The evolutionary history of the sex-determining gene, *doublesex*, in Insecta

昆虫類における性決定遺伝子 *doublesex* の進化史

Chikami, Yasuhiko

千頭　康彦

Doctor of Philosophy

The Graduate University for Advanced Studies, SOKENDAI
School of Life Science
Department of Basic Biology

2022

The evolutionary history of the sex-determining gene, *doublesex*, in Insecta

Chikami, Yasuhiko

Abstract

Alternative splicing is a pervasive mechanism of processing pre-mRNAs among Eukaryotes and underpins the pleiotropic natures of genes due to the functional differences among splicing isoforms. In the last 30 years, there has been evidence of the neo-functionalization of splicing isoforms. How did splicing isoforms acquire new functions? This question has been debated since the discovery of mRNA splicing. There are several hypotheses on the neo-functionalization of isoforms. However, it is still unclear how novel functions opposite to an ancestral role via the alternative isoforms and what changes in genomic space led to the new role of the isoform since there are few examples to trace the evolutionary process of isoform function. Here, I focus on the *doublesex* (*dsx*) gene as a case to consider this problem.

dsx encodes a transcriptional factor essential for sex determination and sex differentiation in Arthropoda. In Holometabola, such as *Drosophila melanogaster*, *dsx* undergoes sex-specific splicing producing either male- or female-specific isoforms. The sex-specific isoforms serve sex-antagonistic roles and promote either male or female differentiation. In contrast, recently, it has been reported from some hemimetabolan species that *dsx* possesses the sex-specific isoforms required only for male differentiation but not for female differentiation. This functional difference suggests the neo-functionalization of genes via alternative splicing. Therefore, *dsx* would provide an example to examine the hypothesis on the neo-functionalization of isoforms. Hence, in this thesis, the subject is to infer the evolutionary history of *dsx* in Insecta. Central questions are: when did the sex-specific splicing of *dsx* occur, how did the female-specific isoform of *dsx* come into use for female differentiation, and what changes are linked to the neo-functionalization of *dsx*.

Over the last decade, some studies have proposed several hypotheses on the functionalizing process of *dsx*. However, its functional evolution has still been ambiguous because of a gap between studied taxa due to the absence of information on the closely related outgroups of winged insects. To fill the gap, I add knowledge of Zygentoma, an

apterygote taxon that is the sister clade of winged insects, to the evolutionary study of *dsx*. I used the firebrat, *Thermobia domestica*, a species of Zygentoma.

First, I found that *T. domestica* has two *dsx* copies and sex-specific isoforms in one of them. Molecular phylogenetic analysis revealed that gene duplication of *dsx* occurred before the emergence of the common ancestor of Zygentoma and winged insects and that the copy with the isoforms is the ortholog of *dsx* of winged insects. This result supports that *dsx* of *T. domestica* is under the sex-specific splicing control and that its splicing control has a single origin in insect evolution. This finding emphasizes the utility of *T. domestica* for examining *dsx* evolution since this species would retain the ancestral state of winged insects in terms of the coexistence of the gene copy number and splicing control.

Then, I investigated the roles of *dsx* of *T. domestica* by the nymphal RNAi system and revealed that *dsx* is required for male morphogenesis during post-embryonic development but is not essential for differentiating female traits. This result and previous information on winged insects indicate that the function in female morphology occurred in the common ancestor of Holometabola except for Hymenoptera, i.e., Aparaglossata. Thus, my result strongly supports the earlier argument that *dsx* was necessary only for male differentiation when the isoforms appeared and later became responsible for female differentiation. I also uncovered that *dsx* has sex-antagonistic roles in the female-specific expression of *vitellogenin* homologs. These results show that *dsx* in females came into use separately among morphogenesis and other biological processes in evolution.

Finally, I found that the C-terminal female-specific region of Dsx is massively different between *T. domestica* and *D. melanogaster*. Then, to trace the evolution of the C-terminal portion, I reconstructed the ancestral sequences of Dsx and compared the exon-intron structure among insects. I then unveiled that the C-terminal motif became longer in the common ancestor of Aparaglossata. This result indicates that the emergence of the C-terminal motif correlates with the appearance of the function of *dsx* in female differentiation of morphogenesis and may be due to accumulating coding mutation rather than the emergence of the female-specific exon.

Overall, I infer the evolutionary history that the female-specific isoform of *dsx* was initially responsible for promoting *vitellogenin* transcription and later acquired its feminizing roles in morphogenesis via the elongation of its C-terminal region. Our model provides insight that isoforms gradually obtain roles opposite to an ancestral function and

that mutations to the C-terminal region might result in the recruitment of isoforms.

Acknowledgements

I would like to thank Prof. Teruyuki Niimi (National Institute for Basic Biology) for providing me the chance to research. I express my deep appreciation to Dr. Toshiya Ando Kyoto University), Dr. Taro Nakamura (National Institute for Basic Biology), Dr. Hiroki Sakai* (National Agriculture and Food Research Organization), Dr. Tatsuro Konagaya* (Nara University of Education), and Dr. Masaki Takenaka* (Shinshu University) for critical discussion on my study. I am also indebted to Dr. Shinichi Morita (National Institute for Basic Biology) for his technical advice on molecular experiments. I want to thank Dr. Takahiro Ohde* (Kyoto University) for teaching me how to keep *Thermobia domestica*. I appreciate the feedback offered by Prof. Mitsuyasu Hasebe (National Institute for Basic Biology), Prof. Shosei Yoshida (National Institute for Basic Biology), and Prof. Shuji Shigenobu (National Institute for Basic Biology) through the "Life Science Progress" program of the SOKENDAI. Computations were performed on the NIG supercomputer at ROIS, National Institute of Genetics, and the Data Integration and Analysis Facility, National Institute for Basic Biology. I thank the Model Plant Research Facility, NIBB Bioresource Center, for providing the network camera system. The Sasakawa Scientific Research Grant from The Japan Science Society supported this work.

*Their addresses are current ones. They helped me when they belonged to National Institute for Basic Biology.

Chikami, Yasuhiko

TABLE OF CONTENTS

Chapter 1

General Introduction

Abstract In this thesis, I serve a question on the exaptation in alternatively spliced isoforms. This question is how genes obtain new functions via alternative isoforms or exons. Some hypotheses have already been put forward, as the question is one of the long-standing issues regarding the evolution of alternative splicing. However, it is still under debate what the exaptive state was before the neo-functionalization and what changes in genomic space led to the new role of the isoform. To examine these questions, I investigate the evolutionary history of the *doublesex* gene, which encodes a transcriptional factor controlling arthropod sex differentiation. Here, I provide a brief history of the notion of "exaptation" at the molecular genetic level and some hypotheses about the exaptation of alternatively spliced isoforms. Finally, I bring out the *doublesex* gene as the research agenda to examine the evolution of alternative isoforms.

1.1. Brief history of exaptation at molecular level

"... structures thus indirectly gained, although at first of no advantage to a species, may subsequently have been taken advantage of by its modified descendants, under new conditions of life ..."

Darwin, 1872, p. 187

In 1872, the British naturalist and evolutionist Charles Richard Darwin tried to refute criticisms to his theory by some authors in the final edition of *On the Origin of Species*. Of his rebuttals, he focused primarily on the criticism to the gradual effects of natural selection by the British naturalist and anti-evolutionist St. George Jackson Mivart. In that statement, Darwin mentioned the above sentence and recognized that functional turnover in existing entities plays a crucial role in evolution (Gould 2002). In other words, evolutionary change is achieved by transforming existing structures without creating

structures *de novo*, i.e., without creating something from nothing. Although a similar idea is expressed in the first edition of *On the Origin of Species* (Darwin 1859), according to Gould (2002), Darwin came to enhance this idea following the criticism by Mivart.

According to McLennan (2008), the idea of functional transformation was named 'preadaptation' by the French biologist Lucien Claude Marie Julien Cuénot in the early 20th century. However, 'preadaptation' and Cuénot's ideas have often been criticized for their teleological connotations (McLennan 2008). Note that some authors, such as Ganfornina and Sanchez (1999) and Casinos (2017), argue that Cuénot's ideas did not include teleology. Then, in the second half of the 20th century, Gould and Vrba (1982) coined the term 'exaptation' to exclude the teleological aspect. However, it is difficult to say that 'preadaptation' has completely replaced 'exaptation' as 'preadaptation' is still frequently used. Note that, although these terms seem synonymous in essentialism, Chipman (2021) defined 'preadaptation' as a condition and 'exaptation' as a process. In this thesis, I use exaptation to express the condition and process to avoid confusion and a misleading interpretation.

Gould and Vrba (1982) defined 'exaptation' as: a trait, previously produced by the adaptation for a particular role, is coopted for a new use, or a trait that originated due not to the direct adaptation, is recruited for a current use. Based on the idea of the exaptation, it can provide a framework for the evolutionary process in which features that have some adaptive function in the environment or that do not have any function in the surrounding condition obtain a novel adaptive role for the new environment. Thus, the exaptation is common to the early stage of the adaptive evolution (Futuyma and Kirkpatrick 2017) and can explain the origin of currently observed adaptive traits. Note that the American paleontologist Stephen Jay Gould was likely to advocate the exaptation to rebuttal the Neo-Darwinism and the adaptationism (Dennett 1995). However, as Simpson (1944) had clearly mentioned, the exaptation (preadaptation in the original term) would not oppose the Neo-Darwinism, but rather complements it. To date, many biological theories have been developed by considering adaptation and exaptation together (Brosius 2019). One excellent example is the sensory exploitation theory (Smith et al. 2004) which explains the origin of mate choice by stickleback females. Thus, the exaptive process has been debated for a long time and has great significance for evolutionary biology, especially macroevolution (Gould 2002).

From the Darwin's notion, the exaptation has well explained complex phenotypic evolution (e.g., McLennan 2008, Futuyma and Kirkpatrick 2017, Alexander 2018, Shubin 2020). Furthermore, the concept of exaptation has been extended to the molecular genetic level. Its sign was already in Gould and Vrba (1982). In the beginning, the exaptation at the molecular genetic level was discussed in the context of the significance of junk DNA, focusing on the functionalization of transposable elements (e.g., Brosius et al. 1991, Brosius and Gould 1992). The integration of evolutionary biology, molecular biology, and developmental genetics at the end of the 20th century and the beginning of the 21st century led to significant progress. During this paradigm, the exaptive processes have been demonstrated in which existing genes or gene regulatory networks obtain new functions (McLennan 2008). This process, called the co-option or the recruitment, is a source of evolutionary innovation (e.g., True and Carroll 2002, Chipman 2010, Glassford et al. 2015).

Currently, there is no doubt that the exaptation has been seen at the genetic level. One issue arises here: the influence on existing (ancestor) function when a gene acquires a new role. Mutations that affect existing gene function are often detrimental and are expected to be subject to strong purifying selection (True and Carroll 2002). Therefore, in many cases, the genetic exaptation will be achieved by changes that avoid effects on existing gene function. How can genes avoid this functional constraint and acquire new roles? In the last 30 years, many hypotheses to this question have been put forward and demonstrated. Examples can be seen in changes in the spatiotemporal expression or dosage of genes due to mutations in cis-regulatory regions and the appearance of new gene copies due to gene duplication (True and Carroll 2002). Alternative splicing would also be the mechanism to explain the question. Some opinions have been asserted since the discovery for splicing. However, the study on recruitment by alternative splicing has remained understudied, compared with focused investigations of transcriptional regulation through gene duplication or diversification of cis-regulatory regions (Schaefke et al. 2018, Verta and Jacobs 2021). How did isoforms acquire new functions in the course of evolution? This question is of the central concern in this thesis.

1.2. Exaptation in alternatively spliced isoforms

Alternative splicing is the pervasive mechanism for processing pre-mRNAs in Eukaryota and underpins transcriptome diversity (Nilsen and Gravely 2010, Blencowe 2017). Alternatively spliced genes can be translated into various protein forms. If one isoform can exert roles of the gene required for cellular activity, other isoforms would be released from purifying selection due to the functional constraint (Keren et al. 2010). Hence, alternative isoforms would be remnants leading to the neo-functionalization of genes. The earliest suggestion for this opinion had already been represented in the essay in 1978 by Walter Gilbert written shortly after the discovery for pre-mRNA splicing from adenovirus (Berget et al. 1977, Chow et al. 1977). Gilbert (1978) looked upon alternative splicing as *"Evolution can seek new solutions without destroying the old"* and drew a new roadmap to neo-functionalization of genes, which, at that time, was regarded as being difficult to achieve unless gene duplication occurred.

Over the 45 years until today, evidence has accumulated that each alternatively spliced isoform serves a specific function and achieves functional diversity in the genome. For example, the *Drosophila* transmembrane protein gene *Dscam* can theoretically produce more than 38000 transcripts, each of which is involved in the recognition of infected antigens and self-recognition of axons during neurogenesis (Nilsen and Graveley 2010). Alternative splicing also gives rise to isoforms that act antagonistically to each other, as in the *ich-1* gene encoding one of the Caspases in the mouse, *Mus musculus* (Wang et al. 1994). There is also much evidence that alternative splicing correlates with organismal complexity (Nilsen and Graveley 2010, reviewed in Bush et al. 2017). In addition, alternatively spliced isoforms are differentially expressed between sexes (Telonis-Scott et al. 2009, Gómez-Redondo et al. 2021, Naftaly et al. 2021, Rogers et al. 2021), among populations (Graveley 2008), among species (Barbosa-Morais et al. 2012, Gueroussov et al. 2015), and among morphs in phenotypic plasticity (Grantham and Brisson 2018, Steward et al. 2022). These facts imply that alternative splicing contributes to phenotypic evolution through the neo-functionalization or loss of function of genes. Indeed, the emergence of alternative isoforms has been linked to evolutionary phenomena such as rapid adaptation in the jaw and dorsal spine morphology in stickleback species (Howes et al. 2017, Singh et al. 2017), the domestication of the sunflower, *Helianthus annuus* (Smith et al. 2018, 2021), caste differentiation in the honeybee, *Apis mellifera* (Lyko et

al. 2010), and niche differentiation in the human lice, *Pediculus humanus* (Tovar-Corona et al. 2015). Examples at present are listed in an excellent review (Verta and Jacobs 2021).

As described above, alternative splicing is likely to act as a mechanism leading to the neo-functionalization of genes. How, then, do alternative isoforms acquire a new role? As suggested by Gilbert (1978), one answer would be the acquisition of new exons or introns in the coding region, which would result in immediate changes in molecular function and may explain rapid adaptation (Singh and Ahi 2022). Further, many expressed isoforms in humans are non-functional or have no visible function (Nielsen and Gravely 2010, Baralle and Giudice 2017). There is also an abundance of noise isoforms that arise from splicing errors (Sorek et al. 2004, Pickrell 2010, Chen et al. 2012, Tress et al. 2017). Hence, additive hypotheses have been proposed that explain the exaptive process by which "non-functional" isoforms came into use in the last two decades.

Common to several hypotheses is that emerging isoforms and exons initially have weak fitness or neutrality and that these 'non-functional' isoforms later increase their fitness by acquiring new functions via accumulating mutations in coding regions (Boue et al. 2003, Xing and Lee 2006, Keren et al. 2010, Verta and Jacobs 2021). These hypotheses also share the assumption that existing roles of a gene are maintained in one isoform, while the effects of mutations that occur in other isoforms are masked in the cell until the function becomes adaptive. The difference between the hypotheses lies in the mechanisms that moderate the effects of the mutations. One hypothesis is that ancestral isoforms with pre-existing roles are highly expressed in the cell, while newly emerged isoforms are expressed much lower in the cell (Boue et al. 2003, Xing and Lee 2006, Kim et al. 2008, Keren et al. 2010, Chen et al. 2012). This ‘neutral’ hypothesis is supported by an experiment using mouse hybrid cells (Zou et al. 2022). On the other hand, the hypothesis has recently been put forward that both isoforms are highly expressed, but the non-sense mediated mRNA decay (NMD) inactivates isoforms without the pre-existing role (Verta and Jacobs 2021). The latter hypothesis regards the NMD as an evolutionary capacitor. The high expression of isoforms may also be interpreted in a balancer model where alternative splicing tunes up the balance between isoforms (Kim et al. 2008, Keren et al. 2010). According to the balancer model, alternative isoforms are merely a by-product of balancing isoforms. Note that, as the balancer model explains the significance

of splicing producing the non-functional isoforms and the NMD capacitor model shows the non-functionality of isoforms, they are not mutually exclusive hypotheses.

These theoretical predictions are supported by interspecific comparisons of exon-intron structure at the genomic level (e.g., Modrek and Lee 2003). If a non-functional or neutral isoform acquires a new function through a coding mutation, the isoform should retain its coding region and be maintained during evolution. In contrast, non-functional or neutral isoforms and exons are frequently replaced or lost among species and sometimes become non-sense ones with a very short coding region (Graveley 2001, Xing and Lee 2006, Kelemen et al. 2013). Hence, there seem to be contradictions regarding the functional evolutionary process of isoforms. This inconsistency might be resolved if there is a minor or cryptic function that is challenging to recognize as a phenotype. Seemingly non-functional or neutral isoforms may have some pre-existed function. However, it is still unresolved whether such functions exist and, if so, what nature of functions are involved.

Alternative splicing sometimes provides two opposite functions in a single gene. For example, a mammalian apoptosis-related gene, *Bcl-x*, produces two isoforms (Boise et al. 1993, Stevens and Oltean 2019). These isoforms are characterized by the inclusion or exclusion of 2nd exon. Then, the long isoform with the exon 2 encodes an anti-apoptotic protein and promotes cell survival. In contrast, the short isoform is translated into a pro-apoptotic protein and enhances cell death. Antagonistic roles through alternative splicing are found in many genes and may provide functional diversity in the protein space. If roles of isoforms evolved gradually, such that a novel, less adaptive isoform emerged and subsequently acquired a new, more adaptive function (Boue et al. 2003, Xing and Lee 2006, Keren et al. 2010, Verta and Jacobs 2021), the antagonism in a single gene might also gradually occurred. However, the evolutionary history of the antagonistic nature has not been investigated.

Moreover, according to the hypotheses, the functional evolution of isoforms results from increasing fitness due to coding mutations (Boue et al. 2003). While it is known that the acquisition of a new exon can lead to a new function by changing the amino acid sequence (e.g., Singh et al. 2017), it is not clear what mutations can lead to a new function for an existing isoform. Profiling of isoforms within a species suggests that differences in coding between isoforms are centered on terminal sequences and disordered regions rather than on solid structures such as protein domains (Lareu et al. 2004, Blujan et al.

2013, Reixachs-Solé and Eyras 2021). In addition, there are frequent differences between isoforms in sites responsible for protein-protein interactions (Ule and Blencowe 2019). Given these facts, it is likely that changes in coding regions occur in disordered regions and/or terminal sites. It is not evident whether such mutations correlate with the neo-functionalization of isoforms.

In this thesis, the general subjects are to examine the questions of 1) whether coding mutations cause neo-functionalization of "non-functional" isoforms, 2) what exaptive functions exist before the neo-functionalization, and 3) what kind of coding mutations result in the new roles. To answer these questions about the functional evolution of selective isoforms, the detection of the exaptation would be useful. The exaptive process can be investigated by the methodology used in systematics, which maps information in extant organisms onto phylogenetic relationships and infers their evolutionary history (Gould 2002, McLennan 2008). Applying this methodology to the exaptation of isoforms would not have been easy, as it requires information on functions of alternatively spliced genes in broad taxa. In this thesis, I focus on functional diversity in the *doublesex* (*dsx*) gene as a phenomenon that can resolve this problem.

1.3. *doublesex* gene as model for evolution of alternative splicing

dsx gene encodes a transcriptional factor belonging to the Doublesex and Mab-3 Related Transcriptional factor (DMRT) family (Mawaribuchi et al. 2019). This gene is pervasive among almost all arthropods and acts as a global regulator at the bottom of the genetic cascade to govern sexual differentiation (Kopp 2012, Verhulst and van de Zande 2015) (Figure 1.1). In Pterygota, *dsx* is controlled by sex-specific splicing and produces male- and female-specific proteins (e.g., Burtis and Baker 1989, Wexler et al. 2019). In Holometabola, the sex-specific Dsx proteins are essential for promoting either male or female determination/differentiation (e.g., Hildreth 1965, Burtis and Baker 1989, Ohbayashi et al. 2001, Kijimoto et al. 2012, Shukla and Palli 2012b, Ito et al. 2013, Gotoh et al. 2016, Xu et al. 2017). For example, in the fruit fly *Drosophila melanogaster*, *dsx* is required to realize sex differences in external genitalia and foreleg bristle rows, while *dsx* mutants show an intersexual phenotype in these traits because both male and female differentiation are inhibited (Hildreth and Lucchesi 1963, Hildreth 1965).

The sex-antagonistic nature of *dsx* has been common sense since the discovery of *dsx* mutants in *D. melanogaster* (Hildreth and Lucchesi 1963) for about 50 years. However, for the last five years, the conventional wisdom has been broken since it is reported from hemimetabola and Hymenoptera species that *dsx* has the sex-specific splicing but contributes to only male differentiation (Mine et al. 2017, 2021, Guo et al. 2018, Zhuo et al. 2018, Wexler et al. 2019, Takahashi et al. 2019, 2021). These breakthroughs show that *dsx* of extant insects has diversity in function of its female-specific isoform for female differentiation, i.e., non-functional or functional isoforms. This functional diversity suggests that *dsx* might experience a neo-functionalization in the course of evolution. Hence, I expect that the function of *dsx* isoform will provide a model to empirically examine the hypothesis on the neo-functionalization of alternative isoforms. However, the functional evolution of *dsx* has still been ambiguous because of poor information on

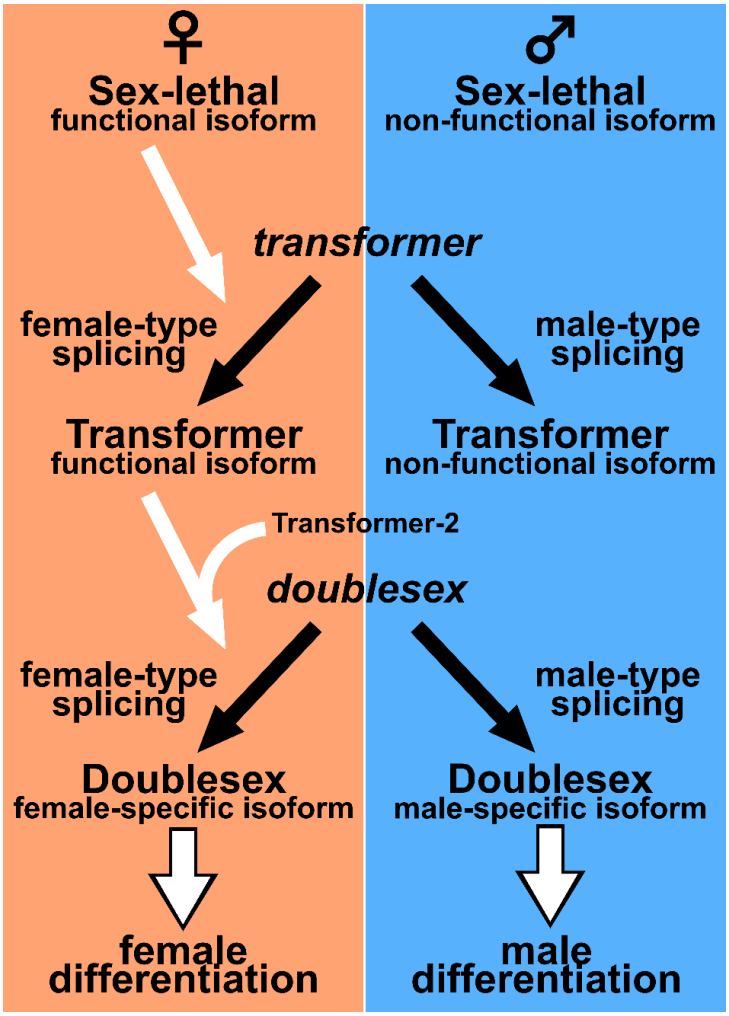


Figure 1.1. Schematic image of sex-determining cascade of *Drosophila melanogaster*.

its phylogenetic relationship and its features in apterygote species and lack of its sequence comparison through insect orders (detailed in the following chapters).

