drosophila melanogaster*', *Science*, 287: 2185-95.

Akbari, O. S., I. Antoshechkin, H. Amrhein, B. Williams, R. Diloreto, J. Sandler, and B. A. Hay. 2013. 'The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector', *G3 (Bethesda)*, 3: 1493-509.

Amaya, E, TJ Musci, and MW Kirschner. 1991. '*Express*ion of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in Xenopus embryos.', *Cell*, 66: 257-70.

An, J. J., K. Gharami, G. Y. Liao, N. H. Woo, A. G. Lau, F. Vanevski, E. R. Torre, K. R. Jones, Y. Feng, B. Lu, and B. Xu. 2008. 'Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons', *Cell*, 134: 175-87.

Andrews, S. 2010. "FastQC." In.: Babraham Bioinformatics.

Ansari, S., N. Troelenberg, V. A. Dao, T. Richter, G. Bucher, and M. Klingler. 2018. 'Double abdomen in a short-germ insect: Zygotic control of axis formation revealed in the beetle Tribolium castaneum', *Proc Natl Acad Sci U S A*, 115: 1819-24.

Anvar, S. Y., G. Allard, E. Tseng, G. M. Sheynkman, E. de Klerk, M. Vermaat, R. H. Yin, H. E. Johansson, Y. Ariyurek, J. T. den Dunnen, S. W. Turner, and P. A. C. t Hoen. 2018. 'Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing', *Genome Biol*, 19: 46.

Baek, Daehyun, Colleen Davis, Brent Ewing, David Gordon, and Phil Green. 2007. 'Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters.', *Genome Res*, 17: 145-55.

Bassett, A., and J. L. Liu. 2014. 'CRISPR/Cas9 mediated genome engineering in Drosophila', *Methods*, 69: 128-36.

Berkovits, B. D., and C. Mayr. 2015. 'Alternative 3' UTRs act as scaffolds to regulate membrane protein localization', *Nature*, 522: 363-7.

Berleth, T., M. Burri, G. Thoma, D. Bopp, S. Richstein, G. Frigerio, M. Noll, and C. Nüsslein-Volhard. 1988. 'The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo', *EMBO J.*, 7: 1749-56.

Bharti, Kapil, Wenfang Liu, Tamas Csermely, Stefano Bertuzzi, and Heinz Arnheiter. 2008. 'Alternative promoter use in eye development: the complex role and regulation of the transcription factor MITF', *Development*, 135: 1169-78.

Bolognesi, R., L. Farzana, T. D. Fischer, and S. J. Brown. 2008. 'Multiple Wnt genes are required for segmentation in the short-germ embryo of Tribolium castaneum', *Curr Biol*, 18: 1624-9.

Bownes, M., and K. Kalthoff. 1974. 'Embryonic defects in Drosophila eggs after partial u.v. irradiation at different wavelengths', *J Embryol Exp Morphol*, 31: 329-45.

Brent, A. E., G. Yucel, S. Small, and C. Desplan. 2007. 'Permissive and instructive anterior patterning rely on mRNA localization in the wasp embryo', *Science*, 315: 1841-3.

Bucher, G., L. Farzana, S. J. Brown, and M. Klingler. 2005. 'Anterior localization of maternal mRNAs in a short germ insect lacking *bicoid*', *Evol. Dev.*, 7: 142-49.

Buljan, M., G. Chalancon, A. K. Dunker, A. Bateman, S. Balaji, M. Fuxreiter, and M. M. Babu. 2013. 'Alternative splicing of intrinsically disordered regions and rewiring of protein interactions', *Curr Opin Struct Biol*, 23: 443-50.

Bullock, S., and D. Ish-Horowicz. 2001. 'Conserved signals and machinery for RNA transport in *Drosophila* oogenesis and embryogenesis', *Nature*, 414: 611-16.

Burglin, T. R., and M. Affolter. 2016. 'Homeodomain proteins: an update', *Chromosoma*, 125: 497-521.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. 'BLAST+: architecture and applications', *BMC Bioinformatics*, 10: 421.

Carroll, SB, JK Grenier, and SD Weatherbee. 2005. *From DNA to diversity. Molecular genetics and the evolution of animal design* (Blackwell: Oxford).

Carter, J. M., M. Gibbs, and C. J. Breuker. 2015. 'Divergent RNA Localisation Patterns of Maternal Genes Regulating Embryonic Patterning in the Butterfly Pararge aegeria', *PLoS One*, 10: e0144471.

