

Trinity College

Trinity College Digital Repository

Senior Theses and Projects

Student Scholarship

Spring 2022

THE ROLE AND REGULATION OF THE CAUDAL GENE IN TRIBOLIUM CASTANEUM SEGMENTATION

Suzanne Nicole Carpe Elias

Trinity College, Hartford Connecticut, scarpe@trincoll.edu

Follow this and additional works at: <https://digitalrepository.trincoll.edu/theses>



Part of the [Biology Commons](#), [Developmental Biology Commons](#), and the [Genetics Commons](#)

Recommended Citation

Carpe Elias, Suzanne Nicole, "THE ROLE AND REGULATION OF THE CAUDAL GENE IN TRIBOLIUM CASTANEUM SEGMENTATION". Senior Theses, Trinity College, Hartford, CT 2022.

Trinity College Digital Repository, <https://digitalrepository.trincoll.edu/theses/998>

TRINITY COLLEGE

THE ROLE AND REGULATION OF THE *CAUDAL* GENE
IN *TRIBOLIUM CASTANEUM* SEGMENTATION

BY

SUZANNE N. CARPE ELIAS

A THESIS SUBMITTED TO
THE FACULTY OF THE DEPARTMENT OF BIOLOGY
IN CANDIDACY FOR THE BACCALAUREATE DEGREE
WITH HONORS IN BIOLOGY

DEPARTMENT OF BIOLOGY

HARTFORD, CONNECTICUT

6 MAY 2022

THE ROLE AND REGULATION OF THE *CAUDAL* GENE
IN *TRIBOLIUM CASTANEUM* SEGMENTATION

BY
SUZANNE N. CARPE ELIAS

Honors Thesis Committee

Approved:

Terri A. Williams, Advisor

Robert J. Fleming

Claire T. Fournier

Date: _____

TABLE OF CONTENTS

ABSTRACT	5
INTRODUCTION.....	6
Segmentation in animals.....	6
Beyond the <i>D. melanogaster</i> paradigm	7
The vertebrate-like segmentation in <i>T. castaneum</i>	8
Importance of the <i>cad</i> gene in different species	9
The <i>cad</i> gene in the <i>T. castaneum</i> segmentation clock	10
Regulation of early <i>cad</i> gene expression.....	11
The role and the regulation of the <i>cad</i> gene in <i>T. castaneum</i>	13
CHAPTER 1: ROLE OF THE <i>CAD</i> GENE	14
INTRODUCTION.....	14
MATERIALS AND METHODS	17
Beetle care and culture.....	17
Primer design for RNAi and qPCR experiments	17
dsRNA synthesis for RNAi.....	19
dsRNA microinjections.....	19
Fixation and staining of embryos.....	20
Mounting and imaging of embryos.....	20
Scoring phenotypes	21
Quantitative PCR (qPCR)	22
Dot blot for protein analysis	23
RESULTS.....	24
Embryonic dsRNA injections effectively reproduce the well-known <i>cad</i> phenotype.....	24
The role of <i>cad</i> changes during the first 24 hours of development.	27
Some segmentation clock and Wnt signaling genes are affected in <i>cad</i> knockdowns. ...	33
Expression levels for Cad protein are also decreased in <i>cad</i> RNAi knockdowns.	36

DISCUSSION.....	38
CHAPTER 2: REGULATION OF THE <i>CAD</i> GENE	45
INTRODUCTION.....	45
MATERIALS AND METHODS	47
DNA sequence and nucleotide frequency for <i>cad</i> gene.....	47
Candidate transcription factors	47
Weight matrix model on the binding sites of a transcription factor from <i>D. melanogaster</i> genes models.....	49
MCAST tool from MEME Suite.....	51
RESULTS.....	52
A final computational output for predicted transcription factor binding clusters was developed.	52
Predicted transcription factor binding clusters can be used to establish putative enhancer regions.....	57
DISCUSSION.....	60
CONCLUSION.....	63
APPENDIX	65
Change in Ct (cycle time) for each of our knockdown qPCRs	65
Weight matrix model on the binding sites of a transcription factor from <i>D. melanogaster</i> genes models.....	66
ACKNOWLEDGMENTS	77
LITERATURE CITED.....	79

ABSTRACT

The embryo of the red flour beetle *Tribolium castaneum* develops sequentially by adding segments in an anterior-to-posterior progression using a “clock”-like mechanism similar to that of vertebrates. Previous studies indicate that the oscillations of this segmentation clock are driven by a gradient of the transcription factor *caudal* (*cad*), which activates and regulates the clock. Knocking down the *cad* gene using parental or early embryonic RNAi leads to animals with only head segments. We hypothesized that progressively later embryonic knockdowns would produce animals with progressively more segments if the function of *cad* does not change during segmentation. To examine this, we knocked down the gene using RNAi at three different timepoints prior to segmentation: 4, 8.5, and 11.5 hours after egg lay (hAEL). We found that segment addition was affected for the two earlier timepoints as expected, but late blastoderm embryos (11.5 hAEL) did not require *cad* to add segments despite having very few segments already patterned. Therefore, our results suggest that *cad* is regulating segmentation in very early development only, and we propose that a different regulatory network is controlling late segmentation. Additionally, it has been shown that the frequency of the clock changes during development, hence we hypothesized that *cad* might be dynamically regulated by various transcription factors during different phases. We performed bioinformatics analyses using the MCAST tool to establish predictions of transcription factor binding clusters that might be regulating *cad* gene expression, and used these predictions as the basis to clone putative enhancer regions for yeast one-hybrid and cross-species transgenics. We infer that a change in *cad* regulation causes its function to change through development as we observed in our knockdowns.

INTRODUCTION

Segmentation in animals

The process of segmentation consists of patterning an organism's body into a series of repeated units—a phenomenon that is considerably widespread in the animal kingdom. There are three major taxa that develop using segments: annelids, arthropods, and chordates (Fortey and Thomas, 1997). A hallmark of all these animal groups is that they are able to generate high morphological diversity, a feature that has been attributed to the ability to specify their body regions from the repetition of building blocks in segmentation. Additionally, developing by means of segmentation allows animals to be flexible with respect to differentiating body regions to perform different tasks, a characteristic that is most evident in arthropods (Tautz, 2004). Arthropoda contains the highest number of animal species on the planet, divided into four major classes. Insects make up the largest class of arthropods, and they represent the largest percentage of animal species on the planet (Fortey and Thomas 1997).

The general body plan of insects is well-conserved among species: adult bodies consist of a head with six segments, a thorax with three segments, and an abdomen with eight to 11 segments. Even though body plans are considerably similar, the segmentation process by which they are achieved can be surprisingly varied and represent a wide spectrum of ancestral and conserved mechanisms (Liu and Kaufman, 2005). The great diversity of insect segmentation has allowed for its categorization into three main types: short, intermediate, and long germ embryogenesis, which describe the length of germ anlage (the group of cells that will become the germ band or the embryo itself) relative to the length of the entire egg. Long germ insects specify almost all segments simultaneously within the blastoderm (prior to gastrulation) while short germ insects specify only head segments in the blastoderm, and the

remaining segments form progressively from a posterior growth zone after gastrulation. The terms “short” and “long” germ embryogenesis were initially coined to represent opposite poles of a continuum, with intermediate embryogenesis representing the mechanism in between (Davis & Patel, 2002). Nonetheless, regardless of the evolving classification of these types of development, the two contrasting mechanisms include patterning almost all segments at once (simultaneous segmentation) or in a sequence (sequential segmentation).

Beyond the *D. melanogaster* paradigm

The best studied model arthropod, the fruit fly *Drosophila melanogaster*, patterns its body using long germband embryogenesis or virtually simultaneous segmentation. This form of development is highly derived and not therefore representative of the way in which most other insect species develop, thus it is necessary to consider other model organisms to understand the ancestral state and the more prevalent mechanism of segmentation in insects. The red flour beetle *Tribolium castaneum* has emerged as a model system to study sequential segmentation, since it uses short germband embryogenesis (Liu and Kaufman, 2005). In this process, only the anterior segments are specified before gastrulation in a development phase known as blastoderm. In this stage, the embryo is a single layer of cells surrounding a central yolk mass without any specified tissues. Later in development, the posterior segments are formed from the growth zone during germband elongation (El-Sherif et al., 2014). Therefore, by understanding segmentation in *T. castaneum*, we can learn more about the current mechanism of segmentation of the vast majority of insects in the planet, as well as the evolution of segmentation within the arthropods and among annelids and vertebrates.

The vertebrate-like segmentation in *T. castaneum*

The process by which short germband insects undergo segmentation is similar to vertebrate somitogenesis (Liao & Oates, 2017), in which a “clock”-like mechanism gives rise to individual somites from a posterior growth zone (Pourquié 2001). It has been proposed that vertebrates possess a molecular oscillator or clock that regulates the temporal periodicity of presomitic mesoderm cells to create a spatial periodicity of the somites in vertebrates, such as the cell cycle model (Stern et al., 1988) or the “clock and wavefront” model (Cooke and Zeeman, 1976). Subsequent experiments showed that a set of interacting genes including *hairy* and components of the Notch signaling produce the temporal oscillations of the vertebrate segmentation clock that give rise to spatial patterns (Palmeirim et al., 1997; Pourquié, 2003)

This type of mechanism has been demonstrated in *T. castaneum* as well: the embryo develops sequentially by adding segments in an anterior-to-posterior progression also using a “clock”-like mechanism. This segmentation clock consists of a molecular oscillator: three pair-rule genes, *even-skipped (eve)*, *runt*, and *odd-skipped (odd)*, regulate one another to produce waves of expression during elongation, which give rise to the posterior segments in the developing germband (El-Sherif et al., 2012; Sarrazin et al., 2012). A frequency oscillation is converted into a spatial pattern by this clock. Therefore, *T. castaneum* develops using a segmentation clock similar to that of vertebrates in which molecular oscillations of the clock specify individual segments. A side-by-side comparison of vertebrate somitogenesis and short germband elongation is shown in Figure 1.

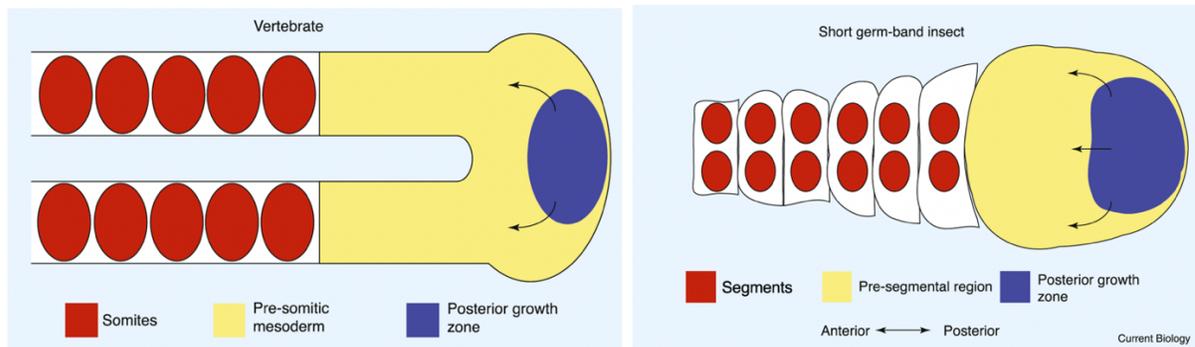


Figure 1. Vertebrate somitogenesis and short germ-band segmentation in insects. In both processes, a posterior growth zone (blue) gives rise to individual segments (red) patterned by a clock-like mechanism (Figure from Martin and Kimelman, 2009).

Importance of the *cad* gene in different species

The *cad* gene has been involved in posterior patterning in several different species. For instance, previous research has shown that *cad* homologs are necessary for normal development in vertebrates: in mice, *Cdx1* is necessary for anteroposterior axial skeletal identity and for putative regulation of the *Hox* genes (Subramanian et al., 1995), while *Cdx2* is involved in cell differentiation in the intestinal epithelium (Beck et al., 1999), normal development and growth (Chawengsaksophak et al., 1997), and the integration of pathways that control embryonic axis elongation and anterior-posterior patterning (Chawengsaksophak et al., 2004). On the other hand, different *cad* genes in the chicken have been shown to establish an anterior-posterior gradient by using temporal and spatial patterns of expression that overlap with one another (Marom et al., 1997), and the *cad* homolog *X-cad2* in *Xenopus laevis* is a key component of the posterior network that divides the early embryo into anterior head and trunk

domains, gives rise to the anterior-posterior axis, and is thought to be involved in regulating *Hox* genes as well (Epstein et al., 1997).

In the nematode *Caenorhabditis elegans*, the maternal contribution of the *cad* homolog *pal-1* is necessary to determine somatic identity of the posterior blastoderm in the 4-cell embryo (Hunter and Kenyon, 1996), plus a zygotic contribution of this gene has been shown to regulate posterior patterning during late embryogenesis (Edgar et al., 2001). In other organisms such as the milkweed bug *Oncopeltus fasciatus* and the cricket *Gryllus bimaculatus*, *cad* has been shown to regulate posterior patterning (Novikova et al., 2020; Shinmyo et al., 2005). Finally, there is plenty of information on the role of *cad* in *D. melanogaster*, including its role in establishing the anterior-posterior axis and activation of *eve* (Macdonald and Struhl, 1986; Moreno and Morata, 1999). Therefore, the *cad* gene is a very important regulator in the posterior development of animals across different phyla.

The *cad* gene in the *T. castaneum* segmentation clock

Previous studies indicate that the oscillations in the *T. castaneum* segmentation clock are driven by a gradient of the transcription factor *caudal* (*cad*), which activates and regulates the clock. More specifically, it had been demonstrated that *cad* regulates the expression of the pair-rule gene *even-skipped* (*eve*) (El-Sherif et al., 2014), which has been shown along with its partner *odd-skipped* (*odd*) to oscillate in waves of a given periodicity (El-Sherif et al., 2012; Sarrazin et al., 2012). A diagram of the current understanding on how the segmentation clock operates in arthropods is shown in Figure 2.

Although it has been documented that the frequency of the clock changes during late segmentation (Nakamoto et al., 2015), not much is known about whether or how the *cad* gene modulates these changes. For instance, knockdown of *cad* has been performed only by parental

and very early embryonic RNAi (Copf et al., 2004; Benton et al., 2013; El-Sherif et al., 2014). Therefore, there is no available data on whether *cad* regulates the clock later in embryogenesis. Even though the function of the segmentation clock has been proven for early development only, the claims made about the function of this clock throughout segmentation have been made based on inferences drawn from these early embryo observations.

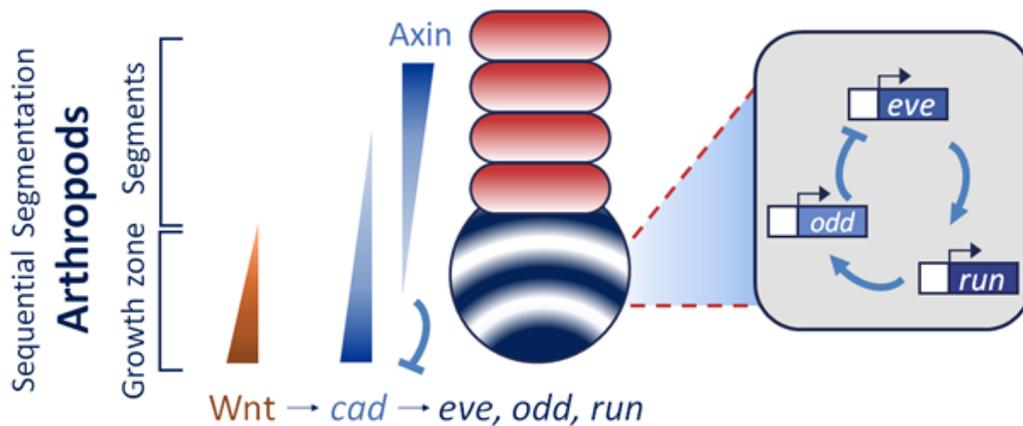


Figure 2. Arthropod segmentation clock and the genetic framework involved. It has been shown that Wnt signaling activates *cad* expression while Axin inhibits it. *cad* activates *eve*, which in turn activates *run* and *run* activates *odd*. The *eve* gene is then inhibited by *odd* expression (Figure modified from Liao and Oates, 2017).

Regulation of early *cad* gene expression

Similarly, it is known that *cad* is being dynamically regulated by different transcription factors such as Axin and components of the Wnt signaling pathway (Liao and Oates, 2017; Ansari et al., 2018). Axin has been proposed to regulate the maternal contribution of the *cad* gene, more specifically by preventing the ubiquitous *cad* mRNA from being translated in the anterior region through regulation of the Wnt signaling pathway. Therefore, the *cad* gradient is localized in the posterior of the embryo alongside Wnt signaling components as shown in

Figure 3. The transcription factor Mex3 has also been proposed to regulate *cad* function by anterior inhibition driven by *zen1* and *homeobrain* (*hbn*) gene activity.

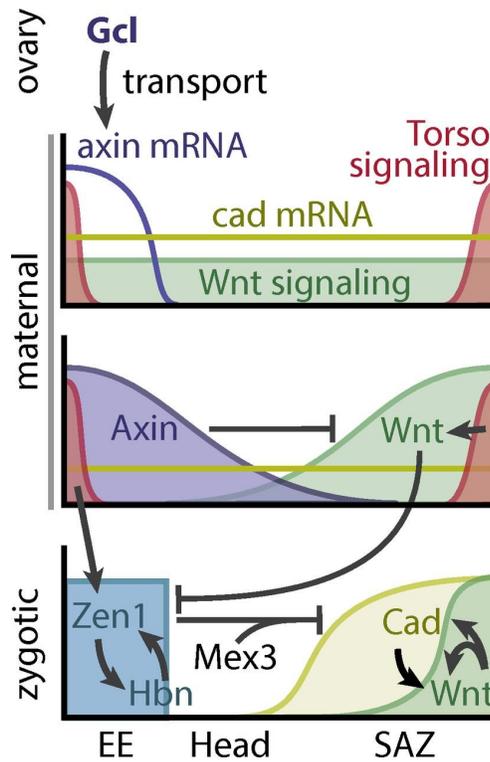


Figure 3. Proposed genetic regulatory network controlling anterior-posterior axis specification. *cad* mRNA starts being ubiquitously expressed in the ovary and the early blastoderm, but by the time the zygotic genes are activated, a *cad* domain is specified in the anterior by Wnt signaling and Mex3 (and indirectly by Axin, Zen1, and Hbn). Figure from Ansari et al., 2018.

However, it has also been shown that there is some redundancy in *cad* gene regulation, as knocking down certain genes such as *mex3* and *zen1* still produce embryos with head structures, while knocking down both produce double abdomen embryos (Ansari et al., 2018). On the other hand, knocking down both Wnt1 and Wnt8 produces embryos with no abdominal (posterior) segments (Bolognesi, 2009). This regulations account for setting up the *cad* gradient, but implies that the regulation of *cad* is kept constant throughout development.

Developmental genes are highly dynamic and their regulation changes constantly as embryos develop, so it becomes necessary to consider not only what establishes the *cad* domain but also how its expression in development.