In this thesis, the specific subject is to infer the evolutionary history of the function of *dsx* in Insecta. The central questions in this thesis are: does the sex-specific splicing of *dsx* have a single origin, did the non-functional isoform of *dsx* come into use, and, if so, what a change linked to the neo-functionalization. To this end, I investigate its phylogenetic relationship containing apterygote insects and its sex-specific isoforms in the apterygote, *Thermobia domestica* (Zygentoma) (Chapter 2). Zygentoma is the sister clade of Pterygota (Misof et al. 2014: Figure 1.2). Then, I look into the function of *dsx* in *T. domestica* for sexually dimorphic morphogenesis using nymphal RNA interference (RNAi) assays (Chapter 3). Finally, I infer the sequence evolution of the female-specific motif of *dsx* (Chapter 4). In the final chapter, the implication of my results for the exaptive process in alternative isoforms is discussed (Chapter 5).

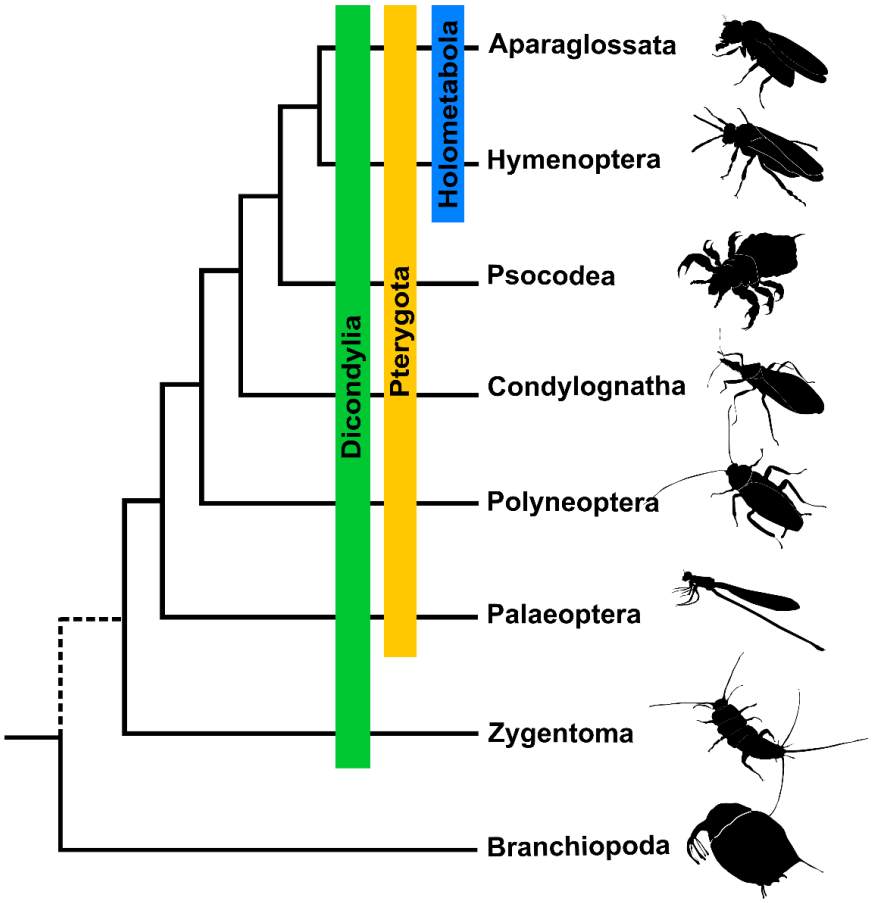


Figure 1.2. Phylogenetic relationship in Insecta. The topology is based on Misof et al. (2014).

Chapter 2

Gene Duplication and Sex-specific Splicing of *dsx*

Abstract The evolution of alternative splicing is generally linked to gene duplication. Recently, the presence of paralogs has been proposed in *doublesex* of Hexapoda. Inferring the evolutionary process of the gene duplication and sex-specific splicing of *doublesex* requires information on the phylogenetic relationship between *doublesex* and its paralogs and the splicing nature of *doublesex* in species that retain both genes. To this end, I investigate the phylogenetic relationship of *doublesex* homologs in insects and the sex-specific splicing of *doublesex* in the apterygote species, *Thermobia domestica* (Zygentoma). Here, I show two copies of *doublesex* in Zygentoma and some pterygote orders and that the paralog locates on the lineage different from crustacean *doublesex*. This result infers that the *doublesex* paralog in insects occurred independently of the duplication of crustacean *doublesex* and before the appearance of the common ancestor of Dicondylia (= Zygentoma + Pterygota). I also find the sex-specific isoforms of *doublesex* and the single form of its paralog in *T. domestica*. These findings suggest the coexist of the sex-specific isoforms and the paralog of *doublesex* during insect evolution. Hence, the loss of *dsx-like* occurred after acquiring the sex-specific isoforms of *dsx*. *dsx-like* and the isoforms of *dsx* might play cooperative or redundant roles in sexual traits. My result further emphasizes that *T. domestica* is a suitable species to investigate the evolutionary history of *dsx*.

2.1. Introduction

Alternative splicing underlies transcriptome diversity and can produce various proteins from one gene (Bush et al. 2017). Similarly, transcriptome and proteome diversity are brought by gene duplication (e.g., True and Carroll 2002, Taylor and Raes 2004). Since both alternative splicing and gene duplication are responsible for functional diversity in

genomic space and organism complexities (Nilsen and Graveley 2010), there has long been debated regarding the relationship between these events. For example, the size of gene families that indicates the frequency of gene duplication is negatively correlated with the number of alternative splicing events (Kopelman et al. 2005, Su et al. 2006, Talavera et al. 2007, reviewed in Bush et al. 2017). Furthermore, the relationship between these events can be seen in functional evolution in splicing isoform and paralog (reviewed in Bush et al. 2017). For instance, a sub-functionalization model has been proposed in which paralogs resulting from gene duplication imitate the function of alternatively spliced isoforms in single-copied genes of ancestors (Bush et al. 2017). This model states that new paralogues are retained by sub-functionalization. Examples of this functional interplay can be seen in the neuromuscular gene *troponin I* in the vase tunica, *Ciona intestinalis* (MacLean et al. 1997), and the transcription factor gene *mitf* in bony fish (Altschmied 2002). Given these inter-relationships between alternative splicing and gene duplication, an accurate understanding of functional evolution in splicing isoforms will require tracing their evolutionary history together with gene duplication.

The *doublesex* (*dsx*) gene is a member of the Doublesex and Mab-3 Related Transcriptional factors (DMRT) family, which is essential for the arthropod sex determination and sex differentiation. Currently, evidence of the sex-specific splicing of the *dsx* has been accumulated from Diptera, Lepidoptera, Coleoptera, Hymenoptera, Hemiptera, Dictyoptera, and Odonata (e.g., Burtis and Baker 1989, Ohbayashi et al. 2001, Kijimoto et al. 2012, Shukla and Palli 2012b, Ito et al. 2013, Gotoh et al. 2016, Xu et al. 2017, Mine et al. 2017, 2021, Zhuo et al. 2018, Wexler et al. 2019, Takahashi et al. 2019, 2021). In contrast, the *dsx* of branchiopods and a chelicerate is controlled by the male-specific expression (Kato et al. 2011, Pomerantz et al. 2015). Thus, the sex-specific isoforms may have appeared from the last common ancestor of Branchiopoda and Pterygota until the divergence of the Pterygota (Kato et al. 2011, Wexler et al. 2019). Further, a recent sequence comparison of all hexapod orders (Price et al. 2015) exhibited alternatively spliced transcripts in the most basal lineage of hexapods and showed evidence of two copies of the *dsx* in insects such as Zygentoma and Ephemeroptera. Paralogs of the *dsx* are present in branchiopodan and cheliceracean *dsx* (Kato et al. 2011, Pomerantz et al. 2015). Hence, these results support that gene duplication in

the *dsx* occurred before the appearance of hexapod ancestors and was lost in the derived taxa (Price et al. 2015).

Moreover, the previous results suggest the emergence of the sex-specific splicing of *dsx* coupled with mutation and loss of its paralog (Price et al. 2015). However, it remains unclear when the paralog and sex-specific isoforms of the *dsx* occurred. There could be three alternative possibilities on the timing of the *dsx* duplication. The *dsx* paralogs occurred before the emergence of the last common ancestor of Branchiopod and Hexapoda, after the common ancestor of the two taxa, or independently at each lineage. Understanding the phylogenetic relationship of the *dsx* and its paralogs is necessary to examine these hypotheses. Previous phylogenetic analyses of the DMRT family (e.g., Volff et al. 2003, Wexler et al. 2014, 2019, Mawaribuchi et al. 2019) have focused on Pterygota, crustaceans, and Chelicerata. They have not included sequences of the DMRT family genes from insect species that retain the *dsx* paralogs. Further, the insight into the splicing profiles of the *dsx* in the apterygote insects would be essential for investigating the timing of the advent of the sex-specific splicing. Since the previous studies focused on the sex-specific splicing of the *dsx* in the winged insects, information from the apterygote insects is absent. Thus, it is still unresolved whether the sex-specific isoforms of the *dsx* exist in hexapods other than Pterygota. In particular, understanding the expression profiles of the *dsx* and its paralog would help examine the relation between the sex-specific isoforms and the paralogs of the *dsx*.

In this chapter, I am subject to infer the evolutionary process of the gene duplication and the sex-specific splicing of *dsx*. To this end, I analyzed the phylogenetic relationship of *dsx* of Hexapoda, including almost all hexapod orders, and searched for the sex-specific isoforms of *dsx* in the firebrat, *Thermobia domestica*, belonging to the sister group of Pterygota, Zygentoma. According to Price et al. (2015), Zygentoma has both *dsx* and its paralog. In addition, based on the results of this chapter, I suggested *T. domestica* belonging to Zygentoma as a suitable model for elucidating the evolutionary history of *dsx*.

2.2. Materials and Methods

2.2.1. Animals

The firebrat, *Thermobia domestica* (Packard 1873), was used as an emerging model for apterygote. *T. domestica* is one of the species belonging to Zygentoma (Lepismatidae). The insects were kept at 37°C in total darkness condition and fed with fish food (TetraFin Goldfish Flakes, Tetra GmbH, Melle, Germany) in my laboratory. Stock colonies were reared in plastic cases of 30 cm×40 cm or 18 cm × 25 cm in length. Eggs were collected from tissue paper in the case and incubated at 37°C.

2.2.2. Molecular phylogenetic analysis

Dsx is a member of the Doublesex and Mab-3 Related transcriptional factors (DMRT) family, and has a DNA binding domain, Doublesex and Mab-3 (DM) domain. Pancrustacea generally has four DMRT family genes, Dsx, Dmrt11, Dmrt93B, and Dmrt99B (Mawaribuchi et al. 2019). Phylogenetic analysis of the *dsx* homologs was performed using the amino acid sequences of the DM domain. I used *dsx* sequences of *D. melanogaster* as a query and obtained 166 metazoan DMRT family proteins from the NCBI and the i5k databases (https://i5k.nal.usda.gov/) and the genome data of *T. domestica* by the BLAST analysis (listed in Table 2.1). I then aligned the sequences using MAFFT version 7 (Katoh and Standley 2013) with the -linsi option to use an accuracy option, L-INS-I, and manually extracted the DM domain, which consisted of 65 amino acids (Figure 2.1). Molecular phylogenetic analysis of the aligned sequences was performed using a maximum likelihood method with the IQ-TREE software (Minh et al. 2020). The substitution model was selected by -MPF and -AIC option of the IQ-TREE. The best-fit model was the WAG+I+G4. The proportion of invariable sites is 0.0645, and the Gamma shape alpha was 0.6664. I used the ultrafast bootstrap (UFBoot) and Simodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) to evaluate the branch reliability according to the manufacturer's recommendation (http://www.iqtree.org/doc/Frequently-Asked-Questions#how-do-i-interpret-ultrafast-bootstrap-ufboot-support-values). I set 1000 replications in each test. Typically, the branch with the SH-aLRT ≥ 80% and the UFboot ≥ 95% would be reliable (Guindon et al. 2010, Minh et al. 2013). Therefore, I regarded the branch with both support values of more than the thresholds as the reliable clade.

Table 2.1. List of taxa used for molecular phylogenetic analysis of DMRT family.

OTU name	gene	accession number	species	Phylum	Subphylum	Class	Order	genome region
Aage_Dsx	dsx	XP_041968930.1	*Aricia agestis*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Aasp_Dsx-like	*dsx-like*	GAZQ02010078.1	*Aretaon asperrimus*	Arthropoda	Hexapoda	Ectognatha	Phasmatodea	
Acha_Dsx	*dsx*	GAUW02033438.1	*Apachyus charteceus*	Arthropoda	Hexapoda	Ectognatha	Dermaptera	
Afra_Dsx3	*dsx3*	AWC26109.1	*Artemia franciscana*	Arthropoda	Crustacea	Brachiopoda	Anostraca	
Afra_Dsx4	*dsx4*	AWC26111.1	*Artemia franciscana*	Arthropoda	Crustacea	Brachiopoda	Anostraca	
Afus_Dmrt11E	*dmrt11E*	CAG7816593.1	*Allacma fusca*	Arthropoda	Hexapoda	Entognatha	Collembola	
Afus_Dsx	*dsx*	CAG7821548.1	*Allacma fusca*	Arthropoda	Hexapoda	Entognatha	Collembola	
Afus_Dmrt99B	*dmrt99B*	CAG7826582.1	*Allacma fusca*	Arthropoda	Hexapoda	Entognatha	Collembola	
Amel_Dsx	*dsx*	NP_001104725.1	*Apis mellifera*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera	
Annu_Dmrt99B	*dmrt99B*	GATX01081132.1	*Annulipalpia* sp.	Arthropoda	Hexapoda	Ectognatha	Trichoptera	
Annu_Dsx	*dsx*	GATX01084595.1	*Annulipalpia* sp.	Arthropoda	Hexapoda	Ectognatha	Trichoptera	
Aros_Dsx	*dsx*	XP_012262263.1	*Athalia rosae*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera	
Atub_Dsx	*dsx*	ATE86739.1	*Asobara tabida*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera	
Baet_Dsx-like	*dsx-like*	GATU02014641.1	*Baetis* sp.	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	
Bdor_Dsx	*dsx*	AAB99948.1	*Bactrocera dorsalis*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Bger_Dsx	*dsx*	PSN43312.1	*Blattella germanica*	Arthropoda	Hexapoda	Ectognatha	Dictyoptera	
Bhye_Dsx	*dsx*	GAYK02032082.1	*Boreus hyemalis*	Arthropoda	Hexapoda	Ectognatha	Mecoptera	
Bmor_Dmrt11E	*dmrt11E*	XP_004930266.1	*Bombyx mori*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Bmor_Dmrt93B	*dmrt93B*	XP_004932028.3	*Bombyx mori*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Bmor_Dmrt99B	*dmrt99B*	XP_004924389.2	*Bombyx mori*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Bmor_Dsx	*dsx*	XP_012544211.1	*Bombyx mori*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Bmut_Dmrt1	*dmrt1*	ELR53308.1	*Bos mutus*	Chordata	Vertebrata	Mammalia	Cetartiodactyla	
Bpyr_Dsx	*dsx*	XP_043582639.1	*Bombus pyrosoma*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera	
Btau_Dmrt2	*dmrt2*	XP_005210039.1	*Bos taurus*	Chordata	Vertebrata	Mammalia	Cetartiodactyla	
Btry_Dsx	*dsx*	AAV85890.1	*Bactrocera tryoni*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Caqu_Dmrt99B	*dmrt99B*	CAQU003464-RA	*Catajapyx aquilonaris*	Arthropoda	Hexapoda	Entognatha	Diplura	
Caqu_Dmrt93B	*dmrt93B*	CAQU000591-RA	*Catajapyx aquilonaris*	Arthropoda	Hexapoda	Entognatha	Diplura	
Caqu_Dmrt11E	*dmrt11E*	CAQU006176-RA	*Catajapyx aquilonaris*	Arthropoda	Hexapoda	Entognatha	Diplura	
Caqu_Dsx	*dsx*	CAQU003748-RA	*Catajapyx aquilonaris*	Arthropoda	Hexapoda	Entognatha	Diplura	
Ccap_Dsx	*dsx*	XP_012158607.1	*Ceratitis capitata*	Arthropoda	Hexapoda	Entognatha	Diptera	
Ccor_Dsx	*dsx*	GATG02018436.1	*Corydalus cornutus*	Arthropoda	Hexapoda	Ectognatha	Megaloptera	
Cdip_Dsx	*dsx*	CAB3378992.1	*Cloeon dipterum*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	

Cdip_Dsx2	*dsx*	CAB3378996.1	*Cloeon dipterum*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	
Cdip_Dsx-like	*dsx-like*	CAB3366989.1	*Cloeon dipterum*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	
Ceut_Dmrt11E	*dmrt11E*	GAUX02031275.1	*Ceuthophilus* sp.	Arthropoda	Hexapoda	Ectognatha	Orthoptera	
Cfel_Dsx	*dsx*	GAYP02016500.1	*Ctenocephalides felis*	Arthropoda	Hexapoda	Ectognatha	Siphonaptera	
Cgal_Dsx	*dsx*	GAWK02011923.1	*Ceratophyllus gallinae*	Arthropoda	Hexapoda	Ectognatha	Siphonaptera	
Cgig_Dsx	*dsx*	XP_043258119.1	*Colletes gigas*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera	
Choo_Dmrt11E	*dmrt11E*	NQII01002646.1	*Clitarchus hookeri*	Arthropoda	Hexapoda	Ectognatha	Phasmatodea	
Choo_Dsx-like	*dsx-like*	NQII01000109.1	*Clitarchus hookeri*	Arthropoda	Hexapoda	Ectognatha	Phasmatodea	
Clec_Dmrt99B	*dmrt99B*	XP_014246101.1	*Cimex lectularius*	Arthropoda	Hexapoda	Ectognatha	Hemiptera	
Cpun_Dsx	*dsx*	-	*Cryptocercus punctulatus*	Arthropoda	Hexapoda	Ectognatha	Dictyoptera	
Ctos_Dsx	*dsx*	GIEL01051748.1	*Cultus tostonus*	Arthropoda	Hexapoda	Ectognatha	Plecoptera	
Dcar_Dsx1	*dsx1*	AIL86779.1	*Daphnia carina*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dcar_Dsx2	*dsx2*	AIL86780.1	*Daphnia carina*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dgal_Dsx1	*dsx1*	BAM33609.1	*Daphnia galeata*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dgal_Dsx2	*dsx2*	BAM33610.1	*Daphnia galeata*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dipl_Dsx	*dsx*	GDCS01037195.1	*Diplatys* sp.	Arthropoda	Hexapoda	Ectognatha	Dermaptera	
Dmag_Dmrt11E	*dmrt11e*	BAG12871.1	*Daphnia magna*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dmag_Dmrt93B	*dmrt93b*	BAG12872.1	*Daphnia magna*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dmag_Dmrt99B	*dmrt99b*	BAG12873.1	*Daphnia magna*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dmag_Dsx1	*dsx1*	BAJ78307.1	*Daphnia magna*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dmag_Dsx2	*dsx2*	BAJ78309.1	*Daphnia magna*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dmel_Dmrt11E	*dmrt11e*	NP_511146.2	*Drosophila melanogaster*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Dmel_Dmrt93B	*dmrt93b*	NP_524428.1	*Drosophila melanogaster*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Dmel_Dmrt99B	*dmrt99b*	NP_524549.1	*Drosophila melanogaster*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Dmel_Dsx	*dsx*	NP_731197.1	*Drosophila melanogaster*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Dpul_Dsx1	*dsx1*	AGJ52190.1	*Daphnia pulex*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Dpul_Dsx2	*dsx2*	BAM33608.1	*Daphnia pulex*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Drer_Dmrt1	*dmrt1*	AAQ04555.1	*Danio rerio*	Chordata	Vertebrata	Actinopterygii	Cypriniformes	
Drer_Dmrt2	*dmrt2*	NP_571027.1	*Danio rerio*	Chordata	Vertebrata	Actinopterygii	Cypriniformes	
Eaff_DM		XP_023327398.1	*Eurytemora affinis*	Arthropoda	Crustacea	Copepoda	Calanoida	
Edan_dmrt11E	*dmrt11E*	EDAN008414-RA	*Ephemera danica*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	
Edan_dmrt93B	*dmrt93B*	EDAN004527-RA	*Ephemera danica*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	
Edan_dmrt99B	*dmrt99B*	EDAN010669-RA	*Ephemera danica*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	
Edan_Dsx	*dsx*	KAF4518127.1	*Ephemera danica*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	
Edan_Dsx-like	*dsx-like*		*Ephemera danica*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera	ephdan_Scaffold23

