Doctoral Dissertation

Sense-overlapping lncRNA as a decoy of translational repressor protein for dimorphic gene expression

(翻訳抑制タンパク質のデコイとしてのセンスオーバーラッピング長鎖非コード RNA による二型遺伝子発現制御の研究)

Perez, Christelle Alexa Garcia

July 2021

Department of Biotechnology
Graduate School of Engineering
Osaka University

TABLE OF CONTENT

CHAPTER 1 GENERAL INTRODUCTION

1.1 Long noncoding RNA

Noncoding RNAs (ncRNAs) are not translated into proteins but they are vastly transcribed in the genome, representing most of the total transcript (Mellor, Woloszczuk, & Howe, 2016). Sequencing of the human genome revealed that there are only ~20,000 protein-coding genes, representing < 2% of the total genomic sequence; while the other 98% does not encode proteins but are likely transcribed, yielding transcripts with little or no protein-coding capacity, the ncRNAs (Figure 1A) (Wilusz, Sunwoo, & Spector, 2009).

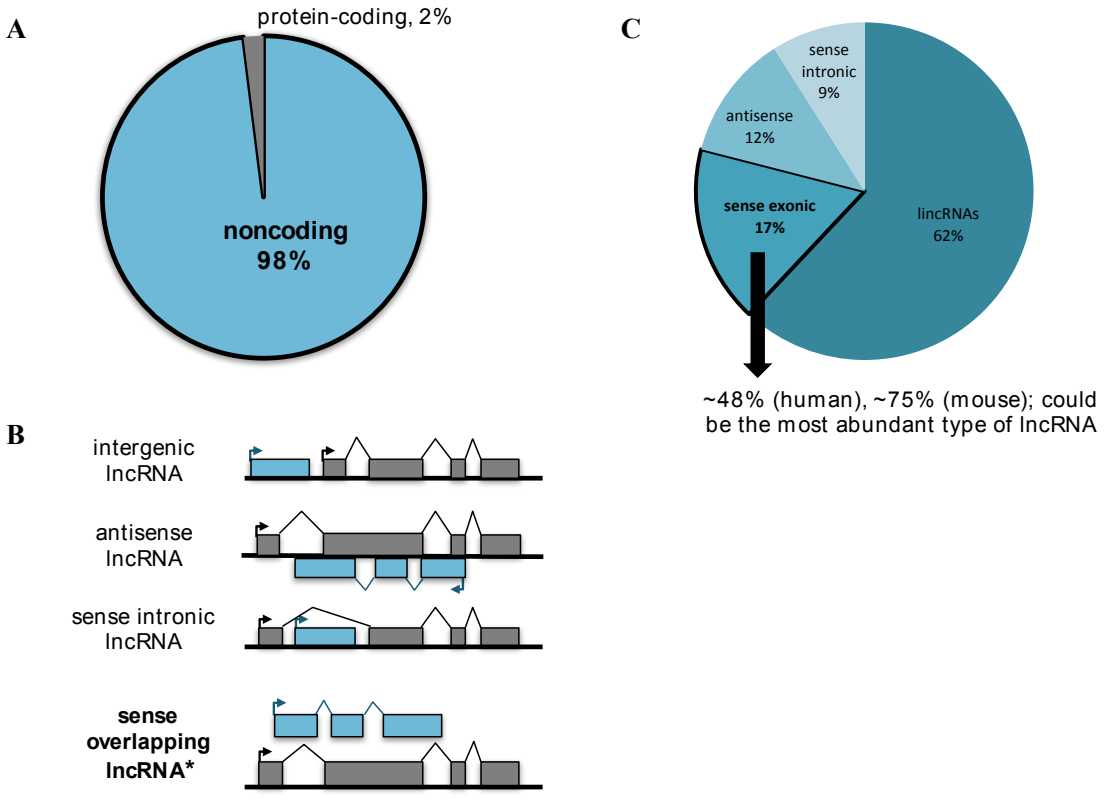

Figure 1. Long noncoding RNAs (lncRNAs).

(A) Proportion of protein-coding, 2% (gray), and noncoding RNAs, 98% (blue), in the total RNA transcript (Mellor et al., 2016). **(B)** Different types of lncRNAs: intergenic, antisense, sense intronic, and sense overlapping). Boxes represent exons while lines are introns. Protein-coding genes are gray while lncRNAs are depicted in blue. **(C)** The proportion of different types of lncRNAs according to the Pacbio annotation (Kuo et al., 2017).

Within the broad category of ncRNAs, a number of transcripts have been grouped according to their length, localization, and function. Classes of ncRNAs include, but are not limited to, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), piwi-interacting RNA (piRNA), microRNA (miRNA), small interfering RNA (siRNA), and long noncoding RNA (lncRNA) (Kashi, Henderson, Bonetti, & Carninci, 2016). While the small noncoding RNAs represent a wide spectrum of functional molecules, lncRNAs have gained widespread attention in recent years. With the advent of sensitive and high-throughput genomic technologies, novel transcripts are being detected and lncRNAs are found involved in a range of disease and developmental processes (Kung, Colognori, & Lee, 2013).

Also, among the noncoding RNAs, 61% are lncRNAs (Kashi et al., 2016). Long noncoding RNA (lncRNA) is the general term for transcripts that have a minimum length of 200 nt and lack protein-coding potential (Quinn & Chang, 2016). With this broad definition, lncRNAs exhibit diverse biogenesis and they are numerously present in the genome in different forms.

The different types of lncRNA were identified as intergenic, antisense, sense intronic, and sense overlapping lncRNA (Figure 1B) (Kung et al., 2013). Long intergenic noncoding RNAs (lincRNAs) are lncRNAs that do not overlap annotated coding genes; while the rest overlap coding genes. However, antisense lncRNAs are transcribed in the opposite direction and the sense intronic group overlaps the introns of their overlapping coding genes. Whereas, sense-overlapping lncRNAs are transcribed in the same direction and overlap the exon/s of protein-coding genes, leaving them poorly studied due to the difficulties in annotating and distinguishing them from their overlapping coding genes. This is despite the projection that sense-overlapping lncRNA transcripts are actually the most abundant type of lncRNA based on the proportions of lncRNA classes in chicken PacBio annotation (Figure 1C) (Kuo et al., 2017). Sense overlapping lncRNAs make up 17% of the total lncRNA transcripts in chicken. Based on this information, it was predicted that there are 24, 385 and 11, 901 sense overlapping

lncRNAs in human and mouse, respectively. These numbers equate to 48 and 57%, in that order. This indicates the probable highest occurrence of sense overlapping lncRNAs in mouse and human transcripts, as well as other vertebrate species, and highlights the importance to study the function and mechanism of this group of lncRNAs.

In general, a variety of screens and expression analyses have provided evidence on the biological functions of lncRNAs despite previously being labeled as transcriptional noises or "junk" (Mercer, Dinger, & Mattick, 2009). lncRNAs have been known to affect cellular processes such as proliferation, differentiation, quiescence, senescence, stress and immune response, development, disease, and aging (Quinn & Chang, 2016).

With their wide biological and functional repertoire, the mode of action of lncRNAs is also diverse. One of the most emergent roles of lncRNAs is their involvement in regulating the expression of protein-coding genes (Mercer et al., 2009). They regulate gene expression at the level of chromatin modification, transcription, and post-transcriptional processing. LncRNAs may mediate epigenetic changes by recruiting chromatin remodeling complexes to genomic loci (Campos & Reinberg, 2009). They may also act as scaffolds onto which protein complexes can assemble, allowing chromatin modifications; or influence epigenetic regulation through modulation of DNA methylation at CpG sites (Kung et al., 2013). To regulate transcription, lncRNAs can form loop structures to recruit transcription factors (TFs) or interfere with Pol II transcription machinery (Yao, Wang, & Chen, 2019). LncRNAs can also regulate various steps in post-transcriptional processing of mRNAs, including splicing, editing, transport, translation, and degradation (Mercer et al., 2009). Another known mechanism of lncRNAs in gene regulation is by acting as decoys to sequester RNA-binding proteins (RBPs) or microRNAs (L. L. Chen, 2016).

RNA-binding proteins (RBPs) are usually involved for lncRNAs to perform their function. LncRNAs interact and form a complex with RBPs, leading to activation or repression of their

target gene (Wang & Chang, 2011). To fully understand the function and mechanism of lncRNAs, it is important to identify the RBPs they associate with and analyze how the lncRNA together with its RBPs regulate their target gene.

1.2 Gene regulation

Gene regulation refers to the mechanism that acts to induce or repress the expression of a gene. This includes structural and chemical changes to the genetic material, binding of proteins to specific DNA elements to regulate transcription, RNA processing, and stability, or mechanisms that modulate translation of mRNAs and the activity of proteins (Hoopes, 2008).

Gene regulation can be positive or negative based on the regulator involved in controlling the expression of the target gene. Gene regulators comprise activators, enhancers, silencers, insulators, repressors, or competitors which range from DNA elements, coding, and noncoding RNAs to proteins (Maston, Evans, & Green, 2006; Reményi, Schöler, & Wilmanns, 2004).

Gene regulation ensures that the appropriate genes are expressed where and when it is needed. It drives cellular differentiation; a process during which different tissues and cell types are produced, and it also helps cells maintain differentiated states of cells or tissues (Ralston & Shaw, 2008). Gene regulation results in different patterns of gene expression; allowing various cell types to have different sets of active protein and making each cell type specialized for its intended form and function despite sharing the exact same set of genetic material. Basically, the particular combination of genes that are expressed or repressed dictates cellular morphology and function.

While some genes can have varying levels of expression or transcriptional activity, others only assume one of the two discrete states: active and inactive (Zhang, Andersen, & Conolly, 2006). With this binary mode of action, gene regulators induce or repress gene expression by affecting the probability by which the gene is turned on or off (Fiering, Whitelaw, & Martin,

2000). This all-or-none binary mechanism generates ultrasensitive responses that play important role in cell-fate differentiation and development (Buchler & Louis, 2008).

1.2.1 Binary sex regulation

Regulation of sex-determining genes is a notable model for binary gene expression. Sex determination is a basic biological mechanism for an embryo to decide which sex-determining pathway will it adopt. It is a binary decision between the male or female pathway that will govern sexual differentiation and regulate the consequent development of the sex-specific parts of the organism that differ between sexes. Aside from their gametes and gonads, sexual individuals also vary in their morphology, physiology, and behavior (Bull, 1985). This is called sex binarism.

While sex is generally binary, the rule of dimorphism does not always prevail in nature. For instance, there have been numerous reports on individuals that cannot be classified as male or female across different animal groups from crustaceans, to insects, birds, fish, and even mammals (Bahamonde, Munkittrick, & Martyniuk, 2013; Ford, 2012; Harper, 2007; Lambeth & Smith, 2012; Pereira, Narita, Kageyama, & Kjellberg, 2010; Villagómez et al., 2009). This event wherein the sexual phenotypes of these individuals deviate off the typical male and female is called “intersex”. Intersex is either a result of genetic mutation or imbalance in gene expression and/or external factors like anthropogenic pollutants which affected the individual’s sexual differentiation and development. Intersex impacts survival as it degrades the organisms’ ability to reproduce. Hence to ensure fitness, sex must be strictly governed by binary regulation; the mechanism of which is unknown. Identifying factor/s that may be responsible for this safekeeping regulation may give light in understanding how population heterogeneity or intersex is kept rare in nature.

1.3 Environmental sex determination (ESD)

The diverse mechanisms animals follow to determine their sex can be divided into two main categories: the genetic and environmental pathways (Bull, 1985). These two pathways are based on the primary causal factor that triggers the differentiation; genetics sex determination (GSD) is established by genetics cues such as chromosomal composition or segregation during fertilization, while environmental sex determination (ESD) is triggered by environmental factors that serve as cues for genetic signaling (Marin & Baker, 1998; Zarkower, 2001).

In environmental sex determination, environmental cues involve temperature, photoperiod, nutrition, population density, and the likes (Korpelainen, 1990). ESD has arisen repeatedly during evolution (Organ & Janes, 2008), which may imply the adaptive significance of this system in nature. It has been reported that selection forces drive the transition between GSD and ESD (Bull, 1985; Bulmer & Bull, 1982). For instance, through a temperature-sensitive mutation gene created artificially in a single control in *Dr. melanogaster* and *C.* elegans, it was revealed that GSD could rapidly evolve into ESD (Epper & Bryant, 1983; Hodgkin, 2002). Furthermore, orthologs of some genes involved in GSD have been found in ESD animals. Some of these GSD sex-determining genes were revealed to be expressed in the gonads of temperature-dependent sex-determining reptiles during the temperature-sensitive period (Shoemaker & Crews, 2009). Also in water flea crustaceans, studies have shown that both ESD and GSD have the same origin and share similar genetic components in their downstream sex-determining cascade, which is the conserved DM-domain transcription factor *Doublesex1 (Dsx1)* as the sexual major effector or the binary switch (Kato, Kobayashi, Watanabe, & Iguchi, 2011; Toyota et al., 2013). However, a detailed ESD mechanism has not been reported yet. It is therefore important to identify the genes involved in the ESD cascade and unravel the molecular binary regulation involved in the sexually dimorphic pathway.

Also, while numerous studies about GSD are available on mammals, worms, and insects to understand its sex regulation process (Cline and & Meyer, 1996; Williams & Carroll, 2009), the molecular mechanism behind ESD is poorly studied; because amidst the diversity of ESD-obligate organisms like rotifers, nematodes, crustaceans, insects, fishes and reptiles, they tend to be poor experimental models (Crews & Bull, 2009; Vogt & Bull, 1982). Hence, a model organism for the extensive study on ESD and binary sex regulation should be utilized.

1.4 The model organism *Daphnia magna*

The crustacean *Daphnia magna* has emerged as a good model for environmental sex determination studies. *Daphnia magna* is a water flea that follows a parthenogenesis life cycle reproducing asexually, but shifts to sexual reproduction when environmental quality becomes worse (Hebert, 1978). This shift from parthenogenetic to sexual reproduction triggers male differentiation (Figure 2).

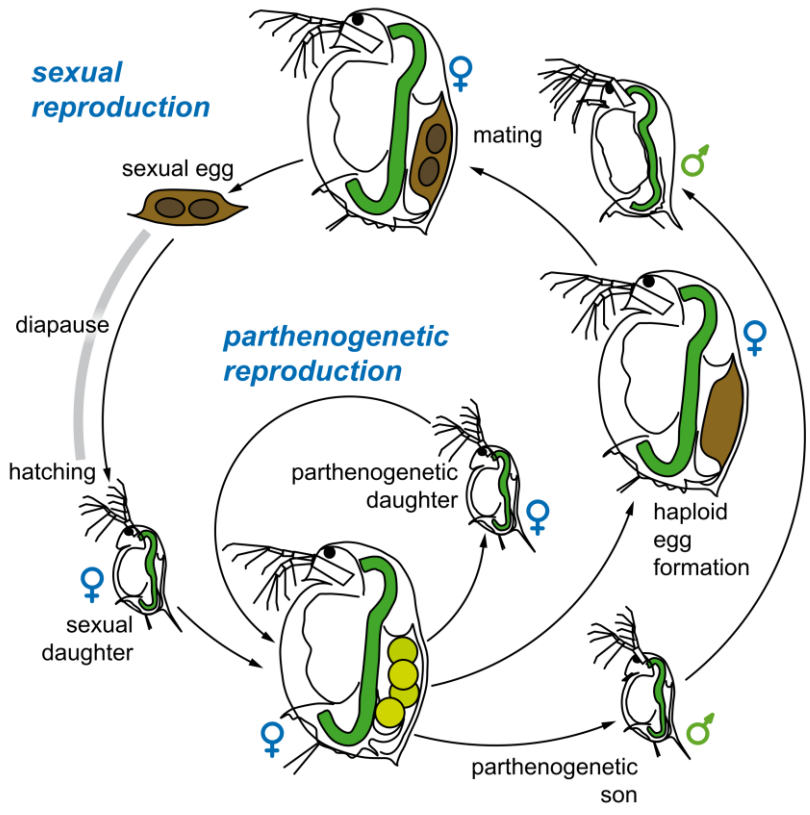

Figure 2. The life cycle of *Daphnia magna.*

The diagram illustrates the parthenogenetic (asexual) and sexual cycle of *Daphnia magna*. Under normal conditions, female *Daphnia* produces diploid eggs that develop to be female offspring through parthenogenesis. When *Daphnia* is stimulated by environmental stress, such as shortened photoperiod, food shortage, and overcrowded population; the female produces male offspring, and later haploid egg that requires male fertilization. The fertilized egg, called the resting egg, is enclosed in a protective hard shell termed ephippium. The resting eggs can endure unfavorable conditions and diapause over a long period of time. Picture source: *https://commons.wikimedia.org/w/index.php?curid=47524211*

Daphnia magna is a fitting model organism because unlike other ESD animals like turtles or crocodiles, it is easily handled in the laboratory. Since it has a small body, it can reproduce fast and can be easily bred. Its sensitivity to environmental changes also makes it advantageous for ESD research. Moreover, genetic information like its expressed sequence tags (EST) and genome sequence are widely available (Colbourne et al., 2011; Watanabe et al., 2005). And currently, genetic manipulation methods have been made possible through this information. RNA interference (Kato, Kobayashi, et al., 2011), ectopic expression (Kato, Matsuura, & Watanabe, 2012), targeted mutagenesis (Naitou, Kato, Nakanishi, Matsuura, & Watanabe, 2015; Nakanishi, Kato, Matsuura, & Watanabe, 2014) and targeted genome modification (Kumagai, Nakanishi, Matsuura, Kato, & Watanabe, 2017; Nakanishi, Kato, Matsuura, & Watanabe, 2015, 2016) are some of the gene manipulation toolboxes that have paved the way to wide gene functional analyses that have aided in understanding molecular mechanisms including of sex determination.

1.5 *Doublesex1*

In light of the advances in researches on sex determination in *Daphnia magna*, it has been discovered that the *Doublesex1* (*Dsx1*) gene is the binary switch that is primarily responsible to determine and regulate male trait development. *Dsx1* encodes a functionally conserved transcription factor, the DM-domain, and exhibits a sexually dimorphic expression pattern during embryogenesis (Kato, Kobayashi, et al., 2011). In the same study, it has been known

that *Dsx1* is alternatively transcribed via two promoters, α, and β, utilizing different transcription start sites as shown in Figure 3. These two promoters lead to the production of two different mRNA isoforms: Dsx1 α and Dsx1 β. The differences in the promoter have led to diverse gene expression, hence the role of the 5′ UTRs of Dsx1 was further investigated (Kato et al., 2018).

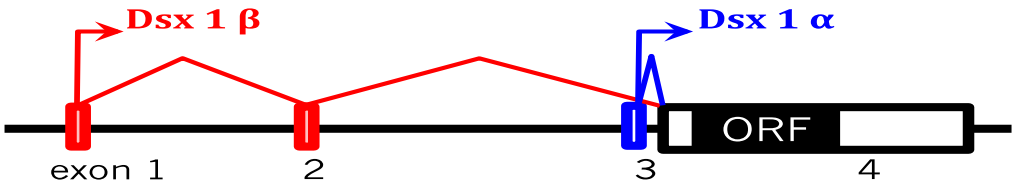

Figure 3. Genomic structure of *Doublesex1* gene in *Daphnia magna*.

Boxes indicate the exons. The red boxes represent the promoter of Dsx1 β, while the blue box represents the promoter of Dsx1 α. The white box is the exon that the two isoforms share which has the *Dsx1* coding sequence or open reading frame (ORF) represented in the black region.

1.6 *DAPALR*

1.6.1 Discovery of *DAPALR*

To investigate the role of the two promoters of Dsx1 α and Dsx1 β, these specific regions were individually introduced into female daphniids (Kato et al., 2018). Injection of the 5′UTRs alone even without the ORF revealed the Dsx1 α 5′UTR can induce the development of the first elongated antennae which a male defining phenotype (Figure 4). While the Dsx1 β 5′UTR and the control did not have any effects. This suggested that the noncoding region of Dsx1 α has a sex-determining ability. Interestingly, a previously annotated noncoding exon encodes this 5′UTR α region in *Daphnia pulex* (Colbourne et al., 2011) and this is orthologous to *Daphnia magna*. This RNA is named DAPALR or the Doublsex1-alpha-promoter-associated-long noncoding-RNA (Kato et al., 2018).

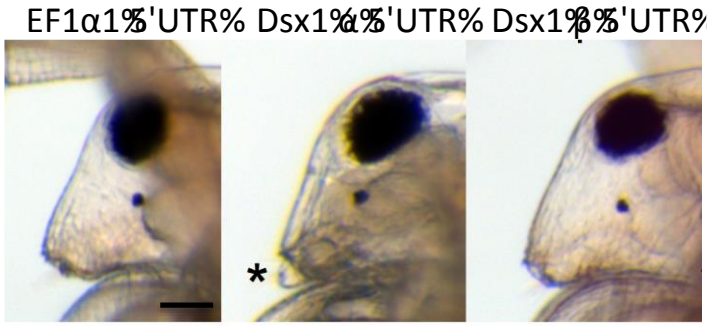

Figure 4. Masculinization of female *Daphnia* injected with Dsx1 α and Dsx1 β 5′ UTRs.

Phenotypic changes induced by dsx1 5′UTRs. Elongated first antennae are represented by an asterisk (*) in the sample injected with Dsx1 α 5′UTR. Scale bar = 100 μm.

1.6.2 Characterization of *DAPALR*

DAPALR is a long noncoding RNA that has a full length of 3,650 bases (Figure 5) and is capped and non-polyadenylated (Kato et al., 2018). It overlaps with *Dsx1 α* 5′UTR in a sense direction.

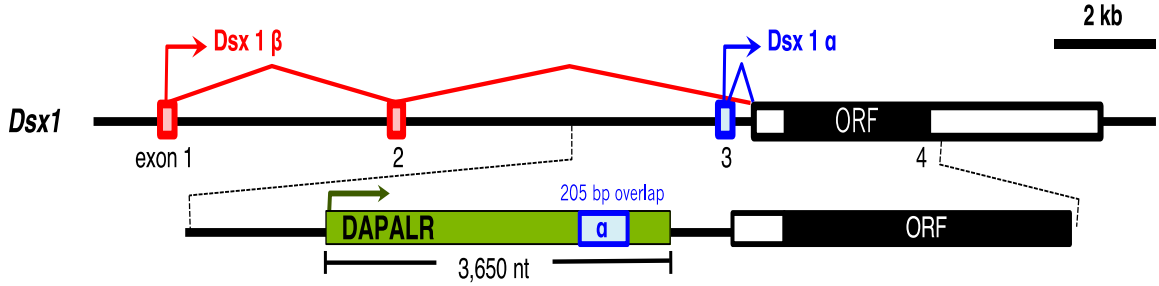

Figure 5. Genomic organization of *Dsx1* gene in *D. magna* showing *DAPALR* position.

Exons are indicated by boxes; red: *Dsx1* β, blue: *Dsx1* α. The black box indicates the ORF of *Dsx1*. The position and orientation of *DAPALR* are indicated by the green box and arrow. The blue box within the *DAPALR* is its overlapping region with the *Dsx1* α 5′UTR.

In the same study, *DAPALR* was further characterized and its expression pattern in female and male embryos shows that its expression is sexually dimorphic (Figure 6). Its expression is male-specific and it peaks from 24 to 72 h post-ovulation, signifying its role in male sex differentiation. To test this, *DAPALR* was knocked down in male eggs and this led to the feminization of male *DAPALR* siRNA-injected samples. As observed in Figure 6, in

contrast to the control, silencing of *DAPALR* in males resulted in the development of shortened first antennae similar to females and the production of viable eggs.

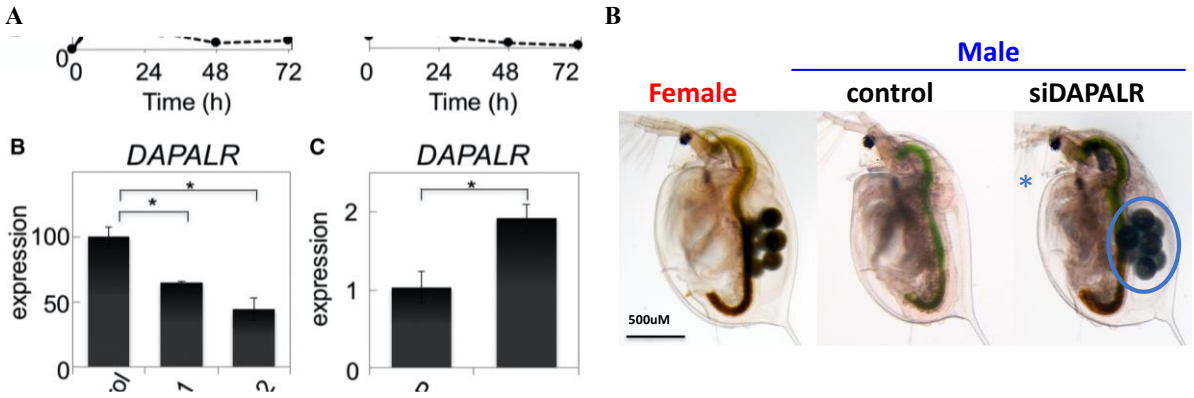

Figure 6. Sexually dimorphic expression of *DAPALR* and feminization of male due to *DAPALR*-silencing (Kato et al., 2018).

(A) Temporal dimorphic expression of *DAPALR* during embryogenesis. Results are shown as expression level relative to the ribosomal protein L32. **(B)** Lateral view of daphniids injected with control and DAPALR siRNAs. DAPALR siRNA-injected developed short first antennae represented by an asterisk (*) and produced eggs in their brood chamber as encircled.

1.7 Objective of this study

In this study, I aim to reveal how *DAPALR* specifically regulates *Dsx1*. Recently, the DAPALR, a sense-overlapping lncRNA, was found involved in the sex determination process in *D. magna*. However, amidst establishing the relationship of *DAPALR* with *Dsx1* in *D. magna*, its molecular mechanism remains unknown.

Long noncoding RNAs, like *DAPALR*, are pervasively transcribed in the genome, representing most of the total transcript (Mellor et al., 2016). Studies have also emphasized the various significant roles lncRNAs play in cell state, differentiation, development, and disease (Kung et al., 2013; Quinn & Chang, 2016). And among the different types of lncRNA, recent findings revealed that sense long noncoding RNAs comprises the majority of lncRNAs present (Kuo et al., 2017); however, this group remains poorly studied because of the difficulty in

distinguishing them from their overlapping protein-coding genes, little is known in their functions and mechanisms.