The role and the regulation of the *cad* gene in *T. castaneum*

The *cad* gene and its homologues have been shown to be very important factors in the development of different organisms across different animal phyla. More specifically in *T. castaneum*, previous analyses suggest *cad* is regulating the segmentation clock that patterns the segments, and therefore the body, of these animals. By understanding the genetic mechanisms in which segmentation occurs in *T. castaneum*, we can better understand short germband or sequential segmentation process in different arthropods, and we can also draw parallels between this process and vertebrate somitogenesis. Therefore, *cad* is a gene that merits further investigation to understand both its function—what exactly it is doing throughout development and not only at the early blastoderm stages, and its regulation—which transcription factors and signaling pathways are involved in these processes. To address such necessity, I examined the role of this gene throughout *T. castaneum* development by knocking it down at later stages using RNA interference (RNAi). We hypothesized that the role of the *cad* gene is continuous during development based on its early functions: progressively later knockdowns would produce embryos with progressively more segments. In addition to this, I have also studied *cad* gene regulation using the bioinformatics analysis tool Motif Cluster Alignment Search Tool (MCAST) to consider predictions of transcription factor binding clusters that might be regulating *cad* gene expression. I used these predictions as the basis to clone putative enhancer regions for yeast one-hybrid and cross-species transgenics.

CHAPTER 1: ROLE OF THE *CAD* GENE

INTRODUCTION

Early embryo evidence demonstrates that *cad* regulates the segmentation clock at the blastoderm stage, and it is inferred based on these results that the *cad* gene is necessary to regulate the entire segmentation process. This function has been proposed based on the ability of changes in *cad* expression to modify *eve* expression, demonstrating that *cad* is actively regulating the frequency of the clock (El-Sherif et al., 2014). However, all of the different experiments knocking down *cad* with RNAi to examine its function have only focused on dsRNA injections at parental or very early blastoderm stages (the phase prior to formation of early embryo) (Copf et al., 2004; Benton et al., 2013; El-Sherif et al., 2014). Therefore, there is no substantial evidence to indicate that the *cad* gene is required throughout segmentation, although RNA-seq data from our lab has shown that the *cad* gene is being expressed at least during the first 24 hours of *T. castaneum* development when segmentation is occurring (Goldman-Huertas et al., in prep).

Based on this, we hypothesized that the role of the *cad* gene does not vary during *T. castaneum* development. To assess this prediction, we performed RNAi experiments at three different stages in segmentation that would allow us to investigate the role of this gene throughout development. We analyzed embryos at 4, 8.5, and 11.5 hours after egg lay (hAEL), which are still very early blastoderms: even at 11.5 hAEL, the cells that will form the embryo have just begun to condense and pattern at most four of the sixteen segments (Nakamoto et al., 2015). Figure 4 illustrates how early these blastoderm stages are: the embryo has still not even finished the process of cellularization by 11.5 hAEL, so most of what is seen in the blastoderm is individual nuclei with no cell membrane—there are no tissues at this stage yet.

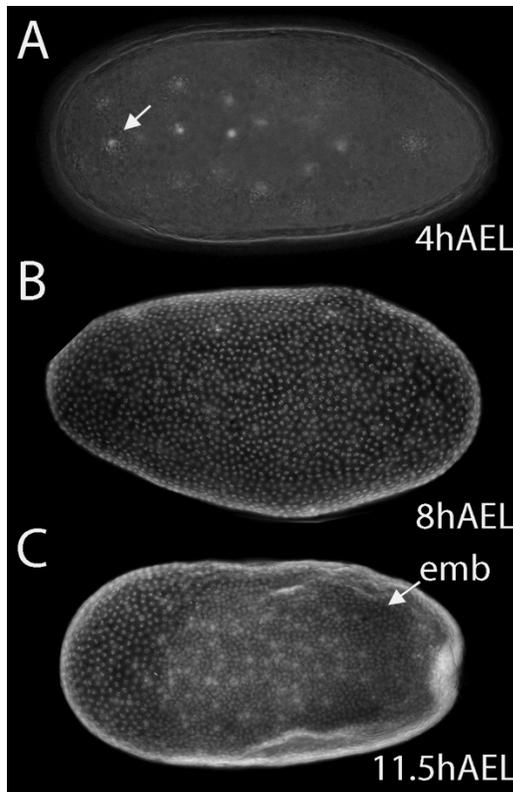


Figure 4. Wild type embryos stained with DAPI (to visualize nuclei) at 4 hAEL (A), 8 hAEL (B), and 11.5 hAEL (C). Arrow in (A) points to individual nuclei, while arrow in (C) points to the location of embryonic tissue.

If our hypothesis is correct, we expect to see progressively more segments being specified with progressively later knockdowns. These results will ultimately expand our understanding of the function of the *cad* gene in *T. castaneum* and in other animal phyla that develop using similar mechanisms. A schematic representation of our experiments next to the published data for parental knockdowns is shown in Figure 5. Besides examining the phenotypes of the gene knockdowns, we were also interested in looking at the mRNA and protein expression of these knockdowns.

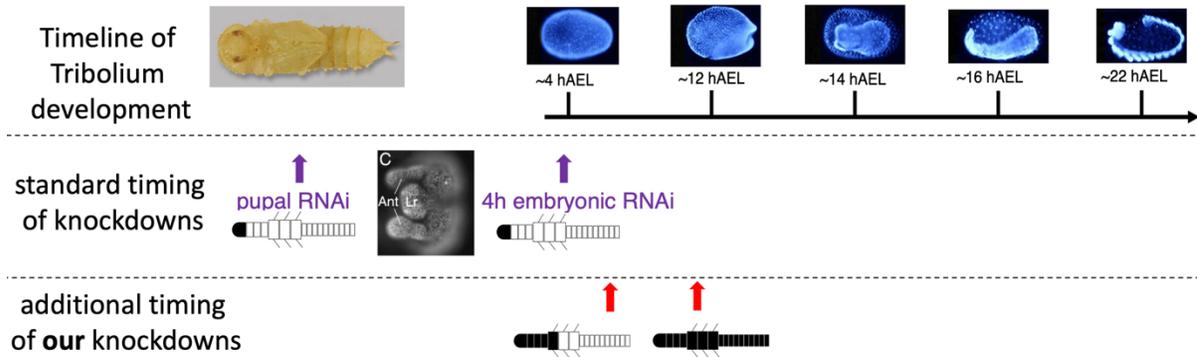


Figure 5. Schematic representation of our embryonic RNAi experiments at 4, 8.5, and 11.5 hAEL, compared to the parental RNAi knockdown usually seen in the literature. The diagrams at the bottom represent our results as to what the effect of knocking down *cad* is at subsequent stages in development (Embryo images from Tautz, 1999).

We used qPCR to quantify the mRNA expression of *cad* after RNAi, not only for the *cad* gene itself but also for other segmentation clock and Wnt signaling genes that could be affected by the knockdowns. For instance, we looked at *eve*, *runt*, and *odd* expression as we hypothesized that there might be feedback regulation in the clock, and considered *Wnt1* and *Wnt8* as they have been shown to be necessary for posterior development and wanted to see how their expression changes after RNAi for each of our timepoints (Bolognesi et al., 2009). Finally, we measured Cad protein expression to establish the translational effect of *cad* RNAi.

We found that knocking down the gene at 4 and 8.5 hAEL produce defects in segmentation, reminiscent to those of parental knockdowns. However, almost all, if not all of the segments are specified in our 11.5 hAEL knockdowns, which indicate that the *cad* is not having an effect later on. We validated our knockdowns using qPCR and dot blot to verify our mRNA and protein levels had gone down. Finally, we also saw that knocking down *cad* does affect some segmentation and Wnt signaling genes in different ways at the three timepoints.

MATERIALS AND METHODS

Beetle care and culture

The beetle embryos used in the present experiments are the offspring of our stock of red flour beetle *Tribolium castaneum* adults (GA-1 strain, originating from Kansas Stock Center in 2010). The animals are reared in jars of whole wheat flour supplemented with 5% brewer's yeast at 30 °C and approximately 30-50% humidity. These beetles have the advantage of ease of culture in controlled conditions, short life cycle, easy mating and breeding, large brood sizes, and a fully sequenced genome (Pointer et al., 2021). We followed the routine care and culture procedures detailed in *The Beetle Book* (Jenkins, 2012).

Primer design for RNAi and qPCR experiments

We obtained the *cad* gene (TC032769) sequence from the database Ensembl Metazoa (<https://metazoa.ensembl.org/index.html>). We designed primers using the National Center for Biotechnology Information (NCBI)'s Primer Blast tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) for a region encompassing the 5'-exon of the *cad* gene. We targeted this region since *cad* has two transcripts that differ in the 3'-exon (Schulz et al., 198)—so we wanted to make sure we knocked down both transcripts through a region that both of them have in common. We used the New England BioLabs T_m calculator (<https://tmcalsculator.neb.com>) to estimate appropriate annealing temperatures for PCR products and the Integrated DNA Technologies Oligo Analyzer Tool (<https://www.idtdna.com>) to find possible hairpins and dimers. The best primers were selected using the following criteria: be around 20 base pairs (bp) in length, have a GC-content close to 50%, the highest T_m (melting temperature) hairpin must be at least 5 °C below the annealing

temperature for the reaction, the self-complementary score and 3' self-complementary score must be as low as possible, the most negative self-dimer and heterodimer ΔG is higher -9, the T_m values between primers in a set cannot be more than 3 °C different, and the primer cannot bind anywhere else in the *T. castaneum* genome. The primer sequences are described in Tables 1 and 2 RNAi and qPCR experiments, respectively.

Table 1. Primer set sequences for *cad* RNAi experiments (Ensembl Sequence ID: TC032769). Both primers in the 5' → 3' direction.

Forward primer	Reverse primer
CGTCAAGTGACAAGTGCGTG	CGGGATTAGGCTGACTCTGG

Table 2. Gene of interest, gene Ensembl sequence IDs, and primer sequences for qPCR experiments (all primers in the 5' → 3' direction).

Gene	Ensembl Sequence ID	Forward primer	Reverse primer
<i>cad</i>	TC032769	GGACCTCCAACGATCGAGT	TTTTGATCTGCCGCTCCGAC
<i>H3</i>	TC005398	GAACAGACCCACGAGGTACG	CTGCCCTTCCAGAGATTGGT
<i>Wnt1</i>	TC030877	ATCGGCGACCTCCTCAAAGA	TGCGGCGATTCTCCCTCTTA
<i>Wnt8</i>	TC010155	ATTCAATCAGGACCTTAACCCT GT	TGGACAATTCCACCGATCCCA
<i>eve</i>	TC009469	CGAGGCTGGAGAAGGAGTTC	TGGCCATTCTTTGGCGTTTG
<i>odd</i>	TC005785	GGCGTCAAAGACCATCTGAGG	CACTTGTGCGGGGATTCTCT
<i>runt</i>	TC006542	CTCGGGAGCCCTACTACCAGA	CGGCCACGTATAGCTCATGT

dsRNA synthesis for RNAi

We cloned a 466 base pair section of the gene targeting the first exon to study the difference in gene function during development. This *cad* gene region was isolated using the primer described in Table 1 through a polymerase chain reaction (PCR) using a genomic DNA template and OneTaq 2X MasterMix with standard buffer as described in the StrataClone PCR cloning kit. The reaction was run on a Bio Rad T00™ Thermal Cycler and gel electrophoresis was used to verify the size of the amplified products. After the PCR products were successfully verified, we ligated them into pSC-A vector plasmids. These plasmids were then transformed into StrataClone *E. coli* bacterial cells and then selected on Amp plates. The inserts were verified using gel electrophoresis and the sequences were confirmed at Genewiz (now Azenta). The verified clones were used as a template for making dsRNA using the ThermoFisher MEGAscript™ RNAi Kit with T7 primers according to company instructions. Product size was then verified using gel electrophoresis, and if product was verified, dsRNA was aliquoted and stored in a -20 °C freezer until used for microinjections.

dsRNA microinjections

Embryos needed for injections were collected from 30-minute egg lays at 4, 8.5, and 11.5 hAEL. The embryos were dechorionated in 5% bleach solution while shaking for 2 minutes, and then were rinsed with distilled water. The embryos were then transferred to glass slides using a paintbrush (approximately 50 embryos per control and experimental groups). The injections were done using a World Precision Instrument PV820 picospritzer, a Nikon stereomicroscope, and a Narishige micromanipulator. For the injections, we used a pulled glass needle which was pulled on a Sutter Instrument Co. Model P-67 Flaming/Brown Micropipette Puller. Experimental embryos were injected with dsRNA at a concentration of ~500 ng/uL (the

dsRNA was diluted with 0.1 M sodium phosphate injection buffer as needed to reach this concentration). The control embryos were injected with the same amount of injection buffer. Injected embryos were then placed in a 30 °C incubator and raised for 24 hours, after which they were fixed.

Fixation and staining of embryos

After injections, embryos were allowed to develop for a certain time (24, 36, or 48 hAEL) and then fixed in 8% formaldehyde solution in PBS/EGTA buffer in a one-to-one volume with heptane as described by Shippy et al. (2009). They were put on to a nutator and rocked for 45 minutes, after which the fix was removed, methanol added, and the tube shaken for two minutes. All the solution was removed, and the animals stored in 100% methanol at -20 °C temperature until use (Shippy et al., 2009). Fixed embryos were placed in subsequent dilutions of methanol until 25% methanol was reached, and then placed in PBT and stained with 1 uL of Hoechst or DAPI nuclear staining for 30-40 minutes in the shaker. This stain allowed for the visualization of cell nuclei. After the embryos were stained, they were washed 2x in PBT and 2x in PBS then placed in 70% glycerol and stored in a -4°C refrigerator before mounting.

Mounting and imaging of embryos

After successful staining of the embryos, we transferred them into glass slides using a 200 uL pipette set at 35 uL (for each slide, we transferred 4-5 embryos for controls and 2-3 embryos for experimentals). If needed, additional 70% glycerol was added to ensure that embryos were properly protected. We then placed a coverslip with clay feet on top of each slide, and carefully placed the slides on a slide book. The slides were imaged on a Nikon

Eclipse E600 epifluorescence microscope under UV light, and images of the embryos on the side were taken at 200X magnification using NIS-Elements D software with the Extended Depth of Focus (EDF) function. This function allows us to take several images with different focus levels depending on the various tissue layers of an embryo, which can then be overlaid into a single, well-focused image. Some of the embryos were rolled for better visualization either at lateral or dorsal planes. Images were stored on our shared Microsoft OneDrive folder.

Scoring phenotypes

Each embryo resulting from a knockdown experiment was scored using the phenotype scoring key described in Figure 6 (adapted from Novikova 2020). We categorized embryos into three main classes: Class I consisted of severely truncated embryos only with the most anterior head segments (consistent with the phenotypes described in Copf et al., 2004); Class II consisted of truncated embryos with head segments and additional putative thoracic structures and Class III consisted of almost wild-type embryos which exhibited almost normal length and elongation with most segments being specified and with some obvious defects in the morphology of the segments.

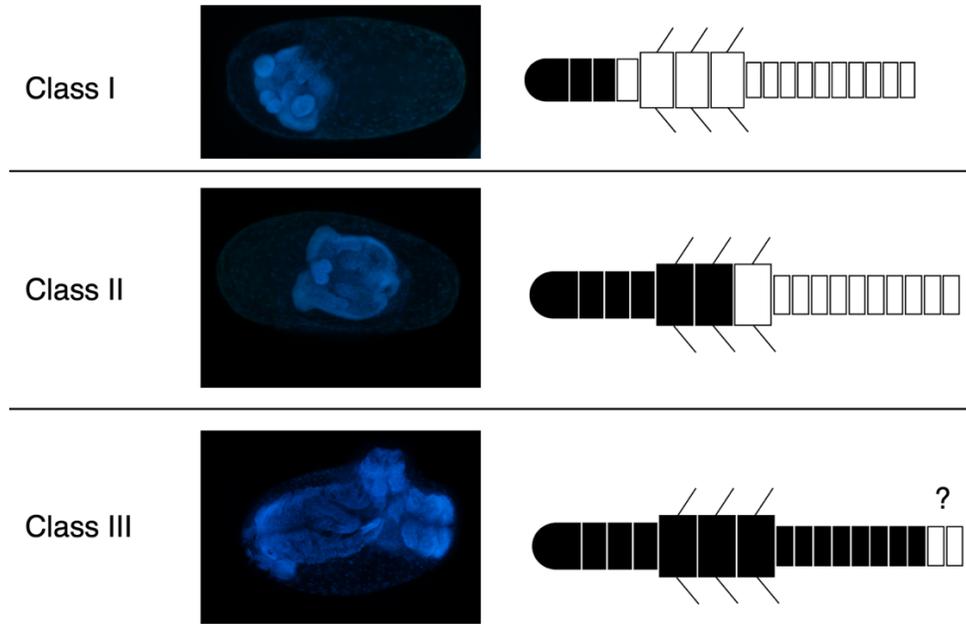


Figure 6. Key for phenotype scoring into three different classes with increasing severity. Class I being the most severe phenotypes with only the most anterior head structures and Class III being the least severe phenotypes with almost all of the segments but with some morphological defects. Mild phenotypes which included some or all thoracic segments were classified as the intermediate phenotype, referred to as Class III. Anterior is to the left, and embryos are imaged at 200X magnification. Black segments on the diagrams signify that a given segment is morphologically visible and the question mark indicates that we don't know for sure how many segments are being specified.

Quantitative PCR (qPCR)

T. castaneum wild type embryos were ground up into NEB Monarch® DNA/RNA Protection Reagent and mRNA was extracted using the NEB Monarch® Total RNA Miniprep Kit #T2010S according to manufacturer's instructions. The mRNA was used for qPCR analyses to evaluate the expression of genes known to be directly or indirectly activated by *cad* in RNAi knockdowns (for all three knockdowns). The genes examined are described in Table 2: *Wnt1*, *Wnt8*, and the segmentation clock genes *eve*, *odd*, and *runt*. We additionally used a histone 3 (*H3*) gene as a reference gene because data from our wild type transcriptome showed

that it is highly stable during this period of segmentation (Goldman-Huertas et al., in prep). We used the NEB Luna® Universal Probe One-Step RT-qPCR Kit with ThermoFisher according to manufacturer's instructions with the primers described in Table 2, and reactions were run on a BioRad CFX Connect Real-Time PCR System. Results were collected and analyzed on CFX Maestro and in Microsoft Excel to calculate the foldchange ($\Delta\Delta CT$ value) of the knockdowns compared to the buffer injected controls. We subtracted the cycle number for *H3* from each of our genes of interest to account for individual variability between data sets. Buffer controls were also used as reference points, as the measured buffer expression was normalized to 1.0 to calculate the foldchange. Results were graphed using the Prism software.