Cha, B. J., B. S. Koppetsch, and W. E. Theurkauf. 2001. 'In vivo analysis of Drosophila bicoid mRNA localization reveals a novel microtubule-dependent axis specification pathway', *Cell*, 106: 35-46.

Chan, S. K., and G. Struhl. 1997. 'Sequence-specific RNA binding by bicoid', *Nature*, 388: 634.

Chen, H., Z. Xu, C. Mei, D. Yu, and S. Small. 2012. 'A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes', *Cell*, 149: 618-29.

Chen, W., Q. Jia, Y. Song, H. Fu, G. Wei, and T. Ni. 2017. 'Alternative Polyadenylation: Methods, Findings, and Impacts', *Genomics Proteomics Bioinformatics*, 15: 287-300.

Choe, C. P., F. Stellabotte, and S. J. Brown. 2017. 'Regulation and function of odd-paired in Tribolium segmentation', *Dev Genes Evol*, 227: 309-17.

Ciolli Mattioli, Camilla, Aviv Rom, Vedran Franke, Koshi Imami, Gerard Arrey, Mandy Terne, Andrew Woehler, Altuna Akalin, Igor Ulitsky, and Marina Chekulaeva. 2018. 'Alternative 3′ UTRs direct localization of functionally diverse protein isoforms in neuronal compartments', *Nucleic Acids Res*, 47: 2560-73.

Clark, E., and M. Akam. 2016. 'Odd-paired controls frequency doubling in Drosophila segmentation by altering the pair-rule gene regulatory network', *Elife*, 5.

Consortium, Fantom, Riken Pmi the, Clst, A. R. Forrest, H. Kawaji, M. Rehli, J. K. Baillie, M. J. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jorgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. A. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse,

S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. M. Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drablos, A. S. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furino, J. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. Ho Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. P. Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-Sim, R. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, S. Noma, T. Noazaki, S. Ogishima, N. Ohkura, H. Ohimiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. Prendergast, O. J. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. t Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyodo, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verado, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci, and Y. Hayashizaki. 2014. 'A promoter-level mammalian expression atlas', *Nature*, 507: 462-70.

Cranston, P. S., N. B. Hardy, and G. E. Morse. 2011. 'A dated molecular phylogeny for the Chironomidae (Diptera)', *Systematic Entomology*, 37: 172–88.

Crombach, A., M. A. Garcia-Solache, and J. Jaeger. 2014. 'Evolution of early development in dipterans: reverse-engineering the gap gene network in the moth midge Clogmia albipunctata (Psychodidae)', *Biosystems*, 123: 74-85.

Datta, R. R., J. Ling, J. Kurland, X. Ren, Z. Xu, G. Yucel, J. Moore, L. Shokri, I. Baker, T. Bishop, P. Struffi, R. Levina, M. L. Bulyk, R. J. Johnston, Jr., and S. Small. 2018. 'A feed-forward relay integrates the regulatory activities of Bicoid and Orthodenticle via sequential binding to suboptimal sites', *Genes Dev*, 32: 723-36.

Davis, W, and R M Schultz. 2000. 'Developmental change in TATA-box utilization during preimplantation mouse development', *Developmental Biology*, 218: 275-83.

Davuluri, R. V., Y. Suzuki, S. Sugano, C. Plass, and T. H. Huang. 2008. 'The functional consequences of alternative promoter use in mammalian genomes', *Trends Genet*, 24: 167-77.

de Klerk, E., and P. A. t Hoen. 2015. 'Alternative mRNA transcription, processing, and translation: insights from RNA sequencing', *Trends Genet*, 31: 128-39.

Derti, Adnan, Philip Garrett-Engele, Kenzie D Macisaac, Richard C Stevens, Shreedharan Sriram, Ronghua Chen, Carol A Rohl, Jason M Johnson, and Tomas Babak. 2012. 'A quantitative atlas of polyadenylation in five mammals.', *Genome Res*, 22: 1173-83.

Driever, W. 1993. 'Maternal control of anterior development in the *Drosophila* embryo.' in M. Bate and A. Martinez-Arias (eds.), *The development of Drosophila melanogaster* (Cold Spring Harbor Laboratory Press: Cold Spring Harbor).

Driever, W., and C. Nusslein-Volhard. 1988a. 'The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner', *Cell*, 54: 95-104.

———. 1988b. 'A gradient of bicoid protein in Drosophila embryos', *Cell*, 54: 83-93.

Driever, W., V. Siegel, and C. Nusslein-Volhard. 1990. 'Autonomous determination of anterior structures in the early Drosophila embryo by the bicoid morphogen', *Development*, 109: 811-20.