Hence, this research aims to give light on the unique function and significant role of a lncRNA in biological processes like sex determination, and also pave a deeper understanding of the biogenesis and function of the poorly studied group of sense-overlapping lncRNAs. To accomplish these goals, I conducted this research which is mainly divided into three parts:

1. The functional analysis of *DAPALR* was conducted. Specifically, I confirmed that *DAPALR* activates *Doublesex1* (*Dsx1*) expression. I also demonstrated that its transactivation element, its overlapping region with *Dsx1* α 5′ UTR, can alone activate *Dsx1*. Finally, using this transactivation element, RNA pulldown was conducted to identify two associating proteins of DAPALR: Alan shepard (Shep) and CGUBP1. This part will be described in Chapter 2.
2. I then focused on one associating protein, the *Shep*, and characterized its function by performing knockdown, mutagenesis, and overexpression experiments. These revealed that Shep represses *Dsx1* and regulates it at the post-transcription level. This will be described in Chapter 3.
3. Lastly, the mechanism of Shep and *DAPALR* in regulating the binary expression of *Dsx1* was elucidated by post-transcription assays. *In vivo* and *in vitro* translation experiments showed that Shep binds to and represses the *Dsx1* mRNA and *DAPALR* sequesters Shep to activate the *Dsx1* translation. This will be described in Chapter 4.

CHAPTER 2 *DAPALR* activates *Doublesex1*

2.1 Introduction

Sense overlapping lncRNAs are predicted to be the most abundant type of lncRNA but they remain poorly studied because of the difficulty of distinguishing them from the protein-coding genes they overlap with. lncRNAs under this group have not been extensively studied yet and little is known about their mechanisms and functions.

In light of this, a sense overlapping lncRNA was recently discovered in *D. magna*. The capped and polyadenylated *DAPALR* or *Doublsex1-alpha-promoter-associated-long noncoding-RNA* was found to be involved in the process of sex determination (Kato et al., 2018). Preliminary studies showed that *DAPALR*, like *Dsx1*, the gene it overlaps with and the major effector of sexual differentiation in *D. magna*, is sexually dimorphic and is only expressed in males.

DAPALR silencing on male *Daphnia* embryos also resulted in feminization (Kato et al., 2018), suggesting that *DAPALR* may play a role in *Dsx1* regulation. In accordance with the previous reports, I hypothesized that overexpression of *DAPALR* in females can de-repress *Dsx1*. In addition, since the silencing of the *DAPALR* thru RNAi may also result in heterochromatin formation of the neighboring gene *Dsx1* as reported in *Drosophila melanogaster* and *Caenorhabditis elegans* (Martienssen & Moazed, 2015), the ectopic expression of the *DAPALR* full-length is necessary to ensure that *DAPALR* expression is sufficient and indeed vital to trigger *Dsx1* activation, consequently male development in females.

Since the full region of *DAPALR* spans a length of 3,650 bases, it would also be helpful to determine its core element or the region required for its transactivation. The minimal element of *DAPALR* that initiate its transcription and function may lead to better understand its mode of action. The transactivation element could also encompass the binding site/s of the RNA-

binding proteins (RBPs) of *DAPALR*. Mechanism of lncRNAs often involves RBPs that they interact and form a complex with to activate or repress their target gene (Wang and Chang, 2011).

In this chapter, I investigated the effect of overexpression of the full region of *DAPALR* in the *Dsx1* expression. I also determined the transactivation element of *DAPALR* by ectopic expression of its partial regions.

2.2 Materials and Methods

2.2.1 Wildtype *Daphnia magna* and *Doublesex1* reporter strain culture condition

Wildtype *Daphnia magna* strain (NIES clone) was used to extract the cell lysate for the identification of the associating proteins of DAPALR. This strain was obtained from the National Institute of Environmental Studies (NIES, Tsukuba, Japan) and cultured for many generations. While the transgenic *Daphnia magna* generated in the laboratory was used for most parts of the research (Nong, Mohamad Ishak, Matsuura, Kato, & Watanabe, 2017). This *Daphnia* line serves as a reporter of the *Doublesex1* gene. It contains the mCherry gene in one of the alleles of the *Doublesex1* gene; hence, mCherry fluorescence reflects the activation of the *Dsx1* gene. It allows easy detection of male production and observation of the male-specific organs where *Dsx1* is highly localized. This line also has GFP gene fused to *D. magna* histone H2B gene under the control of the elongation factor 1α1 (EF1α1) gene promoter permitting visualization of the embryo developmental stages.

Daphniids were cultured in Aachener Daphnien Medium (ADaM) (Klüttgen, Dülmer, Engels, & Ratte, 1994) under the following conditions: culturing temperature of 22-24°C in a constant light/dark photoperiod of 16h/8h. Furthermore, the daphniids were fed once a day with 120 µl of $8x10^9$ cells/mL *Chlorella vulgaris* (Nikkai Center, Tokyo, Japan) and 10 µl 0.15 g/mL baker's yeast (Marusan Pantry, Ehime, Japan).

2.2.2 Artificial male production by Fenoxycarb exposure

Male embryos were needed for the overexpression of *DAPALR* expressing-plasmid in males. For that, 2-3 weeks old adult daphniids with 42 h old embryos in their back were exposed to synthetic juvenile hormone agonist, fenoxycarb (Wako Pure Chemical Industries, Japan), a critical time point previously identified (Tatarazako et al, 2003). 1 mg/L of fenoxycarb was prepared using N-N-Dimethyl Formamide (Nacalai teque Inc, Kyoto, Japan) as solvent. It was then added to ADaM medium so the final concentration of the fenoxycarb was 1 µg/L. Selected daphniids were exposed to this medium for 16-18 h until they start ovulating, then the newly-ovulated eggs in the brood chamber were collected for their intended use.

2.2.3 Ectopic expression of *DAPALR*-expressing plasmids

Plasmid expressing *DAPALR* was constructed by flanking the full-length sequence of DAPALR with the EF1α1 5′UTR and 3′UTR. And to identify the transactivation element of *DAPALR*, its region that overlaps with the Dsx1 α 5′UTR was also overexpressed. This 205bp region was inserted to replace the DAPALR full sequence in the abovementioned plasmid. The effect of the other regions of DAPALR was also examined by the removal of the 205 bp overlapping region from the plasmid having the full region of DAPALR and overexpressing this to compare with the other expression plasmids. On the other hand, the control plasmid has the DAPALR/ Dsx1 α 5′UTR overlap/ DAPALR other region sequence removed. Al the plasmid constructs are shown in Figure 7. These plasmids (200 ng/µl) were each injected into female and male eggs of the *Dsx1*-reporter strain. To check the change in *Dsx1* or *mCherry* expression levels, injected eggs were observed 30 h after injection and collected after 48 h for

RNA extraction, cDNA synthesis, and Real Time- quantitative PCR (RT-qPCR) to check for the change in *Dsx1* expression level.

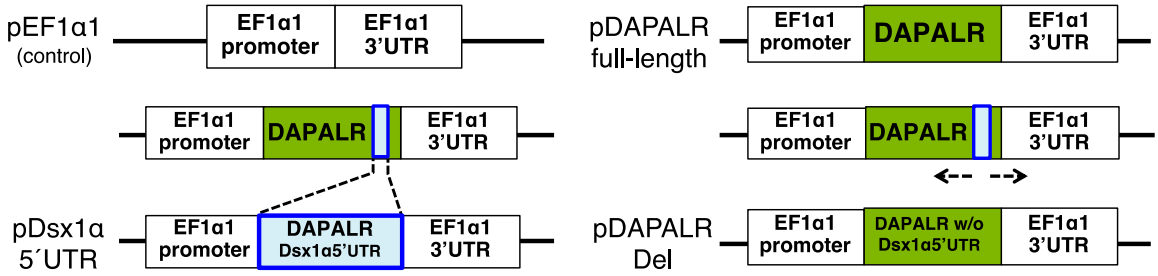

Figure 7. Variants of the *DAPALR*-expressing plasmids used for ectopic expression.

Plasmid harboring the EF1α1 5´UTR and 3´UTR was used as the vector and control. pDAPALR full-length was sued to expressed the 3.65kb sequence of DAPALR. pDsx1α 5´UTR has the full sequence of DAPALR replaced with its partial region overlapping with Dsx1α 5´UTR. Lastly, pDAPALR Del expresses the other regions of DAPALR without the 205 bp overlapping region with Dsx1α 5´UTR.

2.2.4 Microinjection

Following the established protocol for microinjection (Kato, Kobayashi, et al., 2011), embryos of the *Dsx1* reporter strain were collected from the brood chamber of 2-3 week-old daphniids right after ovulation and were transferred to ice-chilled M4 medium with 80 mM sucrose (Elendt & Bias, 1990). The injection needle made of 0.6 mm ID glass capillary (Narishige, Tokyo, Japan) was prepared using a capillary puller machine (Sutter Instrument, Novato, USA). The injection samples were loaded from the un-pulled end of the capillary after centrifugation at 15,000 rpm for 5 min. And the tip of the needle was manually cut under the microscope using a sharp, insect pin (Shiga, Tokyo, Japan). To confirm the right size of the needle opening, the injected sample volume was normalized by measuring its droplet in oil at approximately 0.2 nL after pressing the injection pedal for 3-5 sec. Nitrogen gas was used as a pressure source and was set within the range of 300-450 kPa. Balance pressure of 16 kPa was also set to avoid the entry and contamination of the M4 medium into the capillary. Microinjection was performed within 1 h after ovulation to prevent the hardening of the egg membrane. After which, the surviving eggs were transferred into each well of 96-well plates

(BD Falcon, USA) which has 100 µL of M4-sucrose medium and were then kept in an incubator at 23 °C.

2.2.5 Fluorescence photography and quantification

Samples were observed and their photos were taken using Leica DC500 CCD Digital Camera mounted on Leica M165FC fluorescence microscope (Leica Microsystem, Mannheim, Germany). Fluorescence photography of embryos was taken using GFP and mCherry filters under the settings listed in Table 1.

Table 1. Camera settings for fluorescence observation.

Filter	GFP	mCherry
Exposure	1.0 s	2.0 s
Gain	3.0 x	8.0 x
Saturation	1.0	1.0
Gamma	1.0	1.6

2.2.6 Total RNA extraction

Samples from different experiments intended for RNA extraction were collected into a 2.0 mL tube (TOMY, TM-625S). Zirconia beads of 1.0 ⌀ and 3.0 ⌀ sizes were added to each tube and homogenized together with Sepasol-RNA I Super G reagent (Nacalai Tesque Inc., Kyoto, Japan) using Micro Smash machine (TOMY, MS-100) followed. Total RNA was isolated according to the Sepasol manufacturer's protocol; followed by phenol/chloroform extraction and washing of RNA pellet using 70% cold ethanol. The dried RNA pellet was then re-solubilized with 18 µL of ultra-pure distilled water. Lastly, the amount of purified total RNA was measured using Nanodrop 2000 (Thermo Fisher Scientific).

2.2.7 First-strand cDNA synthesis by Reverse Transcription

Following the Reverse-Transcriptase method, 1 μg of the purified total RNA extract was used to synthesize the cDNA using PrimeScript III Reverse Transcriptase (Takara Bio, Shiga, Japan). The RNA templates mixed with 25 ng of Random Primer (Invitrogen, Carlsbad CA, USA), 4 μL of 2.5 mM dNTPs mix (Takara Bio, Shiga, Japan), and ultra-pure distilled water were heated at 65 °C for 5 min. After incubation on ice for at least 1 min, 4 μL of 5X First-Strand Buffer (Takara Bio, Shiga, Japan), 1 μL of RNaseOUT Recombinase Ribonucelase Inhibitor (Invitrogen, Carlsbad CA, USA) and 200 units of PrimeScript III Reverse Transcriptase (Takara Bio, Shiga, Japan) were added into the denatured sample. To ensure that there is no genomic DNA contamination in the RNA samples, the same reaction mixture was prepared for each sample but replacing the Reverse Transcriptase enzyme with ultra-pure distilled water. All mixtures were incubated in the following conditions: 30 °C for 10 min, 42 °C for 60 min, 70 °C for 15 min using Takara Thermal Cycler PCR machine.

To confirm the synthesized cDNA and check the purity of the samples from genomic DNA contamination, the β-actin gene was amplified by PCR using the following primers: forward primer (5′-GGCAAGGAATAGTTCGATAC-3′) and reverse primer (5′-CACCGACGTACGAATCCTTCTGACC-3′). Amplified products were then run on gel electrophoresis.

2.2.8 Quantitative real time-PCR

To measure the expression level of *Dsx1* in the overexpression experiments, the cDNAs of each sample were prepared in three replicates for RT-qPCR analysis. mRNA transcripts were measured using Mx3005P Real-Time QPCR System (Agilent Technologies) under the following conditions: 50 °C for 2 min, 95 °C for 10 min, 40 cycles of 95 °C for 15 sec and 60 °C for 1 min and using SYBR Green qPCR SuperMix (Invitrogen, Carlsbad CA, USA) and specific primers designed to amplify short PCR products (<150 bp). Expressions based on the

Ct value during amplification were calculated and normalized by quantitating the expression level of the ribosomal protein, *L32*, as the reference gene. The normalized expression levels of the treated samples are then relatively compared to the expression levels of the control to get the final values. Lastly, gel electrophoresis and dissociation curve analysis were performed to confirm the correct amplicon size and the absence of non-specific bands. Primers used for this experiment are shown in Table 2.

Table 2. Primer sequences for RT-qPCR for the ectopic expression experiment.

Gene name	Primer sequence (5′-3′)
Dsx1	Forward: AAGTTTGGTGTAGGGGAGGATGAG Reverse: CCATTCATCATTACCAAATCCCTTC
L32	Forward: GACCAAAGGGTATTGACAACAGA Reverse: CCAACTTTTGGCATAAGGTACTG

2.3 Results

2.3.1 Activation of *Doublesex1* by *DAPALR*

To confirm the hypothesis that overexpression of *DAPALR* can de-repress *Dsx1*, the 3.65 kb full region of *DAPALR* driven by the *EF1α1* promoter/enhancer was ectopically expressed in female eggs of the *Dsx1* reporter strain. This transgenic strain expresses a red fluorescence protein mCherry under the endogenous *Dsx1* promoter/enhancer and it also expresses the H2B-GFP to allow visualization of the phenotypes of the cell and tissue influenced by the Dsx1 activity (Nong et al., 2017).

In this experiment, results showed that *Dsx1* was activated in injected embryos in contrast to those injected with the control plasmid which is the *EF1α1* promoter/enhancer vector without the *DAPALR* sequence (Table 3).

Table 3. Ratio of *Dsx1*-activated to survived embryos after *DAPALR* overexpression.

	survived embryos	fluorescent embryos	Dsx1 activation (%)

pEF1α1	17	0	0%
pDAPALR full-length	38	36	95%
pDsx1 α 5′ UTR	19	16	84%
pDAPALR Del	26	0	0%

Figure 8A shows the difference between the fluorescence signals of the samples injected with the two plasmids. Ubiquitous mCherry fluorescence signals were observed in *DAPALR* sequence-injected samples, while the control plasmid did not result in any changes in the mCherry fluorescence signals. Moreover, comparing the resulting mCherry expression pattern on the sample injected with the full sequence of *DAPALR* with the GFP fluorescence as shown in the merged images of GFP and mCherry, it can be deduced that mCherry was highly expressed in most parts of the embryo but less in the thoracic appendages and the midline where GFP signal is high.

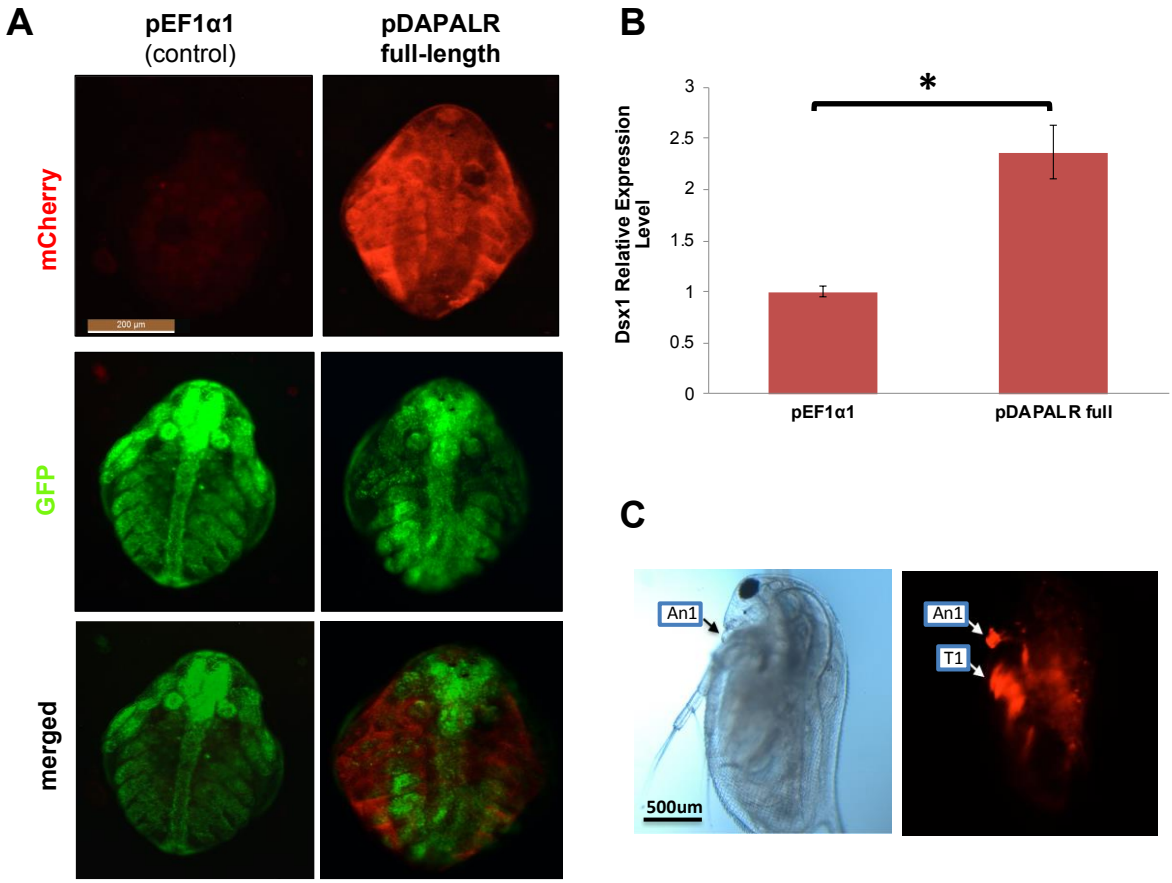

Figure 8. Overexpression of the full region of *DAPALR* in female.

(A) Ventral view of female embryos of *Dsx1* reporter strain injected with *DAPALR*-expressing plasmid and control plasmid (*EF1α1* vector) observed at 30h after injection. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body structures. The merged images of mCherry and GFP were used to understand the localization pattern of mCherry expression. **(B)** Gene expression profile of *Dsx1* in female embryos injected with pDAPALR and pEF1α1. Error bars indicate the standard error of the mean (n=3). *p<0.05 (Student's T-test). **(C)** Lateral view of a female juvenile after *DAPALR* overexpression. An1: first antennae; T1: first thoracic appendage.

In addition, the observed increase in mCherry fluorescent signals was confirmed to represent *Dsx1* activation after checking the expression level of *Dsx1* by RT-qPCR. Significant activation of the endogenous *Dsx1* was observed on *DAPALR* plasmid-injected embryos (Figure 8B). Ectopic expression of the full sequence of *DAPALR*, increased *Dsx1* expression 2.5x higher than the expression of *Dsx1* in the control treatment.

It should also be noted that at 72 h after injection, one of the female samples injected with the *DAPALR* full-length plasmid developed the elongated antennae and has mCherry fluorescence localized in male-specific parts such as the first antennae and first thoracic appendage which are all male-specific phenotypes (Figure 8C). These results demonstrate that the full region of *DAPALR* can trigger male trait development as it can activate *Dsx1* expression in *Dsx1*-silenced females.

The effect of ectopic expression of the full sequence of DAPALR was also tested on males (Figure 9). Similar results with the female were observed in *DAPALR* sequence-injected male samples. Unlike in the control samples, where mCherry signals could only be observed in its male-specific organs such as the regions where the first antennae, the first thoracic appendages, and the genitalia will develop, ectopic expression of *DAPALR* resulted in ubiquitous expression of the mCherry signals in the whole body of the embryo (Figure 9A). qRT-PCR results also tallied with the phenotype observed as significant Dsx1 enhancement was observed in *DAPALR* sequence-injected male samples relative to the control (Figure 9B).

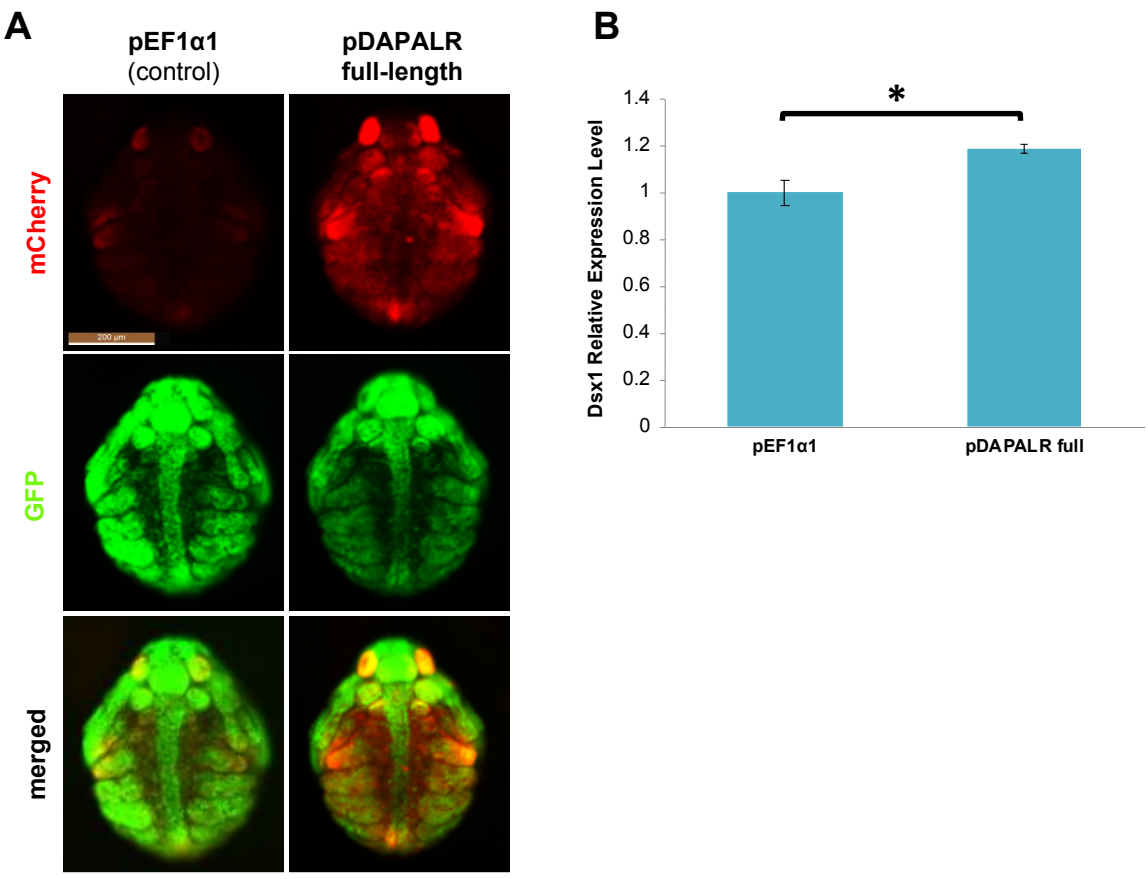

Figure 9. Overexpression of the full region of *DAPALR* in male.

(A) Ventral view of male embryos of *Dsx1* reporter strain injected with *DAPALR*-expressing plasmid and control plasmid (*EF1α1* vector) observed at 30h after injection. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body structures. The merged images of mCherry and GFP were used to understand the localization pattern of mCherry expression. **(B)** Gene expression profile of *Dsx1* in male embryos injected with pDAPALR and pEF1α1. Error bars indicate the standard error of the mean (n=3). *p<0.05 (Student's T-test).

2.3.2 *Dsx1 α* 5′ UTR overlapping region as the transactivation element of *DAPALR*

To determine which region of *DAPALR* is sufficient for activation of *Dsx1*, the *DAPALR* full sequence was replaced with its partial region that overlaps with *Dsx1α* 5′UTR. This 205 bases overlapping sequence was expressed in the same vector plasmid driven by the *EF1α1* promoter/enhancer and was injected into female eggs of the *Dsx1* reporter strain.

Injected samples of this plasmid showed a similar phenotype as the samples injected with the plasmid expressing the full region of *DAPALR*. Results showed that the overexpression of the *DAPALR* overlapping region with *Dsx1α* 5′UTR also led to ubiquitous mCherry expression

(Figure 10A). It can also be observed that the expression pattern of mCherry after overexpression of this partial region of *DAPALR* is broader and more enhanced than the effect of the full region. This result matched the RT-qPCR results wherein the overexpression resulted in an almost 3-fold activation of the endogenous *Dsx1* in contrast to the control (Figure 10B).