Dot blot for protein analysis

Immun-Blot® Low Fluorescence PVDF membranes were cut into squares of approximately 1.25 cm in length, one for each protein sample to test. The membranes were activated with methanol, distilled water, and TBST (Tris-buffered saline with 0.1% Tween® 20 Detergent), and then 10 ug of protein were applied to each membrane. The membrane was allowed to dry completely, and it was re-activated again with methanol, distilled water, and TBST. Then, it was blocked in 5% goat serum in TBST, and the primary antibody was added in a 1:1000 dilution. The next day, the membranes were washed with TBST, and the secondary antibody was also added in a 1:1000 dilution. The membranes were washed with TBST and TBS, and then mixed with Bio-Rad Clarity Western ECL Substrate #1705061 in a 1:1 ratio (or around 300 uL per dot). The membranes were then put on a piece of Immun-Blot® Low Fluorescence filter paper covered in plastic wrap to avoid solution absorption. Results were visualized in an Azure c300 imager. Relative intensity was measured in Fiji (Schindelin et al., 2012) and the results were graphed using the GraphPad Prism software.

RESULTS

Embryonic dsRNA injections effectively reproduce the well-known *cad* phenotype.

All of the *T. castaneum cad* RNAi phenotypes previously described in the literature have been obtained through either parental dsRNA injections or very early embryonic dsRNA injections (Copf et al., 2004; Benton et al., 2013; El-Sherif et al., 2014). These phenotypes consist of severely truncated embryos with only with the most anterior head structures (Copf et al., 2004). We attempted to reproduce these characteristic phenotypes by injecting dsRNA into 4 hour-old *T. castaneum* embryos—comparable to the early embryo knockdowns in El-Sherif et al. (2014). The phenotypes we obtained are shown in Figure 7.

The established *cad* knockdown first described by Copf et al. (2004) (Figure 7E) was replicated using our dsRNA embryonic injections: embryos fixed at both 24 hAEL (Figure 7B) and 48 hAEL (Figure 7D) showed the truncated phenotype with only the most anterior head structures. Comparison with respective buffer-injected controls (Figures 7A and 7C), illustrates the extent of the RNAi disruptions. For instance, wild type embryos possess several appendages including the antenna, labrum, mandible, maxilla, and labia (Figures 7A and 7C), yet dsRNA injected embryos only possess putative labrum and antenna plus an abundance of yolk (Figures 7B and 7D) as described in the literature (and shown in Figure 7E).

The fact that our dsRNA injected embryos at 4 hAEL show severe truncations that are characteristic of the *cad* knockdown phenotype indicates that our phenotypes parallel those used in published results. This is important as we used the same technique and the same model organism as Copf et al. (2004), but we tested a later timepoint in development. By proving that we can use RNAi and still successfully obtain the same phenotypes, we have verified the effectiveness of this technique to move forward with later timepoint experiments.

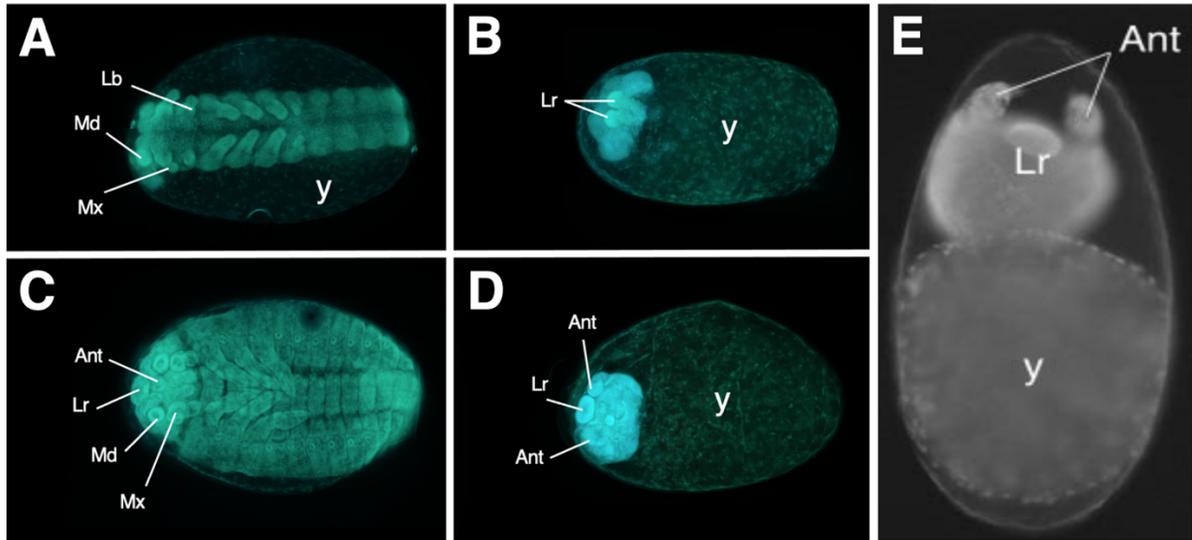


Figure 7. Wild type and *cad* RNAi knockdown phenotypes. Control embryos injected with buffer were fixed at 24 hAEL (A) and 48 hAEL (C) for comparison. Embryos with severe truncations that possess only the most anterior head structures are shown in 24-hour (B) and 48-hour (E) dsRNA injected embryos, replicating the *cad* phenotype first described in Copf et al., 2004 (E, Figure from Copf et al., 2004). Anterior is to the left in all embryos except in (E), where anterior is at the top. Different structures are labeled: antenna (Ant), labrum (Lr), yolk (y), mandible (Md), maxilla (Mx), and labia (Lb). Magnification is 200X.

Even though the phenotypes described in Figure 7 recapitulate the well-known *cad* phenotype, we wanted to further validate our results using molecular techniques to prove that the *cad* gene had been successfully knocked down with dsRNA injections in embryos at 4 hAEL. In addition to this, we wanted to examine how long it takes for knockdown of the mRNA to occur after dsRNA injections. Therefore, we performed qPCR experiments to examine gene expression one hour and two hours after dsRNA injection. We found that *cad* gene expression is significantly reduced one hour after RNAi (at 5 hAEL), and that this decrease in expression persisted two hours after RNAi (at 6 hAEL) when compared to buffer controls (Figure 8). The expression levels in our buffer controls are normalized as 1.0 for a reference value, and the change in the knockdown expression is described as the foldchange in relation to this value. *cad* expression decreased to about 25% of controls in both the 5-hour old

and 6-hour old embryos, which show that the *cad* gene has been significantly reduced. After our knockdowns described in Figure 1 and the present molecular validation, we were confident in using RNAi as a technique to evaluate the function of the *cad* gene in *T. castaneum* segmentation. Moreover, we showed that the mRNA is being knocked down as soon as one hour after dsRNA injection, and that this knockdown persists for at least two hours later.

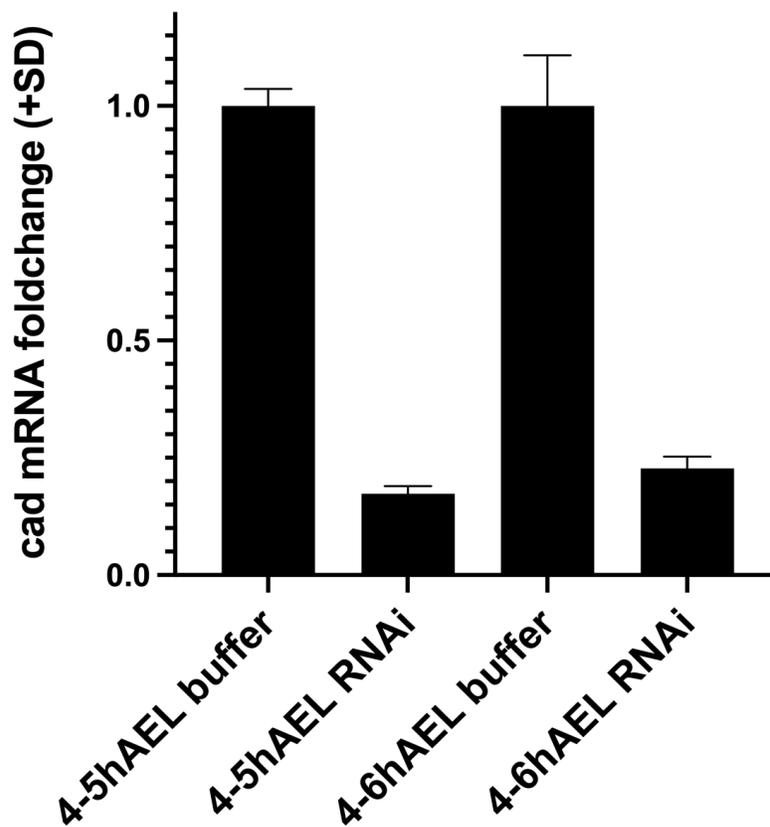


Figure 8. *cad* gene expression after RNAi. When embryos at 4 hAEL were injected with dsRNA, the expression of the *cad* gene significantly decreased after one hour of development (5 hAEL) when compared to buffer controls. Similarly, the *cad* gene levels maintained significantly lower levels two hours after injection (6 hAEL) when compared to buffer controls at the same time. Gene expression is measured by qPCR. Error bars represent standard deviation.

The role of *cad* changes during the first 24 hours of development.

After validating our embryonic RNAi experiments in 4-hour embryos, we investigated the role of the *cad* gene at different stages during development, in particular after maternal-to-zygotic transition (that is to say, the time where the embryo's own genetic material starts being translated instead the maternal transcriptional contribution), which has been shown to occur between 3 and 6 hAEL for *T. castaneum* (Ribeiro et al., 2017). We were interested to see if there is a difference in maternal *cad* function (approximately before 6 hAEL) compared to zygotic *cad* function (approximately after 6 hAEL). Our hypothesis was that zygotic *cad* function should mimic the maternal function, since it has been shown that 1) there is *cad* transcript present during the first 24 hours of development and 2) that the segmentation clock of *T. castaneum*, which *cad* regulates, is present during this time as well (El-Sherif et al., 2012; Sarrazin et al., 2012). We also expected our results to show segment specification as a function of time: that is, that the later the knockdown, the more segments we would see as *cad* would be able to pattern more and more segments as time passes.

To test our hypotheses, we performed dsRNA injections in the same way as our previously described injections, this time at 8.5 and 11.5 hAEL. These are two timepoints after maternal-to-zygotic transition, but it is important to note that they are both still very early in development, when the embryo has not even finished cellularization and has only specified at most four of the sixteen segments (Nakamoto et al., 2015). I found variation in the knockdown phenotypes with some embryos having extreme, moderate, or mild segmental defects, which is consistent with our hypothesis that *cad* regulates segment specification as a function of time. The different phenotypes were characterized into three different classes, which are described in Figure 9.

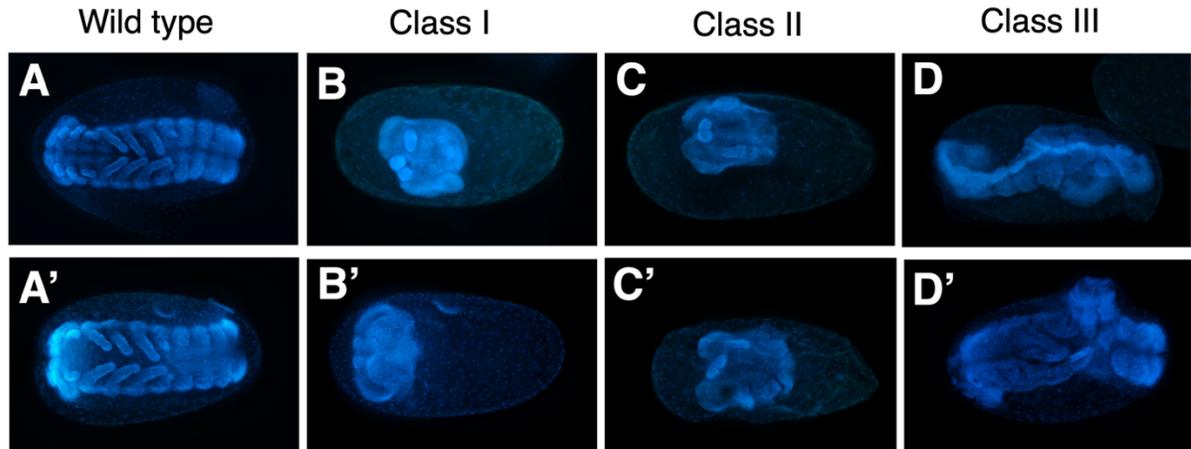


Figure 9. Classification of *cad* RNAi phenotypes. Wild type embryos were obtained from buffer controls (A-A'), and possessed all the expected segments. The well-studied *cad* truncated phenotype with only the most anterior head structures was described as Class I (B-B'), while more moderate phenotypes with head segments plus some putative thoracic structures were described as Class II (C-C'). Finally, embryos with mild segmentation defects but that possessed most segments were described as Class III (D-D'). Anterior is to the left of all embryos, and all images are in 200X magnification.

I then examined the embryos for each of our three dsRNA injection time points (at 4, 8.5, and 11.5 hAEL), and classified all embryos in a given data set based on the phenotypes shown in Figure 6. We additionally included buffer controls for reference. The results of this phenotypic scoring are summarized in Table 3. In addition to this, the phenotype distributions that we found across the three different injection time points are represented in Figure 9.

Table 3. *cad* phenotype scoring distribution. Embryos were visually characterized as either Class I, II, III, or wild type based on the number of segments present as previously described. The total number of embryos for each data set is also shown.

Experiment	Class I	Class II	Class III	Wild type	Total embryos
4-24 hAEL control	-	-	-	42	42
4-24 hAEL RNAi	38	-	-	1	39
8.5-24 hAEL control	-	-	-	23	23
8.5-24 hAEL RNAi	26	5	4	1	36
11.5-36 hAEL control	-	-	-	42	42
11.5-36 hAEL RNAi	-	-	13	37	50

The phenotypic distribution showed that all of our buffer controls produced only wild type organisms with the appropriate number of segments. For the 4-24 hAEL data set (embryos injected with dsRNA at 4 hours and examined at 24 hours), I found that the vast majority of our embryos (38/39) possess the truncated Class I phenotype. The one embryo that appeared wild type was likely caused due to problems with the dsRNA injection, such as a defective needle or the dsRNA going outside of the embryo instead of inside. The phenotypes of the 8.5-24 hAEL embryos were more varied, with most embryos belonging to the Class I phenotype (26/36), while some embryos were classified to be either Class II (5/36) or Class III (4/36). I also found a single embryo (1/36), that appeared wild type, likely due to problems with injection as previously described. These results seem to support our hypothesis that *cad* regulates segmentation as a function of time, at least at 4 and 8.5 hAEL.

On the other hand, the 11.5-36 hAEL (embryos injected with dsRNA at 11.5 hours and examined at 36 hours), embryos belong in either the Class III (13/50) or wild type (37/50) categories. Interestingly, even for the 26% of embryos that show some abnormalities in development, there are no obvious effects in segment addition number (segments are being specified even if not in the proper orientation or with the proper shape). These results do not completely support our hypothesis, since an embryo at the time of injection, 11.5 hAEL, has only specified at most two to four of the sixteen segments (Nakamoto et al., 2015). Therefore, *cad* expression seems to be correlated with segment patterning in the two earlier timepoints, but because the segment number does not change for the later timepoint, *cad* is likely not responsible for patterning the subsequent segments. These results suggest that *cad* might lose its fundamental role in patterning segments around five hours after maternal-to-zygotic transition, since not having *cad* does not seem to affect the number of segments being patterned. It is possible that *cad* is still having some residual effect since Class III embryos do have some morphological defects although not in the number of segments they possess. Therefore, the *cad* gene might not be regulating the *T. castaneum* segmentation clock for the entirety of segmentation as has been assumed.

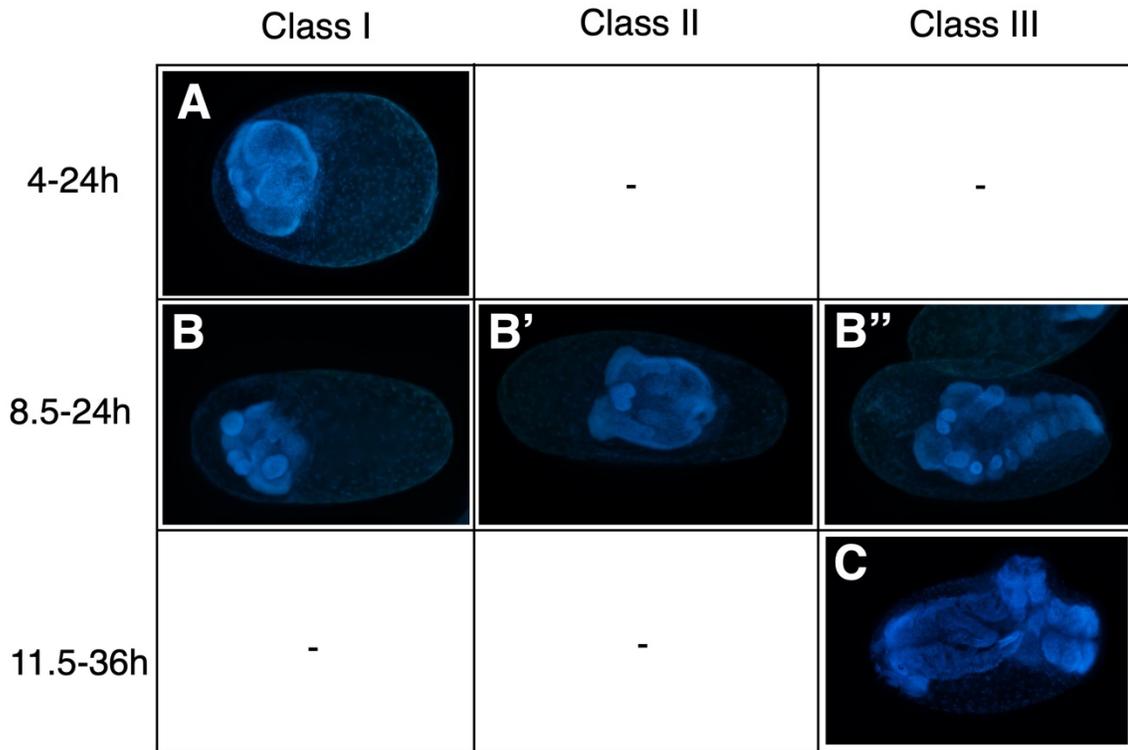


Figure 10. Representation of the phenotype class distribution for the three different injection timepoints. For 4-24h embryos, only Class I phenotypes were found (A), while 11.5-36h embryos only showed the Class III phenotype (C). However, the 8.5-24h embryos had a wider distribution all across the three classes (B-B'').

We wanted to show that, for our knockdowns at 11.5 hAEL, the levels of gene transcription had decreased after our dsRNA injections to verify that we were getting proper knockdown despite our unusual phenotypes. Therefore, we repeated the experiment shown in Figure 8, where we measured the expression of *cad* one hour after injection for the 11.5 hAEL dsRNA injections and the 4.5 hAEL dsRNA injections as a comparison. We found that gene expression significantly goes down for the 11.5 hAEL dsRNA injections as also replicated in the 4 hAEL dsRNA injections (Figure 11). Therefore, we were able to show that *cad* was effectively knocked down at this later timepoint, so our phenotypes are the result of the biological effect of knocking down the gene and not due to problems with our dsRNA

injections. The fact that we didn't observe severe phenotypes in this data set (as we did in the 4 and 8.5 hAEL dsRNA injections) supports our conclusion that *cad* gene function is changing during segmentation, since knocking down the gene later in development does not impede the animal from patterning a normal number of segments.