Ewen-Campen, B., E. E. Schwager, and C. G. Extavour. 2010. 'The molecular machinery of germ line specification', *Mol Reprod Dev*, 77: 3-18.

Feng, Guihai, Man Tong, Baolong Xia, Guan-Zheng Luo, Meng Wang, Dongfang Xie, Haifeng Wan, Ying Zhang, Qi Zhou, and Xiu-Jie Wang. 2016. 'Ubiquitously expressed genes participate in cell-specific functions via alternative promoter usage.', *EMBO reports*, 17: 1304-13.

Ferrandon, D., L. Elphick, C. Nusslein-Volhard, and D. St Johnston. 1994. 'Staufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner', *Cell*, 79: 1221-32.

Frohnhöfer, Hans Georg, and Christiane Nüsslein-Volhard. 1986. 'Organization of anterior pattern in the Drosophila embryo by the maternal gene bicoid', *Nature*, 324: 120-25.

Fu, J., N. Posnien, R. Bolognesi, T. D. Fischer, P. Rayl, G. Oberhofer, P. Kitzmann, S. J. Brown, and G. Bucher. 2012. 'Asymmetrically expressed axin required for anterior development in Tribolium', *Proc Natl Acad Sci U S A*, 109: 7782-6.

Garcia-Solache, M., J. Jaeger, and M. Akam. 2010. 'A systematic analysis of the gap gene system in the moth midge Clogmia albipunctata', *Dev Biol*, 344: 306-18.

Gavis, E. R., and R. Lehmann. 1992. 'Localization of *nanos* RNA controls embryonic polarity', *Cell*, 71: 301-13.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, and J. Zhang. 2004. 'Bioconductor: open software development for computational biology and bioinformatics', *Genome Biol*, 5: R80.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nat Biotechnol*, 29: 644-52.

Grimaldi, D., and M. S. Engel. 2005. *Evolution of insects* (Cambridge University Press: Cambridge).

Gusev, O., Y. Suetsugu, R. Cornette, T. Kawashima, M. D. Logacheva, A. S. Kondrashov, A. A. Penin, R. Hatanaka, S. Kikuta, S. Shimura, H. Kanamori, Y. Katayose, T. Matsumoto, E. Shagimardanova, D. Alexeev, V. Govorun, J. Wisecaver, A. Mikheyev, R. Koyanagi, M. Fujie, T. Nishiyama, S. Shigenobu, T. F. Shibata, V. Golygina, M. Hasebe, T. Okuda, N. Satoh, and T. Kikawada. 2014. 'Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge', *Nat Commun*, 5: 4784.

Hannon, C. E., S. A. Blythe, and E. F. Wieschaus. 2017. 'Concentration dependent chromatin states induced by the bicoid morphogen gradient', *Elife*, 6.

Holt, C. E., and S. L. Bullock. 2009. 'Subcellular mRNA localization in animal cells and why it matters', *Science*, 326: 1212-6.

Houtmeyers, R., J. Souopgui, S. Tejpar, and R. Arkell. 2013. 'The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis', *Cell Mol Life Sci*, 70: 3791-811.

Ihaka, R., and R. Gentleman. 1996. 'R: A Language for Data Analysis and Graphics', *J. Comput. Graph. Stat.*, 5: 299-314.

Ikeshima-Kataoka, H., J. B. Skeath, Y. Nabeshima, C. Q. Doe, and F. Matsuzaki. 1997. 'Miranda directs Prospero to a daughter cell during Drosophila asymmetric divisions', *Nature*, 390: 625-9.

Jackman, S. D., B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S. A. Hammond, G. Jahesh, H. Khan, L. Coombe, R. L. Warren, and I. Birol. 2017. 'ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter', *Genome Res*, 27: 768-77.

Jan, C. H., R. C. Friedman, J. G. Ruby, and D. P. Bartel. 2011. 'Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs', *Nature*, 469: 97-101.

Janssens, H., K. Siggens, D. Cicin-Sain, E. Jimenez-Guri, M. Musy, M. Akam, and J. Jaeger. 2014. 'A quantitative atlas of Even-skipped and Hunchback expression in Clogmia albipunctata (Diptera: Psychodidae) blastoderm embryos', *Evodevo*, 5: 1.

Jereb, S., H. W. Hwang, E. Van Otterloo, E. E. Govek, J. J. Fak, Y. Yuan, M. E. Hatten, and R. B. Darnell. 2018. 'Differential 3' Processing of Specific Transcripts Expands Regulatory and Protein Diversity Across Neuronal Cell Types', *Elife*, 7.