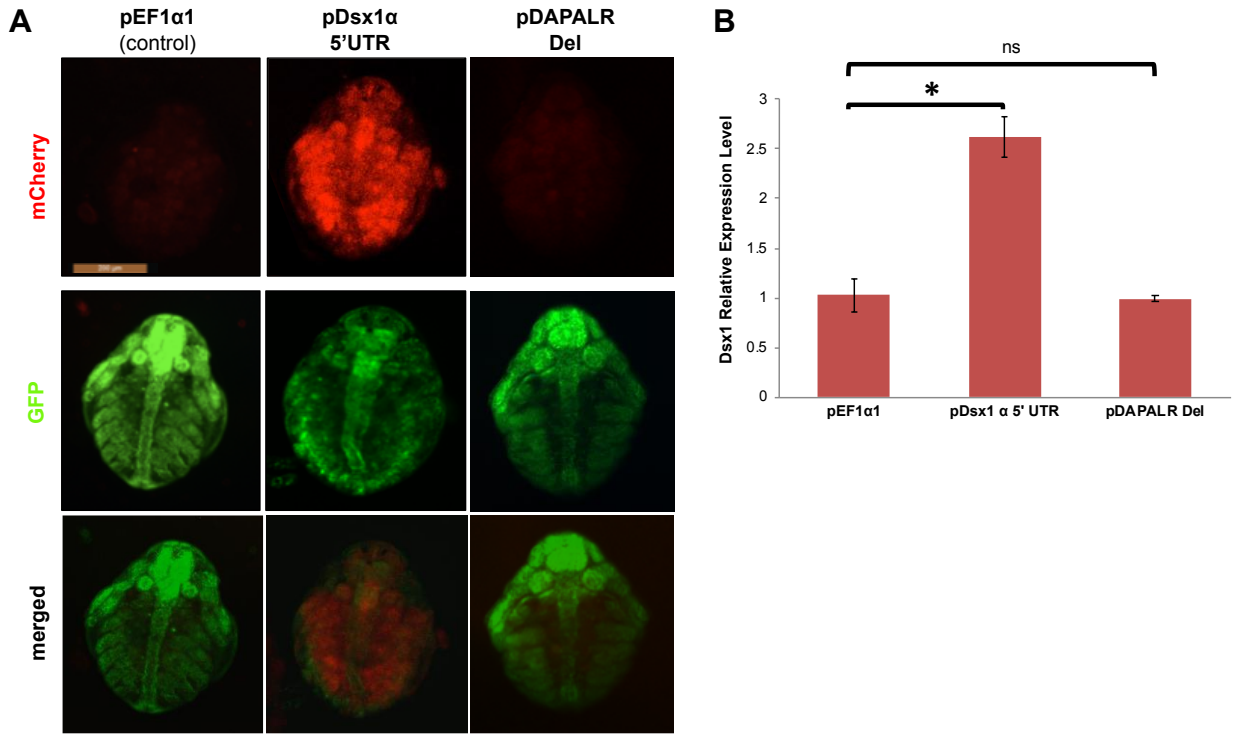

Figure 10. Overexpression of the 205 bp region of *DAPALR* overlapping with *Dsx1 α* 5′ UTR.

(A) Ventral view of female embryos of *Dsx1* reporter strain injected with plasmids expressing the partial region of DAPALR overlapping with *Dsx1 α* 5′ UTR (p*Dsx1 α* 5′ UTR) and its other regions without the 205 bp overlapping sequence (pDAPALR Del) and control plasmid (*EF1α1* vector) observed at 30h after injection. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body structures. The merged images of mCherry and GFP were used to understand the localization pattern of mCherry expression. **(B)** Gene expression profile of *Dsx1* in female embryos injected with pDsx1α 5′ UTR, pDAPALR Del, and pEF1α1. Error bars indicate the standard error of the mean (n=3). *p<0.05, ns: not significant (Student's T-test).

To confirm the importance of the overlapping region with *Dsx1α* 5′UTR on *Dsx1* activation, this region was removed from the *DAPALR*-expressing plasmid to make the pDAPALR Del. This plasmid containing the other regions of *DAPALR* but the 205 bases that

overlap with *Dsx1α* 5′UTR was injected and results were compared with that of the ectopic expression of *DAPALR* and of the *Dsx1α* 5′UTR overlapping region. It can be observed in Figure 10A that the samples injected with the other regions of *DAPALR* did not show any increase in mCherry signals same as with the control, in contrast to the bright signals observed on samples expressing the *Dsx1α* 5′UTR overlapping region. RT-qPCR results reflected the same results, wherein *Dsx1* mRNAs levels of the pDAPALR Del-injected samples were the same as the control.

All these results demonstrated that the *Dsx1α* 5′UTR overlapping region is responsible for *Dsx1* activation. Hence, this partial region overlapping with *Dsx1α* 5′UTR was regarded as the transactivation element of *DAPALR*.

2.4 Discussion

To understand the role of *DAPALR* in sex determination, specifically in relation to *Dsx1* regulation, the full sequence of *DAPALR* was overexpressed in *Dsx1*-silenced female eggs of the *Dsx1* reporter strain. Results showed ubiquitous mCherry fluorescence and a significant increase in *Dsx1* expression level. These results indicated that ectopic expression of *DAPALR* can de-repress *Dsx1*. Since *DAPALR* is only expressed in males (Kato et al., 2018), *DAPALR* could be the one responsible for activating *Dsx1* to launch the male-determining pathway.

Moreover, in males, ectopic expression of the full sequence of *DAPALR* also resulted in a significant increase in *Dsx1* expression level and enhanced and ubiquitous mCherry fluorescence signals not only in its male-specific parts but in the entire body. These results suggest that *Dsx1* can also be transcribed in non-sexually dimorphic tissues. However, this stochastic expression of *Dsx1* is not observed in nature; implying that another factor may be responsible for this control.

Furthermore, the same activation of Dsx1 was also observed when the partial region of *DAPALR* was over-expressed. The *Dsx1* α 5′UTR-overlapping region of *DAPALR* also enhanced Dsx1, showing ubiquitous mCherry expression in injected eggs relative to the control and the samples injected with the other region of *DAPALR* without it. It was also observed that the expression pattern of mCherry after overexpression of this partial region of *DAPALR* is broader and more enhanced than the effect of the full region. This may be due to the other elements present in the full region of *DAPALR* that possibly control or limit its function in activating *Dsx1*. Hence, this region was regarded as the transactivation element of *DAPALR*.

Although current results cannot demonstrate the extent of *DAPALR*'s mechanism, it showed that the sense-overlapping lncRNA *DAPALR* can activate *Dsx1* expression in *trans*. Whether *DAPALR* activates *Dsx1* at the chromatin, transcription or post-transcription level is still unknown. As suggested by a previous study that the *Dsx1* locus is in a closed chromatin state in females, it is possible that *DAPALR* might recruit chromatin-remodeling factors to trigger a change in the chromatin structure of *Dsx1*, leading to its higher expression. Or like other extra-coding RNAs, *DAPALR* can also activate *Dsx1* by regulating its methylation dynamics (Savell et al., 2016). Several lncRNAs have also been reported to activate the transcription of its target gene through enhancer elements or its transcription itself (Ard, Allshire, & Marquardt, 2017). In addition to possible roles in transcription and nuclear organization, *DAPALR* can also function s post-transcriptional, translational or post-translational regulator of *Dsx1* (Statello, Guo, Chen, & Huarte, 2021). Since the mechanism and function of lncRNAs mostly depend on its interacting partners, the key to confirm which regulation level *DAPALR* is involved in to activate *Dsx1* is through identifying its binding proteins.

CHAPTER 3 *DAPALR*'s associating RBP, Shep, represses *Dsx1* at post-transcription level

3.1 Introduction

LncRNAs exert their roles in cell functions through different mechanisms. Their versatile features depend on several reasons, one of which is hinged on their interacting partners; they mainly act through their ability to establish interactions with DNA, RNA, or proteins (Marin & Baker, 1998).

LncRNAs associate with RNA-binding proteins (RBPs) and the mechanism of lncRNAs often involves RBPs that they interact and form a complex with to activate or repress their target gene (Wang and Chang, 2011). Many lncRNAs have been reported to interact with different types of RBPs, including DNA methyltransferases, transcription factors, and splicing factors (Li et al., 2015)(Hentze, Castello, Schwarzl, & Preiss, 2018).

RBPs are well-known for their roles in regulating RNA fate from synthesis to decay and they are also involved in protein translation by guiding or blocking RNAs. A wide range of RBPs has been discovered and investigated over the years to regulate gene expression during chromatin remodeling, transcription, and post-transcription levels. These proteins bind to RNA through one or multiple RNA-binding domains and then change the fate or function of the bound RNAs (Briata & Gherzi, 2020). The combination of the versatility of their RNA-binding domains with their structural flexibility enables RBPs to be involved in virtually all the regulatory layers in the cell and to control the metabolism of a large array of transcripts.

In this chapter, to fully understand the mechanism of *DAPALR*, I identified the RBPs that interact in its transactivation element. From this proteins, I then explored the functional role of one of *DAPALR's* associating protein, the Shep, in sex determination in *D. magna*. I annotated Shep in *D. magna* and analyzed its evolutionary conservation compared to orthologs in other species. I also checked its temporal expression profile and performed loss-of-function,

mutation, and overexpression experiments to understand how it works with *DAPALR* in regulating *Dsx1*.

3.2 Materials and Methods

3.2.1 Wildtype *Daphnia magna* and *Doublesex1* reporter strain culture condition

The *D. magna* (NIES clone), obtained from the National Institute for Environmental Studies (NIES; Tsukuba, Japan), and the transgenic *Dsx1* reporter strain were cultured under the same laboratory condition mentioned in Chapter 2 (see 2.2.1 Wildtype *Daphnia magna* and *Doublesex1* reporter strain culture condition). The male daphniids were obtained with the same method described in Chapter 2 (see 2.2.2 Artificial male production by Fenoxycarb exposure).

3.2.2 Preparation of bait RNAs and RNA pulldown assay

As shown in Figure 11, for the identification of the associating proteins of *DAPALR*, preparation of Flag peptide conjugated bait RNAs were carried out as described previously (Adachi et al., 2014). Briefly, the T7-tagged cDNA template was amplified by the polymerase chain reaction (PCR), transcribed in vitro using the MEGAscript T7 kit (Invitrogen, Carlsbad CA, USA), and purified with an RNeasy Mini Kit (Qiagen). The 3' end of purified cRNA was dialdehyded with 0.1 M $NaIO_4$, precipitated with 2% $LiClO_4$ in acetone and then washed with acetone. The pellet was dissolved in 0.1 M sodium acetate, pH 5.2, and then mixed with 30 mM hydrazide–Flag peptide. The resulting imine-moiety of the cRNA was reduced by adding 1 M NaCNBH3. The Flag -tagged-RNA was purified with an RNeasy Mini Kit (Qiagen).

For the pulldown assay, 1- or 2-day-old female larvae were lysed with lysis buffer [10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.5), 150 mM NaCl, 50 mM NaF, 1 mM Na3VO4, 5 μg/ml leupeptin, 5 μg ml aprotinin, 3 μg/ml pepstatin A, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mg/ml digitonin] using pre-chilled Dounce

homogenizer (type A pestle) and cleared by centrifugation. One mg of cleared lysate was incubated with five pmol of Flag-tagged bait RNA, anti-FLAG antibody (Sigma) and protein G conjugated magnetic beads (Thermo) rotate for 1h at 4 °C. The magnetic beads were then washed three times with wash buffer [10 mM HEPES (pH 7.5), 150 mM NaCl, 0.1% Triton X-100] and co-immunoprecipitated RNA and proteins were eluted with Flag elution buffer [0.5 mg/ml Flag peptide, 10 mM HEPES (pH 7.5), 150 mM NaCl, 0.05% Triton X-100]. The bait RNA-associated proteins were then precipitated with TCA. Precipitated protein was re-dissolved in guanidine hydrochloride and reduced with TCEP, alkylated with iodoacetamide, followed by digestion with lysyl endopeptidase and trypsin. The digested peptide mixture was applied to a Mightysil-PR-18 (Kanto Chemical) frit-less column (45 3 0.150 mm ID) and separated using a 0–40% gradient of acetonitrile containing 0.1% formic acid for 80 min at a flow rate of 100 nL/min. Eluted peptides were sprayed directly into a mass spectrometer (QSTAR Elite, Sciex). The mass spectrometry and tandem mass spectrometry spectra were obtained in information-dependent acquisition mode and were queried against the *Daphnia magna* protein database (http://arthropods.eugenes.org/EvidentialGene/daphnia/daphnia_magna_new/Genes/earlyaccess/) with an in-house Mascot server (version 2.2.1. Matrix Science; Natsume et al., 2002).

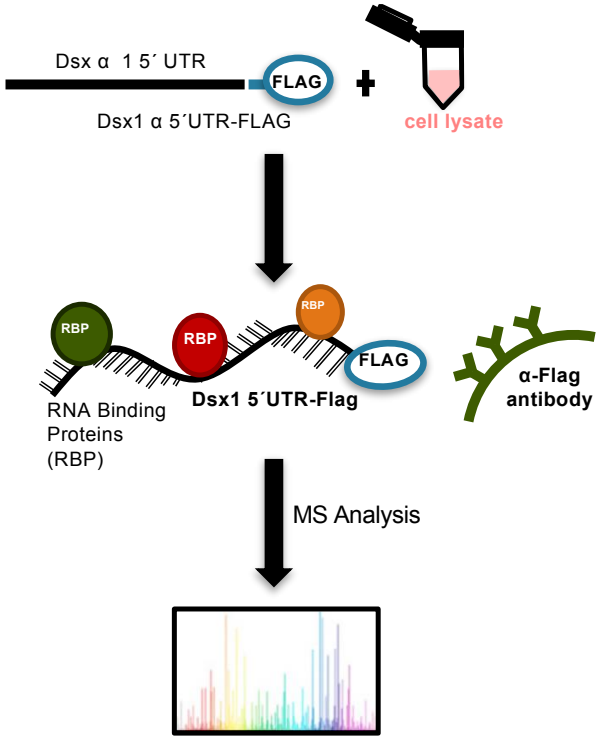

Figure 11. Outline of the RNA pulldown assay for the identification of the RNA binding proteins of *DAPALR*.

3.2.3 Phylogenic analysis

Amino acid sequences of *Shep* family genes were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/) as shown in Table 4; and the conserved sequences, the RNA Recognition Motif (RRM) domains of each protein, were used to construct the phylogenetic tree. Multiple sequence alignments of the amino acid sequences were constructed using the ClustalW (Thompson, Higgins, & Gibson, 1994) in the MEGA program (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). The following settings were used for the analysis: pairwise alignment parameter: gap opening penalty = 6.00, gap extension penalty = 0.21, and identity protein weight; matrix multiple alignment parameter: gap opening penalty = 10.00, gap extension penalty = 10.00, gap extension penalty = 0.24, delay divergent cut-off = 30%, and gap separation distance = 4. The phylogenetic reconstruction was performed using the p-distance algorithm and the neighbor-joining method was implemented in MEGA.

Table 4. Accession numbers of Shep ortholog genes used in this study.

Common name	Scientific name	Gene name (Definition in NCBI)	Accession no.
Water flea	*Daphnia magna*	putative RNA-binding motif, single-stranded-interacting protein	KZS13815.1
Water flea	*Daphnia pulex*	hypothetical protein DAPPUDRAFT_4034	EFX85883.1
Red flour beetle	*Tribolium castaneum*	Protein alan shepard	XP_974237.2
Fruit fly	*Drosophila melanogaster*	alan shepard	NP_729054.3
Eastern bumblebee	*Bombus impatiens*	Protein alan shepard	XP_021207606
Dry wood termite	*Cryptotermes secundus*	Protein alan shepard	PNF18380.1
Roundworm	*Caenorhabditis elegans*	SUPpressor	NP_497856.1
Zebrafish	*Dania rerio*	RNA-binding motif, single-stranded-interacting protein	NP_001070184.1
Australian ghost shark	*Callorhincus milii*	RNA-binding motif, single-stranded-interacting protein	XP_007905195.1
Mouse	*Mus musculus*	RNA-binding motif, single-stranded-interacting protein	XP_06500005.1
Human	*Homo Sapiens*	RNA-binding motif, single-stranded-interacting protein	XP_016860115.1

3.2.4 Expression profile of *Shep*

To analyze the temporal changes in *Shep* expression levels during embryogenesis and adult stages male and female daphniids were collected at different time points: 0, 6, 12, 18, 24, 30, 48, and 72 h after ovulation for embryos and 1, 2, and 3 weeks for adult. These samples were subjected to RT-qPCR using the cDNA synthesized from the total RNA of daphniids at each stage.

3.2.5 Gene knockdown by RNAi

Small interference RNAs were designed using the Block-iT RNAi Designer at http://www. invitrogen.com/rnaidesigner.html. The siRNA targeting Shep gene sequence is as follows: shep_siRNA (5′-GCCTCCTATCAAGCGTCAA-3′). While for the negative control

targeting a random sequence that does not affect the development of the *Daphnia*, this siRNA sequence was used: control_siRNA (5′-GGUUAAGCCGCCUCACAUTT-3′) (Asada, Kato, Matsuura, & Watanabe, 2014a). Two nucleotides dTdT were added to each 3′ end of the siRNAs. The siRNAs were diluted with the injection marker 1 mM Lucifer Yellow dye (Invitrogen, Carlsbad CA, USA) to have the final concentration of 100 μM and were injected into eggs of the *Dsx1* reporter daphnia strain at 2-3 weeks of age which were destined to be male and female. Samples were then observed at 30 h after injection and collected at 48 h for RNA extraction and cDNA synthesis as previously described (Kato et al., 2018). RT-qPCR was then performed to check the expression level changes of the genes of interest (*shep* and *Dsx1)* between the control_siRNA- and shep_siRNA-injected samples.

3.2.6 Mutagenesis by CRISPR/Cas9

Guide RNAs (gRNAs) were designed to recognize sequences that code for any of the two RNA Recognition Motifs (RRMs) of the Shep using the ZiFiT software from the website http://zifit.partners.org/ZiFiT/CSquare9Nuclease.aspx (Sander et al., 2010). The gRNA sequences were as follows: RRM1 (5´-CGACGACCGGCGGCAGTACC-3´) and RRM2 (5´-ACTTGCCGCCGCACATCACC-3´). The gRNAs are specific to Shep and could avoid off-target effects because more than 6 base pair mismatches from these two gRNAs are needed before they match with other genes in the *Daphnia* Genome Database; and the DNA region with up to five base pair mismatches with the gRNA is susceptible to editing by the Cas9/gRNA complex (Fu et al., 2013; Jiang, Bikard, Cox, Zhang, & Marraffini, 2013). These gRNAs were synthesized by the cloning-free method (Gagnon et al., 2014) where sense and antisense synthetic oligonucleotides containing a T7 promoter, the target gRNA sequence, and the first 20 nt of Cas9 binding scaffold sequence were amplified using PrimeSTAR polymerase (Takara Bio, Shiga, Japan). The reaction mixture and PCR conditions are shown in Tables 5 and 6.

Table 5. Reaction mixture for PCR using Primestar Polymerase.

Reagent	Volume
5X PrimeSTAR Buffer	10 μL
2.5 mM dNTP mix	4 μL
Sense, antisense oligos mix (100 μM each)	1 μL
PrimSTAR polymerase	1 μL
MilliQ water	34 μL
Total	50 μL

Table 6. PCR cycle conditions.

Temperature	Duration	Cycle no.
98°C	5 min	1
98°C	10 sec	
55°C	30 sec	30
68°C	15 sec	
78°C	7 min	1

Phenol/Chloroform extraction was performed to purify the amplicons which were then used as templates for *in vitro* transcription using the MEGAscript T7 kit (Life Technologies Inc.). After the RNA transcription, series of purification procedures were followed: column purification using mini Quick Spin RNA gel columns (Roche Diagnostics, Mannheim, Germany), phenol/chloroform extraction, and ethanol precipitation. Finally, the purified RNAs were dissolved in DNase/RNase-free water and their amounts were measured using Nanodrop 2000 (Thermo Fisher Scientific).

The synthesized gRNAs targeting each of the RRMs of Shep were individually mixed with Cas9 protein and were microinjected into female eggs of *Dsx1* reporter strain to induce double-strand breaks as previously described (Kato, Kobayashi, et al., 2011). Somatic mutations of the injected embryos were confirmed by amplification of the target loci from the genomic DNA isolated from each sample. The genomic DNA was extracted by homogenization in 90 μL of 50 mM NaOH with zirconia beads of 1.0 ⌀ size. Samples were heated at 95 °C for 10 min, followed by a neutralization and stabilization step by adding 10 μL

of 1 M Tris-HCl (pH 7.5) and 2 μL of 5 mM EDTA. Centrifugation followed at 13,000 g for 5 min, before the use of the supernatant as a template for PCR amplification of the target sequences. Using Hot Start Ex Taq Polymerase (Takara Bio, Shiga, Japan), RRM1 and RRM2 regions were amplified using the primer sets in Table 7, and amplicons were analyzed through native PAGE gel electrophoresis.

Moreover, screening for germ-line mutagenesis was done by culturing the offspring of the injected embryos until they produced the next generation. The same genotyping procedure mentioned above was then performed until a positive mutant line was found and established.

Table 7. Primer pairs for genotyping.

Target name	Primer sequence (5′-3′)
RRM 1	Forward: AAGGCTACAGCAGCTCGA Reverse: CCGCGAATGTAGAGGTTG
RRM 2	Forward: CCCACTAATTTGTACCTGGC Reverse: CGCATTTCTCTCTGGATTC
B-actin	Forward: GGCAAGGAATAGTTCGATAC Reverse: CACCGACGTACGAATCCTTCTGACC

3.2.7 Overexpression by mRNAs delivery

Chimeric *Shep* cDNA harboring the *EF1α1* 5′ UTR and 3´UTR was designed and subcloned downstream the T7 promoter as previously described (Törner, Nakanishi, Matsuura, Kato, & Watanabe, 2014). The *Shep* CDS of this plasmid was replaced with *GFP* CDS for the template of control mRNA, to investigate the effects of *EF1α1* UTRs on mRNA stability and/or translation efficiency. These plasmids were purified through phenol/chloroform extraction. *In vitro* transcription and poly(A) tail addition for all mRNAs were performed according to the manufacturers' protocols of the commercial kits mMessage mMachine T7 RNA polymerase (Life Technologies Inc.) and Poly(A) Tailing kit (Ambion, Foster City, CA, USA). The synthesized mRNAs, as illustrated in Figure 12, were column purified using RNeasy Mini Kit (Qiagen), followed by phenol/chloroform extraction, ethanol precipitation, and dissolution in

DNase/RNase-free water. The size of synthesized RNAs and length of the attached poly(A) tail were analyzed by denaturing formaldehyde gel electrophoresis and were taken into account for the RNA amounts used for microinjection.

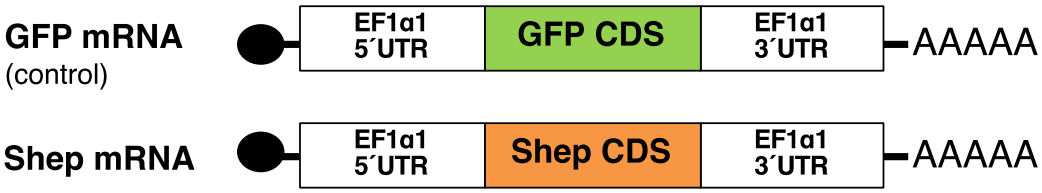

Figure 12. Capped, polyadenylated mRNAs structure constructed for overexpression experiment.

Microinjection was performed into male eggs of *Dsx1* reporter strain. Phenotypes of injected eggs at certain developmental stages were carefully observed under the microscope. The GFP and mCherry fluorescence and the *Shep* and *Dsx1* expressions in mRNAs-injected embryos were observed (see 2.2.6 Fluorescence Photography and Quantification) and validated by qRT-PCR analysis (see 3.2.7 Quantitative real-time PCR), respectively.

3.2.8 Fluorescence photography and quantification

Samples were observed and their photos were taken using Leica DC500 CCD Digital Camera mounted on Leica M165FC fluorescence microscope (Leica Microsystem, Mannheim, Germany) as described in Chapter 2 (See 2.2.5 Fluorescence photography and quantification). Fluorescence signals were calculated using the ImageJ software, following the calculation protocol of a previous study (Törner et al., 2014). The total embryo fluorescence of each sample was normalized by the background fluorescence measurement. In addition, Relative Fluorescence Intensity (RFI) was calculated by dividing the total embryo fluorescence of the injected embryos by the uninjected embryos from the same clutch to nullify the differences in auto-fluorescence between embryos from different mothers. The RFIs of the control samples

were then compared against the RFI of the treated embryos. At least 5 control and treated embryos were used for quantitation of the fluorescence at 30 h and 48 h post-injection.

3.2.9 Quantitative real-time PCR

The cDNA samples of female and male embryos at different embryonic stages (prepared following the same method in Chapter 2; see 2.3.6 Total RNA extraction and 2.3.7 First-strand cDNA synthesis by Reverse Transcription) were used in the qRT-PCR analysis for the temporal expression of *Shep*. qPCR was conducted with the SYBR GreenER qPCR Supermix Universal (Invitrogen) using the Mx3005P real-time (RT)-PCR system (Agilent Technologies). The *Shep* expression was measured using the primer pair shown in Table 7. The *Shep* mRNA expression was normalized with the ribosomal protein *L32* expression level.

For *Shep* and *Dsx1* expression validation in knockdown (3.2.4 Gene knockdown by RNAi), mutagenesis (3.2.5 Mutagenesis by CRISPR/Cas9), and overexpression experiment (3.2.6 Gene overexpression by mRNAs delivery), injected embryos were collected after 48 hrs for each experiment. After extracting the total RNA and synthesizing cDNA for each sample, qPCR was conducted with the SYBR Green qPCR Supermix Universal (Invitrogen) using the Mx3005P (RT)-PCR system (Agilent). Expressions based on the Ct value during amplification were calculated and normalized by quantitating the expression level of several reference genes: the ribosomal protein *L32*, ribosomal *L8* gene, and *Cyclophilin* gene (Sumiya et al., 2016). The geometric mean of the reference genes was calculated for normalization as previously described (Vandesompele et al., 2002). The normalized expression levels of the treated samples are then relatively compared to the expression levels of the control to get the final values. Lastly, gel electrophoresis and dissociation curve analysis were performed to confirm the correct amplicon size and the absence of non-specific bands. All primer pairs used for the RT-qPCR are shown below in Table 7.

Table 7. Primer sequences for RT-qPCR for the temporal expression profile, RNAi, mutagenesis, and overexpression experiments.