Eins_Dsx-like	*dsx-like*	GCCL01024227.1	*Ecdyonurus insignis*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera
Ekue_Dsx	*dsx*	CAG7465062.1	*Ephestia kuehniella*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera
Emex_Dmrt99B	*dmrt99B*	XP_017764520.0	*Eufriesea mexicana*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera
Emex_Dmrt93B	*dmrt93B*	XP_017764520.1	*Eufriesea mexicana*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera
Emex_Dsx	*dsx*	XP_017755508.1	*Eufriesea mexicana*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera
Enos_Dsx	*dsx*	GAXW02019001.1	*Euroleon nostras*	Arthropoda	Hexapoda	Ectognatha	Neuroptera
Epen_Dsx	*dsx*	GAWT02033840.1	*Empusa pennata*	Arthropoda	Hexapoda	Ectognatha	Mantodea
Esin_Dmrt99B	*dmrt99B*	ADH15934.1	*Eriocheir sinensis*	Arthropoda	Crustacea	Malacostraca	Decapoda
Esup_Dsx	*dsx*	GAVW02000373.1	*Epiophlebia superstes*	Arthropoda	Hexapoda	Ectognatha	Odonata
Eury_Dmrt11E	*dmrt11E*	GAZG02011227.1	*Eurylophella* sp.	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera
Eury_Dsx-like	*dsx-like*	GAZG02000044.1	*Eurylophella* sp.	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera
Faur_Dsx	*dsx*	GAYQ02045354.1	*Forficula auricularia*	Arthropoda	Hexapoda	Ectognatha	Dermaptera
Faur_Dsx2	*dsx*	GAYQ02026502.1	*Forficula auricularia*	Arthropoda	Hexapoda	Ectognatha	Dermaptera
Focc_Dsx	*dsx*	FOCC007514-RA	*Frankliniella occidentalis*	Arthropoda	Hexapoda	Ectognatha	Thysanoptera
Fvar_Dmrt99B	*dmrt99B*	XP_043525451.1	*Frieseomelitta varia*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera
Fvar_Dmrt93B	*dmrt93B*	XP_043519811.1	*Frieseomelitta varia*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera
Fvar_Dsx	*dsx*	QEK21873.1	*Frieseomelitta varia*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera
Gbue_Dmrt99B	*dmrt99B*	GBUE000192-RA	*Gerris_buenoi*	Arthropoda	Hexapoda	Ectognatha	Hemiptera
Gcor_Dsx	*dsx*	BAW32683.1	*Gnatocerus cornutus*	Arthropoda	Hexapoda	Ectognatha	Coleoptera
Gdja_Dmrt11E	*dmrt11E*	GDUY01002249.1	*Galloisiana yezoensis*	Arthropoda	Hexapoda	Ectognatha	Grylloblattodea
Gmar_Dsx	*dsx*	GCPI01026896.1	*Gonolabis marginalis*	Arthropoda	Hexapoda	Ectognatha	Dermaptera
Gmar_Dsx2	*dsx*	GCPI01021557.1	*Gonolabis marginalis*	Arthropoda	Hexapoda	Ectognatha	Dermaptera
Gnip_Dsx	*dsx*	GDWI01045959.1	*Galloisiana nipponensis*	Arthropoda	Hexapoda	Ectognatha	Grylloblattodea
Gryl_Dsx	*dsx*	GINB01025122.1	*Grylloblatta* sp.	Arthropoda	Hexapoda	Ectognatha	Grylloblattodea
Harm_Dsx	*dsx*	XP_021192052.1	*Helicoverpa armigera*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera
Hdeu_Dsx	*dsx*	maker-scaffold37size976698-augustus-gene-4.5-mRNA-1	*Holacanthella duospinosa*	Arthropoda	Hexapoda	Entognatha	Collembola
Hdeu_Dmrt93B	*dmrt93B*	maker-scaffold72size687670-augustus-gene-5.10-mRNA-1	*Holacanthella duospinosa*	Arthropoda	Hexapoda	Entognatha	Collembola
Hpal_Dsx	*dsx*	GAZA02093017.1	*Haploembia palaui*	Arthropoda	Hexapoda	Ectognatha	Embioptera
Hvit_Dsx	*dsx*	XP_046674253.1	*Homalodisca vitripennis*	Arthropoda	Hexapoda	Ectognatha	Hemiptera
Hydr_Dsx	*dsx*	GAVM02014074.1	*Hydroptila* sp.	Arthropoda	Hexapoda	Ectognatha	Trichoptera
Ibic_Dsx-like	*dsx-like*	GAXA02007870.1	*Isonychia bicolor*	Arthropoda	Hexapoda	Ectognatha	Ephemeroptera
Icra_Dsx	*dsx*	GAZH02011000.1	*Inocellia crassicornis*	Arthropoda	Hexapoda	Ectognatha	Raphidioptera

Isen_Dsx	*dsx*	-	*Ischnura senegarensis*	Arthropoda	Hexapoda	Ectognatha	Odonata	
Kbie_Dsx	*dsx*	GINP01105830.1	*Karoophasma biedouwense*	Arthropoda	Hexapoda	Ectognatha	Mantophasmatodea	
Kbie_Dmrt11E	*dmrt11E*	GINP01153601.1	*Karoophasma biedouwense*	Arthropoda	Hexapoda	Ectognatha	Mantophasmatodea	
Lcup_Dmrt11E	*dmrt11E*	XP_023291847.1	*Lucilia cuprina*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Lcup_Dmrt93B	*dmrt93B*	XP_023302612.1	*Lucilia cuprina*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Lcup_Dmrt99B	*dmrt99B*	XP_023308885.1	*Lucilia cuprina*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Lcup_Dsx	*dsx*	ADG37648.1	*Lucilia cuprina*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Lful_Dsx	*dsx*	LFUL018497-RA	*Ladona fulva*	Arthropoda	Hexapoda	Ectognatha	Odonata	
Lmig_Dsx	*dsx*		*Locusta migratoria*	Arthropoda	Hexapoda	Ectognatha	Orthoptera	scaffold3427
Lstri_DM		RZF46947.1	*Laodelphax striatellus*	Arthropoda	Hexapoda	Ectognatha	Hemiptera	
Mdom_Dmrt11E	*dmrt11E*	XP_019890834.1	*Musca domestica*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Mdom_Dmrt99B	*dmrt99B*	XP_005186857.1	*Musca domestica*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Mdom_Dsx	*dsx*	AAR23813.1	*Musca domestica*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Mext_Dmrt99B	*dmrt99B*	Medex_00095964-RA	*Medauroidea extradentata*	Arthropoda	Hexapoda	Ectognatha	Phasmatodea	
Mext_Dsx	*dsx*		*Medauroidea extradentata*	Arthropoda	Hexapoda	Ectognatha	Phasmatodea	PNEQ01023967.1 [32614-32324]
Mext_Dsx-like	*dsx-like*	Medex_00099178-RA	*Medauroidea extradentata*	Arthropoda	Hexapoda	Ectognatha	Phasmatodea	PNEQ01097711.1 [2988-3275]
Mfas_Dsx	*dsx*	GCNI01018035.1	*Meroplius fasciculatus*	Arthropoda	Hexapoda	Ectognatha	Diptera	
Mmac_Dsx	*dsx*	BAM33613.1	*Moina macropaeneus*	Arthropoda	Crustacea	Brachiopoda	Diplostraca	
Mmol_Dsx	*dsx*	JP074048.1	*Mengenilla moldrzyki*	Arthropoda	Hexapoda	Ectognatha	Strepsiptera	
Mmol_Dmrt99B	*dmrt99B*	JP103704.1	*Mengenilla moldrzyki*	Arthropoda	Hexapoda	Ectognatha	Strepsiptera	
Mmus_Dmrt1	*dmrt1*	AAO41736.1	*Mus musculus*	Chordata	Vertebrata	Mammalia	Rodentia	
Mrel_Dsx	*dsx*	GASW02021994.1	*Mantis religiosa*	Arthropoda	Hexapoda	Ectognatha	Mantodea	
Msex_Dsx	*dsx*	XP_037293921.1	*Manduca sexta*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Mviol_Dsx	*dsx*	GATA02010186.1	*Meloe violaceus*	Arthropoda	Hexapoda	Ectognatha	Coleoptera	
Nlec_Dsx	*dsx*	GEES01058869.1	*Neodiprion lecontei*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera	
Obru_dsx	*dsx*	KOB69684.1	*Operophtera brumata*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Ofur_Dsx	*dsx*	AHF81635.1	*Ostrinia furnacalis*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Olig_Dsx	*dsx*	XP_034176725.1	*Osmia lignaria*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera	
Osca_Dsx	*dsx*	BAJ25850.1	*Ostrinia scapulalis*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Otau_Dsx	*dsx*	AEX92938.1	*Onthophagus taurus*	Arthropoda	Hexapoda	Ectognatha	Coleoptera	
Paeg_Dsx	*dsx*	XP_039758070.1	*Pararge aegeria*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Pcat_Dsx	*dsx*	GDBY01045014.1	*Ptilocerembia catherinae*	Arthropoda	Hexapoda	Ectognatha	Embioptera	
Pcat_Dsx2	*dsx*	GDBY01045015.1	*Ptilocerembia catherinae*	Arthropoda	Hexapoda	Ectognatha	Embioptera	
Phum_Dsx	*dsx*	MK919539.1	*Pediculus humanus*	Arthropoda	Hexapoda	Ectognatha	Psocodea	

Pmac_Dsx	*dsx*	*XP_014372190.1*	*Papilio machaon*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Ppra_Dsx	*dsx*	GAVV02027199.1	*Pseudomallada prasinus*	Arthropoda	Hexapoda	Ectognatha	Neuroptera	
Psch_Dsx-like	*dsx-like*	GAWJ02028457.1	*Peruphasma schultei*	Arthropoda	Hexapoda	Ectognatha	Phasmatodea	
Ptar_Dsx	*dsx*	GHPW01032505.1	*Peltoperla tarteri*	Arthropoda	Hexapoda	Ectognatha	Plecoptera	
Pxyl_Dsx	*dsx*	XP_037963445.1	*Plutella xylostella*	Arthropoda	Hexapoda	Ectognatha	Lepidoptera	
Rnub_Dsx	*dsx*	GGRG01005123.1	*Rhyacophila nubila*	Arthropoda	Hexapoda	Ectognatha	Trichoptera	
Rpro_Dsx	*dsx*	QGB21099	*Rhodnius prolixus*	Arthropoda	Hexapoda	Ectognatha	Hemiptera	
Rvir_Dmrt11E	*dmrt11E*	GDBX01015771.1	*Rhagadochir virgo*	Arthropoda	Hexapoda	Ectognatha	Embioptera	
Same_Dsx	*dsx*	GHQR01005598.1	*Skwala americana*	Arthropoda	Hexapoda	Ectognatha	Plecoptera	
Smel_Dsx	*dsx*	GAZM02017191.1	*Stylops melittae*	Arthropoda	Hexapoda	Ectognatha	Strepsiptera	
Spar_Dmrt99B	*dmrt99B*	QJD20741.1	*Scylla paramamosain*	Arthropoda	Crustacea	Malacostraca	Dacapoda	
Tcal_DM		TRY61712.1	*Tigriopus californicus*	Arthropoda	Crustacea	Copepoda	Harpacticoida	
Tcas_Dmrt93B	*dmrt93B*	XP_008199135.1	*Tribolium castaneum*	Arthropoda	Hexapoda	Ectognatha	Coleoptera	
Tcas_Dmrt99B	*dmrt99B*	XP_975675.1	*Tribolium castaneum*	Arthropoda	Hexapoda	Ectognatha	Coleoptera	
Tcas_Dsx	*dsx*	NP_001345539.1	*Tribolium castaneum*	Arthropoda	Hexapoda	Ectognatha	Coleoptera	
Tcri_Dsx-like	*dsx-like*	GAVX02010884.1	*Timema cristinae*	Arthropoda	Hexapoda	Ectognatha	Phasmatodea	
Tdic_Dsx	*dsx*	BAM93340.1	*Trypoxylus dichotomus*	Arthropoda	Hexapoda	Ectognatha	Coleoptera	
Tdom_Dmrt11E	*dmrt11E*	this study	*Thermobia domestica*	Arthropoda	Hexapoda	Ectognatha	Zygentoma	scaffold42162_cov39 [159031-158906]
Tdom_Dmrt93B	*dmrt93B*	this study	*Thermobia domestica*	Arthropoda	Hexapoda	Ectognatha	Zygentoma	scaffold1624327_cov47 [121390-121265]
Tdom_Dmrt99B	*dmrt99B*	this study	*Thermobia domestica*	Arthropoda	Hexapoda	Ectognatha	Zygentoma	scaffold21840_cov24 [214972-214787]
Tdom_Dsx	*dsx*	this study	*Thermobia domestica*	Arthropoda	Hexapoda	Ectognatha	Zygentoma	
Tdom_Dsx-like	*dsx-like*	this study	*Thermobia domestica*	Arthropoda	Hexapoda	Ectognatha	Zygentoma	scaffold27567_cov49 [365805-281362]
Tger_Dsx-like	*dsx-like*	GASO02037568.1	*Tricholepidion gertschi*	Arthropoda	Hexapoda	Ectognatha	Zygentoma	
Tgla_Dsx		GDVU01042054.1	*Tyrannophasma gladiator*	Arthropoda	Hexapoda	Ectognatha	Mantophasmatodea	
Tpal_Dmrt99B	*dmrt99B*	XP_034232916.1	*Thrips palmi*	Arthropoda	Hexapoda	Ectognatha	Thysanoptera	
Tpal_Dsx	*dsx*	XP_034237507.1	*Thrips palmi*	Arthropoda	Hexapoda	Ectognatha	Thysanoptera	
Tsub_Dsx	*dsx*	GASQ02027559.1	*Tetrix subulata*	Arthropoda	Hexapoda	Ectognatha	Orthoptera	
Xalp_Dsx	*dsx*	GBVH01020704.1	*Xyela alpigena*	Arthropoda	Hexapoda	Ectognatha	Hymenoptera	
Xant_Dsx	*dsx*	GAUI02048130.1	*Xanthostigma* sp.	Arthropoda	Hexapoda	Ectognatha	Raphidioptera	
Xlae_Dmrt1	*dmrt1*	NP_001089969.1	*Xenopus laevis*	Chordata	Vertebrata	Amphibia	Anura	
Xlae_Dmrt4	*dmrt4*	AAH70678.2	*Xenopus laevis*	Chordata	Vertebrata	Amphibia	Anura	
Xlae_Dmrt5	*dmrt5*	AAI70166.1	*Xenopus laevis*	Chordata	Vertebrata	Amphibia	Anura	
Xves_Dsx	*dsx*	GEAJ01011590.1	*Xenos vesparum*	Arthropoda	Hexapoda	Ectognatha	Strepsiptera	

2.2.3. Transcriptome analysis

To search for *dsx* homologs, I performed RNA-seq analysis. Adults of 15 ♀♀ and 15 ♂♂ of *T. domestica* were sampled 1440 minutes after a molt in December, 2019. The fat bodies of the individuals were removed using tweezers in a phosphated buffered saline (PBS; pH=7.2). Three adults were used per sample. Total RNA was extracted from 10 samples (5♀♀, 5♂♂) using RNeasy Micro kits (QIAGEN K.K., Tokyo, Japan) following the manufacturer's instructions. The concentration of purified RNA was measured using a Qubit 4 fluorometer (QIAGEN K.K., Tokyo, Japan) with Qubit RNA BR Assay kits (QIAGEN K.K., Tokyo, Japan). Paired-end libraries were constructed from 100 ng of the total RNAs using TruSeq RNA Library Prep kits v2 (Illumina K.K., Tokyo, Japan) following the manufacturer's instructions. The libraries were run on a sequence (Hiseq, Illumina, Tokyo, Japan). The library preparation and sequencing were performed by Genewiz Strand-Specific RNA-seq service. Low quality reads and adapter sequences were eliminated from the short-reads using Cutadapt v1.15 (Martin 2011). Then, *de novo* assembly of trimmed short-read sequence was performed using Trinity-v4.2.0 (https://github.com/trinityrnaseq/trinityrnaseq/) (Grabherr et al. 2011). Information about the samples can be obtained from the National Center for Biotechnology Information (NCBI) BioSample database (Accession number: SAMN18175012–SAMN18175021).

2.2.4. Full-length cDNA sequence and exon-intron structure

To elucidate the exon-intron structures of Dsx and Dsx-like, I determined the full-length cDNA sequences using a Rapid Amplification of cDNA Ends (RACE) method and performed a BLAST analysis for my genome database of *T. domestica*. I extracted total RNA from eggs, whole bodies, fat body, and gonads of nymphs and adult females and males of *T. domestica* using TRI reagent (Molecular Research Center Inc., Ohio, USA) following the manufacturer's instructions. The total RNAs were treated with RNase-Free DNase I (New England BioLabs Japan Inc., Tokyo, Japan) to exclude remaining genomic DNA and purified by phenol/chloroform extraction and ethanol precipitation. For 5′ -

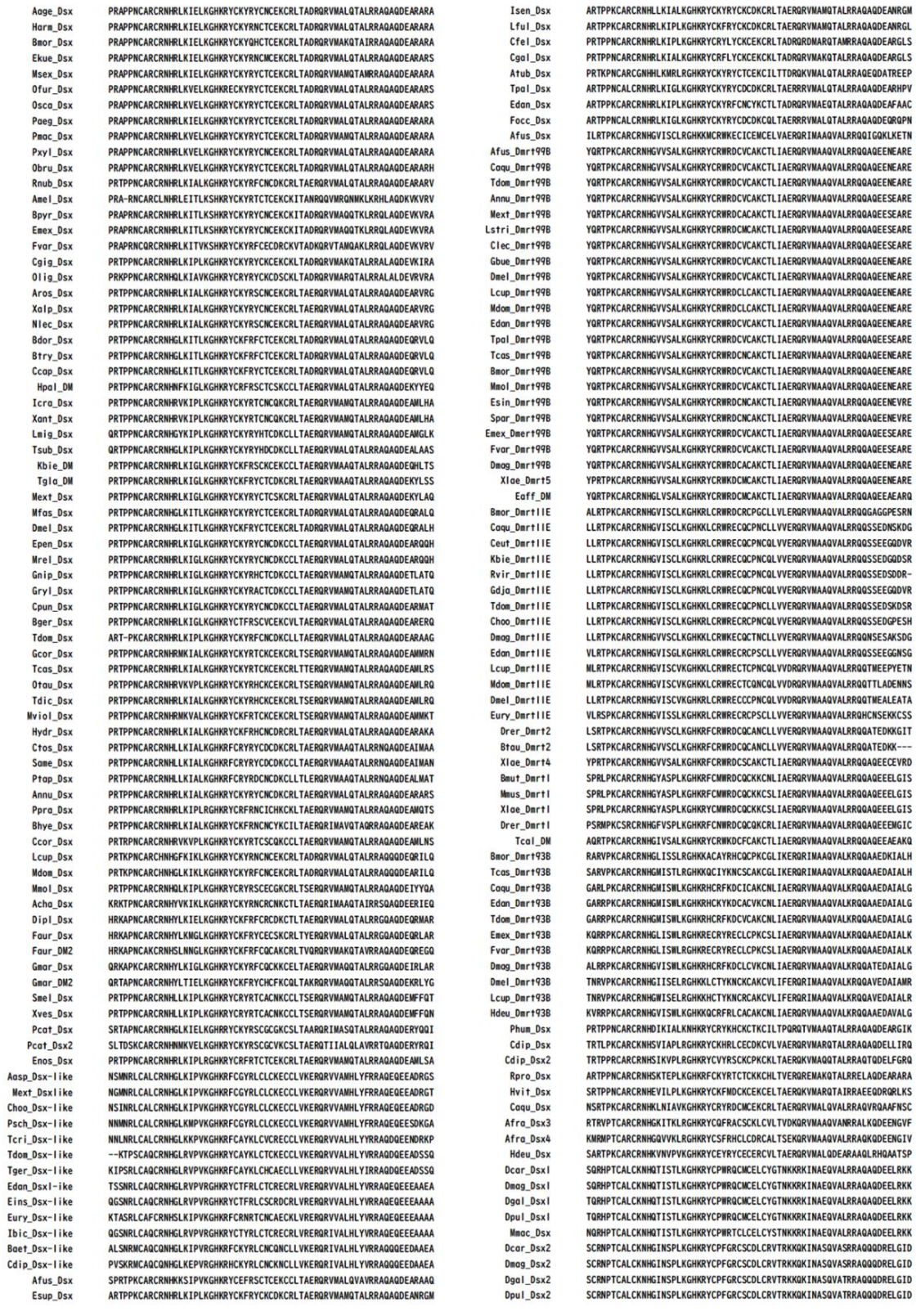


Figure 2.1. Multiple sequence alignment of DM domain of DMRT family proteins for molecular phylogenetic analysis. The 65 amino acids of 166 DMRT proteins were used for the molecular phylogenetic analysis. The left alignment is the continuation of the right one.

RACE analysis, mRNAs were purified from 75 µg of the total RNAs using Dynabeads mRNA Purification kit (Thermo Fisher Scientific K.K., Tokyo, Japan) following the manufacturer's instruction. I then ligated an RNA oligo at the 5'-end of the mRNA using GeneRacer Advanced RACE kits (Thermo Fisher Scientific K.K., Tokyo, Japan). For 3′-RACE analysis, I ligated an RNA oligo of the SMART RACE cDNA Amplification Kit (Takara Bio Inc., Shiga, Japan) at 3′-end of the total RNA during reverse transcription. First stranded (fs-) cDNA was generated from the RNAs using SuperScript II reverse Transcriptase (Thermo Fisher Scientific K.K., Tokyo, Japan). I used primers specific to the RNA oligos and performed RACE analysis by nested RT-PCR using Q5 High-Fidelity DNA polymerase (New England BioLabs Japan Inc., Tokyo, Japan). The primers specific to *dsx* and *dsx-like* were made from sequences of the relevant genomic regions and are listed in Appendix 1. The amplicons were separated using the agarose gel-electrophoresis and cloned using TOPO TA Cloning Kit for Sequencing (Thermo Fisher Scientific K.K., Tokyo, Japan) following the manufacture's protocol. I used a DH5α *Escherichia coli* strain (TOYOBO CO., LTD., Osaka, Japan) as the host cell. Plasmids were extracted using the alkaline lysis and purified by phenol-chloroform and ethanol precipitation. The nucleotide sequences of the cloned amplicons were determined from the purified plasmids by the Sanger Sequencing service of FASMAC Co. Ltd. (Kanagawa, Japan).