Jimenez-Guri, E., K. R. Wotton, B. Gavilan, and J. Jaeger. 2014. 'A staging scheme for the development of the moth midge Clogmia albipunctata', *PLoS One*, 9: e84422.

Juhn, J., and A. A. James. 2006. 'oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti', *Insect Mol Biol*, 15: 363-72.

———. 2012. 'Hybridization in situ of salivary glands, ovaries, and embryos of vector mosquitoes', *J Vis Exp*.

Juhn, J., O. Marinotti, E. Calvo, and A. A. James. 2008. 'Gene structure and expression of nanos (nos) and oskar (osk) orthologues of the vector mosquito, Culex quinquefasciatus', *Insect Mol Biol*, 17: 545-52.

Jürgens, G., E. Wieschaus, C. Nüsslein-Volhard, and H. Kluding. 1984. 'Mutations affecting the pattern of the laral cuticle in *Drosophila melanogaster*. - II. Zygotic loci on the third chromosome', *Roux Arch. dev. Biol.*, 193: 283-95.

Kaiser, T. S., B. Poehn, D. Szkiba, M. Preussner, F. J. Sedlazeck, A. Zrim, T. Neumann, L. T. Nguyen, A. J. Betancourt, T. Hummel, H. Vogel, S. Dorner, F. Heyd, A. von Haeseler, and K. Tessmar-Raible. 2016. 'The genomic basis of circadian and circalunar timing adaptations in a midge', *Nature*, 540: 69-73.

Kalthoff, K. 1971. 'Position of targets and period of competence for UV-induction of the malformation "Double Abdomen" in the egg ofSmittia spec. (Diptera, Chironomidae)', *Wilhelm Roux Arch Entwickl Mech Org*, 168: 63-84.

Kalthoff, Klaus, and Klaus Sander. 1968. 'Der Entwicklungsgang der Mißbildung „Doppelabdomen“ im partiell UV-bestrahlten Ei von Smittia parthenogenetica (Dipt., Chironomidae)', *Wilhelm Roux' Archiv für Entwicklungsmechanik der Organismen*, 161: 129-46.

Kanamori-Katayama, M., M. Itoh, H. Kawaji, T. Lassmann, S. Katayama, M. Kojima, N. Bertin, A. Kaiho, N. Ninomiya, C. O. Daub, P. Carninci, A. R. Forrest, and Y. Hayashizaki. 2011. 'Unamplified cap analysis of gene expression on a single-molecule sequencer', *Genome Res*, 21: 1150-9.

Kandler-Singer, I., and K. Kalthoff. 1976. 'RNase sensitivity of an anterior morphogenetic determinant in an insect egg (Smittia sp., Chironomidae, Diptera)', *Proc Natl Acad Sci U S A*, 73: 3739-43.

Kelley, J. L., J. T. Peyton, A. S. Fiston-Lavier, N. M. Teets, M. C. Yee, J. S. Johnston, C. D. Bustamante, R. E. Lee, and D. L. Denlinger. 2014. 'Compact genome of the Antarctic midge is likely an adaptation to an extreme environment', *Nat Commun*, 5: 4611.

Kim, S., M. Oh, W. Jung, J. Park, H. G. Choi, and S. C. Shin. 2017. 'Genome sequencing of the winged midge, Parochlus steinenii, from the Antarctic Peninsula', *Gigascience*.

Klomp, J., D. Athy, C. W. Kwan, N. I. Bloch, T. Sandmann, S. Lemke, and U. Schmidt-Ott. 2015. 'Embryo development. A cysteine-clamp gene drives embryo polarity in the midge Chironomus', *Science*, 348: 1040-2.

Krueger, F. 2012. "Trim Galore!" In.: Babraham Bioinformatics.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23: 2947-8.

Lau, Anthony G, Hasan A Irier, Jiaping Gu, Donghua Tian, Li Ku, Guanglu Liu, Mingjing Xia, Brita Fritsch, James Q Zheng, Raymond Dingledine, Baoji Xu, Bai Lu, and Yue Feng. 2010. 'Distinct 3′UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF)', *Proc Natl Acad Sci U S A*, 107: 15945-50.

Lazzaretti, D., K. Veith, K. Kramer, C. Basquin, H. Urlaub, U. Irion, and F. Bono. 2016. 'The bicoid mRNA localization factor Exuperantia is an RNA-binding pseudonuclease', *Nat Struct Mol Biol*, 23: 705-13.

Lehmann, R. 2016. 'Germ Plasm Biogenesis-An Oskar-Centric Perspective', *Essays on Developmental Biology, Pt A*, 116: 679-+.