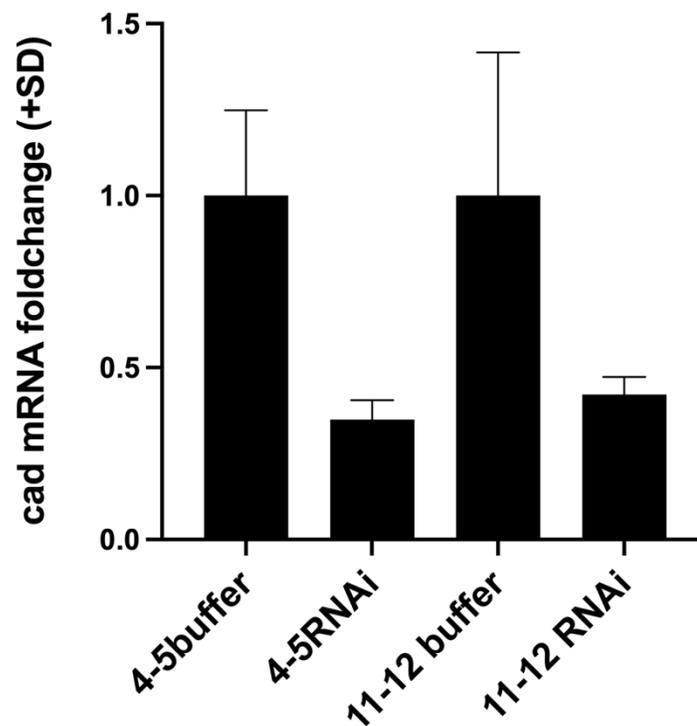


Figure 11. *cad* gene expression after RNAi. When embryos at 4 hAEL were injected with dsRNA, the expression of the *cad* gene significantly decreased after one hour of development (5 hAEL) when compared to buffer controls. Similarly, the *cad* gene levels also decreased levels after one hour (12.5 hAEL) when injected with dsRNA at 11.5 hAEL. Gene expression is measured by qPCR. Error bars represent standard deviation.

Some segmentation clock and Wnt signaling genes are affected in *cad* knockdowns.

To further characterize how knocking down *cad* affects development at the three different timepoints, I examined other gene expression using qPCR. I particularly focused on the segmentation clock genes *eve*, *runt*, and *odd*, which have been described to be downstream of *cad* and to be activated by this transcription factor (Sarrazin et al., 2012; El-Sherif et al., 2014). My results show that the levels of gene expression for all these genes change during the three different timepoints (Figure 12). I also examined two Wnt signaling components (*Wnt1* and *Wnt8*), since this pathway is a known activator of the *cad* gene, and it has been suggested that feedback regulation might be present in this gene network (McGregor et al., 2009).

I found that both *eve* and *odd*, the first and third gene in the *T. castaneum* segmentation clock, were significantly decreased in each of our knockdowns (Figure 12A-C, blue and pink columns). Interestingly, the second gene in the segmentation clock *runt*, showed a slight decrease for both the 4 and 8.5 hAEL dsRNA injections, but a significant decrease was only present at the 11.5 hAEL timepoint (Figure 12A-C, yellow columns). Given that the current understanding of the *T. castaneum* segmentation clock is that *eve* activates *runt* and *runt* activates *odd* (Liao and Oates, 2017), the fact that *runt* is still active for the most part could suggest that there are other factors individually controlling the segmentation clock genes.

On the other hand, I saw some variability in the Wnt signaling genes. For instance, *Wnt1* was significantly increased in our 4 hAEL knockdowns, and it was slightly decreased (although not significantly) in our 8.5 and 11.5 hAEL knockdowns (Figure 12A-C, green columns). These results suggest that Wnt signaling might indeed have some type of feedback regulation interaction with *cad* at least early on in development, as a decrease in *cad* expression correlated with an increase in *Wnt1* expression—consistent with Ansari et al. (2018).

Additionally, the results for *Wnt8* were a bit variable, with significant decreases in the early 4 and late 11.5 hAEL experiments, but not in the intermediate 8.5 hAEL experiment (Figure 12A-C, orange columns). There is a slight decrease in the 8.5 hAEL embryos, but not as comparable as with the other time points. Therefore, these results suggest that *cad* might also be having some type of feedback regulation with Wnt signaling, since a decrease in *cad* also causes a decrease in *Wnt8* expression. However, we only have one biological replicate for this qPCR data, so we don't know the extent to which these results vary from one replicate to another.

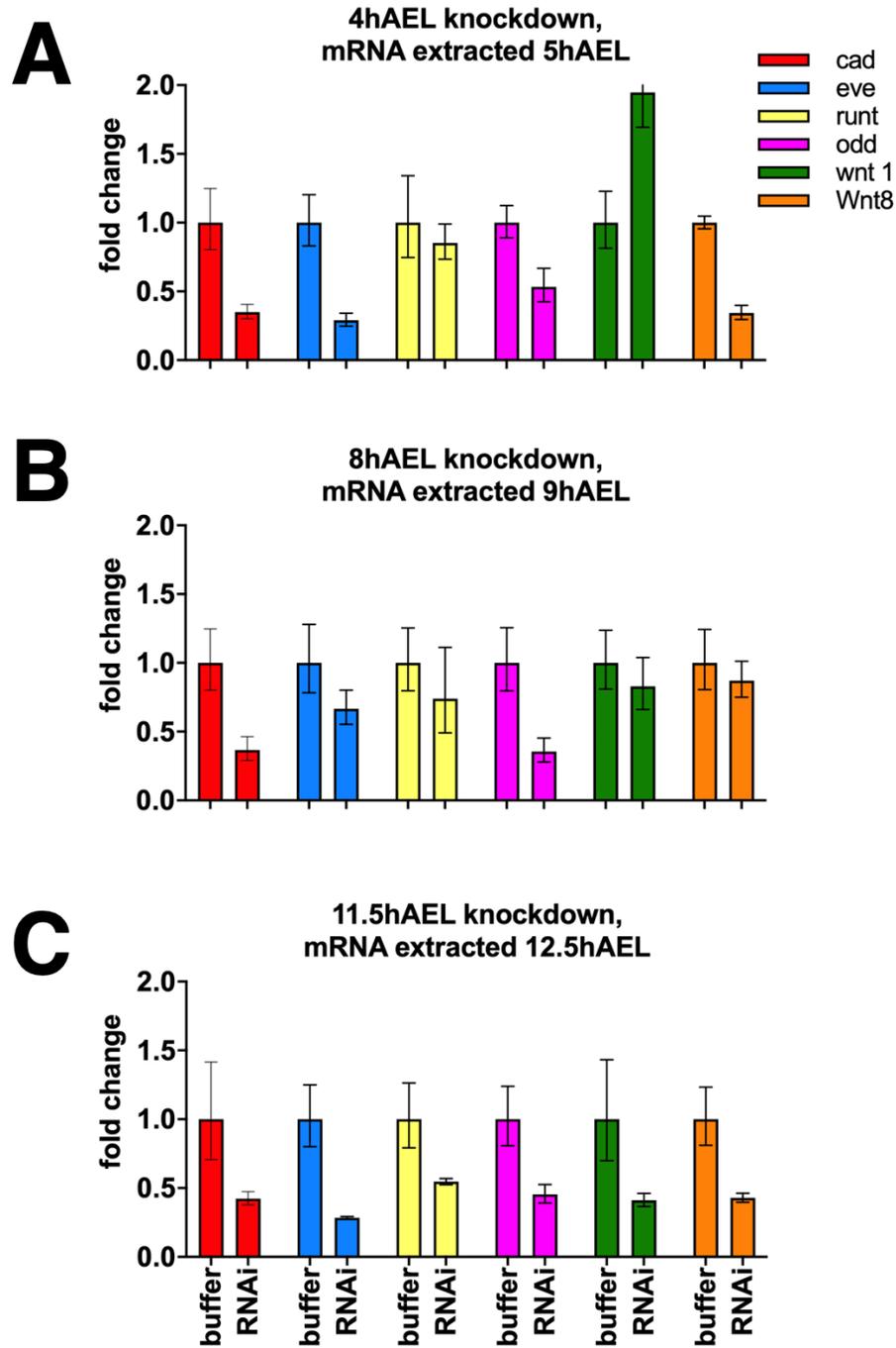


Figure 12. *cad*, Wnt, and segmentation clock gene expression after RNAi at three different timepoints. Samples were analyzed one hour after dsRNA injection. We validated our *cad* knockdowns by showing that *cad* expression (in red) decreases after RNAi for 4h, 8.5h, and 11.5h injections (A-C). The expression of two Wnt genes are shown in the green (Wnt1) and orange (Wnt 8) columns, and the expression of the segmentation clock genes *eve*, *runt*, and *odd* are shown in the blue, yellow, and pink columns. Error bars represent standard deviation.

Expression levels for Cad protein are also decreased in *cad* RNAi knockdowns.

We have shown that the gene expression levels (mRNA) goes down after our *cad* dsRNA injections at least one and two hours after injection for the 4 hAEL timepoint (Figure 8) and at least one hour after injection for the 11.5 hAEL timepoint (Figure 11). However, we know that *cad* is a transcription factor that regulates the segmentation clock at the protein level, so after looking at the later phenotypes with almost normal segmentation after RNAi, we hypothesized that it might be possible for the Cad protein to be stabilized after early translation, and that *cad* mRNA is no longer necessary after a given timepoint to maintain the function of this gene and regulate the segmentation clock. To test this hypothesis, we decided to measure Cad protein expression relative to buffer controls in our 4 and 11.5 hAEL knockdowns. The results are described in Figure 13.

The relative expression of Cad protein in the early 4 hAEL embryos is significantly reduced to around 25% of the buffer control, while the Cad protein level at 11.5 hAEL is also reduced, but only to around 40% of the buffer control levels (Figure 13). This change in CAD expression could indicate that our RNAi knockdowns are not completely effective since there is a considerable amount of protein still present that could have some function regulating the segmentation clock. It is unknown how much relative Cad expression is necessary to regulate downstream gene expression. However, the fact that there is not even half of the wild type levels of protein does not seem to correlate with the fact that we see normal segmentation in almost all of our RNAi embryos. Therefore, it still seems like our phenotypes are being caused by a change in *cad* gene function at later stages in development, not to effect of the dsRNA injections.

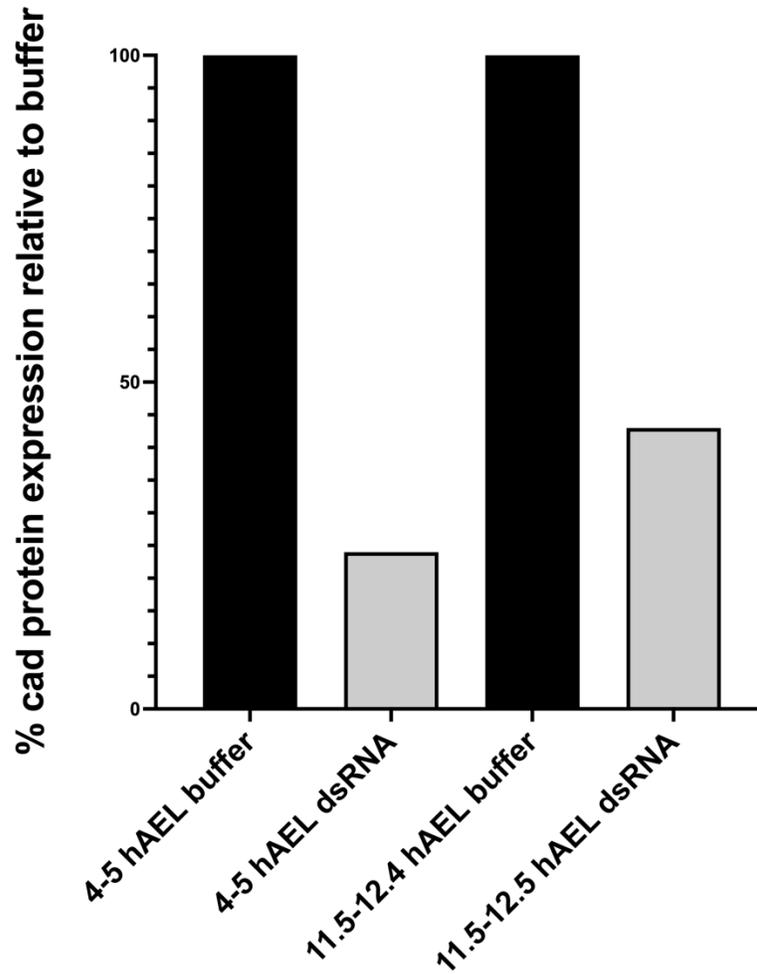


Figure 13. Cad protein expression levels for 4 and 11.5 hAEL dsRNA injections measured one hour after injection (at 5 and 12.5 hAEL, respectively). Buffer control levels have been normalized to 1.0. There is approximately only 25% of Cad protein expression for the 4 hAEL injections, while there is around 40% of Cad protein expression for the 11.5 hAEL injections.

DISCUSSION

Through a comparison of phenotypes of our *cad* dsRNA knockdowns injected at 4 hAEL to the previously described parental and early embryonic RNAi phenotypes (Figure 7), we can affirm that our clone replicates the known *cad* phenotype. Our phenotypic results were further validated with our qPCR data, which demonstrated that the *cad* gene is significantly decreased as soon as one hour after dsRNA injection, and that these levels are maintained two hours after injection when compared to buffer controls (Figure 8). In addition, we also compared the cycle data for our qPCR experiments (normalized only to the reference gene) between the knockdown and buffer treatments and showed that the number of cycles (the amplification that is necessary to detect the transcript) is significantly different for each of our knockdowns (Appendix, Figure 18). This leads to the conclusion that the amount of transcript is indeed different between knockdowns and controls.

After our RNAi experiments were validated, we decided to use this technique to study the role of *cad* at different stages of development. We were particularly interested in examining progressively later timepoints in the blastoderm (as shown in Figure 5), as we hypothesized that these subsequent times would produce embryos with progressively more segments. In addition, we also wanted to consider maternal-to-zygotic transition, which occurs at around 3 to 6 hAEL (Ribeiro et al., 2017). Interestingly, we obtained a wide range of knockdowns described in Figures 9 and Table 3, with the early 4 hAEL timepoint making up most of the Class I severe phenotypes, 8.5 hAEL embryos ranging all across the three classes, and the 11.5 hAEL embryos belonging to Class III. However, we noted that more than half of the 11.5 hAEL embryos had wild type segmentation and were comparable to our buffer controls (Table 3), which suggests that zygotic *cad* is not having the same fundamental role as maternal *cad* in

giving rise to segments. It could also suggest that the *cad* gene might not actually be necessary for segment addition after around the 10 hour-mark, which is not the current consensus in the field.

Nonetheless, as previously described, all of the information we know about *cad* function in *T. castaneum* has been inferred from experiments done in the parental generation or very early embryos as it is the standard in the field (Copf et al., 2004; Benton et al., 2013; El-Sherif et al., 2014, and some papers have cited difficulties in examining truncated phenotypes that arise from knockdown of developmental genes later in development (El-Sherif et al., 2014). To our knowledge, this is the first effort to directly characterize the role of *cad* after 4 hAEL, which could mean that the inferences that have been made based on early blastoderm experiments do not apply after maternal- to-zygotic transition. We propose that there is either a different gene network is regulating the segmentation clock, or that the instructions for making all the segments in the organism are somehow all provided before the end of maternal to zygotic transition. After all, *cad* is a transcription factor that works at the protein level, so if there is enough *cad* protein present at the time of our knockdowns, it might still be enough to regulate segment addition. We measured Cad protein expression one hour after dsRNA injections for our 4 and 11.5 hAEL embryos, and we saw that while both still have less than half of the regular dose of Cad, the later embryos do have a higher amount of protein expression.

On the other hand, it might be possible that *cad* expression is able to return to normal or near-normal expression levels a few hours after dsRNA injections and that is why we don't see severe phenotypes, particularly in our third timepoint. However, this would be highly unexpected based on the fact that RNAi in *T. castaneum* has been shown to have a very robust

systemic effect (Miller et al., 2012). We have shown that *cad* expression remains low at least two hours after dsRNA injection for the 4 hAEL timepoint (Figure 8) and at least one hour after dsRNA injection for the 11.5 hAEL timepoint (Figure 11), but more experiments are needed after these two hours after injection. The possibility exists that *cad* mRNA might be going up a few hours after dsRNA injections and regaining its normal function at any or all of our timepoints, for which we also need to evaluate gene expression at least a few hours after dsRNA injections. However, the consensus in the field is to assume that the knockdown persists—which is shown in parental RNAi such as in Copf et al. (2004) and Choe et al. (2006)—but we will be looking at this in future experiments. We have now successfully extracted the mRNA three hours after dsRNA injections, and future qPCR experiments will be aimed at verifying our knockdowns at this additional timepoint in development. Additionally, it is important to note that the 11.5 hAEL RNAi knockdown coincides with the time of cellularization (Benton et al., 2013). One possible interpretation for our data is that when cellularization occurs, there is a refractory period in which the dsRNA is not able to get into the cells. However, we do not believe this is the case based on the robust systemic RNAi effect previously mentioned (Miller et al., 2012), plus our qPCR and dot blot results where we see that the *cad* mRNA and protein are reduced, respectively. Miller et al. (2012) injected embryos at the larval stage and investigated whether some tissues were more dsRNA-resistant than others, and they did not find evidence for this.

Moreover, I also examined the segmentation clock genes *eve*, *runt*, and *odd*, and the Wnt signaling genes *Wnt1* and *Wnt8* to study how knocking down *cad* affects genes known to be interacting with *cad*. My results further reinforced the claim that *cad* is indeed regulating segmentation clock genes as all of them also decreased their expression (although it is

important to note that this decrease in expression was not significant for *runt* at all time points) (Figure 12).

The fact that *runt* expression is only significantly different in the late 11.5h hAEL timepoint could support our hypothesis that there are other genes regulating the segmentation clock before maternal to zygotic transition, and that these genetic network changes by 11.5 hAEL. However, it was very surprising to us to notice that *eve* levels are also reduced in the *cad* knockdown—after seeing our phenotypes, we made the preliminary conclusion that *cad* might simply not be regulating *eve*, since the current segmentation model establishes that *eve* is necessary to pattern segments. The fact that embryos with decreased *eve* expression are still able to segment normally for the most part suggests that it might be *eve*, or the segmentation clock altogether, that doesn't have a function patterning segments after around 11.5 hAEL. We are currently planning experiments to stain 11.5 hAEL dsRNA injected embryos with *eve* antibodies to see if there is protein expression that might be giving rise to the segments that we see. Our multiple results on *cad* gene function and expression support our hypothesis that *cad* is changing its role throughout development, whether through a difference in regulation or to a difference in the function of its downstream targets such as the genes in the segmentation clock.