To investigate the exon region of *dsx* and *dsx-like*, I searched the genomic region of the full-length cDNA sequences of *dsx* and *dsx-like* via local blastn analysis. The genome of *T. domestica* was obtained the DNA Data Bank in Japan (Accession number: DRA005797; Bioproject: PRJDB5781).

2.2.5. RT-qPCR

To quantitative mRNA expression levels, I performed RT-qPCR analysis. For investigating the sex-specific expression profile of *dsx* and *dsx-like*, I used the fat body of adults of *T. domestica* since the sexes can be distinguishable by the external morphology at this stage. Fat bodies also exhibit sex-specific physiological functions in adults. I dissected the individuals in PBS and collected their fat body in 2 ml tubes containing TRI reagent (Molecular Research Center Inc., Ohio, USA). The fat bodies then were disrupted using a TissueLyser LT small beads mill (QIAGEN K.K., Tokyo, Japan). These disrupted samples were preserved at −80°C until used. Total RNA was extracted from the samples according to the manufacture's protocol for the TRI reagent. Extracted

RNA was treated with 2% RNase-free DNase I (New England BioLabs Japan Inc., Tokyo, Japan) at 37°C for 40 minutes and purified by phenol/chloroform extraction and ethanol precipitation. I measured the concentration of the total RNA using a spectrophotometer (DS-11+, Denovix Inc., Wilmington, USA). fs-cDNA was synthesized from 350 ng of the total RNA using SuperScript II reverse Transcriptase (Thermo Fisher Scientific K.K., Tokyo, Japan). I diluted the fs-cDNA to 1:2 with MilliQ water and preserved it at −30°C until it was used in RT-qPCR assay. The RT-qPCR assays were performed using a LightCycler 96 instrument (Roche, Basel, Switzerland) according to the manufacture's protocol with the THUNDERBIRD SYBR qPCR Mix (TOYOBO Co. Ltd., Osaka, Japan). The reaction volume was 10 µl. I used 1 µl of the fs-cDNA as templates. The preparation of the RT-qPCR solution proceeded on ice. The protocol of the RT-qPCR was as follows:

preincubation at 95°C for 600 seconds and 45 cycles of three-step reactions, such as denaturation at 95°C for 15 seconds, annealing at 60°C for 15 seconds and extension at 72°C for 45 seconds. I used *ribosomal protein 49* (*rp49*) as a reference gene, as described by Ohde et al. (2011). I designed primer sets of the target genes by the Primer3Ib version 4.1.0 (Untergasser et al. 2012) following the manufacture's recommended condition of the THUNDERBIRD SYBR qPCR Mix. I confirmed the primers' specificity using melting curves ranging from 65°C to 95°C. I selected primer sets exhibiting a single peak. The primers are listed in Appendix 1. Each RT-qPCR was technically replicated three times. Some samples were excluded before analyzing the data when the Ct value of any genes was not detected in one or more replicates or when the Ct value of the reference gene deviated from that of other samples. These removed data would be a technical error. I calculated the expression level of target genes by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) and performed the Brunner–Munzel (BM) test for ΔCt value.

Table 2.2. Results of RT-qPCR assay and Brunner–Munzel test.

experiment	treatment	sample size (*N*)	median		proportion of $2^{-\Delta Ct}$	95% confidence interval		Brunner-Munzel test			significance
			⊿Ct	$2^{-\Delta Ct}$				statistics	df	*P*-value	
dsx male-type expression level	male	12	8.59	2.66.E-03	1.00	0.83	1.07	8.25	12.82	1.75.E-06	***
	female	8	13.89	6.64.E-05	0.02						
dsx female-type expression level	male	11	9.98	9.90.E-04	1.00	-0.01	0.01	-92.63	20.95	2.20.E-16	***
	female	12	5.77	1.83.E-02	18.47						
dsx-like expression level	male	12	3.62	8.13.E-02	1.00	0.58	1.02	2.86	21.84	9.24.E-03	**
	female	12	4.71	3.83.E-02	0.47						

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. n.s. means non-significance.

The BM test was carried out using R-v4.0.3. with the *brunnermuzel.test* function of the *brunnermuzel* package (https://CRAN.R-project.org/package=brunnermunzel). Holm's method was used for multiple comparison analyses between the control and treatments. The data are listed in Table 2.2.

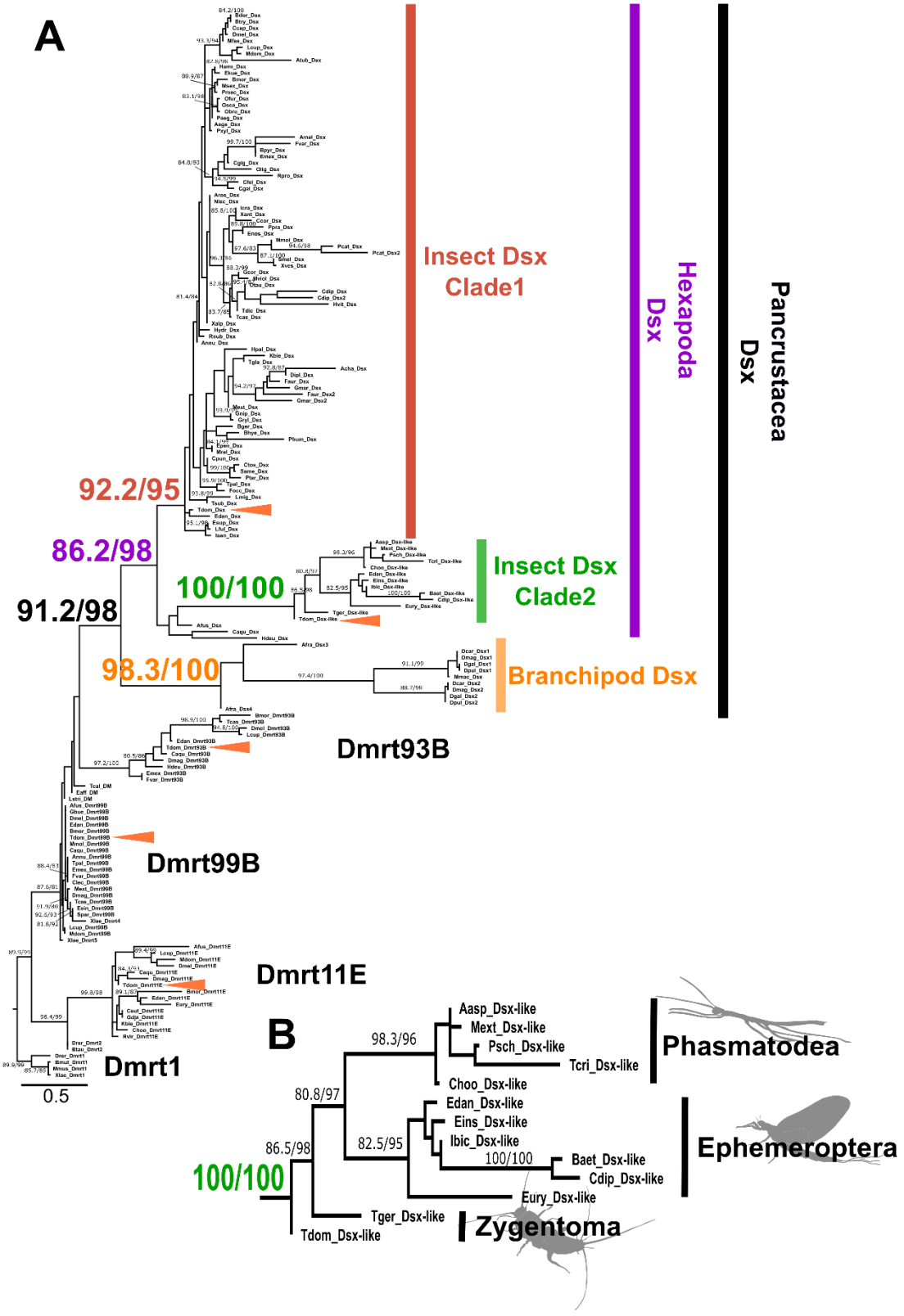
2.3. Results

2.3.1. Phylogenetic analysis of DMRT family

To elucidate the divergence process of DMRT family genes in insects, I searched public transcriptome/genome/protein databases and my RNA-seq data of *T. domestica* for the pancrustaceans (= hexapods + crustaceans) and vertebrate DMRT genes (Table 2.1). Then, I used 166 sequences, including 29 of 32 insect orders, and performed a phylogenetic analysis based on the amino acid sequences of their DM domains. The results revealed that the pancrustacean DMRT family transcription factors were divided into five clusters, including Dmrt1 in Vertebrate and Dmrt93B, Dmrt11E, Dmrt99B, and Dsx in *Drosophila* (Figure 2.2A). The support values of each cluster were indicated by the Shimodaira-Hasegawa approximal likelihood ration test (SH-aLRT) and ultrafast bootstrap (UFBoot) and were 89.9/99.0% (Dmrt1), 97.2/100.0% (Dmrt93B), 96.4/99.0% (Dmrt11E), 20.0/66.0% (Dmrt99B) and 91.2/98% (Dsx). In addition, three subclusters were found within the Dsx cluster: Insect Dsx Clade1, Insect Dsx Clade2, Branchiopoda Dsx Clade. These subclusters were robustly supported by the SH-aLRT/UFBoot values of 92.2/95.0% (Insect Dsx Clade1), 100.0/100.0% (Insect Dsx Clade2), 98.3/100.0% (Branchipod Dsx Clade). Three sequences in entognathan species were clustered with the Insect Dsx Clade2. The UFBoot value of this cluster showed a robust reconstruction (96.0%), but its SH-aLRT moderately or lowly supported this cluster (78.1%). The Insect Dsx Clade1, Clade2, and the entognathan sequences were integrated into a robustly reconstructed clade (SH-aLRT/UFBoot = 86.2/98.0%). Due to low support values, I cannot resolve the relationships among the clusters and within the subclusters.

The Dmrt93B, Dmrt11E, and Dmrt99B clusters contained single genes in *T. domestica*. The Dmrt11E and Dmrt99B clusters also contained some of the vertebrate DMRT family

genes used in this analysis. For example, Dmrt2 of *Danio rerio* and Dmrt4 of *Xenopus laevis* were found in Dmrt11E and Dmrt99B, respectively.



Insect Dsx Clade1 contained *dsx* of Pterygota previously analyzed, including *dsx* of *Drosophila melanogaster*. The clade also contained one of the *dsx* homologs of *T. domestica*. The Branchiopoda Dsx Clade contained *dsx* in *Daphnia*. *Daphnia* is known to have two *dsx* paralogs (Kato et al. 2011, Toyota et al. 2013), which were placed in this clade. Significantly, the Insect Dsx Clade2 included *dsx* homologs of the Zygentoma, Ephemeroptera, and Phasmatodea (Figure 2.2A, B). Some species of these orders also had the *dsx* ortholog in the Insect Dsx Clade1. These facts provide evidence for gene duplication of *dsx* before the divergence of Dicondylia (= Zygentoma + Pterygota). Hereafter, I refer to the *dsx* paralogs in Insect Dsx Clade2 as *dsx-like*.

2.3.2. Gene structure and isoform of dsx *and* dsx-like *in* Thermobia domestica

To examine whether *dsx* and *dsx-like* of *T. domestica* are regulated by alternative splicing, I determined the full-length cDNA sequences of *dsx* and *dsx-like* by RNA-seq and RACE. Mapping these sequences to the *T. domestica* genome revealed that *dsx* and *dsx-like* have four and three exons, respectively. In addition, I found two isoforms of *dsx* with different sequence lengths (Figure 2.3A): a long one (951 bp) and a short one (756 bp). These isoforms differed in the exon contained in the most 3'- side. The short isoform specifically had a region extending to the 3'- side of exon 2 (exon 2*). The long isoform specifically included exon 3. In contrast to *dsx*, only a single transcript was detected from *dsx-like* (Figure 2.3B).

I predicted the open reading frame (ORF) to investigate coding sequence differences among the *dsx* isoforms. Dsx isoforms consisted of putative peptide sequences of 316 amino acids and 251 amino acids. Both isoforms included the DNA binding domain, DM domain. The C-terminal sites of both genes encode the oligomerization domain (OD). Notably, these isoforms differ in their C-terminal coding sequences (Figure 2.3A).

Figure 2.2. Molecular phylogeny of Doublesex in Pancrustacea and Vertebrata. (*A*) Molecular phylogeny of Doublesex and Mab-3 related transcriptional factors. The maximum-likelihood phylogenetic analysis was based on amino acid sequences of DM domain of the DMRT family and performed by IQ-TREE following multiple sequence alignment using MAFFT software. I list all 166 operational taxonomic units in Table 2.1. Orange arrowheads indicate the DM domain-containing genes of *T. domestica*. (*B*) Enlarged view of insect Dsx Clade2 (*dsx-like* clade). The numerical value on each node indicates SH-aLRT/UFBoot values. Both values < 50 are not shown. The larger numbers are the supporting values on the branches interested in this study.

It is known that one of the paralogs generating from gene duplication often becomes a pseudogene (e.g., Endo et al. 2004, Xiao et al. 2016). Therefore, to predict whether *dsx-like* is a pseudogene or not in *T. domestica*, I performed the ORF prediction of *dsx-like*. The results showed that *dsx-like* has a putative ORF consisting of 150 aa. Furthermore, a domain search using Hmmscan revealed that *dsx-like* has a DM domain (Figure 2.3B). The DM domain of *dsx-like* contained the intertwined zinc-finger structure (Figure 2.3C)

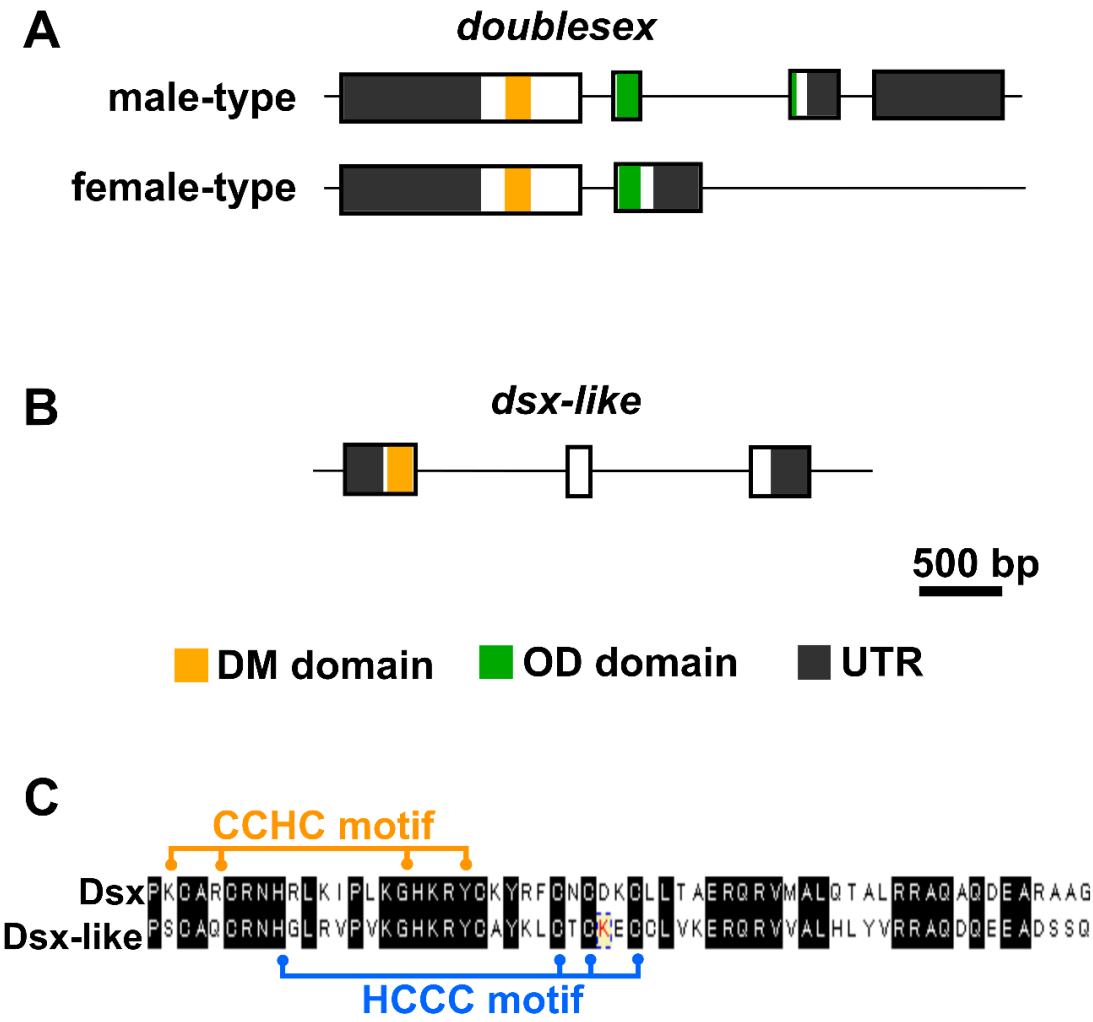
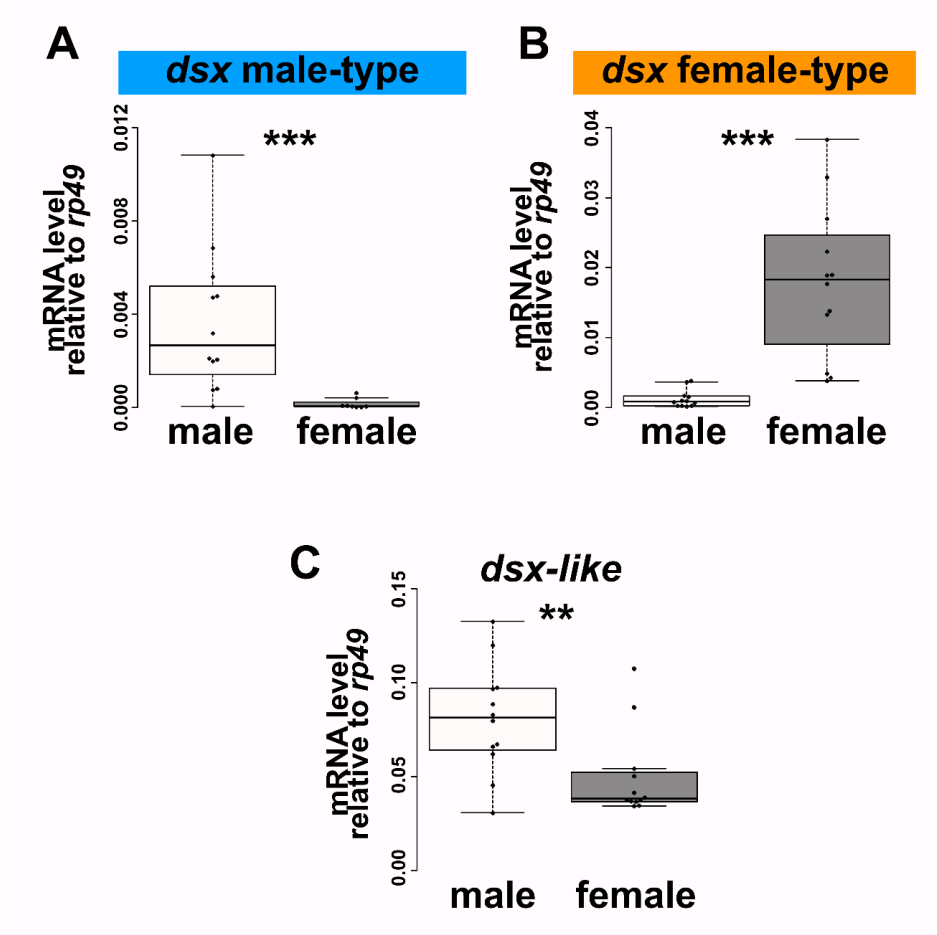


Figure 2.3. The exon-intron structure of *dsx* and *dsx-like*. (*A*) Exon-intron structures of *dsx* in *Thermobia domestica*. The upper and lower schematic images show the gene structure of *dsx* male-type and female-type, respectively. (*B*) Exon-intron structures of *dsx-like* in *T. domestica*. The exon-intron structure is determined by mapping the mRNA sequence of each gene to the genome of *T. domestica*. (*C*) The sequences of the DM domain in *dsx* and *dsx-like* of *T. domestica*. Both genes possess the intertwined structure consisted of two motifs, i.e., the CCHC and HCCC motifs.

that underlies the function of *dsx* (Zhu et al. 2000). These results suggest that *dsx-like* has the potential to function as a protein rather than a pseudogene. In contrast, the C-terminal dimerization domain (OD domain) found in *dsx* was not hit in *dsx-like* (Figure 2.3B), which raises the possibility of functional differences between *dsx* and *dsx-like*.

2.3.3. Expression profile of dsx *and* dsx-like *in* Thermobia domestica

To investigate whether the isoforms in *dsx* are sex-specifically expressed, I quantified the expression of these isoforms by RT-qPCR. Then, I found that the long isoform is highly expressed in males and the short isoform is highly expressed in females (Figure 2.4A). Consistent with this result, differentially expressed exon analysis using RNA-seq revealed that the exon specific to the long isoform had a high inclusion level in males, while the exon specific to the short isoform were highly expressed in females (Figure 2.4B). These facts indicate that *dsx* of *T. domestica* is regulated in a sex-specific splicing



manner. Hereafter, I refer to the male-biased isoform of *dsx* as *dsx* male-type and the female-biased isoform as *dsx* female-type.

Finally, to determine whether *dsx-like* with a single transcript shows a sex-specific expression or not, I investigated the expression levels of *dsx-like* between sexes. The RT-qPCR in the fat bodies revealed that *dsx-like* is expressed 2-fold higher in males than in females (Brunner-Munzel test, $P < 0.001$) (Figure 2.4C). Thus, *dsx-like* shows slightly male-biased expression in the fat bodies.

2.4. Discussion

2.4.1. Gene duplication of dsx *and its evolutionary history*

Both *dsx* and *dsx-like* can be found in Zygentoma, Ephemeroptera, and Phasmatodea. This result indicates that the gene duplication of *dsx* occurred before the emergence of the common ancestor of Zygentoma and Pterygota (= Dicondylia). Price et al. (2015) searched for the sequences of *dsx* among Hexapoda and suggested that *dsx* experienced a duplication in the common ancestor between Zygentoma and Ephemeroptera. This study successfully supports the hypothesis by providing the phylogenetical evidence on the duplication event in the basal lineage of Insecta.