Lemke, S., S. E. Busch, D. A. Antonopoulos, F. Meyer, M. H. Domanus, and U. Schmidt-Ott. 2010. 'Maternal activation of gap genes in the hover fly Episyrphus', *Development*, 137: 1709-19.

Lemke, S., and U. Schmidt-Ott. 2009. 'Evidence for a composite anterior determinant in the hover fly *Episyrphus balteatus* (Syrphidae), a cyclorrhaphan fly with an anterodorsal serosa anlage', *Development*, 136: 117-27.

Lemke, S., M. Stauber, P. J. Shaw, A. M. Rafiqi, A. Prell, and U. Schmidt-Ott. 2008. 'Bicoid occurrence and Bicoid-dependent hunchback regulation in lower cyclorrhaphan flies', *Evol Dev*, 10: 413-20.

Lianoglou, Steve, Vidur Garg, Julie L Yang, Christina S Leslie, and Christine Mayr. 2013. 'Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression', *Genes Dev*, 27: 2380-96.

Liao, Y., G. K. Smyth, and W. Shi. 2013. 'The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote', *Nucleic Acids Res*, 41: e108.

Linnen, Catherine R, Yu-Ping Poh, Brant K Peterson, Rowan D H Barrett, Joanna G Larson, Jeffrey D Jensen, and Hopi E Hoekstra. 2013. 'Adaptive evolution of multiple traits through multiple mutations at a single gene.', *Science*, 339: 1312-16.

Liu, Q., P. Onal, R. R. Datta, J. M. Rogers, U. Schmidt-Ott, M. L. Bulyk, S. Small, and J. W. Thornton. 2018. 'Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development', *Elife*, 7.

Lynch, J. A., A. E. Brent, D. S. Leaf, M. A. Pultz, and C. Desplan. 2006. 'Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp Nasonia', *Nature*, 439: 728-32.

Magoc, T., and S. L. Salzberg. 2011. 'FLASH: fast length adjustment of short reads to improve genome assemblies', *Bioinformatics*, 27: 2957-63.

Mallarino, R., T. A. Linden, C. R. Linnen, and H. E. Hoekstra. 2017. 'The role of isoforms in the evolution of cryptic coloration in Peromyscus mice', *Mol Ecol*, 26: 245-58.

Martin, M. 2011. 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17: 10-12.

Martinez, F. J., G. A. Pratt, E. L. Van Nostrand, R. Batra, S. C. Huelga, K. Kapeli, P. Freese, S. J. Chun, K. Ling, C. Gelboin-Burkhart, L. Fijany, H. C. Wang, J. K. Nussbacher, S. M.

Broski, H. J. Kim, R. Lardelli, B. Sundararaman, J. P. Donohue, A. Javaherian, J. Lykke-Andersen, S. Finkbeiner, C. F. Bennett, M. Ares, Jr., C. B. Burge, J. P. Taylor, F. Rigo, and G. W. Yeo. 2016. 'Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System', *Neuron*, 92: 780-95.

Mayr, C., and D. P. Bartel. 2009. 'Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells', *Cell*, 138: 673-84.

Morgan, M., H. Pages, V. Obenchain, and N. Hayden. 2013. "Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import." In.: Bioconductor.

Murgan, S., W. Kari, U. Rothbacher, M. Iche-Torres, P. Melenec, O. Hobert, and V. Bertrand. 2015. 'Atypical Transcriptional Activation by TCF via a Zic Transcription Factor in C. elegans Neuronal Precursors', *Dev Cell*, 33: 737-45.

Nakao, H. 2012. 'Anterior and posterior centers jointly regulate Bombyx embryo body segmentation', *Dev Biol*, 371: 293-301.

Nashchekin, D., A. R. Fernandes, and D. St Johnston. 2016. 'Patronin/Shot Cortical Foci Assemble the Noncentrosomal Microtubule Array that Specifies the Drosophila Anterior-Posterior Axis', *Dev Cell*, 38: 61-72.

Ochoa-Espinosa, A., G. Yucel, L. Kaplan, A. Pare, N. Pura, A. Oberstein, D. Papatsenko, and S. Small. 2005. 'The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila', *Proc Natl Acad Sci U S A*, 102: 4960-5.

Pal, Sharmistha, Ravi Gupta, and Ramana V Davuluri. 2012. 'Alternative transcription and alternative splicing in cancer', *Pharmacology & Therapeutics*, 136: 283-94.

Pal, Sharmistha, Ravi Gupta, Hyunsoo Kim, Priyankara Wickramasinghe, Valérie Baubet, Louise C Showe, Nadia Dahmane, and Ramana V Davuluri. 2011. 'Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development.', *Genome Res*, 21: 1260-72.