It would also be ideal to repeat the qPCRs with multiple biological replicates to verify our results. We repeated the 8.5 hAEL qPCR twice and obtained comparable expression levels (only one qPCR is shown in Figure 12B). However, the repetition of this experiment was done with the same biological sample in three technical replicates for each qPCR, so there could also be some experiment-to-experiment variation that can't be accounted for. Additionally, for the 4 hAEL and 11.5 hAEL qPCRs, we also only ran a single biological sample with three

technical replicates, and each qPCR was only run once (Figures 12A and 12C). Therefore, the results of these qPCRs might be representative of the expression levels for a given biological sample, but not for all. At the same time, Wnt signaling genes were also affected, which also supports the claim that Wnt and *cad* could possess a feedback regulation mechanism (McGregor et al., 2009). Based on the fact that *Wnt1* increased its expression and then decreased it after *cad* knockdowns (although not significantly), and that *Wnt8* also had some decrease in expression, we propose that the interactions of Wnt signaling and *cad* could be changing with time, and maybe establishing a negative feedback regulation loop early in development that turns into a positive feedback loop later on. This interpretation could also support our claim of a change in genetic regulation after maternal-to-zygotic transition, but further experiments are needed to validate or refute this idea.

It has been shown that in *D. melanogaster*, primary pair-rule genes (*hairy*, *eve*, *runt*, *odd*, and *ftz*) turn on earlier, and some or all of these genes are also expressed in the posterior growth zone in sequentially segmenting arthropods. In both cases, secondary pair-rule genes are turned on later in the anterior (paired and sloppy paired) (Clark et al., 2019). The fact that secondary pair-rule genes can start being expressed later on in similar species might suggest that they behave similarly in *T. castaneum*, and that they might take on different roles (Choe et al., 2006; Janssen et al., 2020). Even if they are only expressed in the anterior, they could still be regulating other genes that control segment addition in the growth zone. Therefore, more research is necessary to examine these pair-rule genes as well—perhaps the segmentation clock switches its regulations at some point in development from *cad* to a pair-rule gene, or the segmentation clock stops adding segments after a certain point and other genes take over the segment addition function.

On the other hand, an early RNAi screen in *T. castaneum* showed that *eve*, *runt*, and *odd* knockdowns resulted in severe truncations (Choe et al., 2006) as has been shown with *cad* (Copf et al., 2004). However, this early RNAi showed that knocking down *hairy*, a gene that is known to be involved with segmentation in *D. melanogaster* and other species, only leads to mild head defects in *T. castaneum* (Choe et al., 2006). However, it is possible that *hairy* might be having a function in *T. castaneum* that is redundant with another gene and unmasked with the *cad* knockdowns, for which more experiments should aim at characterizing this gene. In addition, previous experiments describe the fact that the rate of segment addition is not constant throughout development in *T. castaneum* (Brena & Akam, 2013; Nakamoto et al., 2015), which has been hypothesized to be caused by stage-specific variations in the oscillation period or the dynamics of tissue maturation in the posterior growth zone (Clark et al., 2019). It might be that any of these variations are caused by a change in the genetic regulatory framework that is yet to be elucidated.

The function of the *eve* gene also needs to be further examined. For instance, it has been proposed that *eve* might only be necessary for establishing and/or maintaining the growth zone in sequentially segmenting species such as *T. castaneum* (Liu and Kaufman, 2005; Mito et al., 2007; Cruz et al., 2010; Xiang et al., 2017), and that knocking it down causes severe truncation phenotypes because of this important role, which might be independent of a potential role in the segmentation clock (Clark et al., 2019). Therefore, it becomes necessary to distinguish between these two possible functions for *eve*—and maybe even considering the fact that these functions can be dynamic. For instance, it is possible that *eve* does have a role in the segmentation clock early in development, but that this role later switches to other genes and that *eve* exclusively becomes important for maintaining the growth zone only. Therefore,

even if *cad* is still regulating *eve* with the same function throughout development, it might be that the function of its downstream targets such as *eve* might be changing later, which could account for the results we obtained in our experiments.

Moreover, the possibility of segmentation being regulated by more than one single circuit has been considered before, and Notch signaling has been proposed as a candidate to examine as it has been shown to be involved in development in some arthropods (Williams & Nagy, 2017). Finally, it has been proposed that genes known as “timing factors”, including *Dichaete* and *opa*, might have a function in regulating *cad* and establishing and maintaining the growth zone (Clark et al., 2019), so variations in these genes could also be affecting *cad* expression in some way that is yet to be examined.

In conclusion, there are multiple possibilities as to why the *cad* gene might be having different effects throughout development—which can be caused by a change in either the genes involved in its regulation, its downstream targets, or its own function. More research is necessary to elucidate these mechanisms, and to determine if the proposed segmentation clock in *T. castaneum* is the only genetic framework that regulates segment addition. Our results support the previously proposed hypothesis that more than one genetic circuit might be involved in segmentation, which is not surprising as the development of any organism is such a complex and dynamic process. Therefore, further understanding the role of the *cad* gene in *T. castaneum* merits additional consideration and research as it could help us learn more about the development and genetics of this species and of arthropods, which are the most abundant phyla in the planet. Moreover, sequentially segmenting organisms are also present in annelids and chordates—some of the lessons we learn from arthropods can influence our understanding of our own species.

CHAPTER 2: REGULATION OF THE *CAD* GENE

INTRODUCTION

We currently know that *cad* is part of an intricate regulatory network of genes that regulate anterior-posterior axis patterning early in the embryo (Ansari et al., 2018), while also contributing to elongation (Benton et al., 2013). However, more information is needed about its mechanism of action. The regulation of the posterior by Wnt and *cad* signaling is known to be important across a wide range of metazoans, so knowing more information about *T. castaneum* can help us understand development in other species (Martin & Kimelman, 2009).

Previous research has suggested that *T. castaneum* develops using a segmentation clock in which gene expression waves oscillate to give rise to various segments at different frequencies. It is believed that in this clock, the gene *eve* activates *runt*, which activates *odd*, which inhibits *eve* (Figure 2). Additionally, evidence indicates that these oscillations are driven by a gradient of the transcription factor *cad*, which can both activate and regulate the frequency of clock oscillators (El-Sherif 2014). Since it is known that the frequency of the clock changes during development (Nakamoto et al., 2015), it is possible that such changes are being regulated by the *cad* gene. Therefore, we decided to look for transcription factors that might regulate embryonic *cad* expression and therefore drive these segmentation changes in *T. castaneum*.

Transcription factors are proteins that regulate gene expression at the level of transcription by binding to an enhancer region of a gene as the mRNA transcript is being produced (Spitz and Furlong, 2012). Through this regulation, a gene can be expressed at high

or low levels depending on the combination of transcription factors present in a given sequence. When transcription factors bind to an enhancer sequence, they allow for the enhancer to bind to the promoter, and then recruit RNA polymerase to start transcription. These enhancer regions are usually considered to be located upstream of the gene promoter. However, there is also evidence to suggest that these enhancers can be found in other regions of the genes that aren't transcribed, such as intronic and downstream regions (Venables, 2007; Xiang 2009). We have hypothesized that regulation of the *cad* genes is mediated by enhancer regions located in the upstream region and elsewhere, for which we have proposed to establish putative enhancer regions in other parts of the gene, namely intronic and downstream regions. I decided to conduct a more extensive literature review and look for transcription factors that could be regulating *cad* expression, including those shown in Figures 2 and 3.

To determine putative enhancer regions in the *cad* gene sequence, I used the Motif-Cluster Alignment and Search Tool MCAST (Bailey & Noble, 2003). This is a tool that allows us to predict where groups of transcription factors are likely to bind in a given gene sequence allowing us to hypothesize these transcription factors could be regulating that gene there. Therefore, we can use these predicted clusters to determine putative enhancer regions. I conducted MCAST analyses within the entire *cad* gene sequence.

MATERIALS AND METHODS

DNA sequence and nucleotide frequency for *cad* gene

I obtained its full gene sequence in FASTA format for *cad* gene (TC032769), from the database Ensembl Metazoa (<https://metazoa.ensembl.org/index.html>). For the upstream boundary, we considered around halfway up the next gene in the sequence. I determined the frequency of each nucleotide in our DNA sequence using the EMBOSS word count tool (<https://www.bioinformatics.nl/cgi-bin/emboss/wordcount>): the word size was set to 1, and the DNA sequence was pasted into the input sequence option 2. The output was the total number for each of the four nucleotides in DNA (A, T, G, C) which was then divided by the total nucleotide length of our sequence. The nucleotide frequency is necessary information for our bioinformatics analyses, as it improves the accuracy of the predictions of the binding algorithm.

Candidate transcription factors

My selection of possible transcription factors was drawn from regulators of *cad* based on genetic studies. I considered transcription factors such as *hunchback*, components of the Wnt signaling, and *homeobrain* (Schulz & Tautz 1995; McGregor et al., 2009; Ansari et al., 2018). Additionally, I considered studies from *D. melanogaster* where it has been shown that several transcription factors such as *DRE/DREF* and are involved in regulating the *cad* homologue (Choi et al., 2004;). I also considered timing factors such as *Dichaete* which have been hypothesized to be involved in *cad* regulation, (Clark et al., 2019) and pioneer factors like *STAT92E* and *Zelda/vfl* I expected to be present early in development since they are essential activators of zygotic genes (Tsurumi et al., 2011). Moreover, I also selected other

genes based on the fact that one of our previous bioinformatics searches had identified them as having high probability of binding in the *cad* gene region. I decided to include *cad* itself since we hypothesized that self-regulation might be an aspect of *cad* regulation, and also looked at the segmentation clock genes *eve*, *runt*, and *odd* to look for feedback regulation. The full list of transcription factor candidates we considered based on our literature search and previous data from collaborators are described in Table 4.

Table 4. List of transcription factors examined in our bioinformatics analysis, including the biological process they help regulate and support for inclusion.

Biological process	Transcription factor	Support for inclusion
Wnt signaling pathway	pan2	Literature search (Gaunt et al., 2003)
early dorso-ventral patterning	dl	Literature search (Choi et al., 2008)
general regulators	Stat92E	Pioneer factor (Tsurumi et al., 2011)
	vfl	Pioneer factor (Tsurumi et al., 2011)
anterior-posterior axis specification	cad1	Literature search / hypothesized self-regulation (Ansari et al., 2018)
	D	Timing factor (Clark et al., 2019)
	eve	Segmentation clock gene (El-Sherif et al., 2014; Sarrazin et al., 2012)
	hbn	Literature search (Ansari et al., 2018)
	hb	Literature search (Schulz & Tautz 1995)
	oc	Timing factor (Clark et al., 2019)
	prd	Literature search (Zhao et a., 2014)
segment identification	abd-B	High probability of binding based on previous bioinformatics results
	Antp	High probability of binding based on previous bioinformatics results

	Ubx	High probability of binding based on previous bioinformatics results
other transcription factors	Dref	Literature search (Choi et al., 2004)
	B-H1	High probability of binding based on previous bioinformatics results
	br	High probability of binding based on previous bioinformatics results
	br(var.3)	High probability of binding based on previous bioinformatics results
	br(var.4)	High probability of binding based on previous bioinformatics results
	CG11617	High probability of binding based on previous bioinformatics results
	exd	High probability of binding based on previous bioinformatics results
	fkh	High probability of binding based on previous bioinformatics results
	HHEX	High probability of binding based on previous bioinformatics results
	kni	High probability of binding based on previous bioinformatics results
	nub	High probability of binding based on previous bioinformatics results
	onecut	High probability of binding based on previous bioinformatics results
	slp1	High probability of binding based on previous bioinformatics results
	su(Hw)	High probability of binding based on previous bioinformatics results

Weight matrix model on the binding sites of a transcription factor from *D. melanogaster* genes models

After establishing the transcription factors of interest, I obtained their motif binding sites models from JASPAR (<https://jaspar.genereg.net/>). For each transcription factor, there is a specific JASPAR model that gives its binding specificity. Since no models exist for *T.*

castaneum, we used *D. melanogaster* models for these transcription factors that have been used in over 700 research articles examining gene regulation, including in other species. These are binding motifs that have been experimentally verified, and even though they are *T. castaneum* binding motifs, it has been shown that transcription factor specificities are highly conserved in bilaterians (Nitta et al., 2015), hence we decided to move forward with our analyses using these models. JASPAR allowed us to obtain information related to binding probability given a specific genetic sequence in a MEME format. I then added the *cad* nucleotide frequency and the JASPAR model information for each of the genes of interest into a weight matrix model formatted as text file as shown in the Appendix. Dr. Lisa Nagy and Dr. Benjamin Goldman-Huertas, our collaborators in Arizona, kindly provided a template for this weight matrix to which we could add our data to. Once I had the information about the gene sequence and the transcription factor candidates to test, I was able to proceed with the bioinformatics tool MCAST as described in Figure 14.

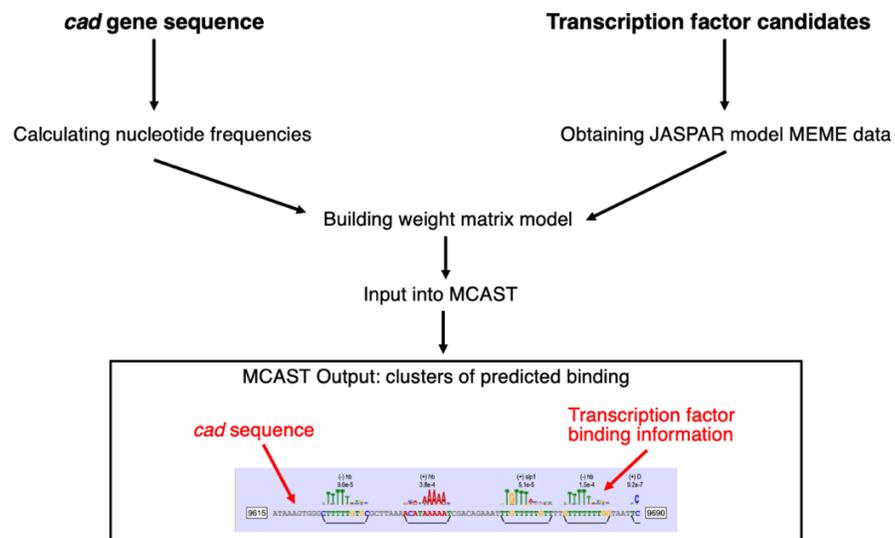


Figure 14. Pipeline of our bioinformatics analysis describing the different steps we followed.

MCAST tool from MEME Suite

I used the bioinformatics tool MCAST (<https://meme-suite.org/meme/tools/mcast>) to predict where clusters of transcription factors are more likely to bind, which would suggest the location for a putative enhancer region. MCAST has been used for almost 20 years as a tool to scan for *cis*-regulatory motif clusters (Bailey & Noble 2003; Grant et al., 2016), and it has been shown to be the best computational model to predict putative enhancer regions (Jayaram et al., 2016). To look for binding, I submitted the entire *cad* gene sequence. MCAST predicted the regions where transcription factors were most likely binding to the *cad* sequence, which were then organized into clusters that we could examine.

RESULTS

A final computational output for predicted transcription factor binding clusters was developed.

I first calculated the nucleotide frequencies for the *cad* gene as previously described (Table 5) and then constructed a weight matrix model on the binding sites of a transcription factor from *D. melanogaster* genes JASPAR models (Appendix). I conducted my own MCAST runs with different sets of transcription factors in order to identify regions that were predicted to have a high number of binding clusters.

Table 5. Calculated nucleotide frequencies (A, C, G, and T) for the *cad* gene sequence.

Nucleotide	Individual count out of 33001	Calculated frequency
A	10992	0.3330808157
C	5427	0.1644495621
G	5361	0.1624496227
T	11221	0.3400199994
Total	33001	1

The transcription factor models we used for each of our MCAST runs are listed in Table 6. The first MCAST that we ran included a total of 16 JASPAR models (corresponding to 16 transcription factors), all of which we obtained exclusively from our literature review (described in Table 4). However, MCAST gives more accurate predictions if 25-30 models are considered since there is more information with which to build the clusters. So, we decided to consider other groups of transcription factors that had high probability of binding based on a previous bioinformatics analysis (also described in Table 4). For our second MCAST run, we considered a total of 32 models as described in Table 6. These JASPAR models were the same

for the transcription factors we had already considered—we simply added other models for additional transcription factors.

Table 6. Description of each of our MCAST runs.

Gene	JASPAR number	MCAST 1	MCAST 2	MCAST 3	MCAST 4
		TF based on Tc and Dm cad regulators from the literature	TF from MCAST1 plus unreported TF with high frequency hits	Repeat of MCAST1, omitting 4 TF not supported by the details of motif binding	Repeat of MCAST3, omitting 4 TF not supported by the details of motif binding
cad1	MA0216.1	Yes	Yes	Yes	Yes
cad2	MA0216.2	Yes	Yes	NO	NO
D	MA0445.1	Yes	Yes	Yes	Yes
Dref	MA1456.1	Yes	Yes	Yes	Yes
eve	MA0221.1	Yes	Yes	Yes	Yes
hbn	MA0226.1	Yes	Yes	Yes	Yes
hb	MA0049.1	Yes	Yes	Yes	Yes
dl	MA0022.1	Yes	Yes	Yes	Yes
oc	MA0234.1	Yes	Yes	Yes	Yes
pan1	MA0237.1	Yes	Yes	NO	NO
pan2	MA0237.2	Yes	Yes	Yes	Yes
prd	MA0239.1	Yes	Yes	Yes	Yes
Stat92E	MA0532.1	Yes	Yes	Yes	Yes
vfl	MA1462.1	Yes	Yes	Yes	Yes
zen1	MA0256.1	Yes	Yes	NO	NO
zen2	MA0257.1	Yes	Yes	NO	NO
abd-B	MA0165.1	NO	Yes	NO	Yes
Antp	MA0166.1	NO	Yes	NO	Yes
B-H1	MA0168.1	NO	Yes	NO	Yes
br	MA0010.1	NO	Yes	NO	Yes
br(var.3)	MA0012.1	NO	Yes	NO	Yes
br4(var.4)	MA0013.1	NO	Yes	NO	Yes
CG11617	MA0173.1	NO	Yes	NO	Yes
exd	MA0222.1	NO	Yes	NO	Yes
fkh	MA0446.1	NO	Yes	NO	Yes
HHEX	MA0183.1	NO	Yes	NO	Yes
kni	MA0451.1	NO	Yes	NO	Yes

nub	MA0197.2	NO	Yes	NO	Yes
onecut	MA0235.1	NO	Yes	NO	Yes
slp1	MA0458.1	NO	Yes	NO	Yes
su(Hw)	MA0533.1	NO	Yes	NO	Yes
Ubx	MA0094.2	NO	Yes	NO	Yes

I considered all of the predicted binding clusters with an E-value of less than ten, an arbitrary value suggested by the program to have the highest probability of a cluster actually occurring in nature (the lower the number, the higher the probability). My MCAST 1 run produced a total of 13 predicted binding clusters all along the *cad* gene sequence as described in Figure 15 (dark green), while our MCAST 2 run (in pink) predicted 11 binding clusters for the same region. Interestingly, there was an overlap of four of the MCAST 1 clusters with MCAST 2 clusters, which suggests that the predictions do change as a function of the transcription factor models being considered. It is important to note that each of the clusters in MCAST 2 extended a longer distance in the gene sequence, which meant that there are more transcription factor binding predictions per cluster (this was also verified by counting individual transcription factors in a predicted cluster). Ten of the 13 predicted clusters in the MCAST 1 run overlapped with our five putative enhancer regions that were already established, while only six out of the 11 predicted clusters in the MCAST 2 run overlapped with these regions. Those putative enhancer regions were originally established with over 60 gene models that were not handpicked as in our literary review, which could suggest that our predictions are slightly more accurate since we considered at least half of the models for transcription factors that regulate *cad* function.