I uncovered that Phasmatodea species retain *dsx-like*, which is not mentioned in previous studies. This result supports that *dsx-like* is maintained at least until the emergence of the common ancestor of Polyneoptera. Furthermore, currently, *dsx-like* cannot be found in pterygote orders other than Ephemeroptera and Phasmatodea. This fact suggests that *dsx-like* has repeatedly been lost in many pterygote taxa. In this study, *dsx-like* cannot be detected from Eumetabola (= Paraneoptera + Holometabola). This result might be ascribed to insufficient sampling from the taxa. Alternatively, it is possible that *dsx-like* was lost from the divergence of Neoptera to the emergence of the common ancestor of Eumetabola.

Figure 2.4. (*A*) Expression level of *dsx* in males. (*B*) Expression level of *dsx* in females of *T. domestica*. (C) Expression level of *dsx-like* in males and females. The expression level was measured by RT-qPCR of *dsx* and *dsx-like* in the adult fat body and is indicated as relative values to expression of the reference gene, *ribosomal protein 49* (*rp49*). Each plot signifies the mRNA expression level of each individual. Total N = 20 (*dsx* male-type), 23 (*dsx* female-type), and 24 (*dsx-like*). Results of Brunner–Munzel tests are indicated by asterisks: $^{**}P < 0.01$; $^{***}P < 0.001$ and are described in Table 2.2.

Some arthropods have two *dsx* copies that exhibit male-specific expression (Kato et al. 2011, Pomerantz et al. 2015). Recently, phylogenetic analysis of *dsx* (Wexler et al. 2019) shows that *dsx* were lineage-specifically duplicated in a chelicerate and branchiopods. However, the relationship of the *dsx* paralogs among insects and other arthropods was ambiguous since previous phylogenetic reconstructions of the DMRT family did not include sequences of the insect species that retains *dsx* and its paralogs. My result provides phylogenetical evidence of the independent origin of the duplication of *dsx* among insects and other arthropods. Hence, I infer that *dsx-like* occurred after the divergence of Branchiopoda until the emergence of the common ancestor of Dicondylia.

In this chapter, more precise timing is not clear when *dsx* and *dsx-like* were duplicated. Since I found only one copy of *dsx* in Collembola and Diplura used in this study, the duplication event might have occurred from the divergence of Ectognatha (= Archaeognatha + Zygentoma + Pterygota) to the emergence of the common ancestor of Dicondylia. This hypothesis is consistent with evidence on the large-scale gene duplication in the common ancestor of Dicondylia (Li et al. 2018). However, the lack of the paralog in collembolan and dipluran species might be attributed to the poor taxonomic collection. More taxon sampling from apterygotes and crustaceans would elucidate more complete history of the gene duplication of *dsx*.

2.4.2. Sex-specific isoform of dsx *and its evolutionary history*

In this chapter, I revealed that *dsx* of *T. domestica* has sex-specific isoforms. This finding is the first experimental evidence on the sex-specific splicing control for *dsx* in arthropod species other than Pterygota.

The previous study (Price et al. 2015) indicates that the sex-specific splicing in *dsx* is shared among Hexapoda by searching for *dsx* sequences from 30 insect orders based on EST and NGS data. Further, evidence on the sex-specific splicing in *dsx* has been accumulating from *Rhodnius prolixus*, the German cockroach, *Blattella germanica*, and the damselfly, *Ischnura senegarensis* (Takahashi et al. 2019, Wexler et al. 2019). These facts suggest a single origin of the sex-specific splicing regulation in *dsx* in Pterygota. My findings using *T. domestica* strongly support that sex-specific splicing of *dsx* was already present in the common ancestor of Pterygota and the single origin of the splicing manner.

Recently, it has been reported from the human louse, *Pediculus humanus* (Wexler et al. 2019), silverleaf whitefly, *Bemisia tabaci* (Guo et al. 2018), and Japanese subterranean termite, *Reticulitermes speratus* (Miyazaki et al. 2021) that *dsx* has alternatively splicing isoforms but lacks its sex-specific manner. Thus, it was possible that the sex-specific splicing in *dsx* could be a trait acquired secondarily in each lineage. According to my inference, it is presumed that *dsx* acquired the sex-specific splicing control before the divergence of the Dicondylia and later became independently regulated in a monosexual splicing manner in some lineages.

To date, there is no evidence of the sex-specific splicing regulation of *dsx* in arthropods other than insects. In a predatory mite and some crustaceans, the *dsx* expression is regulated in a male-specific manner (Kato et al. 2011, Pomerantz et al. 2015, Li et al. 2018, Panara et al. 2019). Based on a recent phylogenetic analysis from phylogenomics (Schwentner et al. 2018), among the crustaceans with information on *dsx*, Branchiopoda such as the water flea, *Daphnia magna*, is the most closely related to insects. These facts and my results from *T. domestica* support that the sex-specific splicing control in *dsx* appeared during the common ancestor of Dicondylia from the common ancestor between Branchiopoda and Dicondylia. According to Misof et al. (2014), the common ancestor between the Branchiopoda and Dicondylia emerged around 500 million years ago. Also. the common ancestor of Dicondylia is inferred to have occurred about 420 million years ago. Therefore, *dsx* is presumed to have been controlled by the sexually dimorphic splicing during about 80 million years.

2.4.3. Relation between gene duplication and alternative splicing of dsx

I show that *dsx-like* coexists with the sex-specific splicing of *dsx* in *T. domestica*. This result indicates that both the duplication and the sex-specific splicing control in *dsx* occurred before the divergence of Pterygota. In other words, the last common ancestor of Pterygota may possess both events in *dsx*. Hence, I infer that the acquisition of the sex-specific isoform of *dsx* preceded the loss of its paralog. Generally, the number of alternative isoforms negatively correlates with that of paralogs (Kopelman et al. 2005, Su et al. 2006, Talavera et al. 2007). Based on the coexistence of the *dsx* paralog and isoforms, *dsx* may deviate from this general concept. Alternative splicing also facilitates the retention of paralogs via its sub-functionalization (Bush et al. 2017). However, I cannot

resolve which event appeared earlier than the other. Thus, the relationship between these events remains ambiguous. Price et al. (2015) suggested that the loss of *dsx* paralogs and the emergence of sex-specific isoforms concomitantly occurred and proposed that the isoforms gained new functions in place of the roles of the lost paralog. Indeed, the loss of *dsx-like* occurred after acquiring the sex-specific isoforms of *dsx*. Considering the expression profile of *dsx-like* and the coexistence of the paralog and isoforms of *dsx*, *dsx-like* and the isoforms of *dsx* might play cooperative or redundant roles in sexual traits. One possibility is that the feminizing roles of *dsx* of holometabolan insects could result from swapping functions of *dsx-like*. Alternatively, the roles and evolution of these two events might be independent of each other. Functional analysis of *dsx* and *dsx-like* in apterygote insects is essential to investigate further the relationship between the paralog and isoforms of *dsx*. This analysis will be conducted in the next chapter.

2.4.4. Utility of **Thermobia domestica** *for examining* **dsx** *evolution*

Based on information on *dsx* in various arthropods in the last decade, several hypotheses have been proposed to explain the molecular and functional evolution of *dsx* (e.g., Kato et al. 2011, Price et al. 2015, Wexler et al. 2019, Hopkins and Kopp 2021; detailed in the next chapter). These hypotheses depend on findings from a small number of species of crustaceans and hemimetabolans and information from NGS data such as genomes and transcriptomes. Therefore, there are gaps in knowledge between the crustaceans and insects studied so far, and the function and regulatory mode of *dsx* have not been experimentally verified. To empirically examine the previous hypotheses, it would be essential to infer the status in the common ancestor and stem group of Pterygota.

In this chapter, I show that *T. domestica* has two paralogs. I also reveal that *dsx* of *T. domestica* is controlled by the sex-specific splicing. Thus, the coexistence of these features may have been retained in the common ancestor and stem group of Pterygota. Since Zygentoma diverged from the common ancestor of Dicondylia about 400 million years ago (Misof et al. 2014), this group may not fully reflect the status in *dsx* of the common ancestor and stem group of Pterygota. However, at least in terms of the gene copy number and regulatory mode, this group may mirror the ancestral state.

Zygentoma such as *T. domestica* has a simple sexually dimorphic morphology, i.e., simple male genitalia that is not aedeagus and female ovipositor. Various features in the

external genital organs of Zygentoma have been interpreted as similar to the condition in the common ancestor of Pterygota (Kristensen 1975, Matsuda 1976, Emeljanov 2014, Beutel et al. 2017, Boudinot 2018). Therefore, Zygentoma is a suitable outgroup for understanding the function of *dsx* for sexual morphology formed during post-embryonic development in the common ancestor of Pterygota in that it encompasses putative states of sex differences in the common ancestor of Pterygota.

Notably, in *T. domestica* (Zygentoma), the tools for analyzing gene function during post-embryogenic development were developed (Ohde et al. 2011). Together with its phylogenetic position, these facts emphasize this species to be a novel model species for elucidating the evolution of *dsx*. Hence, using this species, I will investigate the function of *dsx* and infer its evolutionary history in the next chapter.

Chapter 3

Function of *dsx* in *Thermobia* and Evolution of *dsx* Roles

Abstract In the last two decades, various arthropodan species have accumulated knowledge of the *doublesex*. However, the evolutionary history of its function is not fully understood because of a gap in information between studied taxa. In particular, it has been hard to infer the ancestral roles of the female-specific isoform in the common ancestor of Pterygota. Here, I investigate the roles of *dsx* of the apterygote species, *Thermobia domestica*, to fill the gap. My nymphal RNA interference analysis shows that the *doublesex* is required for male morphogenesis during post-embryonic development but is not essential for female morphogenesis. This result and previous information on Pterygota strongly support that the *doublesex* was necessary only for male morphology when the sex-specific splicing appeared and later became essential for female morphology in the common ancestor of Holometabola except for Hymenoptera (Aparaglossata). In addition, I reveal that the *doublesex* promotes female-specific expression of *vitellogenin* genes. These results suggest that *dsx* in females came into use separately among morphogenesis and other biological processes in insect evolution. Thus, my results strongly support the previous hypothesis that proposed a step-wise evolution of the *doublesex* isoforms and their function. This study also provides insight into an exaptive role of the female-specific isoform of the *doublesex*.

3.1. Introduction

Alternative splicing underpins functional diversity in genomes by producing various mRNAs transcribed from a single gene (e.g., Bush et al. 2017). In this thesis, I focus on the hypotheses proposed that 'non-functional' or neutral isoforms later acquire function and are involved in the neo-functionalization of genes (see Section 1.2.). Recently, the *dsx* gene has been proposed to obtain its role in female differentiation via neo-functionalization of non-functional isoforms in insects.

As discussed in Chapter 2, *dsx* has acquired the sex-specific isoforms before the appearance of the common ancestor of Pterygota and Zygentoma. The sex-specific *dsx* isoforms have sex-antagonistic roles in the transcriptional regulation, promoting either male or female differentiation in sexual dimorphism in Diptera, Lepidoptera, Coleoptera, and Hymenoptera (Burtis and Baker 1989, Ohbayashi et al. 2001, Kijimoto et al. 2012, Shukla and Palli 2012b, Ito et al. 2013, Gotoh et al. 2016, Xu et al. 2017, Roth et al. 2019). In the last five years, *dsx* has sex-specific isoforms and is responsible for male differentiation of morphological traits during postembryonic development, but unnecessary for female differentiation in the sawfly *Athalia rosae* (Mine et al. 2017, 2021), silverleaf whitefly *Bemisia tabaci* (Guo et al. 2018), brown planthopper *Nilaparvata lugens* (Zhuo et al. 2018), German cockroach *Blattella germanica* (Wexler et al. 2019), and damselfly *Ischnura senegalensis* (Takahashi et al. 2019, 2021). The female-specific isoforms in *dsx* do not necessarily promote female differentiation. Therefore, Wexler et al. (2019) proposed a stepwise evolutionary hypothesis in which *dsx* acquired the sex-specific isoforms and later became essential for female differentiation after the emergence of Holometabola.

Generally, alternative isoforms conserved among species have some function (Graveley 2001, Xing and Lee 2006, Kelemen et al. 2013). Thus, considering the female-specific isoform of *dsx* has been retained and possessed an extended coding region in the course of insect evolution, the isoform could take some advantages, i.e., exaptive roles, in females. Currently, such exaptive states remain unclear. Here, I focus on the function of *dsx* to other than female morphogenesis. In *D. melanogaster*, *dsx* plays crucial roles in many aspects in females. For example, *dsx* in *D. melanogaster* contributes to controlling female-specific expressed genes such as a yolk precursor gene *vitellogenin*. Notably, *dsx* in *Be. tabaci* positively regulates the expression of *vitellogenin* in females but remains non-essential for female morphology (Guo et al. 2018). *dsx* of the honeybee *Apis mellifera* also upregulates the *vitellogenin* expression in females but does not contribute to female morphogenesis during postembryonic development (Velasque et al. 2018, see Discussion). These results imply that *dsx* acquired its feminizing roles separately between morphogenesis and other biological processes in females. However, it has not been investigated whether the sex-antagonistic role in transcriptional regulation of female-

biased genes was retained during the insect evolution. Also, the timing of obtaining such function for female gene expression is still unknown.

Information on *dsx* in the outgroup of Pterygota allows for inferring the evolution of *dsx* more rigorously. Currently, the outgroup species with the information on the function of *dsx* belong to some crustaceans (Kato et al. 2011, Li et al. 2018). The crustaceans are phylogenetically distant from the common ancestor of Pterygota and do not have *dsx-like* that presented in the common ancestor. Also, *dsx* of the crustaceans expresses only males and does not have the female-specific isoform (Kato et al. 2011). Filling these knowledge gaps between Pterygota and crustaceans would be essential for inferring the ancestral function of the female-specific isoforms of *dsx* and its evolutionary history during the insect evolution. Therefore, information on outgroups closely related to Pterygota would provide a critical clue to resolving the issue regarding the functional evolution of *dsx* (Wexler et al. 2019). In this chapter, I used the firebrat, *Thermobia domestica*, to infer the process by which *dsx* acquired feminizing roles. This species belongs to Zygentoma, which is the sister group of Pterygota. Further, as discussed in Chapter 2, this species is a novel model for exploring the functional evolution of *dsx* because of some reasons like the retention of two *dsx* copies. Hence, to infer the functionality of *dsx* in the common ancestor of Pterygota and to investigate the exaptive role, I investigated the function of *dsx* in *T. domestica* for sexual differentiation in sexual morphology and *vitellogenin* expression.

3.2. Materials and Methods

3.2.1. Animals

Thermobia domestica (Packard 1873), was kept by the same way in Chapter 2. For examining the roles of *dsx* and *dsx*-like in *vitellogenin* expression, female and male insects were collected from the stock colony and transferred into the plates. For examining the function of *dsx* and *dsx-like* for sexual morphology and gametogenesis, I used firebrats from April to June, 2019, February to April, April to July, and September to December, 2020. For investigating the roles of *dsx* and *dsx-like* in the *vitellogenin* expression, firebrats were manipulated from June to July, 2020.

3.2.2. Estimation of molt timing

Estimating the molt timing of insects is essential for the analysis of developmental processes and the functions of developmental regulatory genes. Its timing of hemi- or holometabolan insects can be inferred using morphological changes such as a wing growth. However, timing is hard to estimate in apterygote insects since they have little change in their morphology during postembryonic development. *T. domestica* forms scales in the fourth instar, and changes the number and length of its styli during the fourth to ninth instar under my breeding conditions. These features can be used to estimate molt timing, but it is difficult to apply these criteria to experiments using adults or a large number of nymphs. To resolve this problem, I used leg regeneration after autotomy and time-lapse imaging to estimate the molt timing of *T. domestica*. Autotomy occurs at the joint between the trochanter and femur in *T. domestica*. An autotomized leg regenerates after one molt (Buck and Edwards, 1990). For the RNAi analysis during postembryonic development, I amputated a right hindleg at the autotomic rift, using tweezers, and observed whether the leg had regenerated. This test enabled us to rapidly estimate the molt timing. For the RNA-seq and the RT-qPCR analysis, the time-lapse imaging was used to determine the precise time of molt. I build a time-lapse imaging system with a network camera system (SANYO, Tokyo, Japan) set in an incubator at 37°C (Figure 3.1A). Photos of insects in the 24-well plate were taken every five minutes. I created a time-lapse movie from the photos every 12 hours using ImageJ 1.52a (https://imagej.nih.gov/ij/) and observed whether the insects molted (Figure 3.1B).

3.2.3. Transcriptome analysis

In the transcriptome analysis, I used same data and methods describing in Chapter 2. I mapped the reads obtained to the assembled genome using the HISAT2 program (Kim et al. 2019) with a default option and counted the mapped reads using the STRINGTie program (Pertea 2015) with default parameter settings. Differential expression gene analysis was performed based on the count matrix using the *edgeR* package (Robinson et al. 2010) in R-v4.0.3 (R Core Team, 2020).

3.2.4. Reverse transcription-quantitative PCR (RT-qPCR)

To quantitative mRNA expression levels, I performed RT-qPCR analysis. Thirteenth instar individuals and adults after molting were sampled for investigating roles of the genes in the sexually dimorphic morphology and the *vitellogenin* expression, respectively. The sample sizes are reported in the figure legends and Table 3.1. The methods and

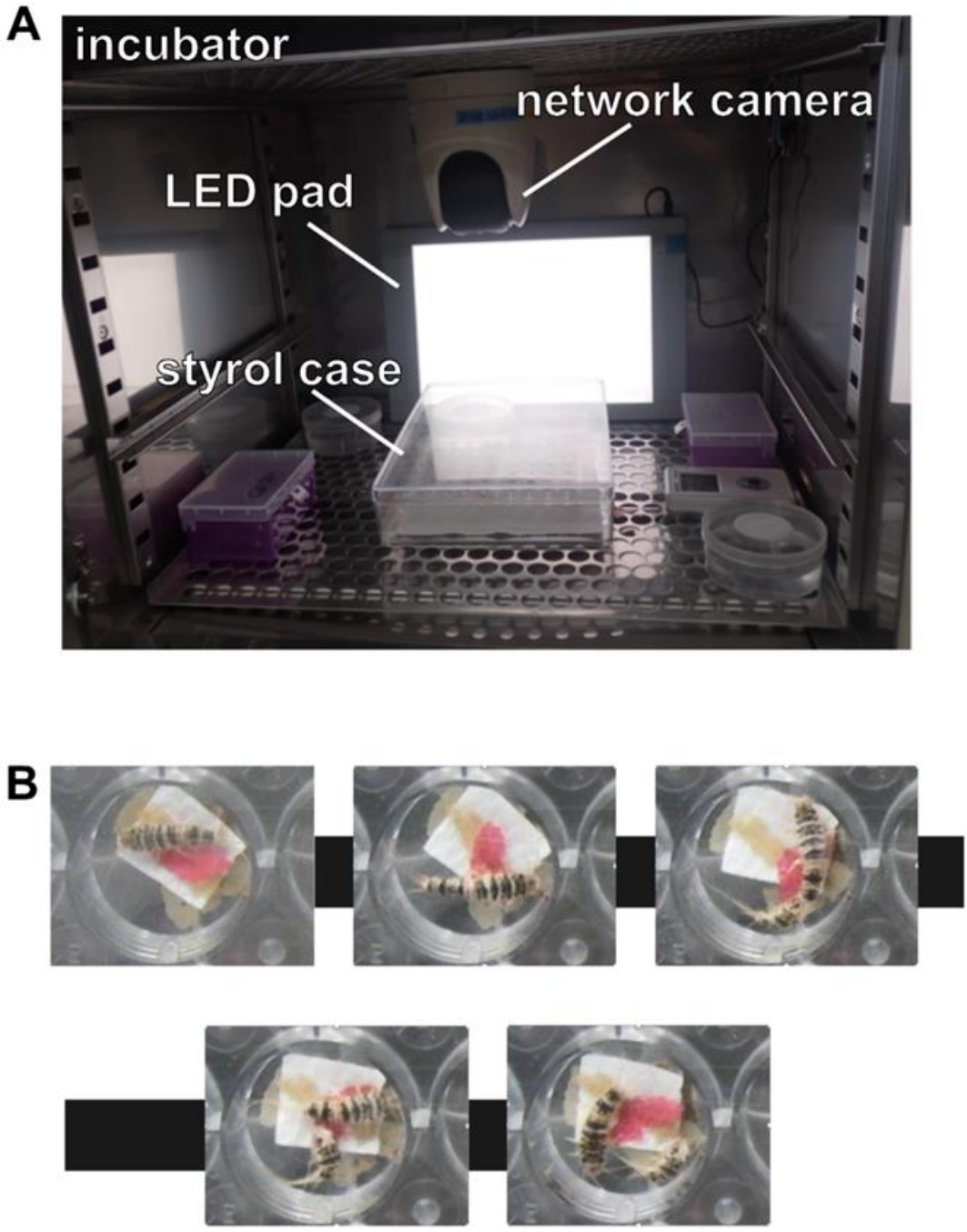


Figure 3.1. Time-lapse imaging system. (*A*) A photo of the time-lapse imaging system used to observe the molt of *T. domestica*. The network camera is located on the floor of the incubator. The insects are put on 24-well plates in the styrol case. The LED pad is used for lighting up the inside of the incubator. The temperature within the incubator is kept at 37°C. (*B*) The time-lapse images during the molt. I set the interval of taking a photo every 5 minutes. The photos are ordered along with the time course from upper-right to lower-left. The firebrat proceeded its molting for 10 minutes.

statistical analyses were described in Chapter 2. In addition, an outlier was detected in the

dsx RNAi male by the Smirnov-Grubbs (SG) test. I repeatedly performed the SG test using the data excluding the outlier. No further outliers were detected. Lastly, I re-analyzed the data, excluding the outlier, using the BM test (Table 3.1).