Gene name	Primer sequence (5′-3′)
Shep	Forward: AACCTCTACATTCGCGGACT Reverse: GTAGCCTTTGCACTTGTTGG
Dsx1	Forward: AAGTTTGGTGTAGGGGAGGATGAG Reverse: CCATTCATCATTACCAAATCCCTTC
L32	Forward: GACCAAAGGGTATTGACAACAGA Reverse: CCAACTTTTGGCATAAGGTACTG
L8	Forward: GGTACTATTGTTTGCAATGTTGAGG Reverse: GTCTTCTTGGTATCGGTATTGTGAC
Cyclophilin	Forward: GACTTTCCACCAGTGCCATT Reverse: AACTTTCCATCGCATCATCC

3.3 Results

3.3.1 Association of RNA binding proteins with *DAPALR*

As the previous results showed that the 205 bp of *DAPALR* fragment overlapping with *Dsx1α* 5′ UTR is the transactivation element for the enhancement of *Dsx1* expression, we then attempted to identify proteins that interact with this region. We used the 205 bp overlapping sequence as bait for the RNA pulldown experiment. Through RNA pulldown using a FLAG-peptide tagged bait RNA incubated with *D. magna* lysate followed by mass spectrometry (Figure 8) (Adachi et al., 2014; Natsume et al., 2002), we identified two candidate proteins: Alan shepard (Shep) and CUG binding protein 1 (CUGBP1). Shep and CUGBP1 have the highest probability for binding to the overlapping sequence of *DAPALR* and they did not associate with the negative bait samples like the *Dsx1ß* 5′ UTR (S1 Table).

3.3.2 Characterization of *Alan shepard* (*Shep*) in *Daphnia magna*

To further understand the mechanism of *DAPALR*, I focused and studied one of its identified RBP, the Shep, and searched for its ortholog in the *D. magna* genome database, *D. magna* Genome BLAST

(http://arthropods.eugenes.org/EvidentialGene/daphnia/daphnia_magna/BLAST/). I found a single *Shep* ortholog consists of 6 exons and codes for 458 amino acids of a polypeptide including two RNA Recognition Motifs (RRMs) (Figure 13).

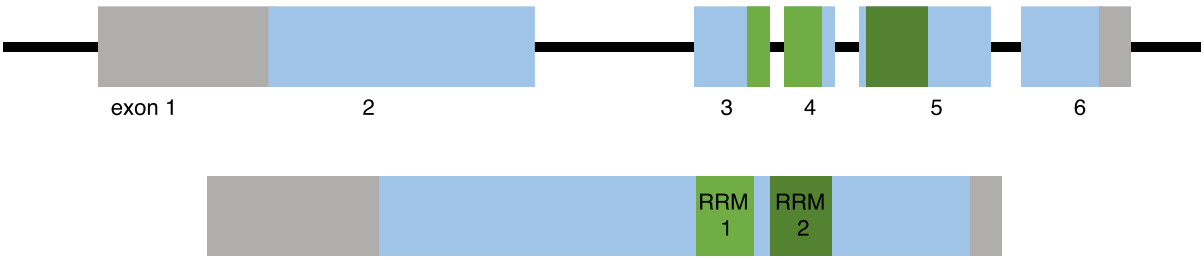

Figure 13. Genomic annotation of *Alan shepard* (*Shep*) gene in *Daphnia magna*.

The exons are indicated by boxes, while the blue boxes represent the shep ORF. The two RRM (RNA Recognition Motif) domains are shown in green boxes within the ORF.

I compared the deduced amino acid sequence obtained from the cDNA sequence of the *D. magna* ortholog to the protein sequence of Shep orthologs in other animals. The multiple sequence alignment revealed that *D. magna* Shep protein contains the conserved RNA Recognition Motif (RRM) domain (Figure 14 and 15).

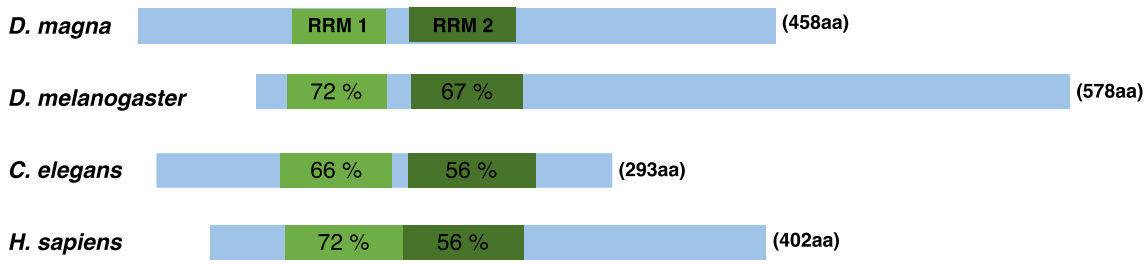

Figure 14. Comparison of the domain structure of *Shep* in *D. magna* and its ortholog in other species.

The blue boxes represent the *Shep* ORF. The two RRM (RNA Recognition Motif) domains are shown in green boxes within the ORF.

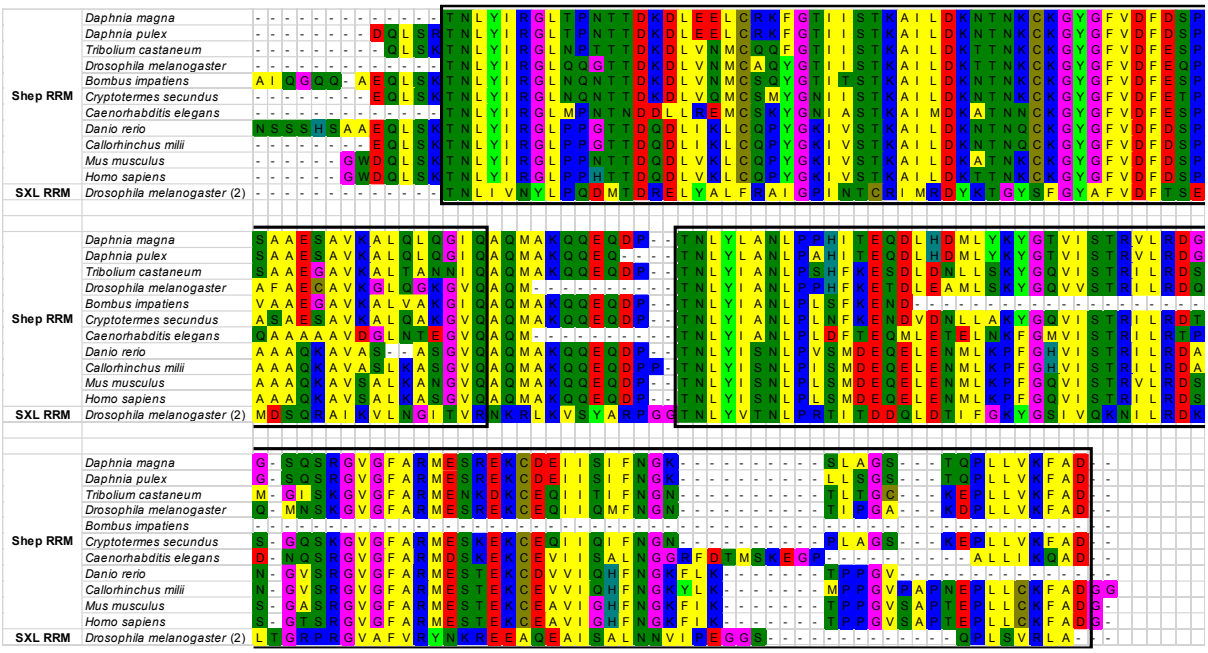

Figure 15. Multiple sequence alignment of the evolutionarily-conserved RRM domains of Shep.

Alignment of the RNA Recognition Motifs (RRMs) of the different shep orthologs from different organisms. The color is based on the physicochemical property of the amino acid-based on ClustalW. The boxes represent the position of the two RRM regions.

To analyze the evolutionary relationship of *D. magna* Shep, a phylogenetic tree with other related proteins was constructed through the neighbor-joining method, using the RRM domain amino acid sequence (Figure 16). The phylogenetic tree revealed that *Daphnia* Shep ortholog is most closely related to the insects' Shep ortholog (Figure 16), which is in good agreement with the taxonomic relationship between crustacean and insects; consistent with the previous hypothesis that insect originated from branchiopod crustaceans (Glenner, Thomsen, Hebsgaard, Sorensen, & Willerslev, 2006).

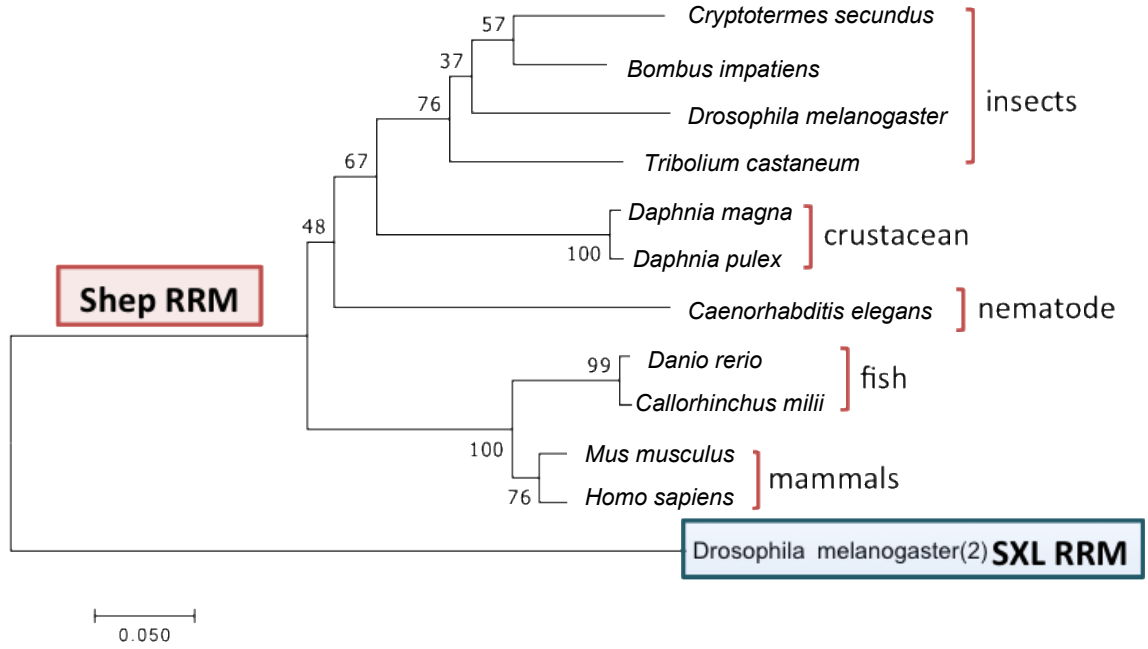

Figure 16. Phylogenetic tree of the RRM domains of the *Shep* orthologs.

Shep RRM orthologs are labeled with red while the *Sex lethal* (*SXL*) RRM is boxed in blue. The percentages of the replicate tree in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The bar indicates branch length and corresponds to the mean number of the differences ($p<0.05$) per residue along each branch. Evolutionary distances were computed using the p-distance method.

These results confirmed that the protein found in the *D. magna* genome is indeed Shep with high similarity to other organism's Shep, specifically in the RRM domains.

3.3.3 *Shep* expression levels during embryogenesis

As I identified the Shep as a *DAPALR* binding protein and *DAPALR* shows sexually dimorphic expression, I examined if *Shep* also shows sexual dimorphism. I analyzed the temporal expression pattern of *Shep* by qRT-PCR during *D. magna* embryogenesis. While *Dsx1* and *DAPALR* both have male-specific expression, *Shep* was expressed both in male and female embryos; it did not exhibit sexual dimorphism (Figure 17). The expression of *Shep* peaked after 30 hours post-ovulation (hpo), mirroring the expression pattern of *Dsx1* in males

(Kato, Kobayashi, et al., 2011). These results indicate that, unlike other sex-related genes, the expression of *Shep* is not sex-specific; it might have a role both in male and female.

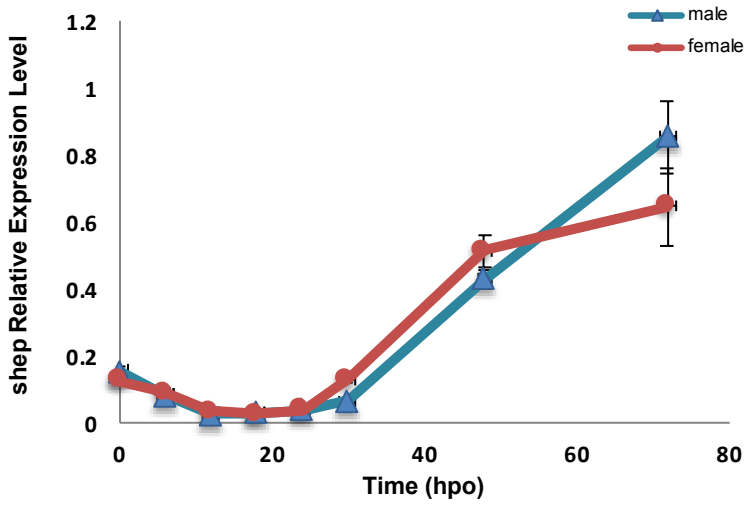

Figure 17. Temporal expression profile of *Shep* in embryonic developmental stages of *D. magna*.

The expression levels of shep in female (red) and male (blue) embryos were determined at 0, 3, 6, 18, 24, 30, 48, and 72 hours post-ovulation (hpo). Results are shown as relative expression normalized with the ribosomal protein L32. Error bars indicate the standard error of the mean (n=3), not significant in all points (Student's T-test between male and female).

3.3.4 Knockdown of *Shep* enhances Dsx1 expression

To examine the roles of *D. magna Shep* gene especially for sex determination, loss-of-function analyses were performed. *Shep* expression was knocked down using the RNAi method (Kato, Shiga, et al., 2011) in male and female embryos of the Dsx1 reporter strain (Nong et al., 2017). In this strain, mCherry fluorescence recapitulates endogenous *Dsx1* expression and H2B-GFP is ubiquitously expressed, which enables us to investigate which tissues or cells express *Dsx1*. We injected the *Shep*-targeting siRNA into the eggs obtained from the reporter strain and found that *Shep* knockdown (Figure 18) resulted in an enhanced mCherry expression pattern that was similar to that of *DAPALR* overexpression in female and male embryos (Figure 9 and 10).

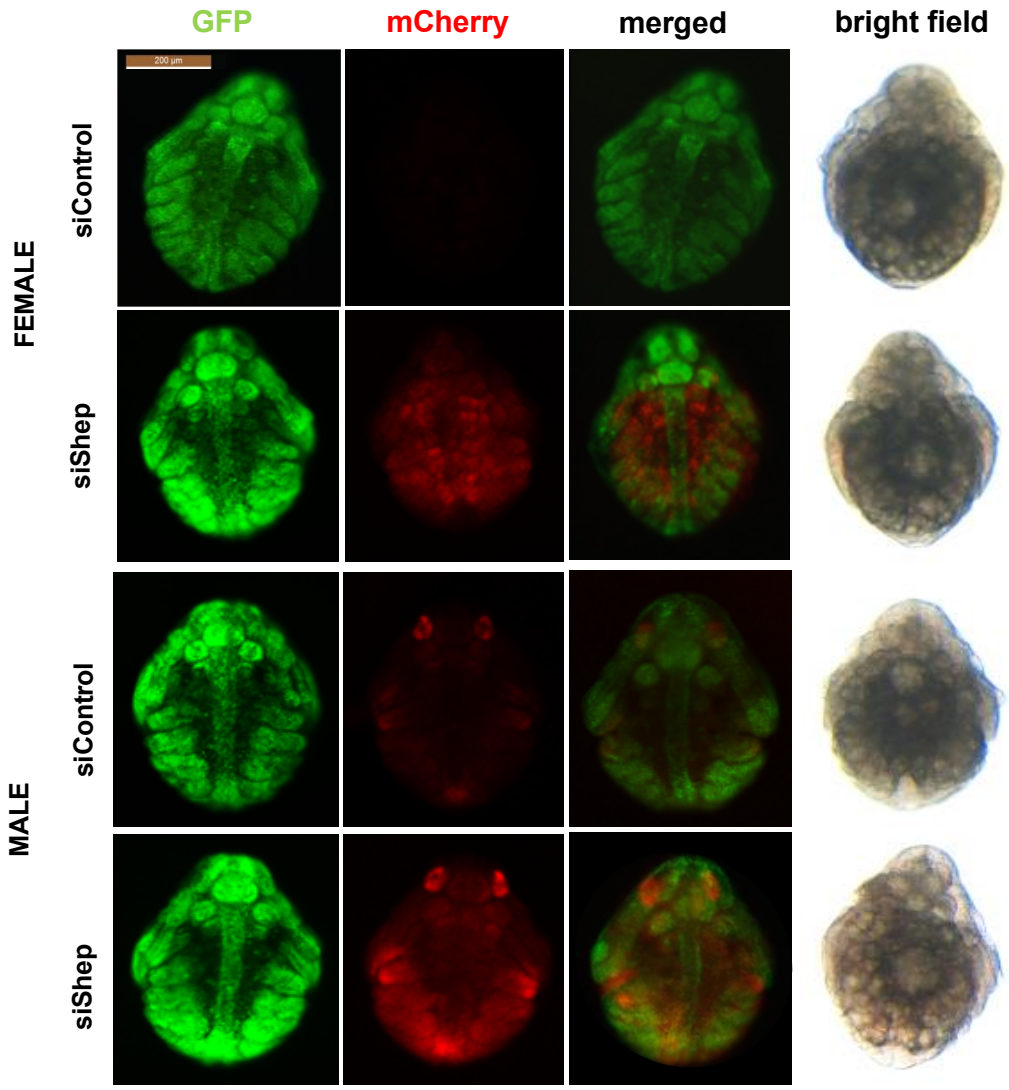

Figure 18. Knockdown of *Shep* by RNAi microinjection after 30 hrs.

Lateral view of female and male embryos of *Dsx1* reporter strain injected with control siRNA and shep siRNA and observed at 30 h after injection. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body structures. The merged images of mCherry and GFP and the bright field images were used to understand the localization pattern of mCherry expression.

Shep RNAi resulted in a 5-fold and 3-fold increase of mCherry fluorescence in female and male embryos, respectively (Figure 19 and 20). In male embryos, the enhanced mCherry signals could be observed not only in its male-specific organs such as the first antennae and its thoracic appendages but ubiquitously in its whole body (Figure 19). While in females, high expression of mCherry signals was observed especially in the yolk region (Figure 19). However,

the mCherry fluorescence recapitulating the *Dsx1* expression was not enhanced in the male-specific organs such as the first antennae, showing that sex reversal did not occur.

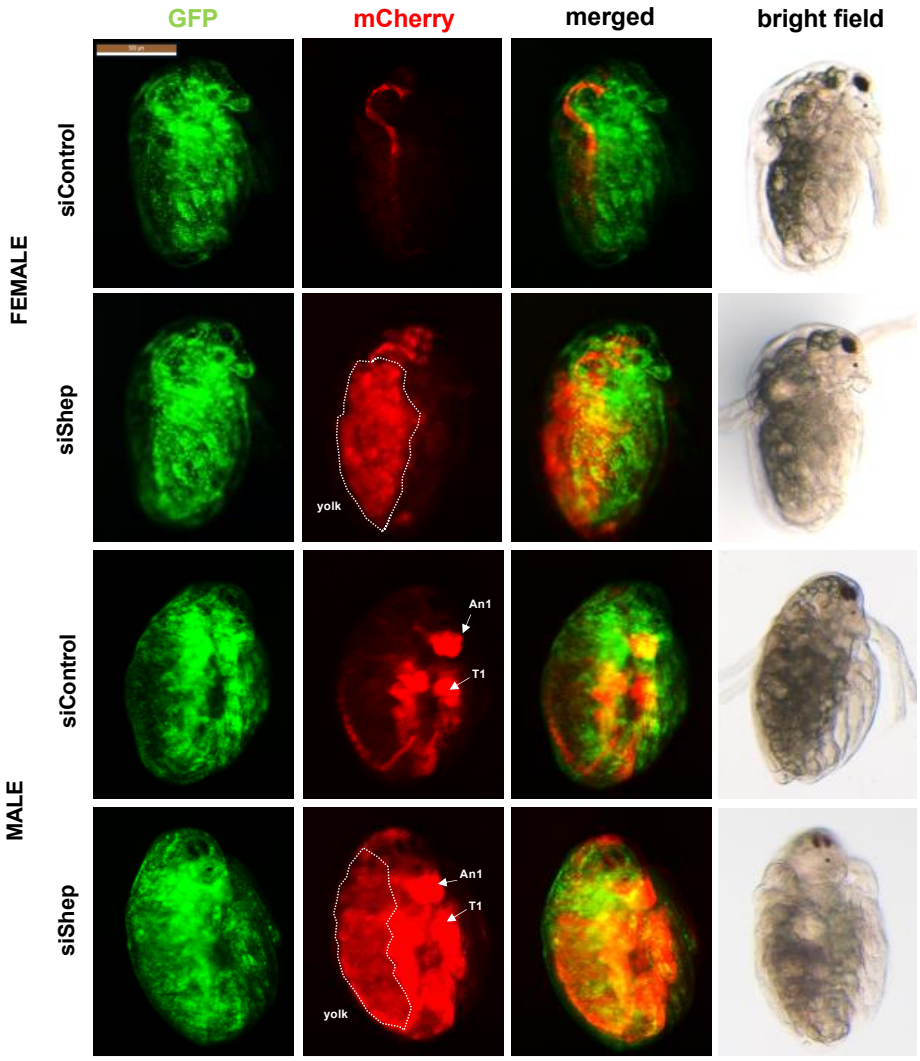

Figure 19. Knockdown of *Shep* by RNAi microinjection after 48 hrs.

Lateral view of female and male embryos of *Dsx1* reporter strain injected with control siRNA and shep siRNA and observed at 48 h after injection. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body structures. The merged images of mCherry and GFP and the bright field images were used to understand the localization pattern of mCherry expression. An1: first antennae, T1: first thoracic leg, dotted lines: yolk area.

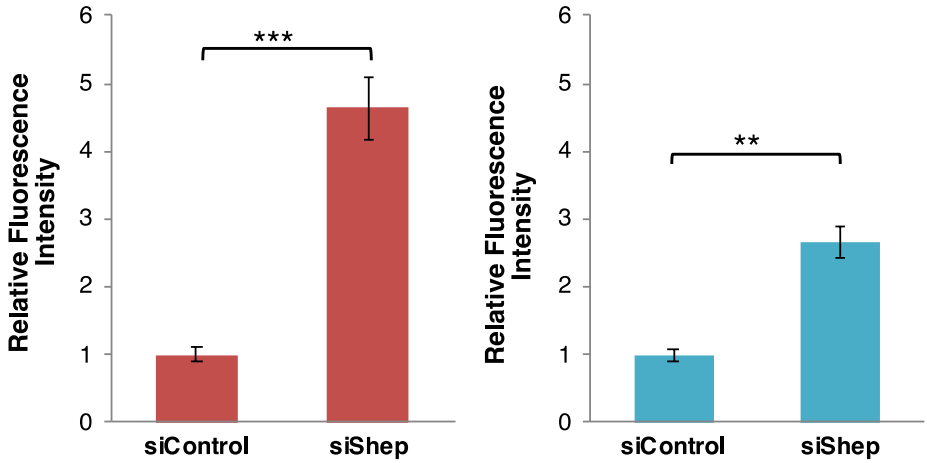

Figure 20. Relative mCherry fluorescence intensity after *Shep* RNAi microinjection.

Relative mCherry fluorescence intensity calculated between *Shep* siRNA- and control siRNA-injected female (red) and male (blue) embryos. Error bars indicate the standard error of the mean (n=5). **p<0.01, ***p<0.001(Student's T-test).

In contrast to the prominent enhancement of the mCherry signals, no significant increase of *Dsx1* transcript was observed in males (Figure 21A) and a two-fold increase of *Dsx1* transcript level was observed in females (Figure 21A). The increase in Dsx1 transcripts did not reflect the amount of protein expressed. However, reduction of the *Shep* mRNA by RNAi relative to the control (Figure 21B) indicated the success of the knockdown experiment; the observed mCherry enhancement was due to the reduction of Shep protein. The finding that the *Dsx1* transcription level does not reflect the enhancement of the mCherry expression suggests the possibility that Shep suppresses the translation of *Dsx1*.

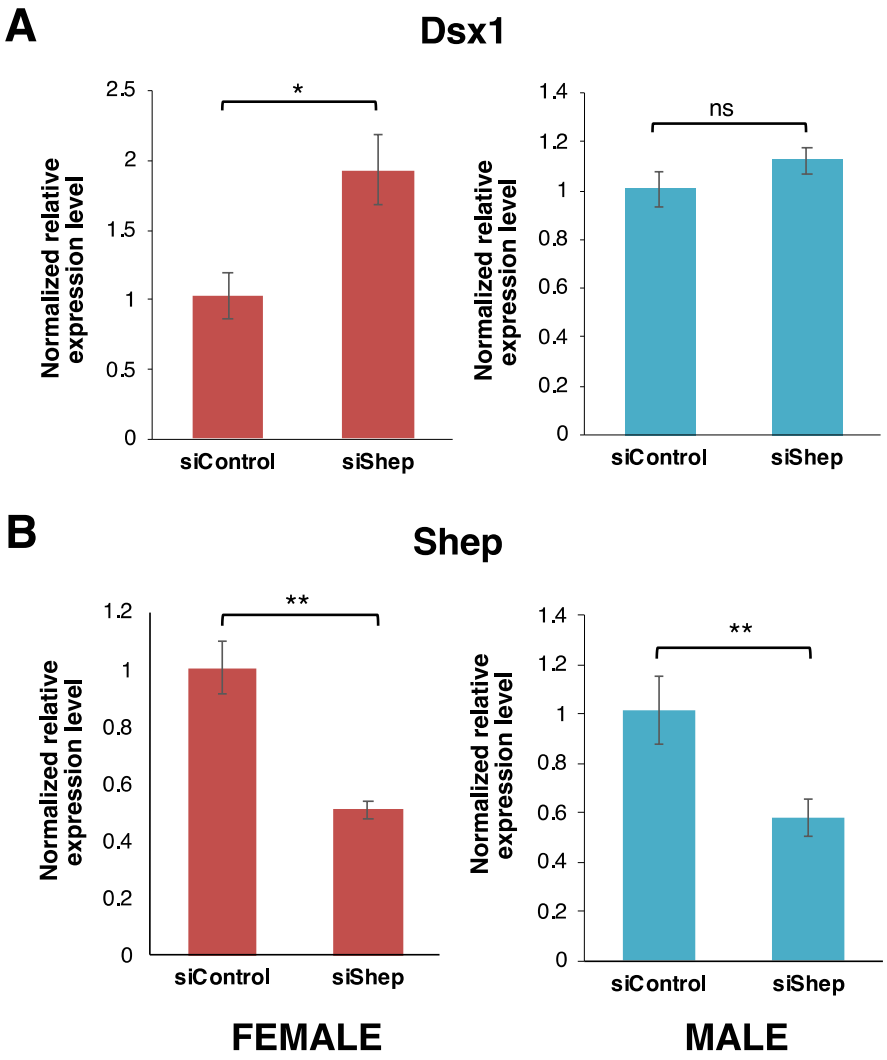

Figure 21. Gene expression profile of *Dsx1* and *Shep* after *Shep* RNAi microinjection.