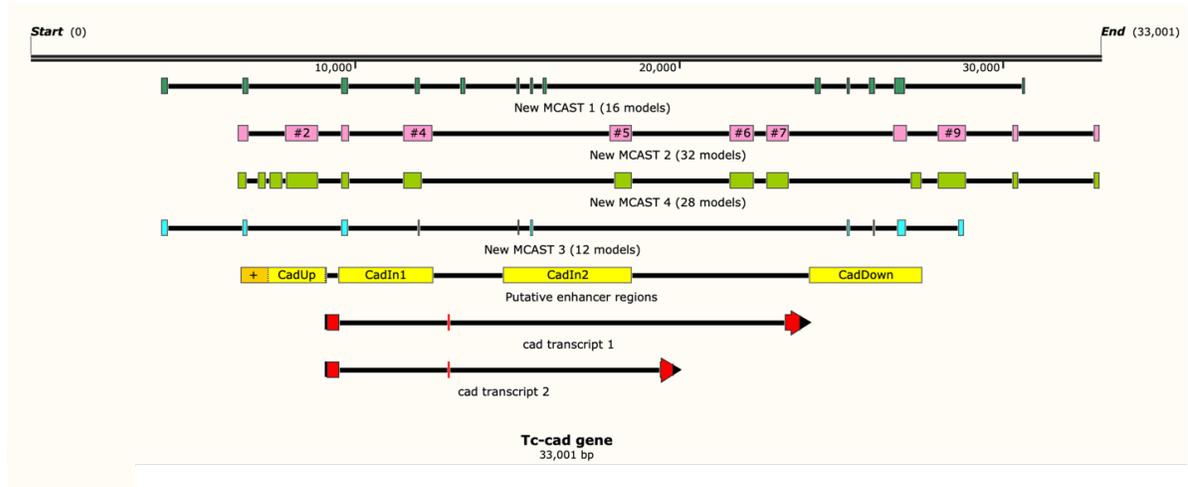


Figure 15. *cad* gene sequence with annotations for its two transcripts (exons in red), our five putative enhancer regions (yellow), and our four MCAST runs: MCAST 1 (16 models, dark green), MCAST 2 (32 models, pink), MCAST 3 (28 models, light blue), and MCAST 4 (12 models, light green).

After further consideration of our predictions and the transcription factor models, I was using, I decided to eliminate some of our models and conduct two further MCAST runs. I first looked at the two *cad* models (*cad1* and *cad2*), which I realized were both models for the *cad* transcription factor but had different binding motifs. A previous study stated that *cad1* was a newer model (Noyes et al., 2008) and was therefore more refined, so we decided to eliminate *cad2* from our subsequent MCAST runs. Similarly, *pan1* corresponds to non-canonical Wnt signaling while *pan2* corresponds to the canonical version. Since it has been shown that only canonical Wnt signaling regulates *cad* function (McGregor et al., 2009), I also removed *pan1* from our future analyses. Finally, we realized that the regulation of *zen* in *T. castaneum* acts through a different gene (Ansari et al., 2018), for which we decided to omit the *zen* models from our analyses. Based on all these new developments, I ran the MCAST analysis two more times: one with the same JASPER models as MCAST 1 without *cad2*, *pan1*, *zen1* and *zen2*

(referred to as MCAST 3, in light blue on Figure 15), and another one with the same JASPER models as MCAST 2 without those four models (MCAST 4, in light green on Figure 15).

MCAST 3 had a total of 10 predicted binding clusters, eight of which overlapped with our putative enhancer regions and nine of which overlapped with MCAST 1 (Figure 15). It was encouraging that this new run mostly overlapped with our putative enhancer regions, as our lab has cloned them and is currently using those regions for cross-species transgenics analysis and yeast-one hybrid experiments.

The predicted binding clusters in MCAST 3 also greatly overlapped with most of the MCAST 1 clusters, but MCAST 1 included three additional clusters. I hypothesized then that MCAST 3 is a more refined version of MCAST 1, and includes models that are likely more accurate to our analysis. On the other hand, MCAST 4 includes 13 predicted binding clusters, eight of which overlap with our putative enhancer regions and 10 of which overlap with MCAST 2. The results of this last run were also a bit encouraging, since there is also a significant overlap with our putative enhancer regions which we are using for other experiments, and also because this is a likely more refined version of MCAST 2. The direct output of this last experiment in MEME Suite website is shown in Figure 16.

I concluded that MCAST 4 would give a good indication of the predicted binding clusters to examine putative transcription factors, since it contains both models backed up by the literature and other models with frequent binding, as well as a total of 28 models which was within the desired range of models to work with.

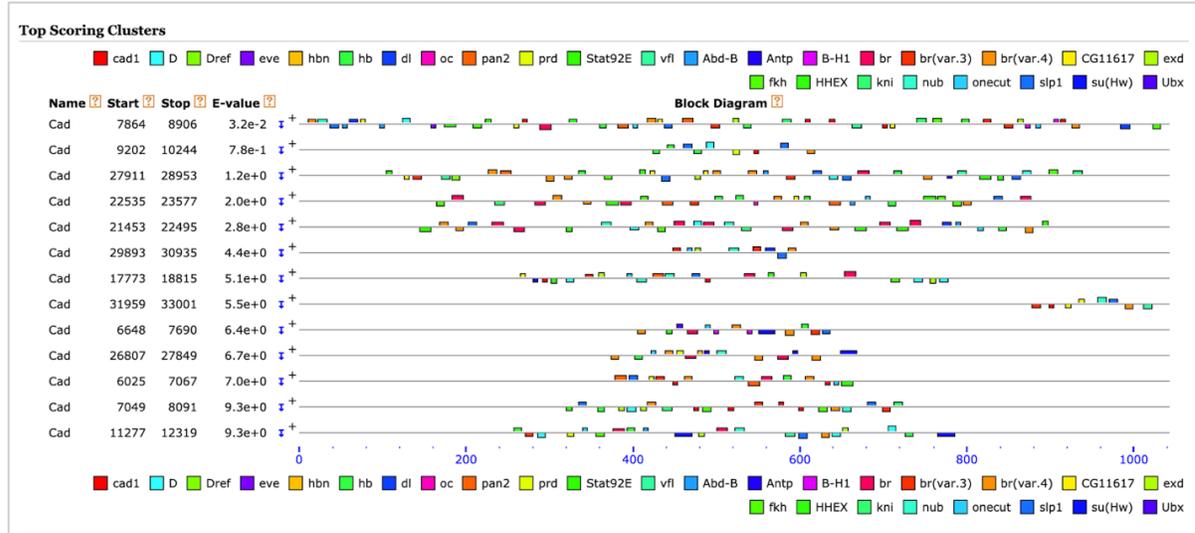


Figure 16. MCAST output with 13 clusters (horizontal black lines), each with a different cluster composition (various transcription factors make up each cluster). E-value indicates the “accuracy” of the cluster as established by the program; values lower than 10 were considered to be the most accurate.

Predicted transcription factor binding clusters can be used to establish putative enhancer regions.

Through my four MCAST runs, I further analyzed the results of previous MCAST runs and verified our putative enhancer regions. It was found that most of the clusters in any of our four runs overlap with these regions we cloned, and we specifically focused our attention at analyzing the gene composition of the clusters in the last run, MCAST 4. I counted each of transcription factor models shown in Figure 3 and constructed a matrix indicating the frequencies of each model per cluster, which has been color-coded and shown in Figure 17.

Biological process	Model	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	Total
Wnt signaling pathway	pan2	2	1	2	2	1	1	2	1	1	1	2	1	1	8
early dorso-ventral patterning	dl	2	1	1	1	2	2	1	1	1	1	1	1	1	4
general regulators	Stat92E	1	1	1	1	1	1	1	1	1	1	1	1	1	8
	vfl	1	1	1	1	1	1	1	1	1	1	1	1	1	5
anterior-posterior axis specification	cad1	3	2	2	2	1	1	2	1	1	2	1	2	1	15
	D	1	1	1	1	2	1	2	1	1	1	1	1	1	13
	eve	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	hbn	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	hb	3	2	2	2	1	1	1	1	1	1	1	1	1	20
	oc	1	1	1	1	1	1	1	1	1	1	1	1	1	0
	prd	1	1	1	1	1	1	1	1	1	1	1	1	1	4
	abd-B	1	1	1	1	1	1	1	1	1	1	1	1	1	6
segment identification	Antp	1	1	1	1	1	1	1	1	1	1	1	1	1	4
	Ubx	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Dref	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	B-H1	1	1	1	1	1	1	1	1	1	1	1	1	1	4
other transcription factors	br	1	1	1	2	2	1	2	1	1	1	1	1	1	21
	br(var.3)	2	1	1	1	1	1	1	1	1	1	1	1	1	15
	br(var.4)	2	1	1	1	1	1	1	1	2	2	1	1	1	35
	CG11617	2	1	1	1	1	1	1	1	1	1	1	1	1	11
	exd	1	1	1	1	1	1	1	1	1	1	1	1	1	11
	fkh	2	1	1	2	1	1	1	1	1	1	1	1	1	14
	HHEX	1	1	2	1	1	1	1	1	1	1	1	1	1	11
	kni	1	1	1	1	1	1	1	1	1	1	1	1	1	7
	nub	1	1	1	1	1	1	1	1	1	1	1	1	1	11
	oncut	1	1	1	1	1	1	1	1	1	1	1	1	1	10
	slp1	2	1	1	1	1	1	1	1	1	1	1	1	1	18
su(Hw)	1	1	1	1	1	1	1	1	1	1	1	1	1	4	
Total		45	9	34	27	29	8	21	8	13	14	14	20	21	263



Figure 17. Number of JASPAR model hits per cluster in MCAST 4. Colors indicate the frequencies of models, with darker blue colors indicating less frequent models and darker red colors indicating more frequent models. Clusters 1-13 are abbreviated as C1-C13.

As possible to see from a first glance, each of the predicted clusters has a unique transcription factor composition. Most of the models are present in lower frequencies (0-2), although there are a couple of models present 3 or more times per cluster (Figure 17). Clusters 1, 9, and 12 are in the upstream gene clone, while cluster 10 is in our downstream gene clone. The fact that the cluster with the higher number of binding predictions in the upstream region is expected, as genes are usually regulated with enhancers located upstream of the gene. However, my predictions also suggest that there might be enhancers located all throughout the intronic and downstream regions.

Interestingly, I saw that *cad1* was one of the transcription factor models that was more frequent in our total cluster count, which supports the hypothesis that *cad* is self-regulation. On the other hand, there was notably a very low frequency of *eve* binding (only once in all 13 clusters), which might mean that segmentation clock genes are not feeding back to regulate *cad* as I had proposed. Moreover, the consistent frequency of models for *hb* and *pan*, two notable regulators of the *cad* gene, suggests that our MCAST predictions are accurate within the limitations of a computer estimation.

However, *hbn*, another known regulator of *cad* at least in early development (Ansari et al., 2018) is present only one time in all of the 13 different clusters; while *br* and its different variations (3 and 4) are present at considerably high frequencies when there is no evidence to suggest that this transcription factor is somehow regulating *cad*. These findings might also point out limitations in our study, since some genes I expect to regulate *cad* are underrepresented, while some genes that have not been proven to regulate *cad* are overrepresented. However, it is impossible to conclude without experimental results whether predicted frequency correlate with the degree of regulation that any given gene exerts on *cad*.

DISCUSSION

Through the use of our MCAST tool, I have been able to establish predictions of binding clusters for transcription factors in the *cad* gene sequence. I considered four different combinations of models that represent transcription factors through my analyses. The models are the same for a given transcription factor in each of the run—what varied was the combination of transcription factors in each analyses as per previous reasoning. All of these models represent transcription factors that have been shown in the literature to regulate the *cad* gene (plus the JASPAR models have been experimentally verified for *D. melanogaster*), and we added some extra models which have been shown to have high frequency binding. MCAST 1 consisted only of models from the literature, and we believe MCAST 3 further refined these models to those that have actually been shown to provide more accurate predictions. On the other hand, MCAST 2 consisted of all these literature modes, plus additional models that have high frequency of binding and were used for the original bioinformatics predictions established by our collaborators at the University of Arizona. We added these other models to use between 25 and 30 models for a more effective analysis. MCAST 4 was simply what we considered a more refined version of MCAST 2 omitting the models that were not useful for our predictions. Therefore, we decided to use MCAST 4 to do a more in-depth analysis of transcription factor composition in each of the predicted clusters. However, this was an arbitrary decision (driven by the fact that this run considered all of the different factors previously mentioned), but there is no evidence to suggest that this one analysis is more accurate than the other ones. Therefore, different conclusions can be drawn from other MCAST runs, which could change a bit the inferences we made on the effect of certain transcription factors in regulating *T. castaneum* development.

Importantly, it is also imperative to note that all of these data are the results of a computer prediction, and not information produced in the laboratory setting. Therefore, we can make inferences using our results, but they need to be verified through experiments to make more substantial conclusions. Additionally, all of the JASPAR models are *D. melanogaster* gene sequences and not *T. castaneum*, yet the *cad* gene sequence that we used belongs to *T. castaneum*. Most of these genes are likely conserved in these different species since they both belong to the Arthropoda phylum which has many genetic similarities (McGregor et al., 2009), but it has been shown that *D. melanogaster* is a highly derived organism that develops differently from most arthropod species. Therefore, the *D. melanogaster* transcription factors might be regulating the *cad* gene in *D. melanogaster* differently than in *T. castaneum*—so our results should also be interpreted with that caveat in mind. The binding motifs are likely the same for both species, but the gene function and regulation might vary.

Nonetheless, it is still interesting to consider the possible implications that our results suggest, such as the fact that *cad* might be self-regulating itself, or that *br* and its different variations might be regulating *cad*, which has not been shown in the literature. At the same time, the fact that our putative enhancer regions overlap with these clusters and that we are making progress in our lab with our cross-species transgenics, and yeast-one hybrid experiments could mean that we might be able to interpret these predictions soon, and be able to validate or refute them. In either case, having these computational predictions mean that we do not have to start with a broad survey encompassing all possible transcription factors, which is time and cost-effective. I found some predicted binding domains in intron 2, between exons 2 and 3, for both MCAST 2 and 3 and have cloned that region for creating transgenic reports. Importantly, we used the results of this MCAST analysis not only to confirm our chosen

transgenic constructs but also to design our yeast one-hybrid bait sequences as the entire putative enhancer constructs are too long to serve as Y1H bait. So, we designed our baits fragments to include predicted MCAST clusters, in hopes of capturing binding in the yeast one-hybrid assay.

Besides the practical applications of these predictions in our on-going experiments, performing these bioinformatics analyses also provided a highly valuable exercise to learn about transcription factors and *cad* regulation through our literature review, and to learn how to use the MCAST tool. This was particularly important as it was a project undertaken at the start of the COVID-19 pandemic where we could not be physically in lab doing experiments, so these analyses allowed us to continue with our research and start to plan and think about the experiments we wanted to conduct later on.

CONCLUSION

Through this project, I was able to expand the current understanding on the role and regulation of the *cad* gene in *T. castaneum* segmentation. My results show that the function of the *cad* gene changes throughout development, as knocking down *cad* with RNAi later in segmentation does not produce the truncated phenotypes described in the literature for early embryos. We verified our knockdowns using qPCR and dot blot experiments, both of which show that the *cad* gene is effectively decreased both at the mRNA and protein levels. Therefore, we believe that the results we see are not likely due to an ineffective RNAi technique or to the effects of cellularization, as *T. castaneum* possesses a very robust RNAi systemic response (Miller et al., 2012), which works in larvae and adults with fully formed cells (Copf et al., 2004; Choe et al., 2006). In addition, the fact that our mRNA and protein levels are significantly decreased after dsRNA injection support this interpretation. RNAi has been proven to be a powerful technique that is widely utilized for understanding gene function, although there are many variables to consider our results and further experiments should address their limitations. Additionally, I was able to show through qPCR that knocking down the *cad* gene does affect the expression of other genes such as those in the Wnt signaling pathway and the segmentation clock, consistent with the role of Cad as a transcription factor involved in *T. castaneum* development. However, these results still do not explain the phenotypes I observed—at least if we think of segmentation in terms of the current model. Therefore, I propose that inferences made on the entire segmentation process based on these early results need to be revised and further experiments need to describe this change in function in more detail.

Moreover, I conducted bioinformatics analyses to further elucidate regions in the *cad* gene sequence where enhancer could be located. These predictions informed further experiments conducted in our lab, particularly the design of bait constructs that are now being used for yeast one-hybrid analysis to identify transcription factors that are interacting with the *cad* gene sequence. Throughout the output I obtained, I was also able to make predictions on the interactions of *cad* with other genes, such as the fact that the gene might possess self-regulation, but not be feedback regulation with other clock genes such as *eve*.

Both the experimental data and the computational predictions demonstrate that *cad* is a fascinating yet complex component of *T. castaneum* development. Even if we only consider its role and regulation in segmentation, my data challenges the current understanding of this gene. Therefore, it is imperative that further experiments dissect this gene from a functional and regulatory point of view to truly understand segmentation in arthropods. We have known about the existence of a segmentation clock in arthropods for around 20 years now, but it has been rather recently where we have been able to understand more about its mechanism of action. With the current advances in developmental techniques in the last few years, it is truly a possibility to embark on a quest to characterize this gene as thoroughly as possible. This would not only help in providing information about basic biology in this species, but it would also allow us to draw inferences for different arthropods, and possibly relate some of that information to other segmenting species such as our own. In conclusion, this project has furthered our understanding of *cad* function and regulation, and future experiments will bring us one step closer to comprehending the wonders of this gene.