3.2.5. RNAi analysis

The RNAi assay can be used to examine the roles of genes during postembryonic development in *T. domestica* (Ohde et al. 2011). The sexual differentiation of insects is generally assumed to be a cell-autonomous mechanism that is independent of systemic hormonal- control (Verhulst and van de Zande 2015) as discussed in De Loof and Huybrechts (1998) and Bear and Monteiro (2013) and progresses during postembryonic development. Therefore, nymphal RNAi is the most effective tool to investigate the roles of genes on sexual trait formation during postembryonic development. To reduce the risk of off-target effects, the dsRNA was designed to avoid the region of the DM domain. I also confirmed that the dsRNA had no contiguous matches of more than 20 bases with other genes on the genome by BLAST (blastn option). To produce templates for the dsRNA, I cloned the regions of *dsx* and *dsx-like* from the fs-cDNA using the same method as the RACE analysis. I amplified the template DNAs from purified plasmids with PCR using Q5 High-Fidelity DNA Polymerase and purified the amplified DNA with the phenol/chloroform extraction and the ethanol precipitation. dsRNA was synthesized from the purified DNA using Ampliscribe T7-Flash Transcription kits (Epicentre Technologies, Co., Wisconsin, USA). I designed the PCR primers using the Primer3Web version 4.1.0 (Untergasser et al. 2012). The PCR primers are listed in Appendix 1. In nymphal RNAi analysis, I injected the dsRNAs repeatedly into the abdomen of the nymphs of *T. domestica* with each molt from the fourth or fifth instar to thirteenth instar to sustain the RNAi effect during postembryonic development. The initial stage was the same within a single experiment. This repeated RNAi treatment was effective in some insects such as *Blattella germanica* (Wexler et al. 2019). I sampled the individuals one, three, and five days after molting, using phenotypic observations, analysis of *dsx* knockdown effects, and the oocyte number. To determine the sex of individuals, I initially observed the gonads: testis and ovary. In my RNAi analysis, the gonads completely formed and there was no difference between the control and *dsx* RNAi individuals in external morphology (see Results). Therefore, individuals with testis were males and those with ovaries were females. *T. domestica* molts throughout its life, even after sexual maturation, and

produces *vtg* during each adult instar (Rousset and Bitsch 1993). To analyze the *vtg* mRNA levels, I also injected the dsRNAs of *dsx* and *dsx-like* repeatedly into the females and males every three days from 12 hours after molting. I sampled the females and males at 720±20 minutes after subsequently molts.

3.2.6. Phenotype observation

I dissected thirteenth instar individuals in PBS using tweezers and removed the thoraxes, reproductive systems, and external genital organs. I took images using the digital microscope system (VHX-5000, KEYENCE, Tokyo, Japan). The thoraxes and external genital organs were fixed with FAA fixative (formaldehyde: ethanol: acetic acid = 15:5:1) at 25°C overnight and then preserved in 90% ethanol. I used the length of the prothorax as an indicator of body size. To measure the prothoracic width, the prothoracic notum was removed from the fixed thorax after treatment with 10% NaOH solution at 60°C for 30 minutes to dissolve the soft tissues. The notum was mounted in Lemosol on a microscope slide. The prepared specimens were imaged using a KEYENCE VHX-5000. With the microscope at 50×, the length of the notum was measured. The ovipositor length was also measured using the microscope at 20× and 50×. To count the sperm number, sperm was collected from seminal vesicles and diluted with 5 ml MilliQ water. 50 µl of the diluted sperm was spotted on a microscope slide and dried overnight. I technically replicated the measurement three times for ovipositor length and six times in sperm number and calculated these means. Measurement was performed by blinding the treatment. I counted the number of oocytes in ovarioles using an optical microscope at 50× (Olympus, Tokyo, Japan). A generalized linear model (GLM) was used to analyze differences in ovipositor length (length data) and sperm and oocyte number (count data) among RNAi treatments. The body size, target genes, and interactions between the target genes were used as explanatory variables. The length was assumed to follow a Gaussian distribution, and the count data to have a negative binomial distribution. I used R-v4.0.3 in these analyses and the *glm* and the *glm.nb* (MASS package) functions for the length and count data, respectively. To analyze the contribution of the explanatory variables, a likelihood ratio test for the result of GLM was performed using the *Anova* function of the *car* package. The statistical results are listed in Tables 3.3 (male) and 3.4 (female).

Table 3.1. Results of RT-qPCR assay and Brunner–Munzel test.

experiment	treatment	sample size (*N*)	median		proportion of $2^{-\Delta Ct}$	95% confidence interval		Brunner-Munzel test			*P*-value adjusted by Holm's method	significance
			ΔCt	$2^{-\Delta Ct}$				statistics	df	*P*-value		
dsx expression level in males	*egfp*	12	7.5	5.52.E-03	1							
	dsx all	8	9.24	1.66.E-03	0.3	0.45	1.03	1.78	13.61	9.68.E-02	1.94.E-01	n.s.
	dsx-like	10	7.14	7.10.E-03	1.29	0.07	0.57	-1.54	17.4	1.43.E-01	1.43.E-01	n.s.
dsx expression level in males excluding outliers	*egfp*	12	7.5	5.52.E-03	1							
	dsx all	7	9.32	1.57.E-03	0.28	0.64	1.05	3.64	13.87	2.73.E-03	5.45.E-03	**
	dsx-like	10	7.14	7.10.E-03	1.29	0.07	0.57	-1.54	17.4	1.43.E-01	1.43.E-01	n.s.
dsx expression level in females	*egfp*	10	5.18	2.76.E-02	1							
	dsx all	17	6.23	1.33.E-02	0.48	0.57	0.99	2.78	16.21	1.33.E-02	2.65.E-02	*
	dsx-like	9	5.17	2.78.E-02	1.01	0.2	0.82	0.08	15.38	9.40.E-01	9.40.E-01	n.s.
dsx-like expression level in males	*egfp*	12	5.72	1.95.E-02	1							
	dsx all	8	6.17	1.39.E-02	0.72	0.5	0.98	2.08	17.99	5.19.E-02	5.19.E-02	n.s.
	dsx-like	10	10.17	8.68.E-04	0.04	1	1	Inf	NaN	2.20.E-16	4.40.E-16	***
dsx-like expression level in females	*egfp*	10	6.08	1.49.E-02	1							
	dsx all	17	6.31	1.26.E-02	0.85	0.47	0.93	1.82	16.18	8.76.E-02	8.76.E-02	n.s.
	dsx-like	9	11.58	3.27.E-04	0.02	1	1	Inf	NaN	2.20.E-16	4.40.E-16	***
vitellogenin-1 expression level in males	*egfp*	5	15	3.06.E-05	1							
	dsx all	9	4.42	4.68.E-02	1530.72	-0.05	0.09	-15.2	10.67	1.44.E-08	2.87.E-08	***
	dsx-like	7	10.18	8.62.E-04	28.18	-0.12	0.41	-3.09	8.42	1.39.E-02	1.39.E-02	*
	dsx+dsx-like	9	4.28	5.14.E-02	1678.94	0	0	-Inf	NaN	2.20.E-16	6.60.E-16	***
vitellogenin-2 expression level in males	*egfp*	3	16.76	9.01.E-06	1							
	dsx all	8	11.24	4.13.E-04	45.89	0	0	-Inf	NaN	2.20.E-16	6.60.E-16	***

	dsx-like	5	15.48	2.19.E-05	2.43	-0.24	0.51	-2.46	5.56	5.24.E-02	6.19.E-01	n.s.
	dsx+dsx-like	8	10.64	6.27.E-04	28.64	0	0	-Inf	NaN	2.20.E-16	6.60.E-16	***
	egfp	10	11.28	4.21.E-04	1							
vitellogenin-3 expression level in males	*dsx* all	10	3.25	1.07.E-01	254.99	-0.09	0.23	-6	9.86	1.40.E-04	2.80.E-04	***
	dsx-like	10	8.73	2.37.E-03	5.64	-0.03	0.39	-3.29	15.01	4.97.E-03	4.97.E-03	**
	dsx+dsx-like	9	2.82	1.42.E-01	336.55	0	0	-Inf	NaN	2.20.E-16	6.60.E-16	***
	egfp	8	-2.67	6.34.E+00	1							
vitellogenin-1 expression level in females	*dsx* all	10	-0.44	1.41.E+00	0.22	0.51	1.02	2.2	15.51	4.33.E-02	4.33.E-02	*
	dsx-like	7	-0.29	1.22.E-00	0.19	0.7	1.08	4.55	10.89	8.51.E-04	2.56.E-03	**
	dsx+dsx-like	8	0.9	5.40.E-01	0.09	0.58	1.1	3	8.82	1.52.E-02	3.05.E-02	*
	egfp	8	1.84	2.80.E-01	1							
vitellogenin-2 expression level in females	*dsx* all	10	4.34	4.96.E-02	0.18	0.65	1.05	3.68	15.57	2.11.E-03	4.22.E-03	**
	dsx-like	7	5.24	2.65.E-02	0.09	0.83	1.06	8.45	12.94	1.27.E-06	3.80.E-06	***
	dsx+dsx-like	8	6.1	1.48.E-02	0.05	0.63	1.15	3.54	7.29	8.92.E-03	8.92.E-03	**
	egfp	8	-3.37	1.03.E+01	1							
vitellogenin-3 expression level in females	*dsx* all	10	-1.98	3.95.E+00	0.38	0.65	1.07	3.75	11.23	3.12.E-03	6.23.E-03	**
	dsx-like	8	0.84	7.82.E-01	0.08	0.81	1.06	7.56	12.6	4.98.E-06	1.49.E-05	***
	dsx+dsx-like	8	0.05	1.09.E+00	0.11	0.57	1.13	2.95	7.56	1.97.E-02	1.97.E-02	*

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. n.s. means non-significance.

3.2.7. Scanning Electron Microscopy (SEM)

The NanoSuit method (Takaku et al. 2013) was used for the SEM analysis. Male penises and female ovipositors preserved in 90% ethanol were washed with distilled water and immersed in 1% Tween20 at 25°C for 10 minutes. The samples were mounted on stubs and imaged using a low-vacuum SEM (DX-500; KEYENCE, Tokyo, Japan).

3.2.8. Histology

The gonads of RNAi individuals were fixed with Bouin's fixative (saturated picric acid: formaldehyde: glacial acetic acid = 15:5:1) at 25°C overnight and washed with 90% ethanol plus Lithium Carbonate (Li2CO3). The ovipositors of RNAi individuals were fixed with FAA fixative at 25°C overnight and then were transferred into 90% ethanol. The samples were dehydrated and cleared with an ethanol-butanol series. The cleared samples were immersed and embedded in paraffin at 60°C. The paraffin blocks were polymerized at 4°C and cut into 5 μm thick sections using a microtome (RM2155: Leica, Wetzlar, Germany). The sections were mounted on microscope slides coated with egg white-glycerin and stained using Delafield's Hematoxylin and Eosin staining. After staining with the hematoxylin, the slides were washed with 1% hydrochloric acid-ethanol for 40 seconds. The stained slides were enclosed with Canada balsam. I observed the slides on an optical microscope (Olympus, Tokyo, Japan) and took photos using a digital single-lens reflex camera (Nikon, Tokyo, Japan).

3.3. Results

3.3.1. RNAi efficacy of dsx *and* dsx-like *in* Thermobia domestica

First, I investigated the efficiency of the knockdown of *dsx* and *dsx-like* by nymphal RNA interference (RNAi). To this end, I selected fat bodies as representative of tissues in which *dsx* and *dsx-like* are robustly expressed during post-embryonic development from RNA-seq analysis and quantified the expression of *dsx* and *dsx-like* in males and females by the RT-qPCR.

The *dsx* RNAi in females and *dsx-like* RNAi treated group in males and females had significantly lower target gene expression than the control (*egfp* RNAi) group (Brunner-Munzel test, $P = 0.0265$ in female *dsx*, 4.40×10-16 in male and female *dsx-like*; Figure 3.2A; Table 3.1). The *dsx* RNAi group in males showed no significant effect on *dsx*

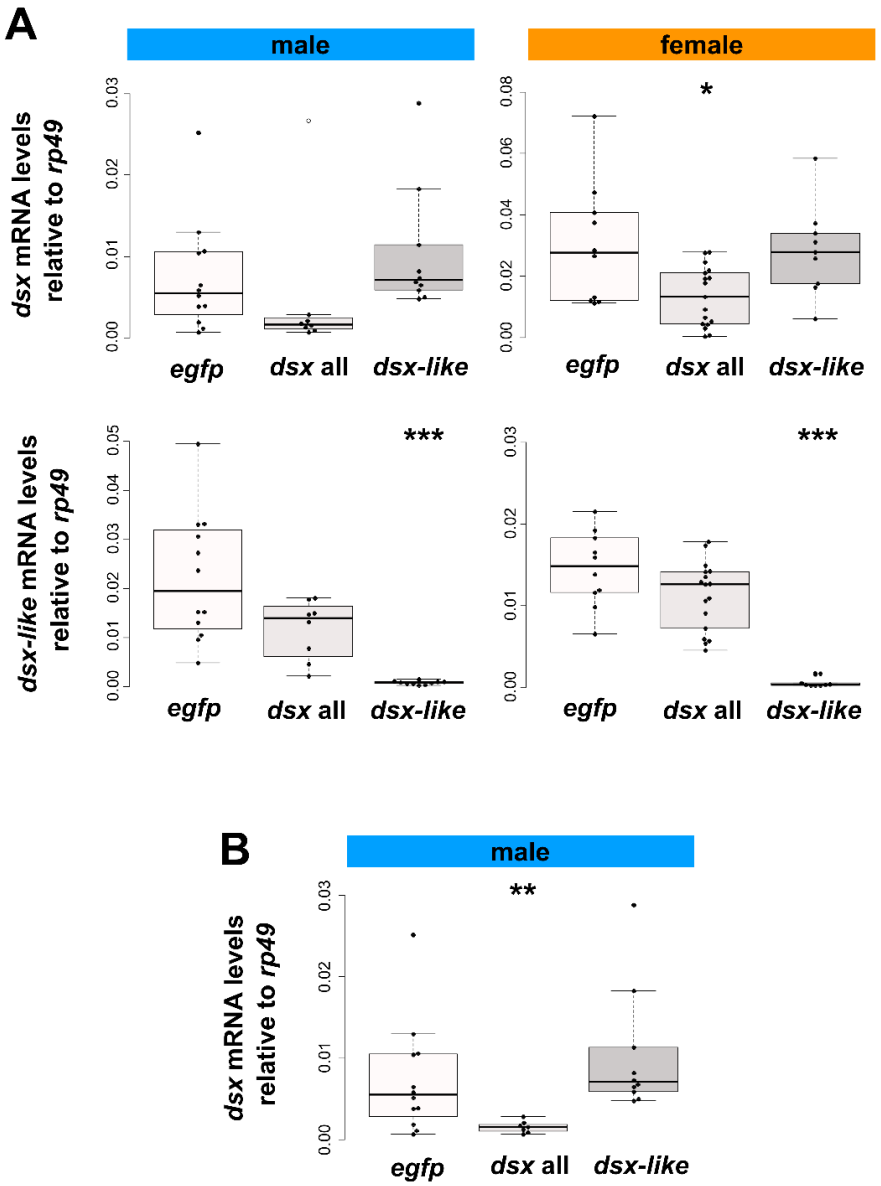


Figure 3.2. Expression of *dsx* and *dsx-like* mRNA in nymphal RNAi individuals. (*A*) Expression level of target genes in RNAi individuals. The mRNA levels of *dsx* and *dsx-like* were analyzed by RT-qPCR assay and are the relative value to the expression of the reference gene, *ribosomal protein 49* (*rp49*). The upper graphs are the expression of *dsx* and the lower ones are that of *dsx-like*. The left column is the result in males and the right one is that in females. White plot suggests the outlier. (*B*) the expression level of *dsx* mRNA in the nymphal RNAi males after excluding an outlier. To test the outlier, the Smirnov–Grubbs' test was performed. The result of the Smirnov–Grubbs' test is shown in Table 3.2. The *egfp*, *dsx* all and *dsx-like* indicates the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group and *dsx-like* dsRNA injected group, respectively. Results of the Brunner–Munzel test are indicated by asterisks: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$ and is also described in Table 3.1. $P \geq 0.05$ is not shown. Each plot indicates the value of each individual. Total $N = 30$ and 36 in males and females.

expression. Since it was suspected that outliers affected this result, I tested for outliers in *dsx* RNAi males and found one outlier. A reanalysis removing the outlier showed that *dsx* expression was significantly reduced in *dsx* RNAi males (Brunner-Munzel test, P = 0.00545; Figure 3.2B; Table 3.2).

Table 3.2. Results of Smirnov–Grubbs' test for expression level of *dsx* mRNA in nymphal RNAi males. The determination of whether a value is an outlier or not is based on the *P*-value and is shown in the "outlier?" column.

data	max/min	value	*G*	*U*	*P*-value	outlier?
egfp	max	10.45	1.8587	0.65738	2.73.E-01	no
	min	5.31	1.59672	0.74715	5.74.E-01	no
dsx	max	10.47	0.97	0.85	1.00.E+00	no
	min	5.23	2.27	0.15688	5.14.E-03	yes
dsx-like	max	7.70	1.01	0.88	1.00.E+00	no
	min	5.12	2.08	0.47	8.35.E-02	no
reanalysis of *dsx*	max	10.47	1.52	0.55	3.52.E-01	no
	min	8.44	1.42	0.61	4.70.E-01	no

Compared to the knockdown efficiency of *dsx* in males (median c.a. 30%), *dsx* RNAi in females seems inefficient (median c.a. 50%) (Figure 3.2; Table 3.1). However, given that half of female individuals in the RT-qPCR analysis in the fat body showed lower *dsx* expression than the minimum control values (Figure 3.2A), I assume that *dsx* expression is suppressed in a certain number of females used in each analysis. Thus, I concluded that *dsx* and *dsx-like* RNAi could knock down the expression of their target genes.

In this analysis, *dsx* RNAi showed no effect on *dsx-like* expression vice versa (Figure 3.2). Therefore, it appears that *dsx* and *dsx-like* do not regulate each other's transcription. It is also unlikely that *dsx* and *dsx-like* dsRNAs affect each other's target gene expression through off-target effects.

3.3.2. Assessment of contribution to sexual size dimorphism

Zygentoma has a few differences between the sexes. One of the sex differences can be seen in body size as noted by Darwin (1871) (Figure 3.3A). However, differences in body size between sexes have not been quantified and statistically verified in Zygentoma.

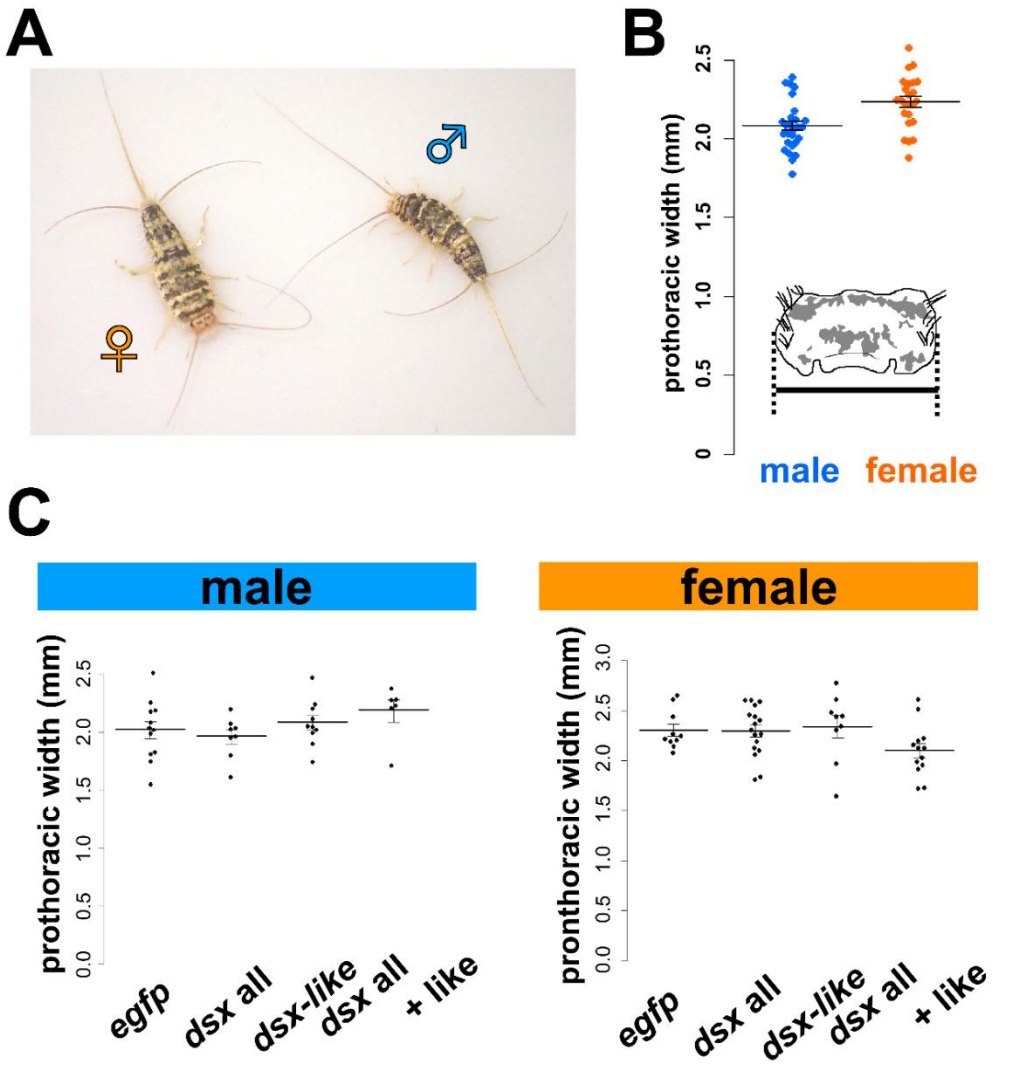


Figure 3.3. Sexual size dimorphism in *T. domestica*. (*A*) A pair of *T. domestica*. The female looks similar to the male. (*B*) Body size of normal individuals. (*C*) Body size of RNAi treatment groups. The pronotum (prothoracic tergum) width was used as an index for body size. The graph shows mean ± SE (standard error). Results of the generalized linear model (GLM) analysis shown in Tables 3.3 (male) and 3.4 (female). Any significant effect can be detected in the RNAi treatments. Total $N = 51$ in normal individuals, 36 in RNAi males, and 49 in RNAi females. Each plot in (*B*) and (*C*) represent the value of each individual. In each panel, the *egfp*, *dsx* all, *dsx-like* and *dsx* + *dsx-like* represent the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group, *dsx-like* dsRNA injected group, and both *dsx* sex-common region and *dsx-like* dsRNAs injected group, respectively.

Therefore, I investigated sexual size dimorphism by measuring the body size of the 13th instar adult of *T. domestica* using pronotal width as an indicator (details in Materials and

Methods). In this analysis, the mean width (mean±sem) was 2082.62±30.24 μm for males and 2235.83±34.32 μm for females (Figure 3.3B). The general linear mixed model analysis detected significant differences in the pronotum width between sexes (random effects: sampling period, P = 0.0017). Thus, it was quantitatively and statistically confirmed that the *T. domestica* has the sexually dimorphic body size.