Perera, Bambarendage P U, and Joomyeong Kim. 2016. 'Sex and Tissue Specificity of Peg3 Promoters', *PLoS One*, 11.

Peter, I. S., and E. H. Davidson. 2011. 'Evolution of gene regulatory networks controlling body plan development', *Cell*, 144: 970-85.

Pinto, P. A., T. Henriques, M. O. Freitas, T. Martins, R. G. Domingues, P. S. Wyrzykowska, P. A. Coelho, A. M. Carmo, C. E. Sunkel, N. J. Proudfoot, and A. Moreira. 2011. 'RNA polymerase II kinetics in polo polyadenylation signal selection', *EMBO J*, 30: 2431-44.

Pozner, Amir, Joseph Lotem, Cuiying Xiao, Dalia Goldenberg, Ori Brenner, Varda Negreanu, Ditsa Levanon, and Yoram Groner. 2007. 'Developmentally regulated promoter-switch transcriptionally controls Runx I function during embryonic hematopoiesis', *BMC developmental biology*, 7.

Prühs, R., A. Beermann, and R. Schröder. 2017. 'The Roles of the Wnt-Antagonists Axin and Lrp4 during Embryogenesis of the Red Flour Beetle *Tribolium castaneum*', *J. Dev. Biol.*, 5: 10.

Reyes, A., and W. Huber. 2018. 'Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues', *Nucleic Acids Res*, 46: 582-92.

Riechmann, V., and A. Ephrussi. 2001. 'Axis formation during Drosophila oogenesis', *Curr Opin Genet Dev*, 11: 374-83.

Rivera-Pomar, R., D. Niessing, U. Schmidt-Ott, W. J. Gehring, and H. Jackle. 1996. 'RNA binding and translational suppression by bicoid', *Nature*, 379: 746-9.

Robinson, M. D., and A. Oshlack. 2010. 'A scaling normalization method for differential expression analysis of RNA-seq data', *Genome Biol*, 11: R25.

Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.', *Bioinformatics*, 26: 139–40.

Rohr, K. B., D. Tautz, and K. Sander. 1999. 'Segmentation gene expression in the mothmidge Clogmia albipunctata (Diptera, psychodidae) and other primitive dipterans', *Dev Genes Evol*, 209: 145-54.

Roper, K., and N. H. Brown. 2004. 'A spectraplakin is enriched on the fusome and organizes microtubules during oocyte specification in Drosophila'.

Sander, Klaus. 1976. 'Specification of the Basic Body Pattern in Insect Embryogenesis.' in J. E. Treherne, M. J. Berridge and V. B. Wigglesworth (eds.), *Advances in Insect Physiology* (Academic Press).

———. 1993. 'Mirror-image anterior patterning in centrifuged eggs of the mothfly Clogmia albipunctata (Psychodidae, Diptera)', *Acta Biologiae Experimentalis Sinica. December;*, 264: 343-51.

Schrankel, Catherine S, Cynthia M Solek, Katherine M Buckley, Michele K Anderson, and Jonathan P Rast. 2016. 'A conserved alternative form of the purple sea urchin HEB/E2-2/E2A transcription factor mediates a switch in E-protein regulatory state in differentiating immune cells', *Developmental Biology*, 416: 149-61.

Schröder, R. 2003. 'The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *Tribolium*', *Nature*, 422: 621-25.

Schupbach, T., and E. Wieschaus. 1986. 'Germline autonomy of maternal-effect mutations altering the embryonic body pattern of Drosophila', *Dev Biol*, 113: 443-8.

Shabalina, S. A., A. N. Spiridonov, N. A. Spiridonov, and E. V. Koonin. 2010. 'Connections between alternative transcription and alternative splicing in mammals', *Genome Biol Evol*, 2: 791-9.

Shabalina, Svetlana A, Aleksey Y Ogurtsov, Nikolay A Spiridonov, and Eugene V Koonin. 2014. 'Evolution at protein ends: major contribution of alternative transcription initiation and termination to the transcriptome and proteome diversity in mammals', *Nucleic Acids Res*, 42: 7132-44.

Shapiro, I. M., A. W. Cheng, N. C. Flytzanis, M. Balsamo, J. S. Condeelis, M. H. Oktay, C. B. Burge, and F. B. Gertler. 2011. 'An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype', *PLoS Genet*, 7: e1002218.

Shepard, Peter J, Eun-A Choi, Jente Lu, Lisa A Flanagan, Klemens J Hertel, and Yongsheng Shi. 2011. 'Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq', *Rna-a Publication of the Rna Society*, 17: 761-72.