Gene expression profile of *Shep* in control siRNA- and *Shep* siRNA-injected female (red) and male (blue) embryos. RT-qPCR results are shown as expression levels normalized with housekeeping genes *L32, L8,* and *Cyclophilin* and relatively compared to the control. Error bars indicate the standard error of the mean (n=3). *p<0.05, **p<0.01, ***p<0.001, ns: not significant (Student's T-test).

3.3.5 *Shep* mutant enhances Dsx1 expression

Next, I tried to introduce a mutation in the *Shep* gene using the CRISPR/Cas system. I used two types of gRNAs targeting each RRM and injected those gRNAs with Cas9 protein into eggs from the *Dsx1* reporter strain. Finally, I obtained a line that has 15 nt insertion just before the RRM1 domain (Figure 22A, 22B, and Table 8). This line could be maintained and they developed normally into adults, producing offspring.

Table 8. Summary of mutagenesis experiment.

injected eggs	unhatched eggs	hatched eggs	survived	knock out	knock out efficiency (%)
37	11	26	11	1	9%

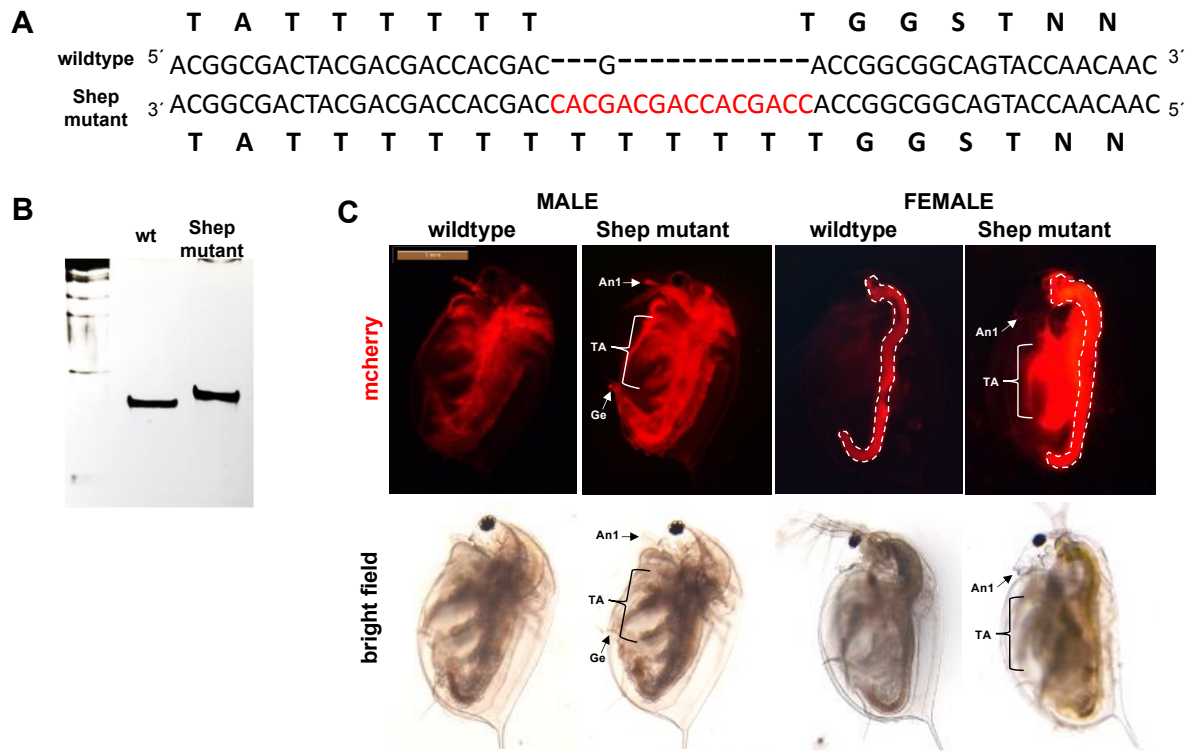

Figure 22. Generated Shep mutant line.

(A) Nucleotide and amino acid sequence comparison between wildtype and shep mutant. **(B)** PAGE analysis of PCR products by genomic PCR to amplify a region in the *Shep* coding sequence with harbors the insertion in the *Shep* mutant line. **(C)** Lateral view of the 1w *Shep* male and female mutant showing an increase in mCherry fluorescence. The red signal from the guts (dotted lines) represents the autofluorescence of *Chlorella*, the main food used in daphniid cultivation. An1: first antennae, TA: thoracic appendages, Ge: genital.

However, there were no noticeable differences between the mutant and wildtype at embryonic stages (Figure 23). Hence, I then observed the mCherry expression of mutant daphniids at the adult stage when sexually dimorphic traits are more evident (Nong et al., 2017). In the *Shep* mutant, both males and females showed significantly higher mCherry fluorescence than the wildtype (Figure 22C). Female daphniids of the *Dsx1* reporter strain do not usually have mCherry fluorescence (Nong et al., 2017), but the *Shep* mutant displayed mCherry signals

in its whole body especially the appendages. High expression of mCherry signal could also be observed in the first antennae that is one of the major male-specific traits. However, the first antennae did not develop elongated like in males, signifying that sex reversal and male differentiation did not occur. On the other hand, the male mutant showed increased mCherry signals not only in male-specific regions such as the first antennae and genital but also in other regions. These suggest that Shep may suppress *Dsx1* in both male and female.

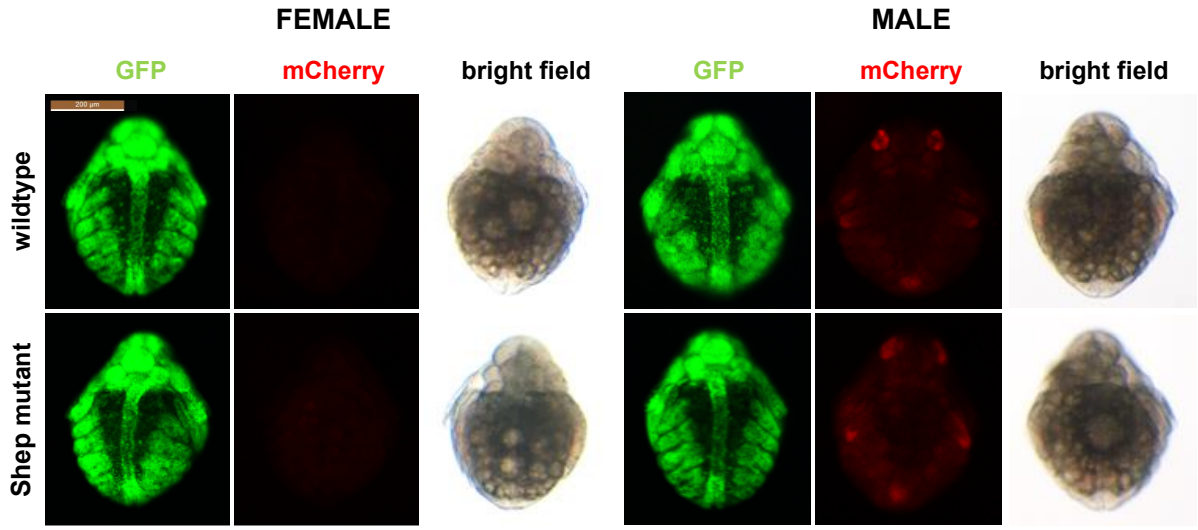

Figure 23. Shep mutant line at 30 hpo.

Lateral view of female and male embryos of Shep mutant line and wildtype observed at 30 h after ovulation. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body structures. The bright-field images were used to understand the localization pattern of mCherry expression.

In contrast to the drastic difference of the mCherry expression between the *Shep* mutant line and wildtype, we could not find a significant difference in *Dsx1* mRNA expression levels between the *Shep* mutant and wildtype (Figure 24) in either male or female. This finding also supports the possibility that Shep controls *Dsx1* expression at post-transcriptional levels.

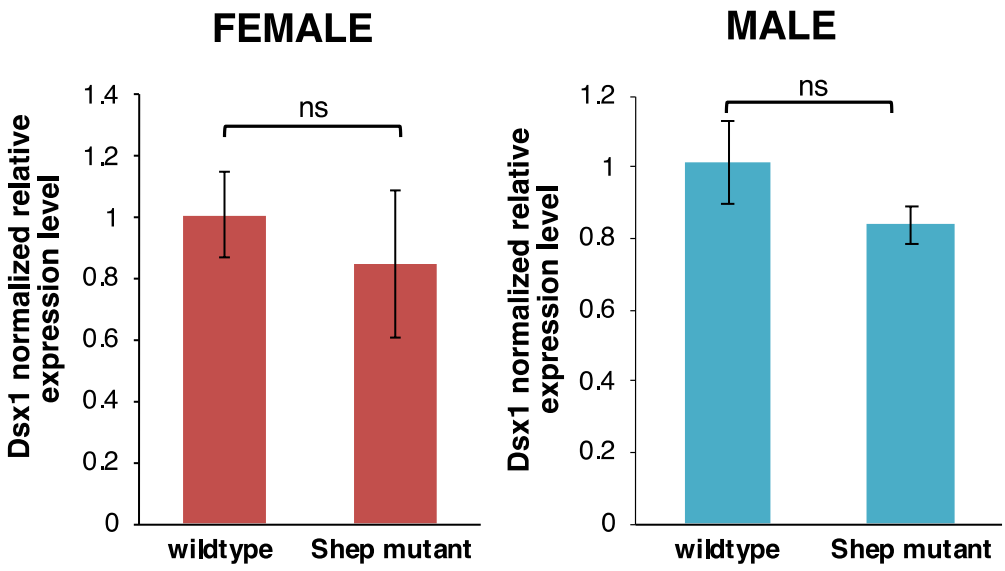

Figure 24. Gene expression profile of the generated Shep mutant line.

Gene expression profile of *Dsx1* in 1w female and male wildtype control and shep-. RT-qPCR results are shown as expression levels normalized with housekeeping genes *L32*, *L8,* and *Cyclophilin* and relatively compared to the wildtype control. Error bars indicate the standard error of the mean (n=3). ns: not significant (Student's T-test).

Using the CRISPR/Cas system, I first aimed to produce *Shep* mutant lines that have deletion mutations in both of the two RRM domains. However, the embryos with indel mutations in both RRM domains could not hatch and they exhibited delayed or deformed phenotypes (Figure 25: delayed development, deformed embryos, and unhatched eggs), suggesting that Shep is also essential for development and morphogenesis.

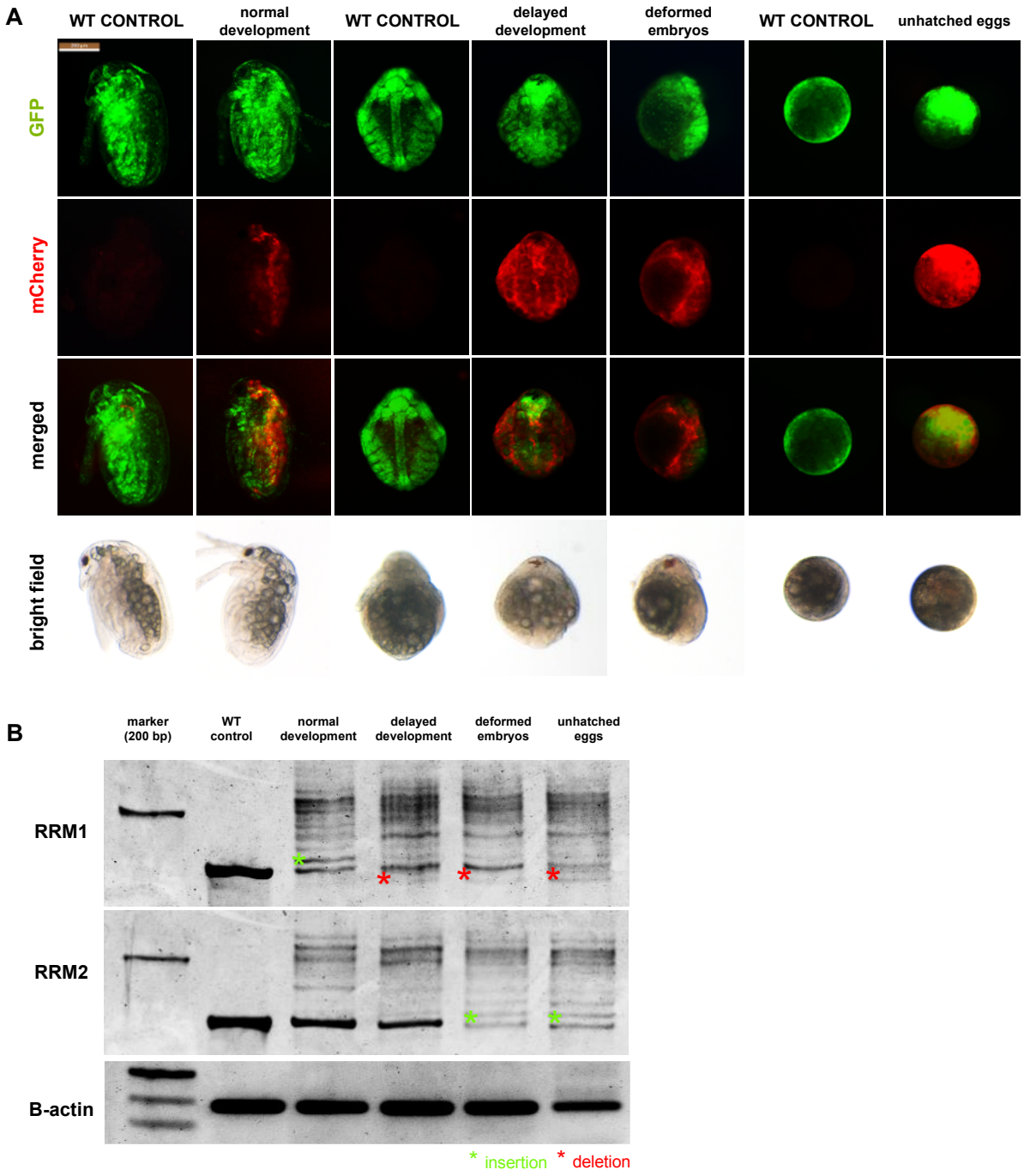

Figure 25. Diverse phenotype of Shep mutants and their genomic mutations.

(A) Ventral and lateral views of the different phenotypes observed after attempt for Shep knock out: (from L to R) normal development, delayed development, abnormal development, and unhatched egg. Embryos showing normal development and phenotype were also shown as a control for each stage. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body structures. The merged images of mCherry and GFP were used to understand the localization pattern of mCherry expression. Bright field showed embryos taken using visible light. Scale bar = 200 µm. **(B)** PAGE analysis of PCR products by genomic PCR to amplify the region targeted by each RRM gRNAs. Asterisks show the genomic mutations in RRM1 and RRM2 of the different phenotypes after shep mutagenesis.

3.3.10 *Shep* overexpression suppresses Dsx1 expression

As the findings suggested that the diminished function of Shep increased the mCherry expression at the post-transcriptional level, I further investigated if *Shep* overexpression can suppress the *Dsx1* expression. I induced transient ectopic expression of *Shep* in males by delivering capped, polyadenylated *Shep* mRNAs into ovulated eggs obtained from the *Dsx1* reporter line via microinjection.

As a result, I found that mCherry fluorescence was reduced in the *Shep* mRNA-injected embryos even on male-specific organs (Figure 26 and 27). However, the transcript level of *Dsx1* (Figure 28) did not show any significant difference from the control despite the success of the enhancement of *Shep* mRNA level as confirmed by qRT-PCR (Figure 28). These results also suggest that Shep does not affect *Dsx1* transcription or change the mRNA stability; rather its effect is at the translational level.

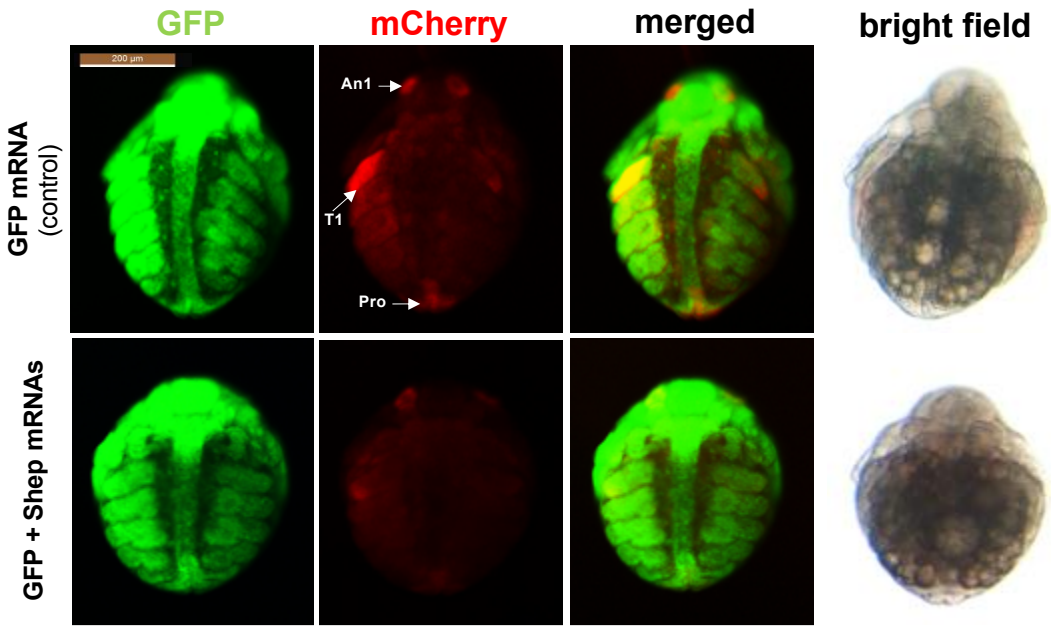

Figure 26. Overexpression of *Shep* in male.

Ventral view of male embryos of *Dsx1* reporter strain injected with GFP mRNA as control and GFP plus *Shep* mRNA observed at 30 h after injection. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body structures. The merged images of mCherry and GFP and the bright field images were used to understand the localization pattern of mCherry expression. An1: first antennae, T1: first thoracic legs, Ge: genital.

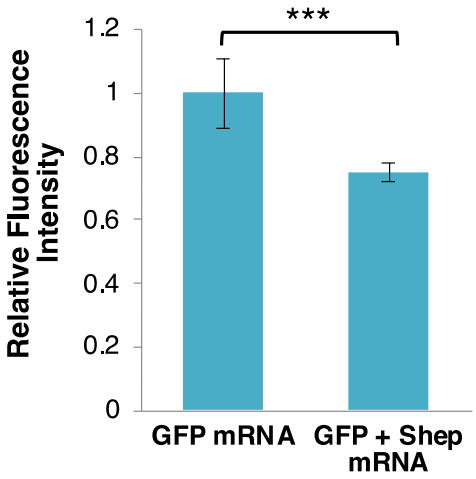

Figure 27. Relative mCherry fluorescence intensity after *Shep* overexpression in male.

Relative mCherry fluorescence intensity calculated between *GFP* mRNA- and GFP plus *Shep* mRNA-injected male embryos. Error bars indicate the standard error of the mean (n=5). ***p<0.001 (Student's T-test).

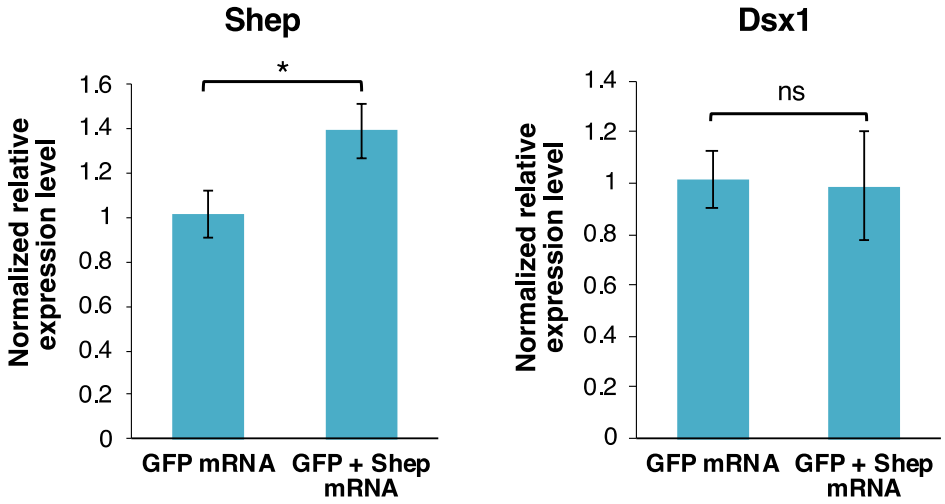

Figure 28. Gene expression profile of *Shep* and *Dsx1* after *Shep* overexpression.

Gene expression profile of *Shep* and *Dsx1* in *GFP* mRNA- and *GFP* plus *Shep* mRNA-injected male embryos. RT-qPCR results are shown as expression levels normalized with housekeeping genes *L32*, *L8,* and *Cyclophilin* and relatively compared to the control. Error bars indicate the standard error of the mean (n=3). *p<0.05, ***p<0.001, ns: not significant (Student's T-test).

3.4 Discussion

The proteins that were identified to associate in the transactivation element of *DAPALR* were Shep and CUGBP1. In previous reports, CUGBP1, also called as CELF1, has been known as an RNA binding protein that regulates gene expression at the post-transcriptional levels and

is involved in mRNA alternative splicing, translation, and stability (Beisang, Bohjanen, & Louis, 2012). While Alan shepard (Shep), also an RNA binding protein, has known functions in both DNA and RNA. Shep in *Drosophila* and its orthologs sup-26 in *Caenorhabditis elegans* and RBMS1, RBMS2, RBMS3 in human, is a known RBP that plays diverse roles in neuronal development (Schachtner et al., 2015). In *Drosophila*, Shep is also a regulator of the chromatin insulator, gypsy gene, as it binds to the transcription start site (TSS) of its target gene and upregulates the expression of the target gene by repressing the insulator activity (Matzat, Dale, Moshkovich, & Lei, 2012). At the same time, it also negatively regulates the transcription of RBP *Dad* (D. Chen, Gu, Pham, Zachary, & Hewes, 2017) and targets mRNAs by interacting with known post-transcriptional regulators such as *Yps*, *traI,* and *Bel* (Olesnicky et al., 2018). Hence, I chose to focus on Shep for further analyses because 1) *DAPALR* overexpression increased *Dsx1* mRNA in female and male embryos, suggesting a potential role in *Dsx1* transcription and 2) Sup-26 regulates translation of the sex-determining gene *tra-2* (Mapes, Chen, Yu, & Xue, 2010) suggesting functional conservation of this family on the regulation of sex-determining genes among animals. So in this chapter, I analyzed the functions of Shep in sex determination, especially its role in the regulation of *Dsx1*.

I found a *Shep* gene ortholog in *D. magna* which has two RRM domains. These RRMs were confirmed to be highly conserved among species and throughout evolution as demonstrated by the multiple sequence alignment and phylogenetic tree of the *Daphnia Shep* and the orthologs from other species.

Moreover, the temporal expression profile of *Shep* showed that its expression peaks at 30h post-ovulation, mirroring the expression pattern of *Dsx1*. However, unlike *Dsx1* and *DAPALR* and other sex-related genes, Shep is expressed on both male and female during embryogenesis. The expression of *Shep* is not sex-specific, indicating that it might have a role both in male and female.

To characterize the function of *Shep*, I performed loss-of-function experiments. *Shep* RNAi phenocopied *DAPALR* overexpression in both male and female. In males, mCherry fluorescent signals were enhanced not only in male-specific organs like the first antennae and first thoracic appendages but ubiquitously in the whole body of the embryo. While in females, mCherry fluorescence was also activated especially in the yolk region, indicating de-repression of *Dsx1*. However, based on the RT-qPCR results, the relative expression levels of *Dsx1* in Shep-silenced male and female samples did not exhibit the same increase in Dsx1 protein expression. The *Dsx1* mRNA level in females only increased 2-fold, in contrast to the 5-fold increase in mCherry fluorescent signals; while in males, no significant change in the *Dsx1* mRNA level was detected as opposed with the mCherry fluorescence that increased 3-fold.

The reason why *Dsx1* was upregulated by Shep RNAi and DAPALR overexpression in female embryos might be related to its expression limited in the yolk region. Because yolk cells have a unique system specialized for energy metabolism that is different from somatic cells (Callaini & Dallai, 1991), *Dsx1* upregulation would occur specifically in yolk cells with an unknown mechanism but not in somatic and germ cells. In male embryos, detection of this perturbation in yolk cells showing lower *Dsx1* expression might be masked by higher *Dsx1* expression in somatic and germ cells. Thus, these results strongly suggested that neither Shep and DAPALR did affect *Dsx1* transcription both in somatic and germ cells.

Also, the expression level of *Dsx1* after *Shep* knockdown in female samples is still not as high compared to the *Dsx1* expression in normal males; hence, no sex reversal could be observed. On the whole, because the effect of *Shep* on the Dsx1 protein signals was not reflected on the mRNA levels detected, this suggested that Shep may function at the post-transcriptional level.

These observations were further confirmed when male and female *Shep* mutants also have higher mCherry fluorescence in their whole body even in non-sex-specific tissues, even

though the *Dsx1* expression levels remain the same between the mutant and control. These also implied the post-transcription regulation of Shep towards *Dsx1*.

On the other hand, the abnormal development resulting from mutation of both the RRMs and total knock out of Shep was not surprising as it was previously known in other organisms, especially *Drosophila,* that Shep has neuronal functions and plays a significant role in metamorphosis or developmental process (D. Chen, Qu, & Hewes, 2014). And since RRMs have similar RNA recognition ability, a single RRM is sufficient for specific binding (Lisbin, Gordon, Yannoni, & White, 2000). Thus, with one intact and functional RRM, the *Shep* mutant line retains its function but with a decrease or change in its activity due to the mutation of the other RRM.