Then, I investigated the effects of *dsx* and *dsx-like* on the sexual size dimorphism. Here, I also performed a double knockdown of *dsx* and *dsx-like* to consider the possibility of redundant effects of the two copies. In this study, RNAi treatment was performed from 5th instar larvae and sampled at 13th instar adults. I quantitatively assessed differences in body size. As a result, the mean width (mean±sem) in males was 2039.03±74.18 μm in the control group, 1962.58±63.57 μm in the *dsx* RNAi group, 2084.17±62.84 μm in the *dsx-like* RNAi group, 2186.67±96.64 μm in the *dsx* and *dsx-like* double RNAi group (Figure 3.3C). Analysis with a generalized linear model (GLM) showed no significant differences between the control and experimental groups (Table 3.3). The mean width in females was 2306.77 ± 62.07 μm in the control group, 2297.39 ± 60.32 μm in the *dsx* RNAi group, 2339.41 ± 113.77 μm in the *dsx-like* RNAi group, and 2103.13 ± 72.50 μm in the *dsx* and *dsx-like* double RNAi group (Figure 3.3C). In the GLM analysis, no statistically significant differences could be detected between the control and experimental groups (Table 3.4). These results suggest that *dsx* and *dsx-like* are not involved in regulating the body size in *T. domestica*.

3.3.3. Contribution to internal reproductive system and gametogenesis

Generally, sex differences in reproductive reproduced animals are found in the internal reproductive system and gametogenesis. These differences are also present in Zygentoma, where sexual dimorphism is scarce. In Pterygota, *dsx* affects the formation of internal reproductive system and gametogenesis (e.g., Hildreth 1965, Zhuo et al. 2018, Wexler et al. 2019). Here, to examine whether *dsx* and *dsx-like* affect the internal reproductive system and gametogenesis, I performed anatomical and histological observations of the organs and measured the number of gametes in RNAi individuals. The internal reproductive system and gametogenesis in Zygentoma are described in Matsuda (1976). Since this previous description is fragmentary, I also described basic morphological features of the internal reproductive system.

Table 3.3. Results of generalized linear model of male traits.

objective variable	sample size (*N*)	explanatory variables	LR Chisq	Df	*P*-value	significance
pronotum width	*egfp*: 12, *dsx* all: *8*, *dsx-like*: 10, *dsx*+*dsx-like*: 6, total: 36	*dsx* RNAi	0.03	1	8.71.E-01	n.s.
		dsx-like RNAi	2.82	1	9.31.E-02	n.s.
		dsx RNAi:*dsx-like* RNAi	1.09	1	2.96.E-01	n.s.
sperm number	*egfp*: 9, *dsx* all: 6, *dsx-like*: 10, *dsx* + *dsx-like*: 4, total: 29	*dsx* RNAi	7.93	1	4.87.E-03	**
		dsx-like RNAi	2.06	1	1.51.E-01	n.s.
		prothoracic width	8.68	1	3.21.E-03	**
		dsx RNAi:*dsx-like* RNAi	0.02	1	8.79.E-01	n.s.

**$P < 0.01$. n.s. means non-significance.

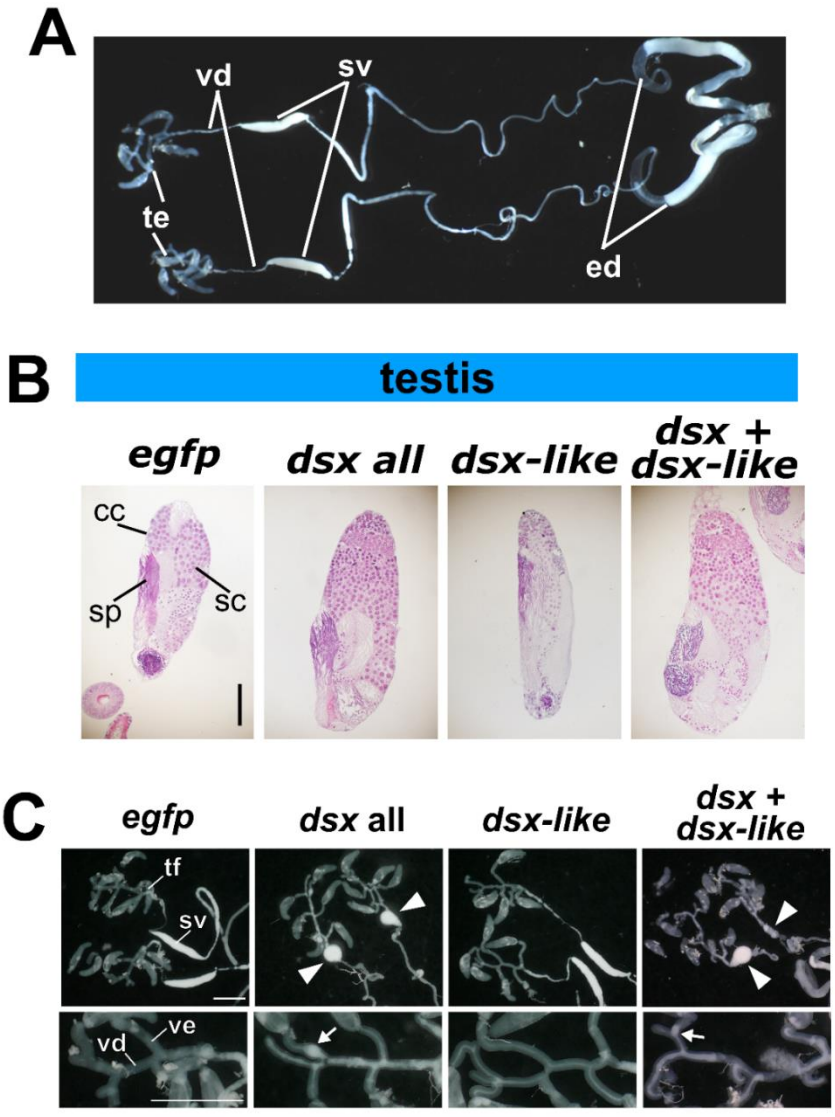
Table 3.4. Results of generalized linear model of female traits.

objective variable	sample size (N)	explanatory variables	LR Chisq	Df	*P*-value	significance
prothoracic width	*egfp*: 10, *dsx* all: 17, *dsx-like*: 9, *dsx+dsx-like*: 13, total: 49	*dsx* RNAi	2.17	1	1.41.E-01	n.s.
		dsx-like RNAi	1.96	1	1.61.E-01	n.s.
		dsx RNAi:*dsx-like* RNAi	2.16	1	1.41.E-01	n.s.
previtellogenic oocyte number	*egfp*: 7, *dsx* all: 16, *dsx-like*: 10, *dsx+dsx-like*: 9, total: 42	*dsx* RNAi	0.08	1	7.83.E-01	n.s.
		dsx-like RNAi	0.23	1	6.31.E-01	n.s.
		prothoracic width	0.97	1	3.25.E-01	n.s.
		dsx RNAi:*dsx-like* RNAi	0.06	1	8.10.E-01	n.s.
Early vitellogenic oocyte number	*egfp*: 7, *dsx* all: 16, *dsx-like*: 10, *dsx+dsx-like*: 9, total: 42	*dsx* RNAi	0.11	1	7.45.E-01	n.s.
		dsx-like RNAi	0.08	1	7.77.E-01	n.s.
		prothoracic width	3.5	1	6.14.E-02	n.s.
		dsx RNAi:*dsx-like* RNAi	0.62	1	4.30.E-01	n.s.
late vitellogenic oocyte number	*egfp*: 7, *dsx* all: 16, *dsx-like*: 10, *dsx+dsx-like*: 9, total: 42	*dsx* RNAi	0.46	1	5.00.E-01	n.s.
		dsx-like RNAi	1.12	1	2.90.E-01	n.s.
		prothoracic width	9.84	1	1.71.E-03	**
		dsx RNAi:*dsx-like* RNAi	2.16	1	1.42.E-01	n.s.
ovipositor length	*egfp*: 10, *dsx* all: 17, *dsx-like*: 9, *dsx+dsx-like*: 13, total: 49	*dsx* RNAi	0.26	1	6.08.E-01	n.s.
		dsx-like RNAi	0.58	1	4.47.E-01	n.s.
		prothoracic width	297	1	2.00.E-16	***
		dsx RNAi:*dsx-like* RNAi	2.67	1	1.02.E-01	n.s.

*P<0.05, **P < 0.01, ***P < 0.001. n.s. means non-significance.

3.3.3.1. Male reproductive system

In males of *T. domestica*, a pair of testes was located on the dorsal side of the abdomen. The testis was consisted of some testicular follicles (Figure 3.4A, B). Each testicular follicle was connected to the vas deferens via the vas efferens (Figure 3.4A). The seminal vesicle lay between the vas deferens and the ejaculatory duct. A pair of the ejaculatory ducts was associated with each other in the front of the gonopore in the penis (Figure 3.4A). The testicular follicles were a bean-like shape and the seminal vesicles were a bean pod-like shape. In the testicular follicle, the spermatogonia was in the antero-most part (Figure 3.4B). The primary and secondary spermatocytes lay in the middle part. In the



posterior part of the testicular follicle, there were some sperm bundles (Figure 3.4B). The wall of the testicular follicle consisted of a single flattened epithelial layer.

I observed the above features of the reproductive system in *dsx* or *dsx-like* RNAi males (Figure 3.4B, C). In the *dsx* knockdown group (*dsx* alone or *dsx* and *dsx-like*), the male seminal vesicle, which is a sperm storage organ and normally has a bean pod shape, became rounded (Figure 3.4C). In contrast, I could not find differences in the morphology of the testicular follicles or spermatogenesis between the RNAi and control males.

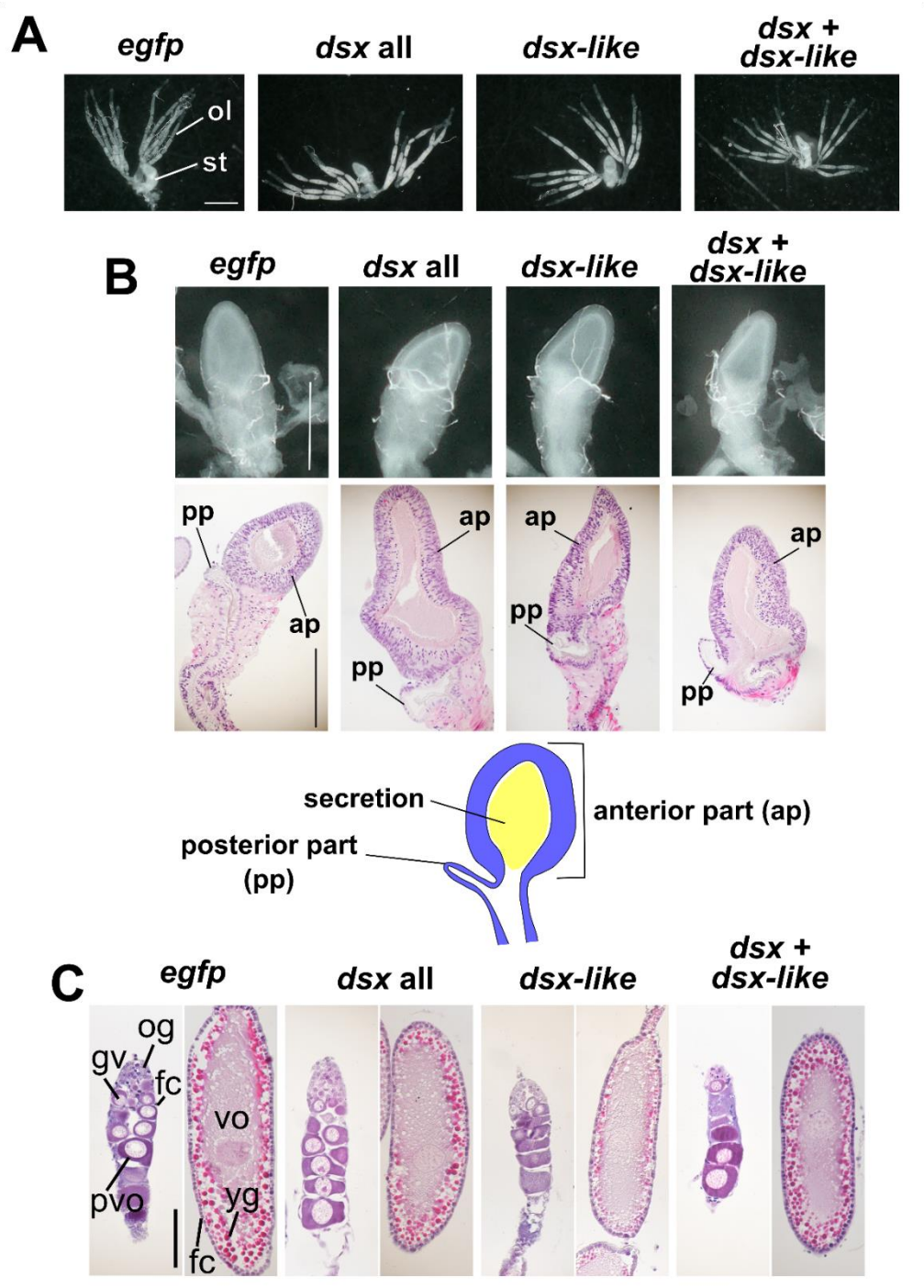
3.3.3.2. Female reproductive system

In females of *T. domestica*, part of the ovary was on the dorsal side of the abdomen. Each ovary consists of five ovarioles and was attached to the anterior part of the abdomen via the terminal tuft (Figure 3.5A). The ovarioles were associated with each other at the lateral oviduct. The lateral oviduct was connected to the common oviduct and subsequently opened at the gonopore in the valvula I. There was no vagina between the gonopore and the oviduct. The spermatheca was located on the branch point of the common oviduct along the midline (Figure 3.5A). The spermatheca was divided into two parts: anterior and posterior (Figure 3.5B). The anterior part consisted of a pseudostratified layer of the columnar epithelial cells that were secretory. The posterior part was surrounded by a single layer of epithelial cells. The ovariole was panoistic-type and was composed of two parts: the germarium and the vitellarium (Figure 3.5C). The germarium contained many oogonia and young oocytes. The vitellarium had previtellogenic and vitellogenic oocytes. The oocytes in the vitellarium were surrounded by a single layer of follicle cells. There were pedicel cells in the terminal of the ovariole. The previtellogenic oocyte had a large germinal vesicle and basophilic cytoplasm. The vitellogenic oocyte was elongated along the anterior-posterior axis of the ovariole and had eosinophilic cytoplasm. Many eosinophilic lipid droplets were present in the

---

Figure 3.4. Male reproductive systems in *T. domestica*. (*A*) Gross morphology of the reproductive systems in the non-treated male. (*B*) Histology of testis in the RNAi groups. Paraffin. Hematoxylin-Eosin staining. (*C*) Effects of RNAi on male internal reproductive system. The lower photos demonstrate the morphology of RNAi males. Arrowheads show rounded seminal vesicle. The lowest photos are focused on the vas efferens. Arrows show clogged sperm in the vas efferens. In each panel, the *egfp*, *dsx* all, *dsx-like* and *dsx* + *dsx-like* represent the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group, *dsx-like* dsRNA injected group, and both *dsx* sex-common region and *dsx-like* dsRNAs injected group, respectively. cc, cystocyte; sc, spermatocyte; sp, sperm; sv, seminal vesicle; tf, testicular follicle; ve, vas efferens; vd, vas deferens. Scales: 50 μm (*B*); 1000 μm (*C*).

peripheral region of the vitellogenic oocytes. The follicle cells were flattened and columnar in shape in the previtellogenesis and the vitellogenesis.



I observed the above features of the reproductive system in *dsx* or *dsx-like* RNAi females (Figure 3.5). I could not detect visible differences in the female reproductive system or oogenesis between the RNAi females and the controls. This result suggests that

dsx and *dsx-like* have no function in the formation of female traits and gametogenesis at the tissue and cellular level.

3.3.4. Contribution to fecundity

It has been reported that *dsx* in holometabolan species contributes to female fecundity (e.g., Kyrou et al. 2018). Does *dsx* in *T. domestica* affect male and female fecundity? *dsx* may be at least indirectly involved in sperm storage in males of this species since the gene contributes to the formation of the male internal reproductive system, especially the sperm storage organ. On the other hand, given that the *dsx* RNAi treatment did not show any effects on the internal reproductive system or the gametogenesis in females, *dsx* might not contribute to fecundity in females. Here, to examine these possibilities, I quantified the number of sperm and oocyte in the *dsx* and *dsx-like* RNAi individuals.

The estimated numbers of sperm (mean±sem) in the seminal vesicles were 25154.63±6730.69 in the controls, 5273.61±1317.51 in the *dsx* RNAi males, 16596.67±3764.38 in the *dsx-like* RNAi males, and 7836.00±72059.18 in the *dsx* and *dsx-like* double RNAi males. The GLM analysis detected significant differences between the control and *dsx* RNAi treated groups (Figure 3.6A; Table 3.3; $P = 0.0049$). On the other hand, the *dsx-like* RNAi treatment did not show a statistically significant effect between the control group and *dsx-like* RNAi treatment. Hence, *dsx* in *T. domestica* contributes to sperm storage in males.

Finally, I measured the numbers of oocytes per ovarioles along with oogenetic stages. The stages were divided into three: the previtellogenesis, which is before yolk formation;

Figure 3.5. Female reproductive systems in *T. domestica*. (*A*) Effects of RNAi on female internal reproductive system. (*B*) Morphology of spermatheca in nymphal RNAi females. The upper photos show the light microscopic images of the spermatheca. The middle ones are paraffin sections of the spermatheca. Hematoxylin-Eosin staining. The lower one is the schematic image of the spermatheca of *T. domestica*. (*C*) Histology of the ovary in the RNAi groups. The left and right panel in each treatment shows germarium/previtellogenesis and vitellogenesis, respectively. Paraffin. Hematoxylin-Eosin staining. In each panel, the *egfp*, *dsx* all, *dsx-like* and *dsx* + *dsx-like* represent the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group, *dsx-like* dsRNA injected group, and both *dsx* sex-common region and *dsx-like* dsRNAs injected group, respectively. fc, follicle cell; gv, germinal vesicle; og, oogonia; ol, ovariole; pvo, previtellogenic oocyte; st, spermatheca; yg, yolk granule; vo, vitellogenic oocyte. Scales: 1000 μm (*A*); 500 μm (*B*); 50 μm (*C*).

the early vitellogenesis, which is the beginning of yolk formation; and the late

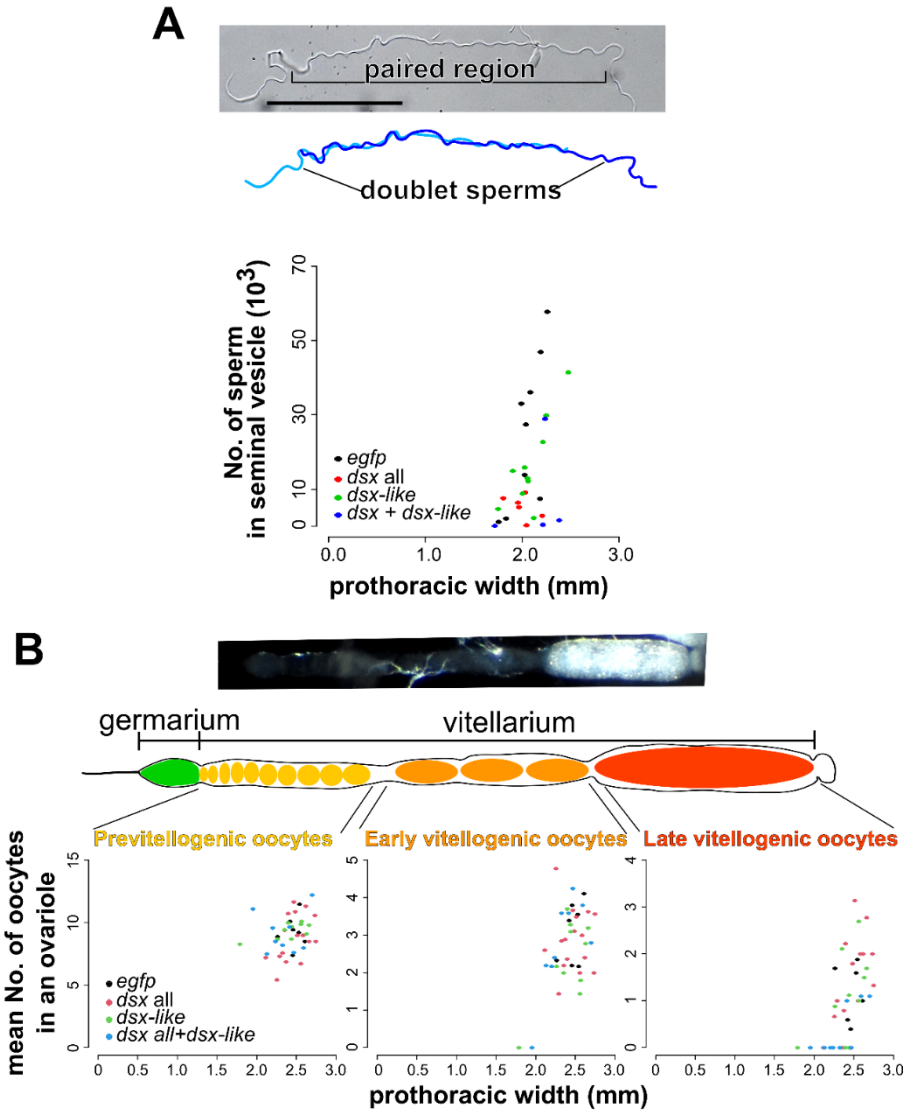


Figure 3.6. Effect of RNAi treatment on fecundity. (*A*) Sperm of RNAi males. The upper photo and figure are sperm morphology in the non-treated male. The sperm forms doublets in the seminal vesicle. The lower figure shows sperm number of RNAi males. The results of the GLM analysis are shown in Table 3.3. A significant effect was detected in the *dsx* RNAi treatment ($P = 0.0049$). Total $N = 29$. (*B*) Effects of the RNAi on oocyte number. The upper photo shows the ovariole of the non-treated female. The lower figures exhibit the number of oocytes in RNAi females along with oogenetic stages. Results of GLM analysis show in Table 3.4. The number of late vitellogenic oocytes was correlated with pronotum width, although any significant effect can be detected in RNAi treatments. Total $N = 42$ n each stage. In each panel, the *egfp*, *dsx* all, *dsx-lik*e and *dsx + dsx-like* represent the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group, *dsx-like* dsRNA injected group, and both *dsx* sex-common region and *dsx-like* dsRNAs injected group, respectively. Each plot in represent the value of each individual.

vitellogenesis, in which the ooplasm is filled with yolk and lipid droplets and the chorion formation begins. The mean numbers of the previtellogenic oocytes were 9.29 ± 0.49 in the control group, 8.80 ± 0.45 in *dsx* RNAi females, 9.38 ± 0.19 in *dsx-like* RNAi females, and 9.16 ± 0.51 in *dsx* and *dsx-like* double RNAi females. The mean numbers of early vitellogenic oocytes were 3.08 ± 0.31 in the controls, 2.88 ± 0.20 in *dsx* RNAi females, 2.35 ± 0.33 in *dsx-like* RNAi females, and 2.76 ± 0.40 in *dsx* and *dsx-like* double RNAi females. The mean numbers of late vitellogenic oocytes were 1.03 ± 0.27 in the controls, 1.30 ± 0.25 in *dsx* RNAi females, 1.10 ± 0.28 in *dsx-like* RNAi females, and 0.36 ± 0.17 in *dsx* and *dsx-like* double RNAi females. The GLM analysis did not detect significant differences between the control and RNAi-treated groups (Figure 3.6B; Table 3.4). Therefore, these results did not support the contribution of *dsx* and *dsx-like* to the oocyte storage in females, consistent with histological observations in gametogenesis.