Shiraki, T., S. Kondo, S. Katayama, K. Waki, T. Kasukawa, H. Kawaji, R. Kodzius, A. Watahiki, M. Nakamura, T. Arakawa, S. Fukuda, D. Sasaki, A. Podhajska, M. Harbers, J. Kawai, P. Carninci, and Y. Hayashizaki. 2003. 'Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage', *Proc Natl Acad Sci U S A*, 100: 15776-81.

Simpson-Brose, M., J. Treisman, and C. Desplan. 1994. 'Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila', *Cell*, 78: 855-65.

Smibert, P., P. Miura, J. O. Westholm, S. Shenker, G. May, M. O. Duff, D. Zhang, B. D. Eads, J. Carlson, J. B. Brown, R. C. Eisman, J. Andrews, T. Kaufman, P. Cherbas, S. E. Celniker, B. R. Graveley, and E. C. Lai. 2012. 'Global patterns of tissue-specific alternative polyadenylation in Drosophila', *Cell Rep*, 1: 277-89.

St Johnston, D., D. Beuchle, and C. Nusslein-Volhard. 1991. 'Staufen, a gene required to localize maternal RNAs in the Drosophila egg', *Cell*, 66: 51-63.

Stauber, M., H. Jäckle, and U. Schmidt-Ott. 1999. 'The anterior determinant *bicoid* of *Drosophila* is a derived Hox class 3 gene', *Proc. Natl. Acad. Sci. USA*, 96: 3786-89.

Stauber, M., S. Lemke, and U. Schmidt-Ott. 2008. 'Expression and regulation of caudal in the lower cyclorrhaphan fly Megaselia', *Dev Genes Evol*, 218: 81-7.

Stauber, M., A. Prell, and U. Schmidt-Ott. 2002. 'A single *Hox3* gene with composite *bicoid* and *zerknüllt* expression characteristics in non-Cyclorrhaphan flies', *Proc Natl Acad Sci U S A*, 99: 274-79.

Stauber, M., H. Taubert, and U. Schmidt-Ott. 2000. 'Function of bicoid and hunchback homologs in the basal cyclorrhaphan fly Megaselia (Phoridae)', *Proc Natl Acad Sci U S A*, 97: 10844-9.

Stern, D. L., and E Sucena. 2000. 'Preparation of larval and adult cuticles for light microscopy.' in W. Sullivan, M. Ashburner and R. S. Hawley (eds.), *Drosophila protocols* (CSHL Press: Cold Spring Harbor).

Struhl, G., K. Struhl, and P. M. Macdonald. 1989. 'The gradient morphogen bicoid is a concentration-dependent transcriptional activator', *Cell*, 57: 1259-73.

Taliaferro, J. M., M. Vidaki, R. Oliveira, S. Olson, L. Zhan, T. Saxena, E. T. Wang, B. R. Graveley, F. B. Gertler, M. S. Swanson, and C. B. Burge. 2016. 'Distal Alternative Last Exons Localize mRNAs to Neural Projections', *Mol Cell*, 61: 821-33.

Thawornwattana, Y., D. Dalquen, and Z. Yang. 2018. 'Coalescent Analysis of Phylogenomic Data Confidently Resolves the Species Relationships in the Anopheles gambiae Species Complex', *Mol Biol Evol*, 35: 2512-27.

Tian, Bin, and James L Manley. 2017. 'Alternative polyadenylation of mRNA precursors', *Nature Reviews Molecular Cell Biology*, 18: 18-30.

Trapnell, C., L. Pachter, and S. L. Salzberg. 2009. 'TopHat: discovering splice junctions with RNA-Seq', *Bioinformatics*, 25: 1105-11.

Tushev, G., C. Glock, M. Heumuller, A. Biever, M. Jovanovic, and E. M. Schuman. 2018. 'Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments', *Neuron*, 98: 495-511 e6.

Vacik, T., J. L. Stubbs, and G. Lemke. 2011. 'A novel mechanism for the transcriptional regulation of Wnt signaling in development', *Genes Dev*, 25: 1783-95.

Valen, E., G. Pascarella, A. Chalk, N. Maeda, M. Kojima, C. Kawazu, M. Murata, H. Nishiyori, D. Lazarevic, D. Motti, T. T. Marstrand, M. H. Tang, X. Zhao, A. Krogh, O. Winther, T. Arakawa, J. Kawai, C. Wells, C. Daub, M. Harbers, Y. Hayashizaki, S. Gustincich, A. Sandelin, and P. Carninci. 2009. 'Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE', *Genome Res*, 19: 255-65.