APPENDIX

Difference in threshold cycle (Ct) between knockdown and buffer controls

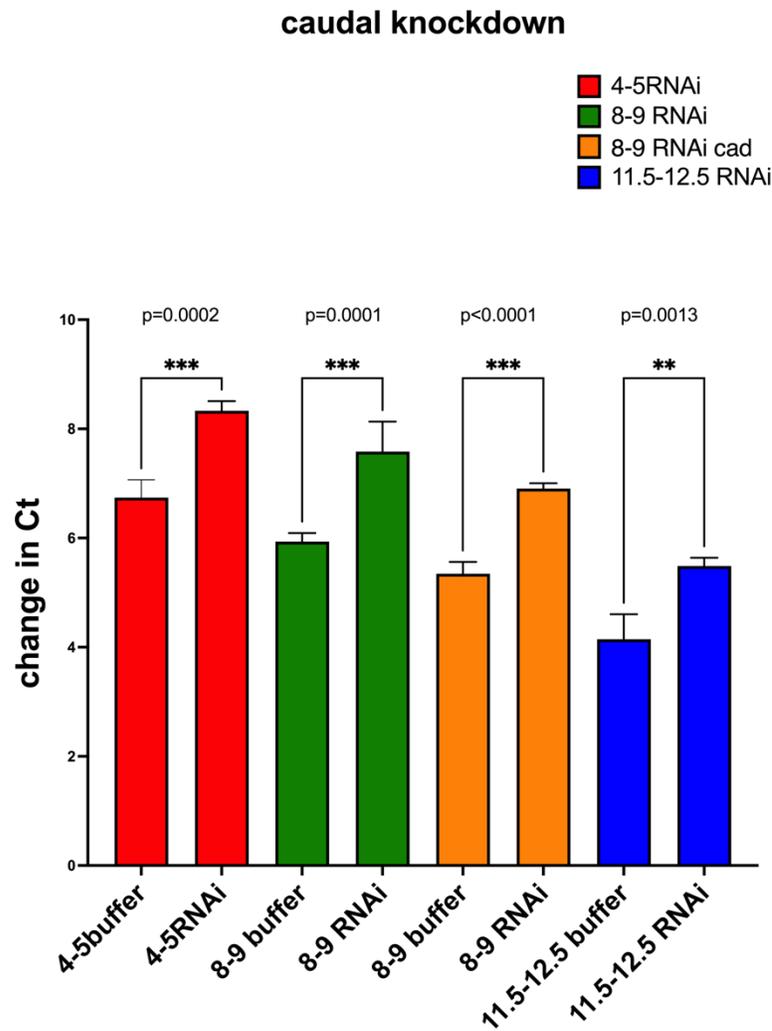


Figure 18. Raw data for the change in qPCR cycles required to obtain product amplification. For each of our knockdowns, a significant difference was found between our buffer injected and dsRNA injected embryos. This shows that the difference in transcript is significantly different from control to experimentals for each case.

Weight matrix model on the binding sites of a transcription factor from *D. melanogaster*

genes models

MEME version 4

ALPHABET= ACGT

strands: + -

Background letter frequencies

A 0.333 C 0.164 G 0.163 T 0.340

MOTIF cad1 MA0216.1

letter-probability matrix: alength= 4 w= 7 nsites= 38 E= 0

0.078947	0.052632	0.078947	0.789474
0.026316	0.000000	0.000000	0.973684
0.210526	0.000000	0.052632	0.736842
0.973684	0.000000	0.026316	0.000000
0.078947	0.000000	0.000000	0.921053
0.078947	0.000000	0.236842	0.684211
0.473684	0.000000	0.500000	0.026316

URL <http://jaspar.genereg.net/matrix/MA0216.1>

MOTIF D MA0445.1

letter-probability matrix: alength= 4 w= 11 nsites= 29 E= 0

0.034483	0.275862	0.241379	0.448276
0.000000	0.862069	0.000000	0.137931
0.000000	0.586207	0.000000	0.413793
0.689655	0.000000	0.000000	0.310345
0.000000	0.000000	0.000000	1.000000
0.000000	0.000000	0.103448	0.896552
0.068966	0.000000	0.931034	0.000000
0.000000	0.000000	0.000000	1.000000
0.034483	0.068966	0.137931	0.758621
0.137931	0.344828	0.206897	0.310345
0.206897	0.034483	0.034483	0.724138

URL <http://jaspar.genereg.net/matrix/MA0445.1>

MOTIF Dref MA1456.1

letter-probability matrix: alength= 4 w= 10 nsites= 3112 E= 0

0.261568	0.196658	0.351864	0.189910
0.254177	0.330334	0.114717	0.300771
0.000000	0.000000	0.000000	1.000000
1.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	1.000000
0.000000	1.000000	0.000000	0.000000
0.000000	0.000000	1.000000	0.000000
1.000000	0.000000	0.000000	0.000000
0.117931	0.053342	0.039203	0.789524
0.586440	0.060411	0.147494	0.205656

URL <http://jaspar.genereg.net/matrix/MA1456.1>

MOTIF eve MA0221.1

letter-probability matrix: alength= 4 w= 7 nsites= 22 E= 0

0.136364	0.454545	0.000000	0.409091
0.045455	0.000000	0.000000	0.954545
1.000000	0.000000	0.000000	0.000000
1.000000	0.000000	0.000000	0.000000
0.000000	0.090909	0.045455	0.863636
0.000000	0.090909	0.500000	0.409091
0.772727	0.000000	0.181818	0.045455

URL <http://jaspar.genereg.net/matrix/MA0221.1>

MOTIF hbn MA0226.1

letter-probability matrix: alength= 4 w= 7 nsites= 17 E= 0

0.117647	0.117647	0.117647	0.647059
0.000000	0.000000	0.000000	1.000000
1.000000	0.000000	0.000000	0.000000
1.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	1.000000
0.000000	0.000000	0.000000	1.000000
0.529412	0.000000	0.411765	0.058824

URL <http://jaspar.genereg.net/matrix/MA0226.1>

MOTIF hb MA0049.1

letter-probability matrix: alength= 4 w= 10 nsites= 16 E= 0

0.062500	0.312500	0.500000	0.125000
0.375000	0.500000	0.125000	0.000000
0.562500	0.187500	0.250000	0.000000

0.250000 0.187500 0.062500 0.500000
0.812500 0.062500 0.000000 0.125000
1.000000 0.000000 0.000000 0.000000
1.000000 0.000000 0.000000 0.000000
0.875000 0.000000 0.125000 0.000000
0.937500 0.062500 0.000000 0.000000
0.562500 0.125000 0.125000 0.187500

URL <http://jaspar.genereg.net/matrix/MA0049.1>

MOTIF dl MA0022.1

letter-probability matrix: alength= 4 w= 12 nsites= 13 E= 0

0.000000 0.384615 0.461538 0.153846
0.000000 0.000000 0.923077 0.076923
0.000000 0.076923 0.846154 0.076923
0.000000 0.000000 0.769231 0.230769
0.076923 0.076923 0.230769 0.615385
0.076923 0.000000 0.153846 0.769231
0.076923 0.000000 0.000000 0.923077
0.000000 0.000000 0.000000 1.000000
0.000000 0.230769 0.000000 0.769231
0.076923 0.692308 0.000000 0.230769
0.076923 0.692308 0.076923 0.153846
0.230769 0.384615 0.384615 0.000000

URL <http://jaspar.genereg.net/matrix/MA0022.1>

MOTIF oc MA0234.1

letter-probability matrix: alength= 4 w= 6 nsites= 19 E= 0

0.000000 0.000000 0.000000 1.000000
1.000000 0.000000 0.000000 0.000000
1.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.105263 0.894737
0.000000 1.000000 0.000000 0.000000
0.000000 0.894737 0.052632 0.052632

URL <http://jaspar.genereg.net/matrix/MA0234.1>

MOTIF pan2 MA0237.2

letter-probability matrix: alength= 4 w= 14 nsites= 71 E= 0

0.000000 0.000000 0.478873 0.521127
0.000000 1.000000 0.000000 0.000000
0.000000 0.000000 1.000000 0.000000

0.000000 0.281690 0.563380 0.154930
0.098592 0.605634 0.042254 0.253521
0.154930 0.140845 0.267606 0.436620
0.098592 0.563380 0.000000 0.338028
0.000000 0.422535 0.211268 0.366197
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.154930 0.845070
0.154930 0.126761 0.577465 0.140845
0.478873 0.000000 0.295775 0.225352
0.309859 0.000000 0.000000 0.690141

URL <http://jaspar.genereg.net/matrix/MA0237.2>

MOTIF prd MA0239.1

letter-probability matrix: alength= 4 w= 9 nsites= 21 E= 0

0.476190 0.190476 0.000000 0.333333
0.047619 0.190476 0.666667 0.095238
0.190476 0.095238 0.095238 0.619048
0.952381 0.000000 0.000000 0.047619
1.000000 0.000000 0.000000 0.000000
0.000000 1.000000 0.000000 0.000000
0.333333 0.238095 0.142857 0.285714
0.000000 0.000000 1.000000 0.000000
0.238095 0.190476 0.095238 0.476190

URL <http://jaspar.genereg.net/matrix/MA0239.1>

MOTIF Stat92E MA0532.1

letter-probability matrix: alength= 4 w= 15 nsites= 118 E= 0

0.127119 0.423729 0.203390 0.245763
0.203390 0.186441 0.500000 0.110169
0.228814 0.296610 0.347458 0.127119
0.771186 0.008475 0.144068 0.076271
0.398305 0.000000 0.228814 0.372881
0.000000 0.008475 0.008475 0.983051
0.000000 0.050847 0.000000 0.949153
0.000000 0.898305 0.000000 0.101695
0.000000 0.567797 0.118644 0.313559
0.322034 0.177966 0.254237 0.245763
0.347458 0.016949 0.593220 0.042373
0.000000 0.008475 0.991525 0.000000

0.957627 0.000000 0.016949 0.025424
1.000000 0.000000 0.000000 0.000000
0.533898 0.076271 0.084746 0.305085
URL <http://jaspar.genereg.net/matrix/MA0532.1>

MOTIF vfl MA1462.1

letter-probability matrix: alength= 4 w= 12 nsites= 11731 E= 0

0.253857 0.264854 0.231182 0.250107
0.249339 0.235018 0.233569 0.282073
0.194613 0.166482 0.442673 0.196232
0.004433 0.909385 0.015344 0.070838
0.974768 0.002387 0.013895 0.008951
0.003666 0.004944 0.989003 0.002387
0.007757 0.004518 0.984145 0.003580
0.007928 0.058733 0.003154 0.930185
0.976899 0.010656 0.008013 0.004433
0.182678 0.148836 0.585031 0.083454
0.278749 0.244480 0.282073 0.194698
0.269116 0.270139 0.201517 0.259228

URL <http://jaspar.genereg.net/matrix/MA1462.1>

MOTIF Abd-B MA0165.1

letter-probability matrix: alength= 4 w= 7 nsites= 21 E= 0

0.047619 0.000000 0.000000 0.952381
0.000000 0.000000 0.000000 1.000000
0.238095 0.000000 0.000000 0.761905
1.000000 0.000000 0.000000 0.000000
0.000000 0.142857 0.000000 0.857143
0.142857 0.000000 0.523810 0.333333
0.619048 0.000000 0.238095 0.142857

URL <http://jaspar.genereg.net/matrix/MA0165.1>

MOTIF Antp MA0166.1

letter-probability matrix: alength= 4 w= 7 nsites= 16 E= 0

0.062500 0.062500 0.000000 0.875000
0.000000 0.000000 0.000000 1.000000
1.000000 0.000000 0.000000 0.000000
1.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.562500 0.437500

0.937500 0.000000 0.062500 0.000000
URL <http://jaspar.genereg.net/matrix/MA0166.1>

MOTIF B-H1 MA0168.1

letter-probability matrix: alength= 4 w= 7 nsites= 21 E= 0

0.190476	0.190476	0.000000	0.619048
0.000000	0.000000	0.000000	1.000000
1.000000	0.000000	0.000000	0.000000
1.000000	0.000000	0.000000	0.000000
0.380952	0.000000	0.000000	0.619048
0.047619	0.333333	0.000000	0.619048
0.047619	0.000000	0.952381	0.000000

URL <http://jaspar.genereg.net/matrix/MA0168.1>

MOTIF br MA0010.1

letter-probability matrix: alength= 4 w= 14 nsites= 9 E= 0

0.333333	0.111111	0.444444	0.111111
0.111111	0.111111	0.111111	0.666667
0.555556	0.222222	0.111111	0.111111
0.777778	0.000000	0.111111	0.111111
0.333333	0.000000	0.000000	0.666667
0.666667	0.000000	0.000000	0.333333
0.444444	0.111111	0.444444	0.000000
0.777778	0.000000	0.111111	0.111111
0.111111	0.888889	0.000000	0.000000
1.000000	0.000000	0.000000	0.000000
0.888889	0.000000	0.111111	0.000000
0.555556	0.000000	0.333333	0.111111
0.444444	0.000000	0.000000	0.555556
0.222222	0.333333	0.222222	0.222222

URL <http://jaspar.genereg.net/matrix/MA0010.1>

MOTIF br(var.3) MA0012.1

letter-probability matrix: alength= 4 w= 11 nsites= 12 E= 0

0.250000	0.083333	0.083333	0.583333
0.750000	0.166667	0.083333	0.000000
0.833333	0.000000	0.000000	0.166667
1.000000	0.000000	0.000000	0.000000
0.000000	0.833333	0.083333	0.083333
0.333333	0.000000	0.000000	0.666667

0.833333 0.000000 0.000000 0.166667
0.500000 0.000000 0.333333 0.166667
0.500000 0.083333 0.166667 0.250000
0.333333 0.250000 0.166667 0.250000
0.166667 0.250000 0.416667 0.166667

URL <http://jaspar.genereg.net/matrix/MA0012.1>

MOTIF br(var.4) MA0013.1

letter-probability matrix: alength= 4 w= 11 nsites= 6 E= 0

0.166667 0.166667 0.000000 0.666667
0.666667 0.000000 0.000000 0.333333
0.333333 0.000000 0.666667 0.000000
0.166667 0.000000 0.000000 0.833333
0.833333 0.000000 0.166667 0.000000
1.000000 0.000000 0.000000 0.000000
0.833333 0.000000 0.166667 0.000000
0.166667 0.500000 0.166667 0.166667
0.500000 0.000000 0.166667 0.333333
0.833333 0.000000 0.000000 0.166667
0.500000 0.166667 0.000000 0.333333

URL <http://jaspar.genereg.net/matrix/MA0013.1>

MOTIF CG11617 MA0173.1

letter-probability matrix: alength= 4 w= 7 nsites= 17 E= 0

0.058824 0.000000 0.000000 0.941176
0.000000 0.000000 0.000000 1.000000
0.588235 0.000000 0.176471 0.235294
1.000000 0.000000 0.000000 0.000000
0.000000 1.000000 0.000000 0.000000
1.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 1.000000

URL <http://jaspar.genereg.net/matrix/MA0173.1>

MOTIF exd MA0222.1

letter-probability matrix: alength= 4 w= 8 nsites= 17 E= 0

0.235294 0.235294 0.235294 0.294118
0.058824 0.000000 0.117647 0.823529
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 1.000000 0.000000

1.000000 0.000000 0.000000 0.000000
0.000000 0.705882 0.000000 0.294118
0.647059 0.000000 0.352941 0.000000
URL <http://jaspar.genereg.net/matrix/MA0222.1>

MOTIF fkh MA0446.1

letter-probability matrix: alength= 4 w= 11 nsites= 27 E= 0

0.111111 0.000000 0.000000 0.888889
0.185185 0.000000 0.814815 0.000000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.037037 0.962963
0.481481 0.000000 0.518519 0.000000
0.148148 0.481481 0.074074 0.296296
0.222222 0.259259 0.111111 0.407407
0.000000 0.407407 0.074074 0.518519
0.851852 0.000000 0.037037 0.111111
0.555556 0.148148 0.148148 0.148148

URL <http://jaspar.genereg.net/matrix/MA0446.1>

MOTIF HHEX MA0183.1

letter-probability matrix: alength= 4 w= 8 nsites= 26 E= 0

0.269231 0.115385 0.076923 0.538462
0.000000 0.269231 0.000000 0.730769
0.000000 0.230769 0.269231 0.500000
0.807692 0.000000 0.076923 0.115385
1.000000 0.000000 0.000000 0.000000
0.000000 0.115385 0.115385 0.769231
0.269231 0.000000 0.038462 0.692308
0.846154 0.000000 0.153846 0.000000

URL <http://jaspar.genereg.net/matrix/MA0183.1>

MOTIF kni MA0451.1

letter-probability matrix: alength= 4 w= 12 nsites= 26 E= 0

0.730769 0.038462 0.076923 0.153846
0.961538 0.038462 0.000000 0.000000
0.615385 0.000000 0.000000 0.384615
0.192308 0.346154 0.230769 0.230769
0.000000 0.153846 0.038462 0.807692
0.807692 0.000000 0.192308 0.000000

0.000000 0.000000 1.000000 0.000000
0.653846 0.000000 0.307692 0.038462
0.038462 0.115385 0.692308 0.153846
0.000000 1.000000 0.000000 0.000000
0.961538 0.000000 0.038462 0.000000
0.192308 0.461538 0.269231 0.076923
URL <http://jaspar.genereg.net/matrix/MA0451.1>

MOTIF nub MA0197.2

letter-probability matrix: alength= 4 w= 12 nsites= 29 E= 0

0.068966 0.034483 0.034483 0.862069
1.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.862069 0.137931
0.000000 0.655172 0.034483 0.310345
0.827586 0.000000 0.000000 0.172414
0.965517 0.000000 0.034483 0.000000
0.965517 0.000000 0.000000 0.034483
0.137931 0.034483 0.000000 0.827586
0.206897 0.172414 0.275862 0.344828
0.655172 0.172414 0.034483 0.137931
0.137931 0.103448 0.517241 0.241379

URL <http://jaspar.genereg.net/matrix/MA0197.2>

MOTIF onecut MA0235.1

letter-probability matrix: alength= 4 w= 7 nsites= 15 E= 0

0.066667 0.133333 0.000000 0.800000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 1.000000 0.000000
1.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.000000 1.000000
0.333333 0.000000 0.266667 0.400000

URL <http://jaspar.genereg.net/matrix/MA0235.1>

MOTIF slp1 MA0458.1

letter-probability matrix: alength= 4 w= 11 nsites= 41 E= 0

0.195122 0.073171 0.341463 0.390244
0.000000 0.000000 0.000000 1.000000
0.073171 0.000000 0.926829 0.000000

0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.024390 0.975610
0.000000 0.000000 0.048780 0.951220
0.658537 0.000000 0.097561 0.243902
0.024390 0.536585 0.170732 0.268293
0.414634 0.195122 0.365854 0.024390
0.097561 0.317073 0.073171 0.512195
0.365854 0.073171 0.097561 0.463415

URL <http://jaspar.genereg.net/matrix/MA0458.1>

MOTIF su(Hw) MA0533.1

letter-probability matrix: alength= 4 w= 21 nsites= 4737 E= 0

0.070931 0.239181 0.592780 0.097108
0.051298 0.868904 0.007389 0.072409
0.164239 0.633523 0.048765 0.153473
0.180494 0.312434 0.206249 0.300823
0.584125 0.123707 0.195271 0.096897
0.725776 0.043910 0.111674 0.118640
0.890226 0.019633 0.062276 0.027866
0.824150 0.004011 0.066709 0.105130
0.086553 0.049609 0.793540 0.070298
0.003589 0.093097 0.020266 0.883048
0.902681 0.004433 0.020055 0.072831
0.001478 0.002322 0.240659 0.755541
0.042010 0.001056 0.955457 0.001478
0.024488 0.969812 0.001267 0.004433
0.595525 0.086764 0.001689 0.316023
0.716065 0.062487 0.150517 0.070931
0.110407 0.512983 0.040321 0.336289
0.535149 0.074731 0.297446 0.092675
0.383154 0.271058 0.162550 0.183238
0.470973 0.094364 0.072831 0.361832
0.421364 0.080008 0.074098 0.424530

URL <http://jaspar.genereg.net/matrix/MA0533.1>

MOTIF Ubx MA0094.2

letter-probability matrix: alength= 4 w= 8 nsites= 20 E= 0

0.150000 0.250000 0.150000 0.450000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.000000 1.000000

0.850000 0.000000 0.000000 0.150000
1.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 1.000000
0.000000 0.000000 0.300000 0.700000
0.700000 0.000000 0.300000 0.000000

URL <http://jaspar.genereg.net/matrix/MA0094.2>

ACKNOWLEDGMENTS

I am immensely grateful to Dr. Terri Williams for all her support and mentorship throughout the last couple of years. Since our first conversation back in 2019 when I expressed interest in joining her lab, she has truly showed me what passion for research looks like. Dr. Williams has taught me how to design and troubleshoot experiments, how to critically analyze data, and how to read and write in the scientific literature. She is always eager to answer questions with kindness and patience, and has always encouraged me to challenge myself. She fosters an amazing collaborative environment in her lab where I have gained not only lab partners but friends, and has also prepared me to thrive as an independent researcher. Dr. Williams has become of my role models and mentors in life—and undoubtedly one of the reasons of why I have decided to pursue research as a professional career. She taught me the research skills that I need to succeed, and has always believed in me. She has made a better scientist, presenter, writer, and person—and for that, I will forever be in her debt.