Moreover, when I overexpressed the *Shep* mRNA in males, the mCherry fluorescent signals decreased as compared to the control. This verified the repressive function of Shep on Dsx1. However, like in the previous knockdown and mutagenesis experiments, the change in Dsx1 protein expression after overexpression of Shep was not observed in the *Dsx1* mRNA level. These results established that Shep does not inhibit *Dsx1* transcription or reduce its mRNA stability; rather its effect is at the post-transcriptional level.

It can only be assumed that the possible post-transcription mechanisms of Shep may be through alternative splicing, mRNA stability, or translation (Antic & Keene, 1997). Alternative splicing is common post-transcriptional processing of RNA regulated by RNA-binding proteins (Brooks et al., 2015). The proteins play key roles in either recruiting or blocking spliceosomes to induce or repress splicing and Shep was one of the splicing regulatory proteins previously identified in *Drosophila*. However, more biochemical studies are still needed to support this.

On the other hand, Shep has also been reported as a translational regulator in other organisms. As mentioned earlier, Sup-26, an isoform of Shep in *C. elegans*, binds and represses

the translation of *tra-2* mRNA (Mapes et al., 2010). Interestingly, *tra-2* is also a key regulator of sex determination like *Dsx1* (Kuwabara, Okkema, & Kimble, 1992).

Identifying which post-transcription mechanism Shep follows to repress *Dsx1* in *D. magna* and also how *DAPALR* enters the picture is the next direction for this research.

CHAPTER 4 *DAPALR* acts as a decoy against Shep for *Dsx1* activation

4.1 Introduction

Multiple studies have shown that a significant number of lncRNAs are emerging as key players in various layers of cellular processes. Increasing evidence suggests that lncRNAs may positively or negatively regulate gene expression at the transcriptional and/or post-transcriptional level. The mechanism of lncRNA-mediated gene regulation is sophisticated as they may serve as 1) signals to induce transcription and/or translation; 2) decoys to titrate regulatory factors; 3) guides for recruitment to target genes and 4) scaffolds to bring together multiple proteins to form functional ribonucleoprotein complexes (Kung et al., 2013; Yao et al., 2019).

At the post-transcriptional level, lncRNAs may interact with a variety of RNA binding proteins (RBPs) to regulate pre-mRNA splicing and polyadenylation, mRNA export, stability, localization, and translation (Briata & Gherzi, 2020). Post-transcription regulation is a crucial aspect in the life of RNAs and is primarily facilitated by RBPs as they influence the structure and interactions of the RNAs and play critical roles in their biogenesis, stability, function, transport,t and cellular localization (Glisovic, Bachorik, Yong, & Dreyfuss, 2008).

Focusing on Shep, one of the identified RBPs of the sense overlapping lncRNA *DAPALR* in *D. magna*, results in the previous chapter showed that endogenous Shep functions as a repressor of *Dsx1* at the post-transcription level. Orthologs of Shep in other organisms have been reported to mediate post-transcription gene regulation too. In *Drosophila*, Shep interacts with other post-transcriptional RBPs like Trailer Hitch (Tra1), Ypsilon schatchel (Yps), Belle (Bel), and Poly(A)-Binding Protein (PABP) to direct dendrite morphogenesis in sensory neurons (Olesnicky et al., 2018). Shep also regulates the bone morphogenetic protein (BMP) signaling by antagonizing the translation of *Dad* mRNA to control neuronal remodeling (D. Chen et al., 2017).

On the other hand, the other Shep ortholog in *C. elegans*, the Sup-26, has been known to function within the sex determination pathway (Mapes et al., 2010). Sup-26 represses the tra-2 protein expression but does not inhibit *tra*-2 transcription or reduce the *tra*-2 mRNA stability. As loss-of-function and overexpression of Shep revealed the same phenomena, Shep may mimic the mechanism of Sup-26 as a translation repressor.

In this chapter, I aim to reveal the post-transcription mechanism of Shep in relation to *Dsx1*. First, I identified how Shep recognizes and binds to *Dsx1* and *DAPALR*. I then confirmed its direct binding and repression to *Dsx1* by *in vivo* and *in vitro* assays. Lastly, through *in vitro* experiment, I explored the mechanism of *DAPALR* and Shep in regulating *Dsx1* and their role in the sex determination process of *D. magna*.

4.2 Materials and Methods

4.2.1 Wildtype *Daphnia magna* and *Doublesex1* reporter strain culture condition

The *D. magna* (NIES clone), obtained from the National Institute for Environmental Studies (NIES; Tsukuba, Japan), was cultured under the same laboratory condition mentioned in Chapter 2 (see 2.2.1 Wildtype *Daphnia magna* and *Doublesex1* reporter strain culture condition).

4.2.2 Ectopic expression of intact and deleted Shep binding site

From pCS-EF1a1::Dsx1 exon3 (Kato et al., 2018), the region of *Dsx1* exon 3 except for the 40 nt sequence which contains the putative binding site of Shep (Shep BS) (Figure 29) was removed for the construct of pCS-EF1a1::Shep BS using the following primer set: Forward (5′- GTGTGTGTGTGTGTTGACGTT -3′) and Reverse (5′- AACACACACACACACACACCCGGGCATTGTGATTG -3′). This plasmid was then used as a template to delete the potential shep binding site using the primer set as follows: Forward

(5′-GTGTGTGTGTTGACGTTTTTCCAATATATAGATGGAGGC-3′) and Reverse (5′-GCCTCCATCTATATATTGGAAAAACGTCAACACACACAC-3′). Embryos injected with each plasmid were compared to embryos injected with pCS-EF1a1::EF1a1 UTR, which only has the *EF1α1* 5′UTR and 3′UTR (Kato et al., 2018). These three plasmids (200 ng/μl) were each injected into female eggs of the *Dsx1*-reporter strain. Injected eggs were observed 30 h after injection to observe and calculate for the fluorescence intensity differences.

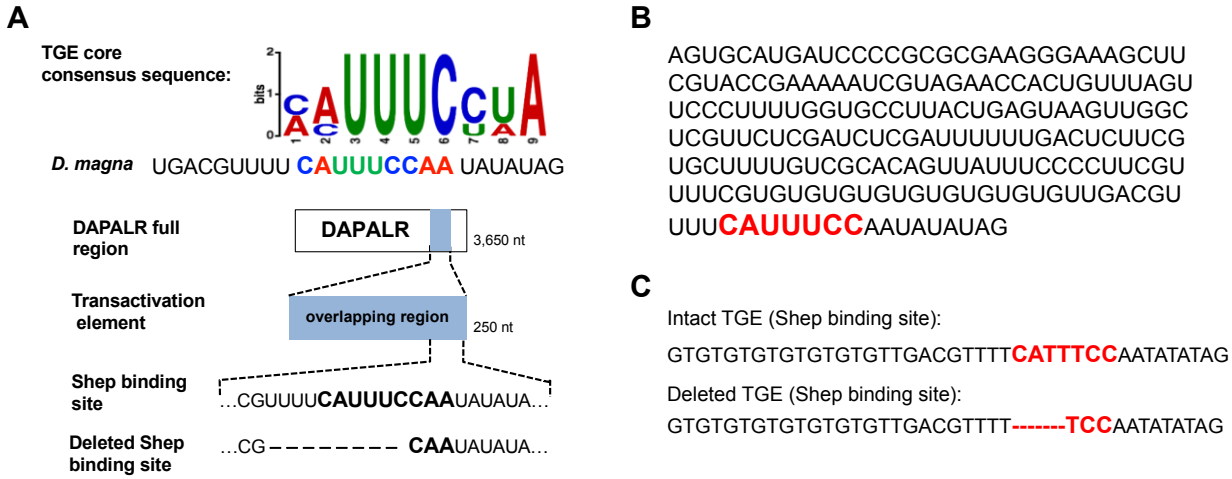

Figure 29. TGE core consensus sequence as potential Shep binding site.

(A) Shep binding site consensus sequence and its similarity with TGE core consensus sequence. **(B)** The position of the potential binding site was also shown located at the 3′end of the transactivation element of *DAPALR*. **(C)** The sequence of the intact and mutated TGE-like element (potential Shep binding site) used for the experiment was also shown.

4.2.3 RNA synthesis

To prepare the GFP reporter mRNA harboring the *Dsx1α* 5′ UTR, the expression plasmid pEX-A2JI that has the EF1α1 3′ UTR and T7 promoter was first synthesized by Eurofins Genomics. Second, the GFP coding sequence of the 4xEcRE-H2B-GFP plasmid (Asada, Kato, Matsuura, & Watanabe, 2014b) was fused with the EF1α1 3′ UTR. Third, *Dsx1α* 5′ UTR was amplified with PCR using the pCS-EF1a1::Dsx1 exon 3 as a template and fused with the GFP harboring EF1α1 3′ UTR to construct the mRNA template plasmid pEX-Dsx1 5′

UTR::GFP. Fourth, using this plasmid, the potential Shep binding site was removed with the same primer set described above, resulting in the generation of the pEX-Dsx1 5′ UTR mutant::GFP.

In vitro transcription and poly(A) tail addition for all mRNAs were performed using T7 RNA polymerase and Poly(A) Tailing kits, respectively (Ambion, Foster City, CA, USA). The size of synthesized RNAs and length of the attached poly(A) tail were analyzed by denaturing formaldehyde gel electrophoresis and were taken into account for the RNA amounts used for microinjection.

4.2.4 Luciferase-based *in vitro* translation assay

Luciferase reporter mRNAs were prepared by using pEX-Dsx1 5′ UTR::GFP and pEX-Dsx1 5′ UTR mutant::GFP and replacing its GFP CDS with the luciferase gene sequence from pG5luc (Promega Corporation, Madison, WI). 0.1 μM of these mRNAs were then transcribed using the nuclease-treated rabbit reticulocytes lysate (RRL) *in vitro* translation system from Promega. Following the manufacturer's protocol, each 25 μL reaction contained 70% v/v of RRL, 0.02 mM amino acid mixture, 0.5 U/μL RNAse Inhibitor (Nacalai Tesque Inc., Kyoto, Japan), and specific concentrations of the mRNAs based on their molecular size. After denaturing at 65 °C for 3 min, luciferase reporter mRNAs were added after pre-incubating the RRL *in vitro* translation mixture at 30 °C for 10 min. The assembled reaction is then further incubated at 30 °C for 90 min and stopped by the addition of 60 μM puromycin. Firefly luciferase activity was then observed using LuminoSkan Ascent where 50 μL Bright-Glo Luciferase assay reagent (Promega Corporation, Madison, WI) was added to 3 μL of the translated reaction. The luminescence data were normalized by subtracting the measurements from the *in vitro* translation reaction without any reporter mRNAs. In the different experiments, the Shep mRNA, Shep-FLAG mRNA, *DAPALR* full RNA (3.6 kb), RNA transcribing the core

element of *DAPALR* harboring the Shep binding site (40 nt), and negative controls (EcR-FLAG and GFP mRNAs) were added together with the reporter mRNAs to test their effect on the translation activity. The full sequence of DAPALR and its core element are shown in S6 Fig.

4.2.5 FLAG pulldown assay

3x FLAG (5′-GACTACAAAGACCACGACGGTGATTACAAAGATCACGACATCGATTACAAGGATGACGATGACAAA-3′) was fused to the C-terminus of the Shep CDS in pCS-EF1a1::Shep to make the mRNA template plasmid pCS-EF1a1::Shep-FLAG. Shep-FLAG mRNA was transcribed *in vitro* and poly(A) was added following the same protocol mentioned above. It was then translated using the nuclease-treated rabbit reticulocytes lysate (RRL) *in vitro* translation system from Promega following the same protocol above. The 25 µL reaction lysate with the translated products was diluted using wash buffer [10 mM HEPES (pH 7.5), 150 mM NaCl, 0.1% Triton X-100]. Diluted lysates were then divided equally into two tubes wherein 10 µg of the RNA with the Shep binding site was added into one tube and the RNA without the Shep binding site was added into the other. Both treatments were incubated to allow interaction at 30°C for 30 min and were irradiated under ultraviolet (UV) light at 200 mJ/cm^2 and were then transferred to a tube containing 50 µL of PVP-treated anti-FLAG M2 Affinity Gel, rotated at 4 °C for 2 h. Washing was done five times using the High-salt wash buffer [50 mM Tris-HCl (pH 7.4), 1 M NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate] and the gel was resuspended using PK buffer [100 mM Tris-HcL (pH 7.4), 50 mM NaCl, 10 mM EDTA] with 200 µg of proteinase K for 20 min at 37 °C as previously described (Huppertz et al., 2014). Total RNA extraction (Kato et al., 2018) was then performed wherein 10 µg of yeast tRNA (Invitrogen, Carlsbad CA, USA) was added as co-precipitant to ensure the collection of a minute amount of RNA, which was followed by cDNA synthesis.

RT-qPCR targeting the bait RNAs and the tRNA as a reference gene was conducted using the primer sets enumerated in Table 9. The geometric mean of the expression levels of the tRNA genes (Met and Phe) was calculated for normalization as previously described (Vandesompele et al., 2002). The wildtype *Shep* mRNA and an unrelated mRNA, EcR-FLAG were used as negative controls. The normalized expression levels of all samples were then relatively compared to the expression level in the EcR-FLAG pulldown experiment with the RNA that has no Shep binding site to get the final values.

Table 9. Primer sequences for RT-qPCR for the FLAG pulldown.

target name	Primer sequence (5′-3′)
Bait RNA for FLAG pulldown	Forward: GCTTTTGTCGCACAGTTATTTC Reverse: CTTCCAGCGGATAGAATGG
tRNA-Met	Forward: GAGCTTGTATAGTTTAATTGGTTAA Reverse: TACTTGTAGAAGGAATTGAACCTTA
tRNA-Phe	Forward: GGACTTAGCTCAGTTGGGA Reverse: GAACTCTGTGGATCGAACA

4.3 Results

4.3.1 TGE element is responsible for the post-transcriptional regulation or *DAPALR* function

Following the idea that *Daphnia* Shep mimics the same mechanism as its ortholog in *C. elegans*, the Sup-26, I performed a literature review regarding its mechanism and found out that Sup-26 has been reported to bind to the target sequence named *tra-2* and *GLI* element (TGE) to regulate the *Tra2* gene translation (Mapes et al., 2010). Hence, I searched for a similar sequence to the TGE in *Dsx1α* 5′ UTR and *DAPALR*. Indeed, a highly conserved sequence with TGE was found in the overlapping region of *Dsx1α* 5′ UTR and *DAPALR* (Figure 29A and 29B).

To prove that the TGE-like motif is essential for the Shep function in *Daphnia*, either 40 nt of the RNA including the potential TGE, or the 30 nt RNA that lacks the potential TGE was overexpressed in female embryos of the *Dsx1* reporter strain (Figure 29C). When the RNA containing the TGE-like motif was expressed, the mCherry expression could be observed (Figure 30). The enhancement of the mCherry expression was the same result as the *DAPALR* overexpression (Figure 9). In contrast, the deleted TGE did not have any effect on the reporter mCherry expression (Figure 30), which was the same result as the injection of the unrelated RNA which only harbors the *EF1α1* 5′UTR and 3′UTR. These results suggest the possibility that the TGE-like motif, like in Sup-26, has a potential role in the function of Shep and *DAPALR* in *Dsx1* regulation in *D. magna*.

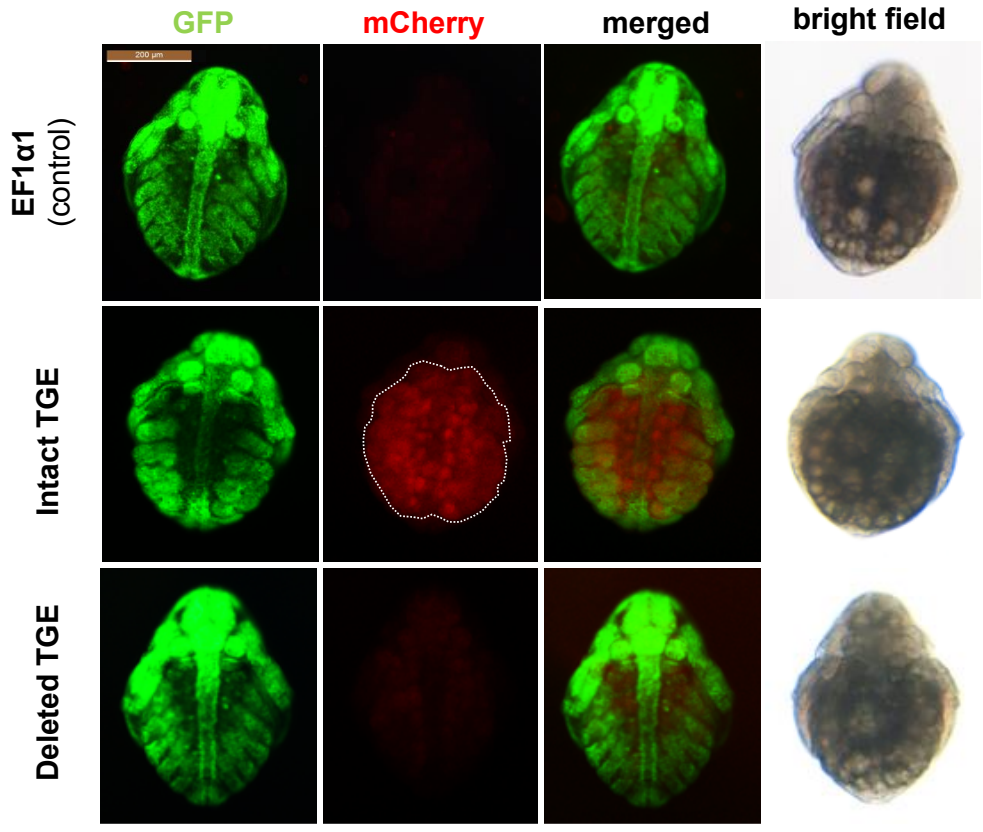

Figure 30. Overexpression of intact and deleted TGE element.

Ventral view of female embryos of *Dsx1* reporter strain injected with the control plasmid, plasmid expressing intact TGE and plasmid expressing deleted TGE. mCherry fluorescence allowed visualization of *Dsx1* expression while GFP fluorescence in the nucleus enabled observation of body

structures. The merged images of mCherry and GFP and the bright field images were used to understand the localization pattern of mCherry expression. dotted lines: yolk area.

4.3.2 Shep binding site (TGE) is a target sequence of translational regulation

To test the hypothesis that the potential Shep binding site located in the 5′ UTR is the target of translational regulation of *DAPALR* and that the Shep functions as a translational suppressor like Sup-26, I examined the translational efficiency in the presence or absence of Shep binding site in the mRNA. I constructed a GFP reporter mRNA harboring the *Dsx1α* 5′ UTR and the same mRNA but lacks the potential Shep binding site. These reporter mRNAs were individually injected into female wild-type eggs.

Results showed that the mRNA lacking the TGE-like motif showed a much higher expression of the GFP than the wildtype mRNA (Figure 31 and 32). This suggests that endogenous Shep may suppress the translation by binding to the *Dsx1α* 5′ UTR.

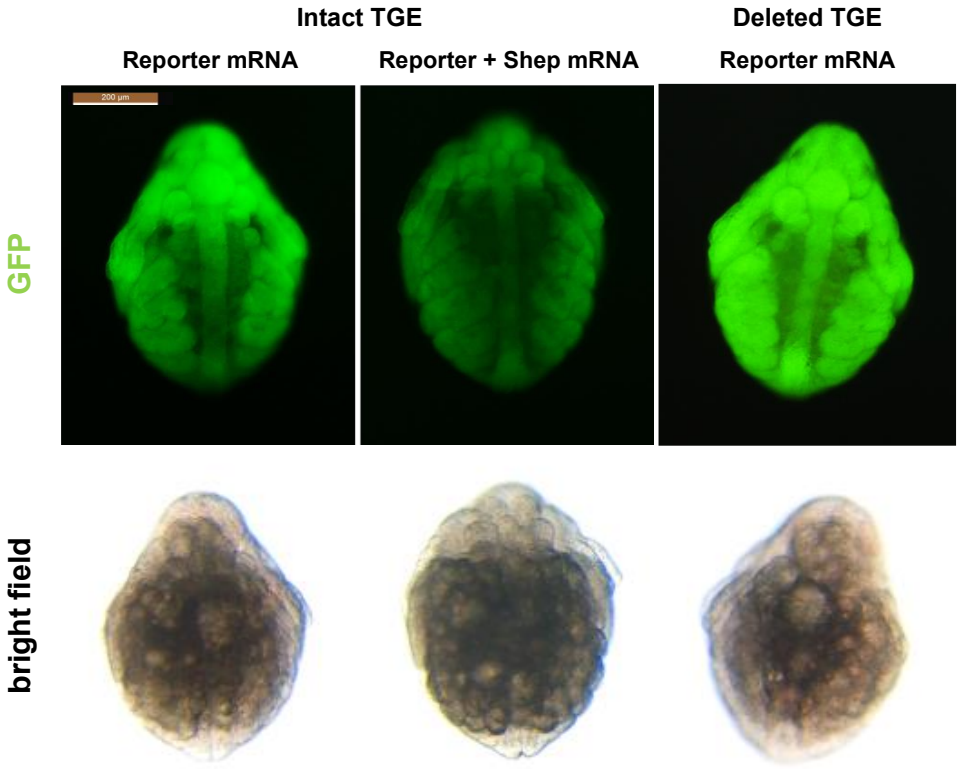

Figure 31. Translation efficiency of *Dsx1* 5′ UTR-Luc reporter mRNA variants with and without the Shep binding site *in vivo*.

Ventral view of female embryos of wildtype strain injected with *Dsx1* 5′ UTR-GFP reporter mRNA and reporter mRNA plus *Shep* mRNA and *Dsx1* 5′ UTR without TGE-GFP reporter mRNA observed at 30 h after injection. GFP fluorescence signals showed the efficiency of translation. The bright field images were used to understand the localization pattern of GFP expression.

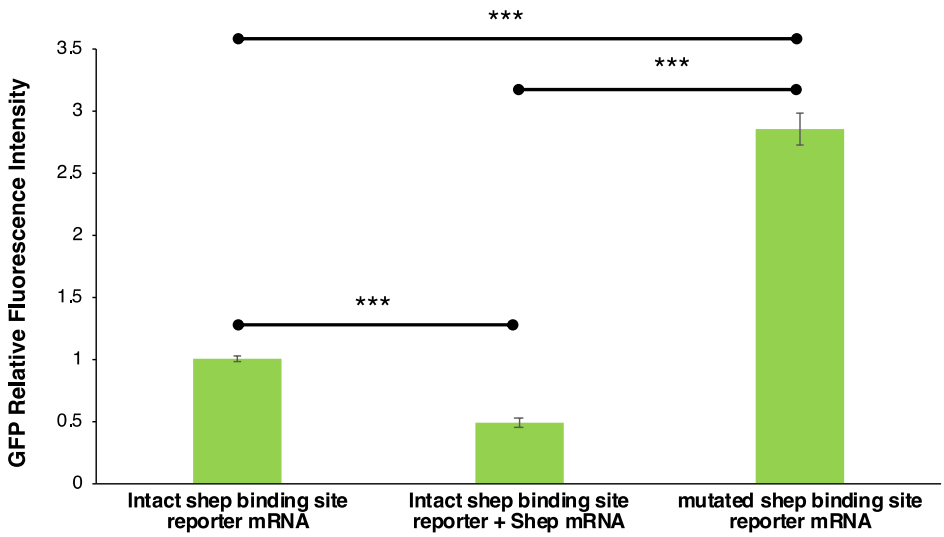

Figure 32. Relative GFP fluorescence intensity of *Dsx1* 5′ UTR-Luc reporter mRNA variants with and without the Shep binding site

Relative GFP fluorescence intensity calculated among the three treatments. Error bars indicate the standard error of the mean, n=5. The endpoints of the line above the bars show which samples were compared statistically. ***p<0.001 (Student's T-test).

Moreover, a significant reduction of the GFP fluorescence was observed when *Shep* mRNA was co-injected with the intact *Dsx1α* 5′ UTR::GFP reporter mRNA (Figure 31 and 32). This indicates that Shep functions at the post-transcriptional level by suppressing translation.

4.3.3 Shep binds the TGE for translational repression of *Dsx1*

To confirm that Shep directly binds to and regulates *Dsx1* translation through the TGE-like motif, I performed the suppression experiment *in vitro*. The luciferase gene was fused to two *Dsx1α* 5′ UTRs, one with an intact TGE-like motif and the other with the deleted TGE

sequence. The mRNAs were synthesized *in vitro* and were translated with or without the *Shep* mRNA.

Results showed that the luciferase activity of the *Dsx1* reporter mRNA which harbors the Shep binding site was significantly reduced by the addition of *Shep* mRNA (Figure 33). In contrast, the translation of the reporter mRNA without the Shep binding site remained unaffected by the presence of Shep, indicating that Shep suppresses the translation of the reporter mRNA through the TGE-like motif.

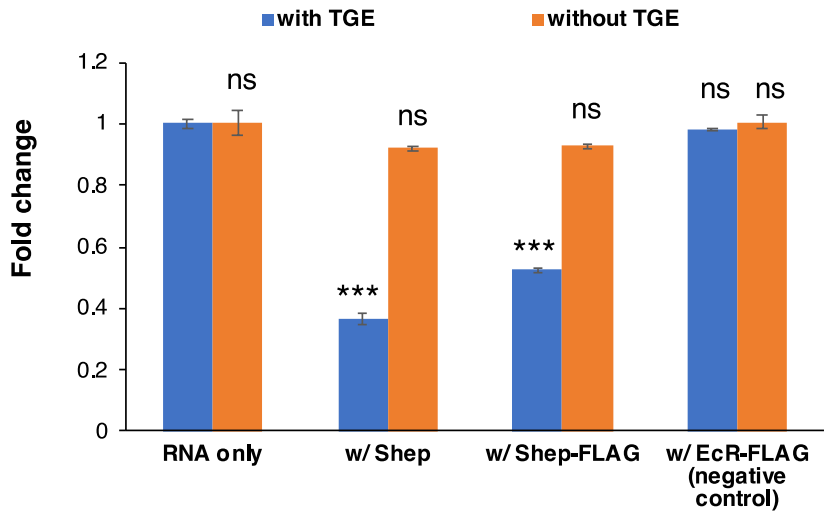

Figure 33. Luciferase activity of *Dsx1* 5′ UTR-Luc reporter mRNA variants with and without the Shep binding site *in vitro*.