Vicoso, B., and D. Bachtrog. 2015. 'Numerous transitions of sex chromosomes in Diptera', *PLoS Biol*, 13: e1002078.

Wang, Xi, Jingyi Hou, Claudia Quedenau, and Wei Chen. 2016. 'Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals.', *Molecular systems biology*, 12: 875.

Wharton, R. P., and G. Struhl. 1991. 'RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos', *Cell*, 67: 955-67.

Wiegmann, B. M., M. D. Trautwein, I. S. Winkler, N. B. Barr, J-W. Kim, C. Lambkin, M. A. Bertone, B. K. Cassel, K. M. Bayless, M. A. Heimberg, B. M. Wheeler, K. J. Peterson, T. Pape, B. J. Sincalir, J. H. Skevington, V. Blagoderov, J. Caravas, S. Narayanan Kutty, U. Schmidt-Ott, G. E. Kampmeier, F. C. Thompson, D. A. Grimaldi, A. T. Beckenbach, G. W. Coutney, M. Friedrich, R. Meier, and D. K. Yeates. 2011. 'Episodic radiations in the fly tree of life', *Proc Natl Acad Sci U S A*, 108: 5690-95.

Wiesner, T., W. Lee, A. C. Obenauf, L. Ran, R. Murali, Q. F. Zhang, E. W. Wong, W. Hu, S. N. Scott, R. H. Shah, I. Landa, J. Button, N. Lailler, A. Sboner, D. Gao, D. A. Murphy, Z. Cao, S. Shukla, T. J. Hollmann, L. Wang, L. Borsu, T. Merghoub, G. K. Schwartz, M. A. Postow, C. E. Ariyan, J. A. Fagin, D. Zheng, M. Ladanyi, K. J. Busam, M. F. Berger, Y. Chen, and P. Chi. 2015. 'Alternative transcription initiation leads to expression of a novel ALK isoform in cancer', *Nature*, 526: 453-7.

Wilhelm, J. E., J. Mansfield, N. Hom-Booher, S. Wang, C. W. Turck, T. Hazelrigg, and R. D. Vale. 2000. 'Isolation of a ribonucleoprotein complex involved in mRNA localization in Drosophila oocytes', *J Cell Biol*, 148: 427-40.

Willer, Martin, Andrew J. Jermy, Barry P. Young, and Colin J. Stirling. 2003. 'Identification of novel protein–protein interactions at the cytosolic surface of the Sec63 complex in the yeast ER membrane', *Yeast*, 20: 133-48.

Wilson, D. S., G. Sheng, S. Jun, and C. Desplan. 1996. 'Conservation and diversification in homeodomain-DNA interactions: a comparative genetic analysis', *Proc Natl Acad Sci U S A*, 93: 6886-91.

Wilson, M. J., and P. K. Dearden. 2011. 'Diversity in insect axis formation: two orthodenticle genes and hunchback act in anterior patterning and influence dorsoventral organization in the honeybee (Apis mellifera)', *Development*, 138: 3497-507.

Wimmer, E. A., A. Carleton, P. Harjes, T. Turner, and C. Desplan. 2000. '*bicoid*-independent formation of thoracic segments in *Drosophila*', *Science*, 287: 2476-79.

Wittkopp, P. J., and G. Kalay. 2011. 'Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence', *Nat Rev Genet*, 13: 59-69.

Wotton, K. R., E. Jimenez-Guri, and J. Jaeger. 2015. 'Maternal co-ordinate gene regulation and axis polarity in the scuttle fly Megaselia abdita', *PLoS Genet*, 11: e1005042.

Xin, D., L. Hu, and X. Kong. 2008. 'Alternative promoters influence alternative splicing at the genomic level', *PLoS One*, 3: e2377.

Yajima, H. 1960. 'Studies on embryonic determination of the harlequin-fly, Chironomus dorsalis. I. Effects of centrifugation and of its combination with constriction and puncturing', *J Embryol Exp Morphol*, 8: 198-215.

———. 1964. 'Studies on Embryonic Determination of the Harlequin-Fly, Chironomus Dorsalis. Ii. Effects of Partial Irradiation of the Egg by Ultra-Violet Light', *J Embryol Exp Morphol*, 12: 89-100.

Zhang, Peter, Emmanuel Dimont, Thomas Ha, Douglas J Swanson, FANTOM Consortium, Winston Hide, and Dan Goldowitz. 2017. 'Relatively frequent switching of transcription start sites during cerebellar development.', *BMC Genomics*, 18: 461.