Additionally, I would like to thank Prof. Robert Fleming and Prof. Claire Fournier for their valuable feedback on my thesis project and their support throughout these last couple of years. Their comments did not only improve the quality of my writing and the analysis of my results, but also allowed me to consider different points of view and open my mind to other opinions. Although my work has been focused on examining *T. castaneum*, I have gained a great deal of respect and admiration for *D. melanogaster* thanks to Prof. Fleming and his passion for this model organism. I am also very grateful for the opportunity to work and learn from him as an SI Leader and first-year seminar mentor. On the other hand, Prof. Fournier has been my academic advisor since I declared my biology major, and has always been a great support system in helping me choose classes and make sure I stay on top of my requirements.

As my lab instructor for BIOL 182, she helped me solidified my desire to venture into biology, and I am very grateful I made that decision. I am extremely grateful for the Biology Department and all the Biology faculty for their encouragement and support throughout these four years. I have greatly enjoyed taking classes in many of the different areas of biology, which have been both interesting and challenging. I am also grateful for their help and feedback in my presentations for seminars, and for their very thoughtful questions. Even though it has not always been easy, being a biology major has been a truly rewarding journey that sparked my passion for science and research.

I owe a debt of gratitude to Jim Camp, our amazing lab tech at the Williams Lab. He is a truly skilled scientist that will learn how to do any protocol we talk about, will happily teach you anything, and knows where everything is in the lab. Doing our 11.5 hAEL knockdowns meant that I had to set up an egg lay very late at night, and he had to come in early that morning to perform the dsRNA microinjections (he's a magician at this!)—which I will forever be grateful for. Similarly, I would like to sincerely thank the Nagy Lab at the University of Arizona, in particular Dr. Lisa Nagy, Bennett Van Camp, Son Tran, and Benjamin Goldman-Huertas for their support in discussions and in teaching me how to use MCAST.

In addition, I am incredibly grateful for many of the friendships I have made at the Williams Lab, including Jeff Sagun '21, Katie Russell '22, Marissa Howlett '22, Archana Adhikari '22, Dimos Sampatakos '24, and Vedika Agrawal '24. A special thanks to Dimos, Marissa, and Vedika for being my *cad* buddies and arduously working with the yeast one-hybrid and the cross-species transgenics experiments. Finally, I would like to thank my family and friends who have supported me throughout this journey, and who are always there to cheer me up when I need it the most.

LITERATURE CITED

- Ansari, S., Troelenberg, N., Dao, V. A., Richter, T., Bucher, G., & Klingler, M. (2018). Double abdomen in a short-germ insect: Zygotic control of axis formation revealed in the beetle *Tribolium castaneum*. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(8), 1819–1824. <http://doi.org/10.1073/pnas.1716512115>
- Bailey, T. L., & Noble, W. S. (2003). Searching for statistically significant regulatory modules. *Bioinformatics*, *19*(suppl_2), ii16–ii25. <https://doi.org/10.1093/bioinformatics/btg1054>
- Beck, F., Chawengsaksophak, K., Waring, P., Playford, R. J., & Furness, J. B. (1999). Reprogramming of Intestinal Differentiation and Intercalary Regeneration in Cdx2 Mutant Mice. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(13), 7318–7323.
- Benton, M. A., Akam, M., & Pavlopoulos, A. (2013). Cell and tissue dynamics during *Tribolium* embryogenesis revealed by versatile fluorescence labeling approaches. *Development*, *140*(15), 3210–3220. <https://doi.org/10.1242/dev.096271>
- Bolognesi, R., Fischer, T. D., & Brown, S. J. (2009). Loss of Tc-arrow and canonical Wnt signaling alters posterior morphology and pair-rule gene expression in the short-germ insect, *Tribolium castaneum*. *Development Genes and Evolution*, *219*(7), 369–375. <https://doi.org/10.1007/s00427-009-0299-3>
- Brena, C., & Akam, M. (2013). An analysis of segmentation dynamics throughout embryogenesis in the centipede *Strigamia maritima*. *BMC Biology*, *11*(1), 112. <https://doi.org/10.1186/1741-7007-11-112>
- Chawengsaksophak, K., James, R., Hammond, V. E., Köntgen, F., & Beck, F. (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature*, *386*(6620), 84–87. <https://doi.org/10.1038/386084a0>
- Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J., Beck, F., & Ruddle, F. H. (2004). Cdx2 Is Essential for Axial Elongation in Mouse Development. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(20), 7641–7645.
- Choe, C. P., Miller, S. C., & Brown, S. J. (2006). A pair-rule gene circuit defines segments sequentially in the short-germ insect *Tribolium castaneum*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(17), 6560–6564. <https://doi.org/10.1073/pnas.0510440103>
- Choi, Y.-J., Choi, T.-Y., Yamaguchi, M., Matsukage, A., Kim, Y.-S., & Yoo, M.-A. (2004). Transcriptional regulation of the *Drosophila* caudal homeobox gene by DRE/DREF. *Nucleic Acids Research*, *32*(12), 3734–3742. <https://doi.org/10.1093/nar/gkh688>

- Choi, Y.-J., Hwang, M.-S., Park, J.-S., Bae, S.-K., Kim, Y.-S., & Yoo, M.-A. (2008). Age-related upregulation of *Drosophila* caudal gene via NF- κ B in the adult posterior midgut. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1780(10), 1093–1100. <https://doi.org/10.1016/j.bbagen.2008.06.008>
- Clark, E., Peel, A. D., & Akam, M. (2019). Arthropod segmentation. *Development*, 146(18), dev170480. <https://doi.org/10.1242/dev.170480>
- Cooke, J., & Zeeman, E. C. (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *Journal of Theoretical Biology*, 58(2), 455–476. [https://doi.org/10.1016/S0022-5193\(76\)80131-2](https://doi.org/10.1016/S0022-5193(76)80131-2)
- Copf, T., Schroder, R., & Averof, M. (2004). Ancestral role of caudal genes in axis elongation and segmentation. *Proceedings of the National Academy of Sciences*, 101(51), 17711–17715. <https://doi.org/10.1073/pnas.0407327102>
- Cruz, C., Maegawa, S., Weinberg, E. S., Wilson, S. W., Dawid, I. B., & Kudoh, T. (2010). Induction and patterning of trunk and tail neural ectoderm by the homeobox gene *eve1* in zebrafish embryos. *Proceedings of the National Academy of Sciences*, 107(8), 3564–3569. <https://doi.org/10.1073/pnas.1000389107>
- Davis, G. K., & Patel, N. H. (2002). Short, Long, and Beyond: Molecular and Embryological Approaches to Insect Segmentation. *Annual Review of Entomology*, 47(1), 669–699. <https://doi.org/10.1146/annurev.ento.47.091201.145251>
- Edgar, L. G., Carr, S., Wang, H., & Wood, W. B. (2001). Zygotic Expression of the caudal Homolog *pal-1* Is Required for Posterior Patterning in *Caenorhabditis elegans* Embryogenesis. *Developmental Biology*, 229(1), 71–88. <https://doi.org/10.1006/dbio.2000.9977>
- El-Sherif, E., Averof, M., & Brown, S. J. (2012). A segmentation clock operating in blastoderm and germband stages of *Tribolium* development. *Development*, 139(23), 4341–4346. <https://doi.org/10.1242/dev.085126>
- El-Sherif, E., Zhu, X., Fu, J., & Brown, S. J. (2014). Caudal Regulates the Spatiotemporal Dynamics of Pair-Rule Waves in *Tribolium*. *PLoS Genetics*, 10(10), e1004677. <https://doi.org/10.1371/journal.pgen.1004677>
- Epstein, M., Pillemer, G., Yelin, R., Yisraeli, J. K., & Fainsod, A. (1997). Patterning of the embryo along the anterior-posterior axis: The role of the caudal genes. *Development*, 124(19), 3805–3814. <https://doi.org/10.1242/dev.124.19.3805>
- Fortey, R. A., & Thomas, R. H. (1997). *Arthropod Relationships*. Springer Science & Business Media.

- Goldman-Huertas, B., Sagun, J., Williams, T.A., Nagy, L.M. Expression of transcription factors during four stages of *Tribolium* embryogenesis. In prep.
- Grant, C. E., Johnson, J., Bailey, T. L., & Noble, W. S. (2016). MCAST: Scanning for cis-regulatory motif clusters. *Bioinformatics*, 32(8), 1217–1219. <https://doi.org/10.1093/bioinformatics/btv750>
- Hunter, C. P., & Kenyon, C. (1996). Spatial and Temporal Controls Target *pal-1* Blastomere-Specification Activity to a Single Blastomere Lineage in *C. elegans* Embryos. *Cell*, 87(2), 217–226. [https://doi.org/10.1016/S0092-8674\(00\)81340-9](https://doi.org/10.1016/S0092-8674(00)81340-9)
- Janssen, R. (2020). The embryonic expression pattern of a second, hitherto unrecognized, paralog of the pair-rule gene *sloppy-paired* in the beetle *Tribolium castaneum*. *Development Genes and Evolution*, 230(3), 247–256. <https://doi.org/10.1007/s00427-020-00660-x>
- Jayaram, N., Usvyat, D., & R. Martin, A. C. (2016). Evaluating tools for transcription factor binding site prediction. *BMC Bioinformatics*, 17(1), 547. <https://doi.org/10.1186/s12859-016-1298-9>
- Jenkins, S. (2012). *The beetle book*. Houghton Mifflin Harcourt.
- Liao, B.-K., & Oates, A. C. (2017). Delta-Notch signalling in segmentation. *Arthropod Structure & Development*, 46(3), 429–447. <https://doi.org/10.1016/j.asd.2016.11.007>
- Liu, P. Z., & Kaufman, T. C. (2005). Short and long germ segmentation: Unanswered questions in the evolution of a developmental mode. *Evolution & Development*, 7(6), 629–646. <https://doi.org/10.1111/j.1525-142X.2005.05066.x>
- Macdonald, P. M., & Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature*, 324(6097), 537–545. <https://doi.org/10.1038/324537a0>
- Marom, K., Shapira, E., & Fainsod, A. (1997). The chicken caudal genes establish an anterior-posterior gradient by partially overlapping temporal and spatial patterns of expression. *Mechanisms of Development*, 64(1), 41–52. [https://doi.org/10.1016/S0925-4773\(97\)00043-9](https://doi.org/10.1016/S0925-4773(97)00043-9)
- Martin, B. L., & Kimelman, D. (2009). Wnt Signaling and the Evolution of Embryonic Posterior Development. *Current Biology*, 19(5), R215–R219. <https://doi.org/10.1016/j.cub.2009.01.052>
- McGregor, A. P., Pechmann, M., Schwager, E. E., & Damen, W. G. (2009). An ancestral regulatory network for posterior development in arthropods. *Communicative & Integrative Biology*, 2(2), 174–176. <https://doi.org/10.4161/cib.7710>

- Miller, S. C., Miyata, K., Brown, S. J., & Tomoyasu, Y. (2012). Dissecting Systemic RNA Interference in the Red Flour Beetle *Tribolium castaneum*: Parameters Affecting the Efficiency of RNAi. *PLoS ONE*, 7(10), e47431. <https://doi.org/10.1371/journal.pone.0047431>
- Mito, T., Kobayashi, C., Sarashina, I., Zhang, H., Shinahara, W., Miyawaki, K., Shinmyo, Y., Ohuchi, H., & Noji, S. (2007). Even-skipped has gap-like, pair-rule-like, and segmental functions in the cricket *Gryllus bimaculatus*, a basal, intermediate germ insect (Orthoptera). *Developmental Biology*, 303(1), 202–213. <https://doi.org/10.1016/j.ydbio.2006.11.003>
- Moreno, E., & Morata, G. (1999). Caudal is the Hox gene that specifies the most posterior *Drosophila* segment. *Nature*, 400(6747), 873–877. <https://doi.org/10.1038/23709>
- Nakamoto, A., Hester, S. D., Constantinou, S. J., Blaine, W. G., Tewksbury, A. B., Matei, M. T., Nagy, L. M., & Williams, T. A. (2015). Changing cell behaviours during beetle embryogenesis correlates with slowing of segmentation. *Nature Communications*, 6(1), 6635. <https://doi.org/10.1038/ncomms7635>
- Nitta, K. R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E. E. M., & Taipale, J. (2015). Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *ELife*, 4, e04837. <https://doi.org/10.7554/eLife.04837>
- Novikova, A. V., Auman T., Cohen M., Oleynik O., Stahi-Hitin R., Gil E., Weisbrod A., Chipman A. D. (2020). The multiple roles of caudal in early development of the milkweed bug *Oncopeltus fasciatus*. *Developmental Biology*, 467(1–2), 66–76. <https://doi.org/10.1016/j.ydbio.2020.08.011>
- Noyes, M. B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M. H., & Wolfe, S. A. (2008). A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one-hybrid system. *Nucleic Acids Research*, 36(8), 2547–2560. <https://doi.org/10.1093/nar/gkn048>
- Palmeirim, I., Henrique, D., Ish-Horowicz, D., & Pourquié, O. (1997). Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell*, 91(5), 639–648. [https://doi.org/10.1016/s0092-8674\(00\)80451-1](https://doi.org/10.1016/s0092-8674(00)80451-1)
- Pointer, M. D., Gage, M. J. G., & Spurgin, L. G. (2021). *Tribolium* beetles as a model system in evolution and ecology. *Heredity*, 126(6), 869–883. <https://doi.org/10.1038/s41437-021-00420-1>
- Pourquié, O. (2001). The vertebrate segmentation clock. *Journal of Anatomy*, 199(Pt 1-2), 169–175. <https://doi.org/10.1046/j.1469-7580.2001.19910169.x>

- Pourquié, O. (2003). Vertebrate somitogenesis: A novel paradigm for animal segmentation? *International Journal of Developmental Biology*, 47(7–8), 597–603. <https://doi.org/10.1387/ijdb.14756335>
- Ribeiro, L., Tobias-Santos, V., Santos, D., Antunes, F., Feltran, G., Menezes, J. de S., Aravind, L., Venancio, T. M., & Fonseca, R. N. da. (2017). Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects. *PLOS Genetics*, 13(7), e1006868. <https://doi.org/10.1371/journal.pgen.1006868>
- Sarrazin, A. F., Peel, A. D., & Averof, M. (2012). A Segmentation Clock with Two-Segment Periodicity in Insects. *Science*, 336(6079), 338–341. <https://doi.org/10.1126/science.1218256>
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–682. <https://doi.org/10.1038/nmeth.2019>
- Schulz, C., & Tautz, D. (1995). Zygotic caudal regulation by hunchback and its role in abdominal segment formation of the *Drosophila* embryo. *Development*, 121(4), 1023–1028. <https://doi.org/10.1242/dev.121.4.1023>
- Schulz, C., Schröder, R., Hausdorf, B., Wolff, C., & Tautz, D. (1998). A caudal homologue in the short germ band beetle *Tribolium* shows similarities to both, the *Drosophila* and the vertebrate caudal expression patterns. *Development Genes and Evolution*, 208(5), 283–289. <https://doi.org/10.1007/s004270050183>
- Shinmyo, Y., Mito, T., Matsushita, T., Sarashina, I., Miyawaki, K., Ohuchi, H., & Noji, S. (2005). Caudal is required for gnathal and thoracic patterning and for posterior elongation in the intermediate-germband cricket *Gryllus bimaculatus*. *Mechanisms of Development*, 122(2), 231–239. <https://doi.org/10.1016/j.mod.2004.10.001>
- Shippy, T. D., Coleman, C. M., Tomoyasu, Y., & Brown, S. J. (2009). Concurrent In Situ Hybridization and Antibody Staining in Red Flour Beetle (*Tribolium*) Embryos. *Cold Spring Harbor Protocols*, 2009(8), pdb.prot5257. <https://doi.org/10.1101/pdb.prot5257>
- Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9), 613–626. <https://doi.org/10.1038/nrg3207>
- Stern, C. D., Fraser, S. E., Keynes, R. J., & Primmitt, D. R. N. (1988). A cell lineage analysis of segmentation in the chick embryo. *Development*, 104(Supplement), 231–244. <https://doi.org/10.1242/dev.104.Supplement.231>

- Subramanian, V., Meyer, B. I., & Gruss, P. (1995). Disruption of the murine homeobox gene *Cdx1* affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. *Cell*, 83(4), 641–653. [https://doi.org/10.1016/0092-8674\(95\)90104-3](https://doi.org/10.1016/0092-8674(95)90104-3)
- Tautz, D. (1999). *Tribolium* early embryonic development. <https://www.sdbonline.org/sites/fly/vdevlhom/trib.htm>
- Tautz, D. (2004). Segmentation. *Developmental Cell*, 7(3), 301–312. <https://doi.org/10.1016/j.devcel.2004.08.008>
- Tsurumi, A., Xia, F., Li, J., Larson, K., LaFrance, R., & Li, W. X. (2011). STAT Is an Essential Activator of the Zygotic Genome in the Early *Drosophila* Embryo. *PLOS Genetics*, 7(5), e1002086. <https://doi.org/10.1371/journal.pgen.1002086>
- Venables, J. P. (2007). Downstream intronic splicing enhancers. *FEBS Letters*, 581(22), 4127–4131. <https://doi.org/10.1016/j.febslet.2007.08.012>
- Williams, T. A., & Nagy, L. M. (2017). Linking gene regulation to cell behaviors in the posterior growth zone of sequentially segmenting arthropods. *Arthropod Structure & Development*, 46(3), 380–394. <https://doi.org/10.1016/j.asd.2016.10.003>
- Xiang, Y., & Garrard, W. T. (2008). The Downstream Transcriptional Enhancer, Ed, Positively Regulates Mouse *Igk* Gene Expression and Somatic Hypermutation. *Journal of Immunology (Baltimore, Md. : 1950)*, 180(10), 6725–6732.
- Xiang, J., Reding, K., Heffer, A., & Pick, L. (2017). Conservation and variation in pair-rule gene expression and function in the intermediate-germ beetle *Dermestes maculatus*. *Development*, 144(24), 4625–4636. <https://doi.org/10.1242/dev.154039>