Relative luciferase activity after *in vitro* translation assay of *Dsx1* 5′ UTR-Luc reporter mRNA with intact Shep binding site and *Dsx1* 5′ UTR-Luc reporter mRNA without the Shep binding site upon addition of *Shep* mRNA, *Shep-FLAG* mRNA, and *EcR-FLAG* mRNA (negative control). Error bars indicate the standard error of the mean, n=3. Samples were compared against the expression of the *Dsx1* 5′ UTR-Luc reporter mRNA with intact Shep binding site without the addition of any other mRNAs. ***p<0.001, ns: not significant (Student's T-test).

Aside from the wildtype Shep, a Shep fused with the FLAG peptide was also added with the *Dsx1* reporter mRNA with and without the TGE-like element. In Figure 33 it was confirmed that the Shep-FLAG functions the same way as the wildtype Shep in repressing the translation of the reporter mRNA with the intact potential Shep binding site. While the

unrelated FLAG-tagged protein (Flag-EcR) which is the negative control showed no effect on the translation of *Dsx1* with or without the TGE. Hence, the FLAG-tagged proteins were used for the next experiment.

To further confirm the direct interaction of Shep and its proposed binding site, I conducted a pulldown experiment using the FLAG-tagged *Shep* using the previously tested FLAG-tagged proteins. The RNA harboring the intact TGE motif and another RNA without the motif were separately incubated to interact with the *in vitro* translated FLAG-tagged Shep and other controls, the Shep and the unrelated FLAG-tagged EcR. After pulldown using the M2 Anti-FLAG Affinity Gel, only the RNA with the intact TGE in the Shep-FLAG treatment showed highly significant enrichment (Figure 34). The RNA without the TGE failed to bind with the Shep-FLAG. And for the other negative controls, the Shep without the FLAG peptide also did not pulldown any RNAs and the unrelated EcR-FLAG did not bind with the RNAs with and without the TGE as well.

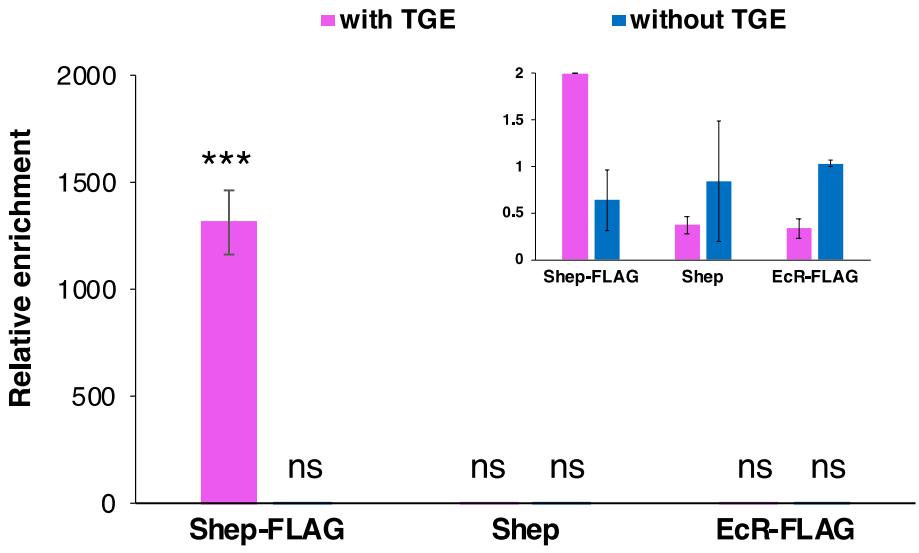

Figure 34. Enrichment of the RNAs with and without the Shep binding site after FLAG pulldown assay.

Samples are compared against the negative control, RNA without the Shep binding site pulled using EcR-FLAG. Error bars indicate the standard error of the mean, n=3. ***p<0.001, ns: not significant (Student's T-test).

These results proved that the Shep exclusively binds with the RNA harboring the TGE, supporting that the TGE-like motif is indeed the Shep binding site and it is through this binding site that Shep regulates the *Dsx1* translation.

4.3.4 *DAPALR* regulates translation efficiency by acting as a decoy

To exhibit the *DAPALR*-Shep regulation at the translational level, I then performed the suppression experiment *in vitro* with the addition of the full region of *DAPALR* and its partial region harboring the core element that has the Shep binding site. Consistent with the *in vivo* experiment results, the addition of either the *DAPALR* or its core element to the *Dsx1* reporter mRNA with the Shep binding site, enhanced the luciferase translation activity even in the presence of Shep (Figure 35).

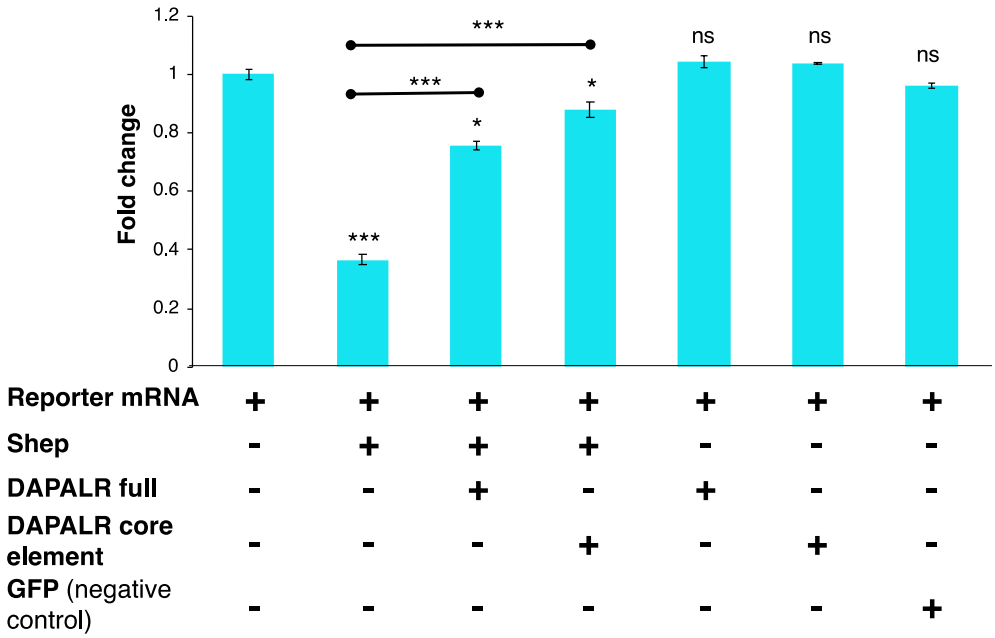

Figure 35. Translation efficiency of the *Dsx1* 5′ UTR-Luc reporter mRNA upon addition of Shep and *DAPALR* full and partial regions.

Relative luciferase activity after *in vitro* translation assay of *Dsx1* 5′ UTR-Luc reporter mRNA with intact Shep binding site upon addition of *Shep* mRNA, *DAPALR* full RNA, *DAPALR* core element, and *GFP* mRNA (negative control). Error bars indicate the standard error of the mean, n=3. Samples were

compared against the expression of the *Dsx1* 5′ UTR-Luc reporter mRNA without the addition of any other mRNAs. The endpoints of the line above the bars show which sample was additionally compared statistically. *p<0.05, ***p<0.001, ns: not significant (Student's T-test).

I also added the GFP mRNA that does not contain any TGE-like element in its sequence as a negative control for the *in vitro* luciferase assay and it showed no effect on the translation efficiency of the *Dsx1* reporter mRNA (Figure35). This proves that the reduced translation observed after the addition of Shep is due to its repression activity and not because of the competition for available ribosomes in the reticulocyte lysate. The *DAPALR* full and partial regions were also individually added with the reporter mRNA to show that the *DAPALR* variants also do not have any adverse effect on its translation efficiency (Figure35).

Finally, I also tested the effect of adding different concentrations of the full and core element of *DAPALR* in the translation of the *Dsx1* reporter mRNA exposed with Shep. Results showed that the effect of the two *DAPALR* versions was not significantly different from one another and their rescue efficiencies were both dose-dependent (Figure 36). The concentration of *DAPALR* needed to be at least 5 times higher than the reporter mRNA and Shep to be able to observe its decoy activity. These results showed the role of *DAPALR* in canceling the suppression of Shep to *Dsx1* translation.

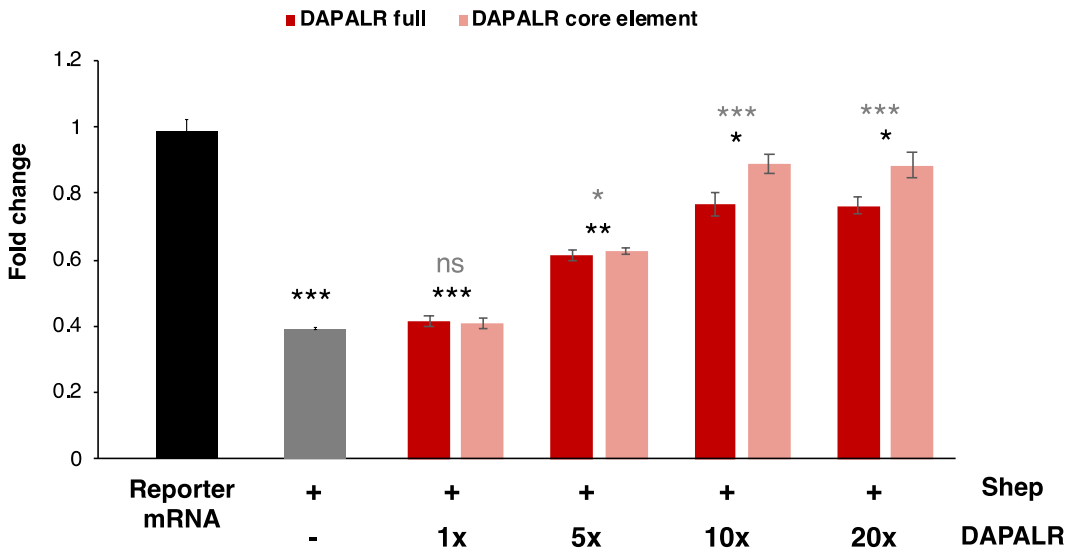

Figure 36. Translation efficiency of the *Dsx1* 5′ UTR-Luc reporter mRNA with different concentrations of the full and partial regions of *DAPALR*.

Relative luciferase activity after *in vitro* translation assay of *Dsx1* 5′ UTR-Luc reporter mRNA with Shep and different concentrations of full region of DAPALR and its core element. Error bars indicate the standard error of the mean, n=3. Black asterisks show significant statistics compared with the expression of the *Dsx1* 5′ UTR-Luc reporter mRNA. Gray asterisks show significant statistics compared with the Reporter mRNA with Shep. *p<0.05, **p<0.01, ***p<0.001, ns: not significant (Student's T-test).

4.4 Discussion

As the previous chapter showed that Shep represses *Dsx1* at the post-transcription level, I presumed that it could have the same function as its ortholog in *C. elegans*, the Sup-26, which acts as a translational repressor in the sex determination pathway. Interestingly, I found the consensus sequence of Sup-26's target sequence, the *tra-2* and *Gli* element (TGE), conserved in *Dsx1α* 5′ UTR and *DAPALR*. This TGE-like element is located in the overlapping region of *DAPALR* with *Dsx1α* 5′ UTR which was identified as the transactivation region of *DAPALR*.

To confirm its function, I induced transient ectopic expression of the RNA expressing the TGE-like element in female eggs of the *Dsx1* reporter strain and this resulted in activation of mCherry fluorescence similar to the overexpression of *DAPALR* and its overlapping region with *Dsx1α* 5′ UTR. This suggests that the TGE-element alone is enough to activate *Dsx1*, that it could be the core element for *DAPALR's* function.

To further test the role of the TGE element in the translational regulation of Shep and *DAPALR*, I performed *in vivo* translation assay using GFP reporter mRNAs which expressions are driven by *Dsx1α* 5′ UTR either with and without the intact potential Shep binding site (TGE). In the result, I observed that the GFP translation of the reporter mRNA with intact TGE was lower than the one without the TGE, suggesting that endogenous Shep represses the translation by associating to its binding site. This was confirmed when further reduction was even observed upon the addition of Shep mRNA with the intact *Dsx1α* 5′ UTR reporter mRNA. These results supported the assumption that Shep functions at the post-transcription level by repressing translation like its ortholog Sup-26. As Sup-26 binds and repress the translation of *tra-2* mRNA (Mapes et al., 2010), Shep in *D. magna* also binds to repress the translation of *Dsx1*; both of which are key regulators of sex determination (Kato, Kobayashi, et al., 2011; Kuwabara et al., 1992).

As there is a possibility in the *in vivo* results that other endogenous factors may have contributed to the repression observed after injection, I performed the translation suppression experiment *in vitro*. Luciferase assay showed that only the reporter mRNA with the intact TGE was repressed by Shep, while the translation of the *Dsx1* reporter mRNA without the TGE remained unaffected by the presence of Shep. The addition of the GFP mRNA which does not recognize and bind to any TGE-like element also showed no effect on the translation of the *Dsx1* reporter mRNAs with and without the potential Shep binding site. These results prove that Shep suppresses the translation of the reporter mRNA through the TGE-like motif and not just because of the competition for available ribosomes in the reticulocyte lysate when two mRNAs are present.

The pulldown experiment using a FLAF-tagged *Shep* further established that the TGE-like element is the binding site of Shep when only the RNA harboring the intact TGE motif showed high enrichment after the pulldown, while the other RNA without the TGE failed to bind with

Shep. This proves that although all RBPs bind to RNA, they do so with different RNA-sequence specificities and affinities; and for Shep, it requires the TGE element to recognize and repress its target gene.

Lastly, with the premise that *DAPALR* activates *Dsx1*, I added the full region of *DAPALR* and its partial region harboring the TGE. The addition of either the *DAPALR* full sequence and its core element, which is the Shep binding site, enabled luciferase translation even in the presence of Shep; the repression caused by Shep towards Dsx1 translation seems to be alleviated through *DAPALR* or its core element. This showed that the effect of *DAPALR* in canceling the suppression of Shep to *Dsx1* translation could be reflected *in vitro*.

Finally, it can be concluded that Shep functions at the post-transcription level by suppressing the translation of its target mRNA by recognizing its binding site and that *DAPALR* relieves this suppression by sequestering Shep through the same binding site.

From these, I proposed the mechanism that Shep associates to its binding site in *DAPALR* and *Dsx1*α 5′ UTR to regulate the sex determination process. Since Shep is non sexually dimorphic, Shep represses *Dsx1* in both male and female. Having *DAPALR* to be only expressed in male (Kato et al., 2018), it can block the negative activity of Shep towards *Dsx1*, allowing *Dsx1* activation which leads to male differentiation. Like other long noncoding RNA that acts as "decoys" (Wang & Chang, 2011), *DAPALR* seems to bind and titrate away Shep from repressing *Dsx1*α 5′ UTR to activate *Dsx1* and to turn on the male sex determination pathway.

As exhibited by *DAPALR*, the transcription itself of lncRNAs that overlap with neighbor-coding genes can serve as regulators of the expression of the gene it overlaps with (Ard et al., 2017). The upstream transcription of *DAPALR* can positively influence downstream gene expression of *Dsx1* by displacing promoter-bound repressors like the Shep (Takemata et al., 2016). This discovery that *DAPALR* displaces Shep from *Dsx1*α 5′ UTR by sequestering

it away also supports the emerging concept that the majority of lncRNAs function as molecular decoys. By serving as a "molecular sink" for RBPs, some lncRNAs are transcribed to bind and titrate away protein targets which can be transcription factors, chromatin modifiers, or regulatory factors, to inhibit their normal functions (Wang & Chang, 2011). Like *DAPALR*, lncRNAs that function as decoys act by negatively regulating an effector like Shep; *DAPALR* inhibits Shep from executing its effector function which is to repress *Dsx1*.

Other lncRNAs have been reported to also function through competitive binding; the lncRNA NORAD and PUM proteins, DHFR lncRNA and TFIIB, PANDA, and NF-YA are just some of the examples (Hung et al., 2013; Lee et al., 2016; Martianov, Ramadass, Serra Barros, Chow, & Akoulitchev, 2007). However the above-mentioned lncRNAs are intergenic, *DAPALR* is the first sense overlapping lncRNA to report such mechanism.

Moreover, being the primary gene responsible to regulate male trait development in *D. magna*, *Dsx1* level in the system needs to be strictly guarded for molecular and phenotype stability (Blake, Kærn, Cantor, & Collins, 2003; Kato, Kobayashi, et al., 2011), hence the significant role of Shep as the repressor of Dsx1 expression. Control of mRNA translation plays a pivotal role in the regulation of gene expression in embryonic and adult tissues. For instance, translational repression mediated by RBPs like Shep has emerged as a major regulatory mechanism for fine-tuning important biological processes, defects in which may be deleterious for development and physiology (Tahmasebi, Khoutorsky, Mathews, & Sonenberg, 2018). Without Shep to patrol stochastic expression of *Dsx1*, it may result in an abnormal phenotype or sexual ambiguity which reduces the fitness or an all-male population that could not reproduce at all.

These findings reveal how critical it is for *Dsx1* expression to be maintained within the right range to safeguard the sex determination pathway and with that, it is regulated by two different factors, the RBP Shep that represses it and the *DAPALR* lncRNA that activates it. The

repression of Shep towards Dsx1 translation is a robust mechanism to silence *Dsx1* in females and limit its expression in males. While *DAPALR* as a decoy ensures Dsx1 activation in males when and where it is needed.

The present study does not only give light to the unique role of *DAPALR*-Shep-*Dsx1* but most importantly this paves a deeper understanding of the function and mechanism of the poorly-studied group of sense-overlapping lncRNAs. With the premise that this group of lncRNAs comprises the majority of the lncRNA present (Kuo et al., 2017) and that Shep is abundant mainly in neurons in addition to ubiquitous tissues (D. Chen et al., 2014; Olesnicky et al., 2018), these suggest that the lncRNA-Shep mechanism is not only for binary sex ultrasensitivity; it may be co-opted in the regulation of other genes for binary cell-fate decisions in various development processes.

CHAPTER 5 GENERAL DISCUSSION AND CONCLUSION

5.1 General discussion

Amidst the increasing knowledge of lncRNAs and recent findings indicating that sense-overlapping lncRNAs are the most abundant type of lncRNA (Kuo et al., 2017), understanding about the function of this group of lncRNA is still lacking. Here we investigated the function and mechanism of the sense overlapping lncRNA *DAPALR* and one of its associating protein, Shep, as key players to harness the gene expression of *Dsx1*.

Initial studies on *DAPALR* showed that it has the potential to regulate *Dsx1*. This was inferred when the expression profile of *DAPALR* revealed that, like *Dsx1*, it is exclusively expressed in males and when silencing of *DAPALR* in males resulted in feminization. However, since the silencing of the *DAPALR* thru RNAi may also result in heterochromatin formation of the neighboring gene *Dsx1,* ectopic expression of the full region of *DAPALR* is necessary to confirm that *DAPALR* expression is sufficient and vital in regulating *Dsx1*. To further understand the mechanism of *DAPALR*, its functional element, and binding proteins were identified and how *DAPALR* functions with its RBP was also explored.

In females where *Dsx1* is transcriptionally silenced, overexpression of *DAPALR* and its transactivation element which is its overlapping region with *Dsx1*α 5′ UTR activated *Dsx1*. Identifying one of the RBPs of *DAPALR*, Shep, we found out that *Shep* loss-of-function increased the *Dsx1* expression too but it was still not as high in the manipulated females compared to that in males. Therefore, the *Shep* loss of function did not lead to sex reversal from female to male. In males, the *Dsx1* expression was enhanced by the *Shep* loss of function throughout the body. Importantly, *DAPALR* overexpression led to a similar change of the *Dsx1* expression pattern both in females and males. To understand the functional relationship on *Dsx1* expression between Shep and *DAPALR*, *in vitro* experiments were performed. The FLAG pulldown experiment showed the exclusive binding of Shep to the TGE and the suppression

experiment showed that Shep inhibits the translation of *Dsx1* in presence of the TGE. Moreover, the addition of *DAPALR* relieved the suppression caused by Shep and activated *Dsx1* translation.

Based on the results, we propose the noise canceling mechanism as a function of the sense overlapping lncRNA and Shep as shown in Figure 37. In females, *Dsx1* transcription is repressed to avoid masculinization. However, due to stochasticity in gene expression (Kærn, Elston, Blake, & Collins, 2005; Ozbudak, Thattai, Kurtser, Grossman, & Van Oudenaarden, 2002), there would be noises in gene expression. In a previous study, we proved that improper expression of *Dsx1* changes the expression profile of its downstream genes resulting in intersex (Nong, Matsuura, Kato, & Watanabe, 2020); which suggested that the stochastic transcription of *Dsx1* causing population heterogeneity, should be avoided. In the presence of the Shep, the *Dsx1* mRNA from the transcriptional noise cannot be translated immediately because of the binding of the Shep at the TGE-like motif. When the sense overlapping lncRNA *DAPALR* is expressed, the Shep is sequestered from the mRNA by the *DAPALR* and the *Dsx1* translation is unlocked. This mechanism may function to avoid the unexpected expression of the *Dsx1* to accomplish sexual dimorphic expression. Quantity of Shep in the cell may define a threshold to cancel the effect of noisy transcription and the stochastic transcript below the threshold may not be translated because of the presence of Shep. Although quantitative estimation of Shep in the cells is difficult, we showed that the copy number of the *DAPALR* is one-tenth of the *Dsx1* mRNA (Kato et al., 2018), which may be a sufficient quantity to unlock the Shep suppression. It is also possible that the *DAPALR* expression was detected to be only 10% of the Dsx1 expression because its region of expression may be more limited than that of the Dsx1 expression and the extracted total RNA was from the whole embryos. In this scenario, *DAPALR* may be expressed in cells only at the early stage of male differentiation and decrease the threshold of Dsx1 expression by sequestering the Shep protein. Then, the translated Dsx1

protein may activate its own promoter by positive feedback loop to maintain Dsx1 expression. And since *Shep* seems to be ubiquitously expressed in different tissues in the whole body, this RBP may be able to silence Dsx1 expression in non-sexually dimorphic tissues that do not express *DAPALR*. This hypothesis is consistent when *DAPALR* was overexpressed ubiquitously or Shep expression was silenced by siRNA and the dsx1 reporter mCherry expression was observed throughout the body even in the non-sexually dimorphic tissues. Further studies to prove this hypothesis could be investigated in the future.

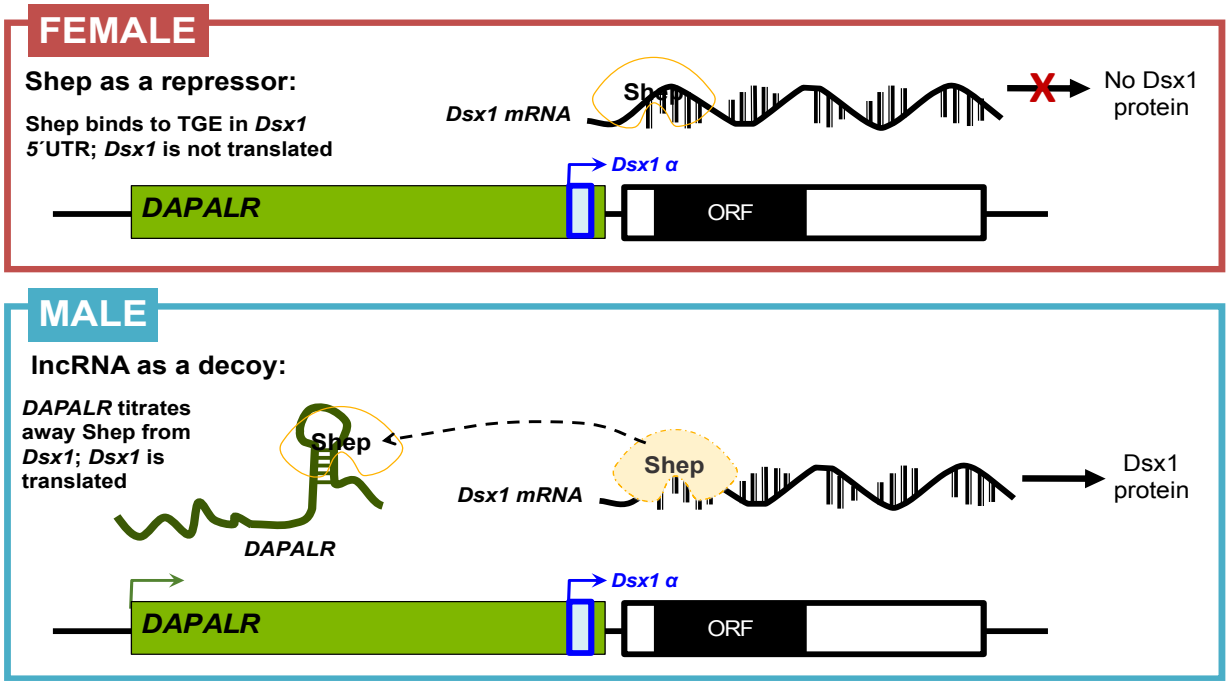

Figure 37. Proposed "noise-canceling" mechanism of the lncRNA *DAPALR* and Shep to regulate *Dsx1* and the sex determination process in *D. magna*.

Both *Dsx1* and *Tra-2* are key regulators of sex determination (Kuwabara et al., 1992) and they should be strictly regulated to avoid sexual ambiguity. Although there is no knowledge about a sense lncRNA at the *Tra-2* locus, a similar mechanism may function in *Drosophila*. The Shep is also known to function in neurogenesis (D. Chen et al., 2014) and the Shep functions at a translational level (Olesnicky et al., 2018). It may also be possible that the Shep

suppresses the translation under the control of unknown lncRNA and the genes whose noisy expression is harmful to the cell may have such kind of noise-canceling system. The tight regulation of *Dsx1* through Shep-dependent suppression and by lncRNA exhibits one mechanism of how nature keeps intersex and sexual ambiguity rare.

Shep has been reported to have many functions such as antagonizing chromatin insulator activity, transcriptional and post-transcriptional control, especially in neurogenesis (D. Chen, Brovkina, Matzat, & Lei, 2019; D. Chen, Dale, & Lei, 2018; Olesnicky et al., 2018). Localization of Shep in the cytoplasm (Mapes et al., 2010) also supports the possibility that Shep functions at a post-transcriptional level.

In this study, we focused on the Shep and found that the Shep functions as a noise canceler, and the *DAPALR* unlocks it. For further understanding of the *DAPALR*, the other RNA binding protein, CGUBP1, should be considered to understand the robust sex-determination system in *D. magna*.

A similar mechanism is known in the lncRNA named *linc-MD1*, in which miRNA is sequestered from the mRNA by the lncRNA (Cesana et al., 2011), and a competing endogenous RNA hypothesis has been proposed (Salmena, Poliseno, Tay, Kats, & Pandolfi, 2011). Our finding suggests that the RBP such as the Shep can be a target of competing endogenous RNA in a broad meaning.

In a previous study enumerating a list of post-transcription regulators that Shep binds to, 5 out 77 Shep targets are noncoding RNAs (Olesnicky et al., 2018). While none of which mechanisms have been studied extensively and that the three-way network of translation regulation involving an mRNA, noncoding RNA, and RBP may be the first involving Shep; this regulation may occur more commonly. It is predicted that the sense-overlapping lncRNAs comprise the majority of the lncRNA present (Kuo et al., 2017). And as Shep is expressed not only in neurons but other tissues (D. Chen et al., 2014; Olesnicky et al., 2018), the unique role

of *DAPALR*-Shep-*Dsx1* may not only be for binary sex ultrasensitivity but also for binary regulation of other genes in various biological processes.

In application, these findings deepened our understanding of the significance and potential function of sense-overlapping lncRNAs in biotechnology studies to regulate gene expression. The lncRNA-RBP network can be used to regulate gene induction and suppress leaky expression in engineered transcription control systems. It could also be applied as a safekeeping tool for binary switches in biosynthetic pathways.

5.2 Conclusion

DAPALR, a sense overlapping long noncoding RNA (lncRNA) was recently discovered to function in the sex determination of *Daphnia magna*. Exploring its functional role and mechanism, in this study, I discovered that

1) *DAPALR* activates *Doublesex1* (*Dsx1*) and that its transactivation element is the overlapping region with *Dsx1α* 5′ UTR; and

2) Two proteins namely Shep and CGUBP1 interact with the transactivation element of *DAPALR* and focusing on Shep, this RBP consists of two RRMs that are evolutionarily conserved among its orthologs from different species and it represses *Dsx1* at post-transcription level; and lastly

3) Shep represses the translation of *Dsx1* by recognizing and binding to a TGE motif in the *Dsx1* 5′ UTR and *DAPALR,* which also bears this TGE motif, acts as a decoy to sequester Shep and activate *Dsx1*.

This study not only gave light to the mechanism of *DAPALR* and Shep in regulating the unintentional translation of *Dsx1* to avoid sexual ambiguity, but it also demonstrated a unique function and significant role of sense-overlapping lncRNAs in dimorphic gene expressions

such as sex determination which may also account for the binary regulation of other genes in various biological processes.

REFERENCES

Adachi, S., Homoto, M., Tanaka, R., Hioki, Y., Murakami, H., Suga, H., … Natsume, T. (2014). ZFP36L1 and ZFP36L2 control LDLR mRNA stability via the ERK-RSK pathway. *Nucleic Acids Research*, *42*(15), 10037–10049. https://doi.org/10.1093/nar/gku652

Antic, D., & Keene, J. D. (1997). Embryonic Lethal Abnormal Visual RNA-Binding Proteins Involved in Growth , Differentiation , and Posttranscriptional Gene Expression. *American Journal of Human Genetics*, *61*, 273–278. https://doi.org/10.1086/514866

Ard, R., Allshire, R. C., & Marquardt, S. (2017). Emerging properties and functional consequences of noncoding transcription. *Genetics*, *207*(2), 357–367. https://doi.org/10.1534/genetics.117.300095

Asada, M., Kato, Y., Matsuura, T., & Watanabe, H. (2014a). Early embryonic expression of a putative ecdysteroid-phosphate phosphatase in the water flea, daphnia magna (Cladocera: Daphniidae). *Journal of Insect Science*, *14*, 1–6. https://doi.org/10.1093/jisesa/ieu043

Asada, M., Kato, Y., Matsuura, T., & Watanabe, H. (2014b). Visualization of ecdysteroid activity using a reporter gene in the crustacean, Daphnia. *Marine Environmental Research*, *93*, 118–122. https://doi.org/10.1016/j.marenvres.2013.11.005

Bahamonde, P. A., Munkittrick, K. R., & Martyniuk, C. J. (2013). Intersex in teleost fish: Are we distinguishing endocrine disruption from natural phenomena? *General and Comparative Endocrinology*, *192*, 25–35. https://doi.org/10.1016/j.ygcen.2013.04.005

Beisang, D., Bohjanen, P. R., & Louis, I. A. V.-S. (2012). CELF1, a Multifunctional Regulator of Posttranscriptional Networks Chapter. *B i n d i n g P r o t e i N*, 181–206. https://doi.org/10.5772/60142

Blake, W. J., Kærn, M., Cantor, C. R., & Collins, J. J. (2003). Noise in eukaryotic gene

expression. *Nature*, *422*(6932), 633–637. https://doi.org/10.1038/nature01546

Briata, P., & Gherzi, R. (2020). Long non-coding RNA-ribonucleoprotein networks in the post-transcriptional control of gene expression. *Non-Coding RNA*, *6*(3). https://doi.org/10.3390/NCRNA6030040

Brooks, A. N., Duff, M. O., May, G., Yang, L., Bolisetty, M., Landolin, J., … Brenner, S. E. (2015). Regulation of alternative splicing in Drosophila by 56 RNA binding proteins. *Genome Research*. https://doi.org/10.1101/gr.192518.115

Buchler, N. E., & Louis, M. (2008). Molecular Titration and Ultrasensitivity in Regulatory Networks. *Journal of Molecular Biology*, *384*(5), 1106–1119. https://doi.org/10.1016/j.jmb.2008.09.079

Bull, J. (1985). Sex determining mechanisms: an evolutionary perspective. *Experientia*, 1285–1296.

Bulmer, M., & Bull, J. J. (1982). Models of polygenic sex determination and sex-ratio control. *Evolution*, *36*, 13–26.

Callaini, G., & Dallai, R. (1991). Abnormal behavior of the yolk centrosomes during early embryogenesis of Drosophila melanogaster. *Experimental Cell Research*, *192*(1), 16–21. https://doi.org/10.1016/0014-4827(91)90151-J

Campos, E. I., & Reinberg, D. (2009). Histones: Annotating chromatin. *Annual Review of Genetics*, *43*, 559–599. https://doi.org/10.1146/annurev.genet.032608.103928

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., … Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, *147*(2), 358–369. https://doi.org/10.1016/j.cell.2011.09.028

Chen, D., Brovkina, M., Matzat, L. H., & Lei, E. P. (2019). Shep RNA-binding capacity is required for antagonism of gypsy chromatin insulator activity. *G3: Genes, Genomes,*

Genetics, *9*(3), 749–754. https://doi.org/10.1534/g3.118.200923

Chen, D., Dale, R. K., & Lei, E. P. (2018). Shep regulates Drosophila neuronal remodeling by controlling transcription of its chromatin targets. *Development (Cambridge)*, *145*(1), dev154047. https://doi.org/10.1242/dev.154047

Chen, D., Gu, T., Pham, T. N., Zachary, M. J., & Hewes, R. S. (2017). Regulatory mechanisms of metamorphic neuronal remodeling revealed through a Genome-Wide modifier screen in drosophila melanogaster. *Genetics*, *206*(3), 1429–1443. https://doi.org/10.1534/genetics.117.200378

Chen, D., Qu, C., & Hewes, R. S. (2014). *Neuronal remodeling during metamorphosis is regulated by the alan shepard (shep) gene in Drosophila melanogaster*. *Genetics* (Vol. 197). https://doi.org/10.1534/genetics.114.166181

Chen, L. L. (2016). Linking Long Noncoding RNA Localization and Function. *Trends in Biochemical Sciences*, *41*(9), 761–772. https://doi.org/10.1016/j.tibs.2016.07.003

Cline and, T. W., & Meyer, B. J. (1996). VIVE LA DIFFÉRENCE:Males vs Females in Flies vs Worms. *Annual Review of Genetics*. https://doi.org/10.1146/annurev.genet.30.1.637

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., … Boore, J. L. (2011). The ecoresponsive genome of Daphnia pulex. *Science*. https://doi.org/10.1126/science.1197761

Crews, D., & Bull, J. J. (2009). Mode and tempo in environmental sex determination in vertebrates. *Seminars in Cell and Developmental Biology*. https://doi.org/10.1016/j.semcdb.2009.02.004

Elendt, B. P., & Bias, W. R. (1990). Trace nutrient deficiency in Daphnia magna cultured in standard medium for toxicity testing. Effects of the optimization of culture conditions on life history parameters of D. magna. *Water Research*, *24*(9), 1157–1167. https://doi.org/10.1016/0043-1354(90)90180-E

Epper, F., & Bryant, P. (1983). Sex-specific control of growth and differentiation in the Drosophila genital disk, studied using a temperature-sensitive transformer-2 mutation. *Dev Biol*, *100*, 294–307.

Fiering, S., Whitelaw, E., & Martin, D. I. K. (2000). To be or not to be active: The stochastic nature of enhancer action. *BioEssays*, *22*(4), 381–387. https://doi.org/10.1002/(SICI)1521-1878(200004)22:4<381::AID-BIES8>3.0.CO;2-E

Ford, A. T. (2012). Intersexuality in Crustacea: An environmental issue? *Aquatic Toxicology*, *108*, 125–129. https://doi.org/10.1016/j.aquatox.2011.08.016

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*(3), 233–239. https://doi.org/10.1038/nbt.2623

Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Ahkmetova, L., Pauli, A., … Schier, A. F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE*, *9*(8), e106396. https://doi.org/10.1371/journal.pone.0098186

Glenner, H., Thomsen, P. F., Hebsgaard, M. B., Sorensen, M. V., & Willerslev, E. (2006). EVOLUTION: The Origin of Insects. *Science*, *314*(5807), 1883–1884. https://doi.org/10.1126/science.1129844

Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters*, *582*(14), 1977–1986. https://doi.org/10.1016/j.febslet.2008.03.004

Harper, C. (2007). *Intersex*. Berg. https://doi.org/10.4324/9781003103554

Hebert, P. D. N. (1978). THE POPULATION BIOLOGY OF DAPHNIA (CRUSTACEA, DAPHNIDAE). *Biological Reviews*, *53*(3), 387–426. https://doi.org/10.1111/j.1469-185X.1978.tb00860.x

Hentze, M. W., Castello, A., Schwarzl, T., & Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology*, *19*(5), 327–341. https://doi.org/10.1038/nrm.2017.130

Hodgkin, J. (2002). Exploring the envelope: Systematic alteration in the sex-determination systemof the nematode Caeraorhabditis elegens. *Genetics*, *162*, 767–780.

Hoopes, L. (2008). Introduction to the gene expression and regulation topic room. *Nature Education*, *1*(1), 160.

Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., … Genet, N. (2013). Panda lncrna, *43*(7), 621–629. https://doi.org/10.1038/ng.848

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., … Ule, J. (2014). iCLIP: Protein-RNA interactions at nucleotide resolution. *Methods*, *65*(3), 274–287. https://doi.org/10.1016/j.ymeth.2013.10.011

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(9), 822–826. https://doi.org/10.1038/nbt.2508

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

Kashi, K., Henderson, L., Bonetti, A., & Carninci, P. (2016). Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1859*(1), 3–15. https://doi.org/10.1016/j.bbagrm.2015.10.010

Kato, Y., Kobayashi, K., Watanabe, H., & Iguchi, T. (2011). Environmental sex determination in the branchiopod crustacean Daphnia magna: Deep conservation of a Doublesex gene in the sex-determining pathway. *PLoS Genetics*, *7*(3), e1001345.

https://doi.org/10.1371/journal.pgen.1001345

Kato, Y., Matsuura, T., & Watanabe, H. (2012). Genomic Integration and Germline Transmission of Plasmid Injected into Crustacean Daphnia magna Eggs. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0045318

Kato, Y., Perez, C. A. G., Mohamad Ishak, N. S., Nong, Q. D., Sudo, Y., Matsuura, T., … Watanabe, H. (2018). A 5′ UTR-Overlapping LncRNA Activates the Male-Determining Gene doublesex1 in the Crustacean Daphnia magna. *Current Biology*, *28*(11), 1811–1817. https://doi.org/10.1016/j.cub.2018.04.029

Kato, Y., Shiga, Y., Kobayashi, K., Tokishita, S. I., Yamagata, H., Iguchi, T., & Watanabe, H. (2011). Development of an RNA interference method in the cladoceran crustacean Daphnia magna. *Development Genes and Evolution*, *220*(11–12), 337–345. https://doi.org/10.1007/s00427-011-0353-9

Klüttgen, B., Dülmer, U., Engels, M., & Ratte, H. T. (1994). ADaM, an artificial freshwater for the culture of zooplankton. *Water Research*, *28*(3), 743–746. https://doi.org/10.1016/0043-1354(94)90157-0

Korpelainen, H. (1990). SEX RATIOS AND CONDITIONS REQUIRED FOR ENVIRONMENTAL SEX DETERMINATION IN ANIMALS. *Biological Reviews*. https://doi.org/10.1111/j.1469-185X.1990.tb01187.x

Kumagai, H., Nakanishi, T., Matsuura, T., Kato, Y., & Watanabe, H. (2017). CRISPR/Cas-mediated knock-in via nonhomologous end-joining in the crustacean Daphnia magna. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0186112

Kung, J. T. Y., Colognori, D., & Lee, J. T. (2013). Long Noncoding RNAs: Past, Present, and Future. *Genetics*, *193*(3), 651–669. https://doi.org/10.1534/genetics.112.146704

Kuo, R. I., Tseng, E., Eory, L., Paton, I. R., Archibald, A. L., & Burt, D. W. (2017). Normalized long read RNA sequencing in chicken reveals transcriptome complexity

similar to human. *BMC Genomics*, *18*(1), 1–19. https://doi.org/10.1186/s12864-017-3691-9

Kuwabara, P. E., Okkema, P. G., & Kimble, J. (1992). tra-2 encodes a membrane protein and may mediate cell communication in the Caenorhabditis elegans sex determination pathway. *Molecular Biology of the Cell*, *3*(4), 461–473. https://doi.org/10.1091/mbc.3.4.461

Lambeth, L. S., & Smith, C. A. (2012). Disorders of sexual development in poultry. *Sexual Development*, *6*(1–3), 96–103. https://doi.org/10.1159/000334059

Lee, S., Kopp, F., Chang, T. C., Sataluri, A., Chen, B., Sivakumar, S., … Mendell, J. T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell*, *164*(1–2), 69–80. https://doi.org/10.1016/j.cell.2015.12.017

Li, J.-H., Liu, S., Zheng, L.-L., Wu, J., Sun, W.-J., Wang, Z.-L., … Yang, J.-H. (2015). Discovery of Protein-lncRNA Interactions by Integrating Large-Scale CLIP-Seq and RNA-Seq Datasets. *Frontiers in Bioengineering and Biotechnology*, *2*(January), 1–11. https://doi.org/10.3389/fbioe.2014.00088

Lisbin, M. J., Gordon, M., Yannoni, Y. M., & White, K. (2000). Function of RRM domains of Drosophila melanogaster ELAV: RNP1 mutations and RRM domain replacements with ELAV family proteins and SXL. *Genetics*, *155*(4), 1789–1798.

Mapes, J., Chen, J.-T., Yu, J.-S., & Xue, D. (2010). Somatic sex determination in Caenorhabditis elegans is modulated by SUP-26 repression of tra-2 translation. *Proceedings of the National Academy of Sciences*, *107*(42), 18022–18027. https://doi.org/10.1073/pnas.1004513107

Marin, I., & Baker, B. S. (1998). The evolutionary dynamics of sex determination. *Science*. https://doi.org/10.1126/science.281.5385.1990

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., & Akoulitchev, A. (2007).

Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, *445*(7128), 666–670. https://doi.org/10.1038/nature05519

Martienssen, R., & Moazed, D. (2015). RNAi and heterochromatin assembly. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a019323

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional regulatory elements in the human genome. *Annual Review of Genomics and Human Genetics*, *7*, 29–59. https://doi.org/10.1146/annurev.genom.7.080505.115623

Matzat, L. H., Dale, R. K., Moshkovich, N., & Lei, E. P. (2012). Tissue-Specific Regulation of Chromatin Insulator Function. *PLoS Genetics*, *8*(11). https://doi.org/10.1371/journal.pgen.1003069

Mellor, J., Woloszczuk, R., & Howe, F. S. (2016). The Interleaved Genome. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2015.10.006

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*, *10*(3), 155–159. https://doi.org/10.1038/nrg2521

Naitou, A., Kato, Y., Nakanishi, T., Matsuura, T., & Watanabe, H. (2015). Heterodimeric TALENs induce targeted heritable mutations in the crustacean Daphnia magna. *Biology Open*. https://doi.org/10.1242/bio.20149738

Nakanishi, T., Kato, Y., Matsuura, T., & Watanabe, H. (2014). CRISPR/Cas-mediated targeted mutagenesis in Daphnia magna. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0098363

Nakanishi, T., Kato, Y., Matsuura, T., & Watanabe, H. (2015). TALEN-mediated homologous recombination in Daphnia magna. *Scientific Reports*. https://doi.org/10.1038/srep18312

Nakanishi, T., Kato, Y., Matsuura, T., & Watanabe, H. (2016). TALEN-mediated knock-in via non-homologous end joining in the crustacean Daphnia magna. *Scientific Reports*.

https://doi.org/10.1038/srep36252

Natsume, T., Yamauchi, Y., Nakayama, H., Shinkawa, T., Yanagida, M., Takahashi, N., & Isobe, T. (2002). A direct nanoflow liquid chromatography - Tandem mass spectrometry system for interaction proteomics. *Analytical Chemistry*, *74*(18), 4725–4733. https://doi.org/10.1021/ac020018n

Nong, Q. D., Matsuura, T., Kato, Y., & Watanabe, H. (2020). Two Doublesex1 mutants revealed a tunable gene network underlying intersexuality in Daphnia magna. *PLoS ONE*, *15*(8 August 2020), 1–19. https://doi.org/10.1371/journal.pone.0238256

Nong, Q. D., Mohamad Ishak, N. S., Matsuura, T., Kato, Y., & Watanabe, H. (2017). Mapping the expression of the sex determining factor Doublesex1 in Daphnia magna using a knock-in reporter. *Scientific Reports*, *7*, 13521. https://doi.org/10.1038/s41598-017-13730-4

Olesnicky, E. C., Antonacci, S., Popitsch, N., Lybecker, M. C., Titus, M. B., Valadez, R., … Killian, D. J. (2018). Shep interacts with posttranscriptional regulators to control dendrite morphogenesis in sensory neurons. *Developmental Biology*, *444*(2), 116–128. https://doi.org/10.1016/j.ydbio.2018.09.022

Organ, C., & Janes, D. (2008). Evolution of sex chromosomes in Sauropsida. *Integ Comp Biol*, *48*(4), 512–519. https://doi.org/10.1093/icb/icn041

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & Van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nature Genetics*, *31*(1), 69–73. https://doi.org/10.1038/ng869

Pereira, R., Narita, S., Kageyama, D., & Kjellberg, F. (2010). Gynandromorphs and intersexes: potential to understand the mechanism of sex determination in arthropods. *Terrestrial Arthropod Reviews*, *3*(1), 63–96. https://doi.org/10.1163/187498310X496190

Quinn, J. J., & Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics*, *17*(1), 47–62. https://doi.org/10.1038/nrg.2015.10

Ralston, A., & Shaw, K. (2008). Gene expression regulates cell differentiation. *Nature Ed*, *1*(1), 127.

Reményi, A., Schöler, H. R., & Wilmanns, M. (2004). Combinatorial control of gene expression. *Nature Structural and Molecular Biology*, *11*(9), 812–815. https://doi.org/10.1038/nsmb820

Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell*, *146*(3), 353–358. https://doi.org/10.1016/j.cell.2011.07.014

Sander, J. D., Maeder, M. L., Reyon, D., Voytas, D. F., Joung, J. K., & Dobbs, D. (2010). ZiFiT (Zinc Finger Targeter): An updated zinc finger engineering tool. *Nucleic Acids Research*, *38*(SUPPL. 2), 462–468. https://doi.org/10.1093/nar/gkq319

Savell, K. E., Gallus, N. V. N., Simon, R. C., Brown, J. A., Revanna, J. S., Osborn, M. K., … Day, J. J. (2016). Extra-coding RNAs regulate neuronal DNA methylation dynamics. *Nature Communications*, *7*(May). https://doi.org/10.1038/ncomms12091

Schachtner, L. T., Sola, I. E., Forand, D., Antonacci, S., Postovit, A. J., Mortimer, N. T., … Olesnicky, E. C. (2015). Drosophila Shep and C. elegans SUP-26 are RNA-binding proteins that play diverse roles in nervous system development. *Development Genes and Evolution*, *225*(6), 319–330. https://doi.org/10.1007/s00427-015-0514-3

Shoemaker, C. M., & Crews, D. (2009). Analyzing the coordinated gene network underlying temperature-dependent sex determination in reptiles. *Seminars in Cell & Developmental Biology*, *20*(3), 293–303. https://doi.org/10.1016/j.semcdb.2008.10.010.Analyzing

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*,

22(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Sumiya, E., Ogino, Y., Toyota, K., Miyakawa, H., Miyagawa, S., & Iguchi, T. (2016). Neverland regulates embryonic moltings through the regulation of ecdysteroid synthesis in the water flea Daphnia magna, and may thus act as a target for chemical disruption of molting. *Journal of Applied Toxicology*, *36*(11), 1476–1485. https://doi.org/10.1002/jat.3306

Tahmasebi, S., Khoutorsky, A., Mathews, M. B., & Sonenberg, N. (2018). Translation deregulation in human disease. *Nature Reviews Molecular Cell Biology*, *19*(12), 791–807. https://doi.org/10.1038/s41580-018-0034-x

Takemata, N., Oda, A., Yamada, T., Galipon, J., Miyoshi, T., Suzuki, Y., … Ohta, K. (2016). Local potentiation of stress-responsive genes by upstream noncoding transcription. *Nucleic Acids Research*, *44*(11), 5174–5189. https://doi.org/10.1093/nar/gkw142

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). ClustalW: improving the sensitivity of progressive multiple sequence aligment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research Acids Res*. https://doi.org/10.1093/nar/22.22.4673

Törner, K., Nakanishi, T., Matsuura, T., Kato, Y., & Watanabe, H. (2014). Optimization of mRNA design for protein expression in the crustacean Daphnia magna. *Molecular Genetics and Genomics*, *289*(4), 707–715. https://doi.org/10.1007/s00438-014-0830-8

Toyota, K., Kato, Y., Sato, M., Sugiura, N., Miyagawa, S., Miyakawa, H., … Iguchi, T. (2013). Molecular cloning of doublesex genes of four cladocera (water flea) species.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., &

Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*(7), research0034.1. https://doi.org/10.1186/gb-2002-3-7-research0034

Villagómez, D. A. F., Parma, P., Radi, O., Di Meo, G., Pinton, A., Iannuzzi, L., & King, W. A. (2009). Classical and molecular cytogenetics of disorders of sex development in domestic animals. *Cytogenetic and Genome Research*, *126*(1–2), 110–131. https://doi.org/10.1159/000245911

Vogt, R. C., & Bull, J. J. (1982). Temperature controlled sex-determination in turtles: ecological and behavioral aspects. *Herpetologica*.

Wang, K. C., & Chang, H. Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2011.08.018

Watanabe, H., Tatarazako, N., Oda, S., Nishide, H., Uchiyama, I., Morita, M., & Iguchi, T. (2005). Analysis of expressed sequence tags of the water flea Daphnia magna. *Genome*. https://doi.org/10.1139/g05-038

Williams, T. M., & Carroll, S. B. (2009). Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2687

Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs : functional surprises from the RNA world Long noncoding RNAs : functional surprises from the RNA world, 1494–1504. https://doi.org/10.1101/gad.1800909

Yao, R. W., Wang, Y., & Chen, L. L. (2019). Cellular functions of long noncoding RNAs. *Nature Cell Biology*, *21*(5), 542–551. https://doi.org/10.1038/s41556-019-0311-8

Zarkower, D. (2001). Establishing sexual dimorphism: Conservation amidst diversity? *Nature Reviews Genetics*. https://doi.org/10.1038/35056032

Zhang, Q., Andersen, M. E., & Conolly, R. B. (2006). Binary gene induction and protein

expression in individual cells. *Theoretical Biology and Medical Modelling*, *3*, 1–15. https://doi.org/10.1186/1742-4682-3-18

LIST OF PUBLICATION

Perez, Christelle, Adachi, S., Nong, Q.D., Matsuura, T., Natsume, T., Kato, Y., and Watanabe H. (2021) Sense-overlapping lncRNA as a decoy of translational repressor protein for dimorphic gene expression. *PLOS Genetics* 17 (7): e1009683.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my professors. To my supervisor, Professor Hajime Watanabe, my utmost thanks for accepting me in the program and for entrusting and encouraging me to do this research. To Associate Professor Kato Yasuhiko, thank you for your helpful teachings and for your patience and guidance in every step of this research. To Professor Tadashi Wada, thank you for all the insightful talks and all your advices. Thank you senseis for your academic support and kindness throughout the years.

I would also like to extend my gratitude to Professor Susumu Uchiyama and Professor Toshiya Muranaka from the International Center for Biotechnology, Graduate School of Engineering, Osaka University, for their valuable comments and suggestions for this dissertation writing and final presentation.

I am also grateful to our collaborators from the Molecular Profiling Research Center for Drug Discovery of the National Institute of Advanced Industrial Science and Technology for helping us with the RNA pull-down experiment.

I also acknowledge the Department of Biotechnology, Osaka University and Monbukagakusho Scholarship (MEXT) for the financial support.

I would also like to thank the staff and students of Watanabe Laboratory for making the laboratory feel like home. Thank you for the friendship and companionship.

Finally, to my family and loved ones, I offer this hard work to you. Thank you for always cheering and believing on me. Your love fuels me to dream big and reach high.

Ad mojorem Dei gloriam!