3.3.5. Contribution to external genital organs

The external reproductive system is representative of the few sexual morphologies of Zygentoma. *T. domestica* has a short unpaired penis in males and a paired long ovipositor in females at the abdominal terminus (Figure 3.7A). Zygentoma is thought to retain various features of the external genital organ that emerged in the common ancestor of the Ectognatha (= Archaeognatha + Zygentoma + Pterygota: Insecta s.str.) (Kristensen 1975, Boudinot 2018). Here, to examine whether *dsx* contributes to the morphogenesis of the external genitalia, I observed the organs in RNAi-treated individuals. In this subsection, I described the morphology in the controls and RNAi-treated individuals. The morphological features in the control group were consistent with previous studies (Snodgrass 1957, Matsuda 1976, Emeljanov 2014, Boudinot 2018).

3.3.5.1. Morphology of male external genitalia

The penis in *T. domestica* male is an unpaired appendix on the abdomen segment IX (Figure 3.7B) and is not aedeagus (copulatory organ) (Matsuda 1976). The penis was sub-segmented into two parts. There were many setae on the left and the right side of the distal tips (Figure 3.7C). The surface of the penis had a reticulated pattern (Figure 3.7C).

In *dsx* RNAi males, a tubular organ was formed instead of the penis (Figure 3.7B, C). This tubular organ consisted of two pairs of appendage-like structures. The inner one is connected to the gonopore and the ejaculatory duct. The outer one had a lot of setae on

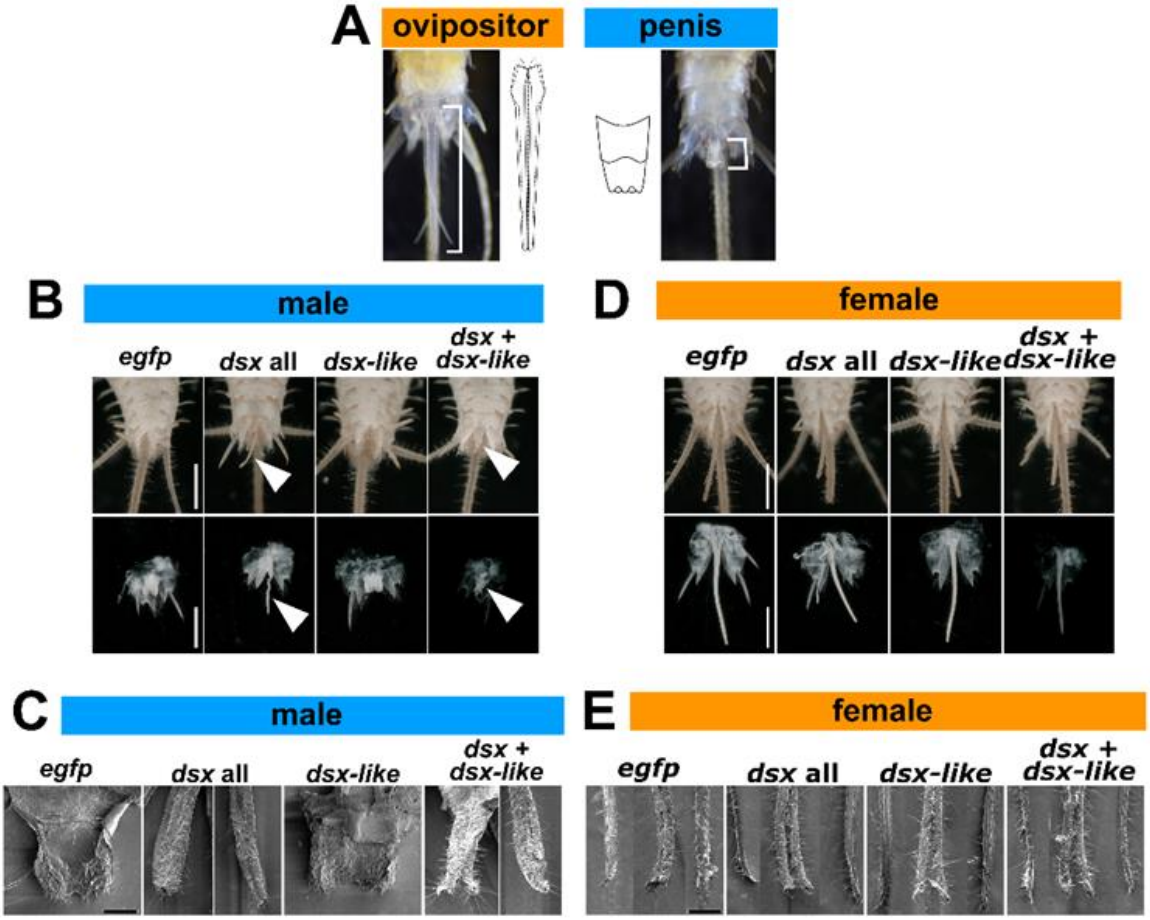


Figure 3.7. Function of *doublesex* and *doublesex-like* for genital organs in *Thermobia domestica*. (*A*) Sexually dimorphic traits of *T. domestica*. Females possess an ovipositor and males have a penis. (*B*) Effects of RNAi treatments on male penial structure. Upper images show the ventral side of the male abdomen. Lower images focus on the male penis. Arrowheads indicate ovipositor-like structure in *dsx* or both *dsx* and *dsx-like* RNAi groups. (*C*) SEM images of male penial structure. In *dsx* and *dsx + dsx-like* RNAi, the two photos are merged into one image. In these images, left panels show an ovipositor valvula II (inner sheath)-like structure. Right panels exhibit the ovipositor valvula I (outer sheath)-like structure. A detail description can be seen in Supplementary Material online. (*D*) Effects of RNAi treatments on female ovipositor. Upper images show the ventral side of the female abdomen. Lower images focus on the female ovipositor. (*E*) SEM images of female ovipositor structure. In each image, the left and right panels show the valvula II and the middle panel shows the valvula I. In each panel, the *egfp*, *dsx* all, *dsx-like*, and *dsx + dsx-like* indicates the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group, *dsx-like* dsRNA injected group, and both *dsx* sex-common region and *dsx-like* dsRNAs injected group, respectively. Scales: 1 cm (*B* and *D*); 50 µm (*C* and *E*).

its tip (Figure 3.7C). Thus, the inner pair was similar to the valvula I of the female ovipositor and the outer one was similar to the valvula II. I could detect sub-segmentation

in both structures (Figure 3.7C). These features indicated that the tubular organ in the *dsx* RNAi males was parallel to the female ovipositor. The same phenotype was found in the *dsx* and *dsx-like* double RNAi males. In contrast, the *dsx-like* males possessed a penis the same as that of the control insects (Figure 3.7B, C). My results indicate that *dsx* is essential for male differentiation of morphological traits in *T. domestica*.

3.3.5.2. Morphology of female ovipositor

T. domestica females has an ovipositor. This ovipositor consists of two pairs of appendices (gonapophysis) and is derived from the retracted vesicles on the abdomen VIII and IX (Matsuda 1976, Emeljanov 2014). The gonapophyses on the abdomen VIII (valvula I) were the ventral part of the ovipositor and a paired structure (Figure 3.7D). The gonapophyses on the abdomen IX (valvula II) were the dorsal side of the ovipositor and were united to form an unpaired structure (Figure 3.8). The distal tip of the valvula remained a paired structure and possessed dense setae (Figure 3.7E), which may play a role in sensory reception. Both valvulae were sub-segmented and have some setae (Figure 3.7E). The valvula I and II were connected through a tongue-and-groove structure (olistheter). The olistheter consisted of an aulax (groove) on the valvula I and a rhachis (tongue) on the valvula II (Figure 3.8). Within the valvulae, the epithelial cells were beneath the cuticular layer. The cuticular layer was thickened and multi-layered in the outer surface of the ovipositor. In contrast, the inner surface (i.e., the side of the egg cavity) of the ovipositor had a thin and single-layered cuticle. Some lumens of the valvulae were extended along the anterior-posterior axis and were hemocoelic cavities.

In females that were treated with RNAi for *dsx*, *dsx-like*, and both genes, the external genital organ was the same as the ovipositor of the control females (Figure 3.7D, E). This genital organ of RNAi females consisted of two pairs of sub-segmented appendage-like structures and possessed dense setae on the tip of the inner pair. The outer pair was connected to the gonopore and the common oviduct. Thus, in the view of histology, the location, and the relation to other elements, the external organ of the RNAi females was not different from the ovipositor of the control ones.

3.3.5.3. Growth of female ovipositor

The external genital organ in the *T. domestica* shows differences in length between the sexes. Although the *dsx* RNAi treatment in females did not affect the morphological characteristics of the ovipositor, it might affect the length of the ovipositor. Therefore, to

examine whether *dsx* affects the growth of the female ovipositor, I measured ovipositor

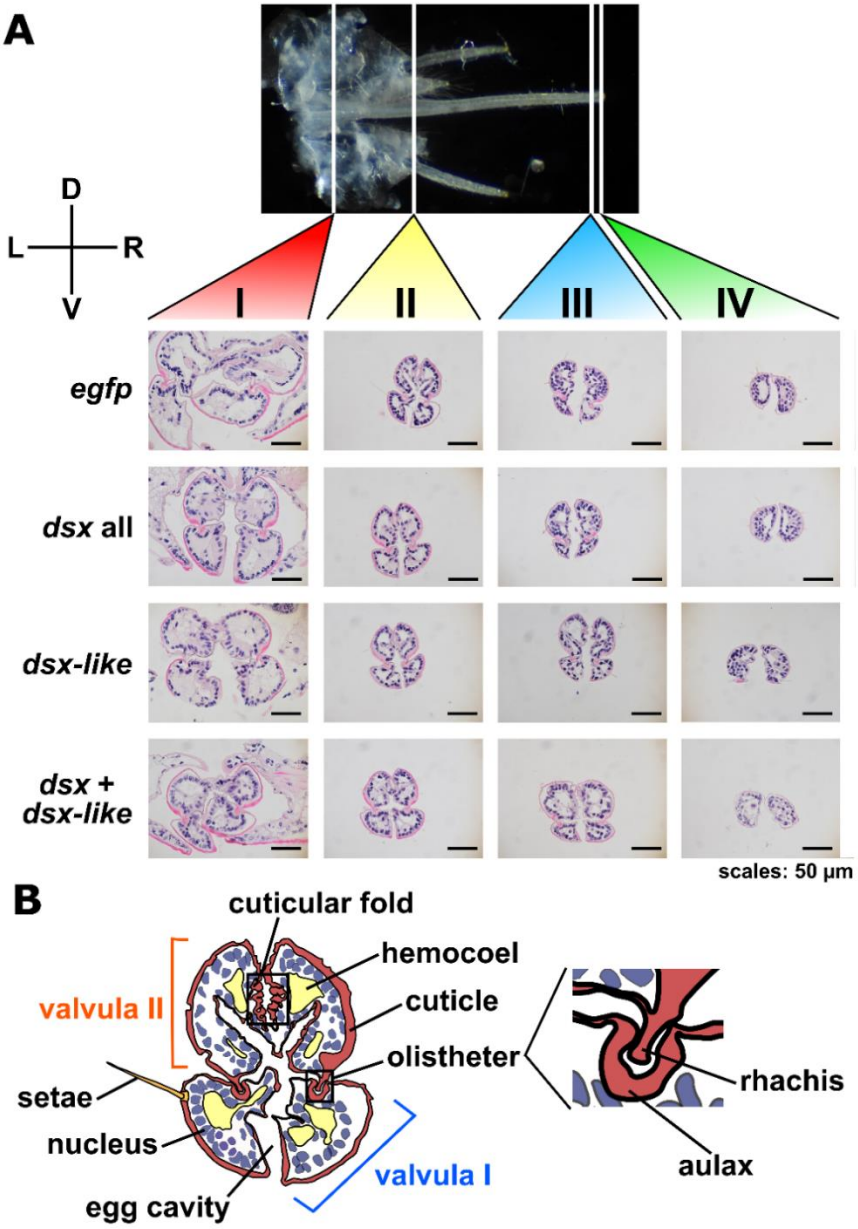
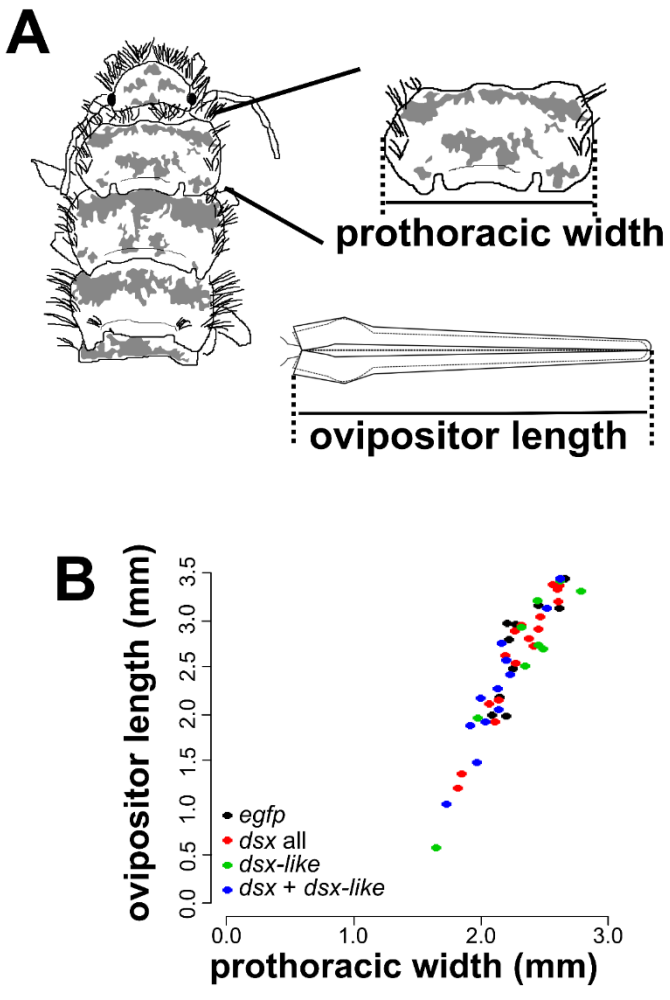


Figure 3.8. Morphology of ovipositor in nymphal RNAi individuals. (A) Cross-section of the ovipositor. The photos show the morphology of the ovipositor in four parts: I (proximal part), II (middle part), III (distal part), and IV (most-distal part). The *egfp*, *dsx* all, *dsx-like* and *dsx+dsx-like* indicates the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group, *dsx-like* dsRNA injected group, and both *dsx* sex-common region and *dsx-like* dsRNAs injected group, respectively. D, dorsal; L, left; R, right; V, ventral. Paraffin. Hematoxylin-Eosin staining. Scales: 50 μm. (B) Schematic figure of the ovipositor morphology. This figure is based on the cross-section of the part II in the control female. The part of ovipositor is constituted of two regions: valvula I and II. These regions are coordinated at the olistheter. The dorsal side of valvula II has folded cuticle.

length in the RNAi-treated females.

I measured the total length of Valvula II of the ovipositor. The length (mean±sem) was 2711.53 ± 163.93 µm for the controls, 2621.67 ± 159.94 µm for *dsx* RNAi females, 2595.89 ± 292.45 µm for *dsx-like* RNAi females, and 2266.14 ± 191.94 µm for *dsx* and *dsx-like* double RNAi females (Figure 3.9). The mean length in the experimental group was shorter than that of the control group. However, The GLM analysis found no significant differences between the control and experimental groups (Table 3.4). Therefore, there was no evidence that *dsx* and *dsx-like* controlled the growth of the ovipositor in *T. domestica*.



3.3.6. Contribution to female-specific expression of genes, vitellogenin *homologs*

dsx in *T. domestica* does not seem to show conflicting functions between sexes in postembryonic morphogenesis. Conversely, other biological processes are open to explore further. I tested whether *dsx* contributes to expression of *vitellogenin* (*vtg*), a yolk

protein precursor gene that is highly expressed in animal females (Byrne et al. 1989, Hayward et al. 2010). Previous studies show *vtg* in pterygote insects is controlled by *dsx* (e.g., Suzuki et al. 2003, Shukla and Palli 2012b, Thongsaiklaing et al. 2018). My RNA-seq analysis showed that three *vtg* homologs, i.e., *vtg1*, *vtg2*, and *vtg3*, expressed female-specificity in the fat body in *T. domestica* (Figure 3.10; Table 3.1). I analyzed expression of *vtg* in fat bodies of *dsx*, *dsx-like*, or both genes RNAi groups by RT-qPCR.

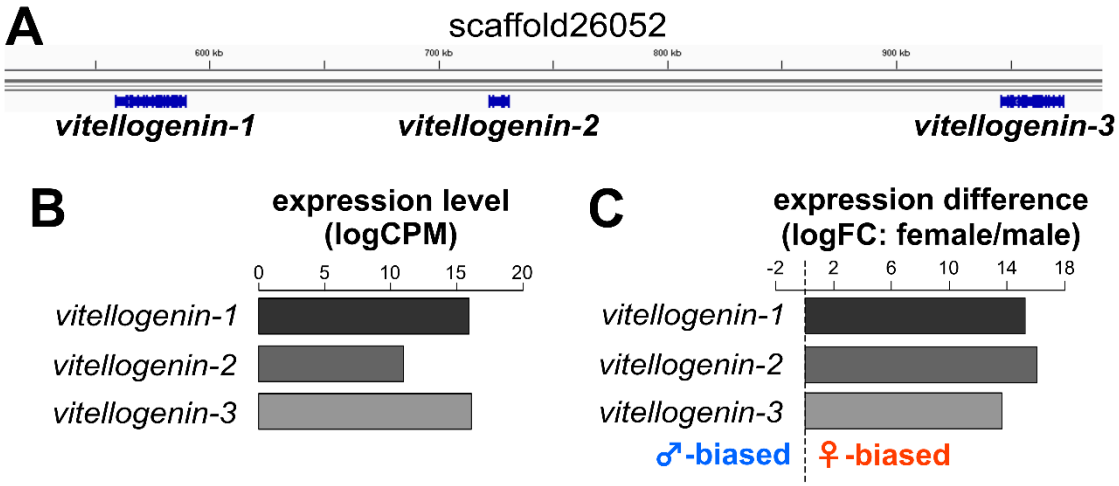


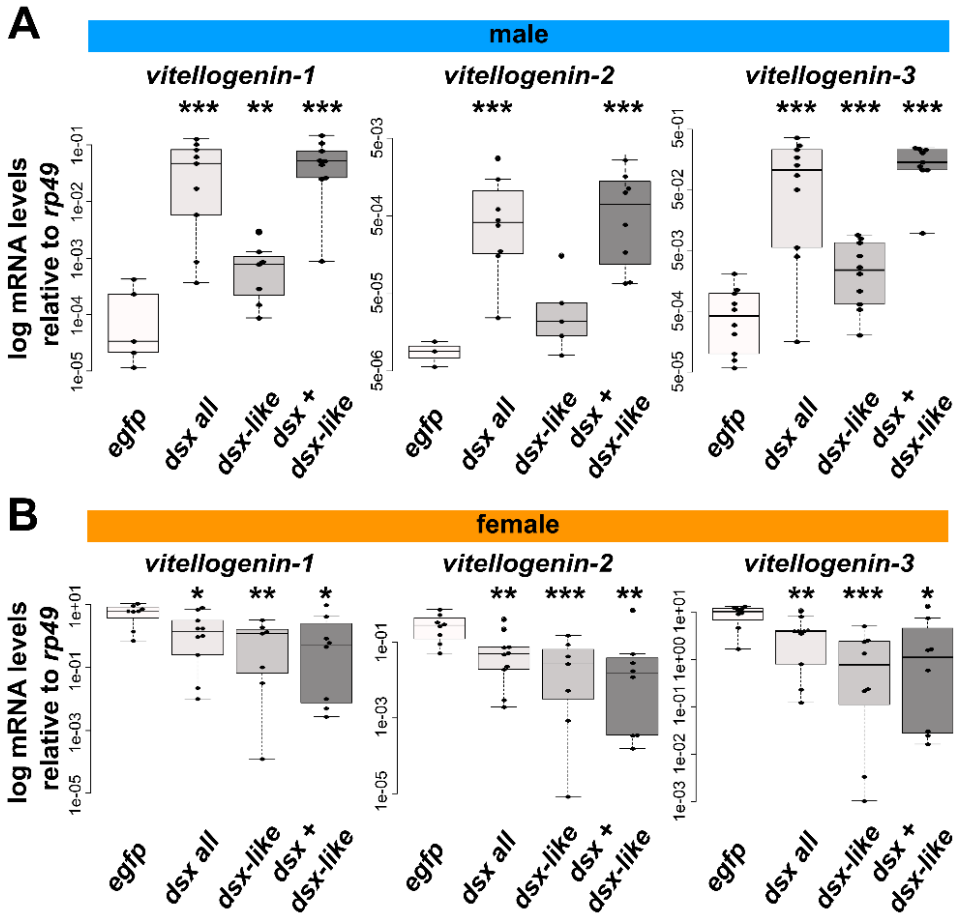
Figure 3.10. Expression of *vitellogenin* homologs in *T. domestica*. (*A*) Genome mapping of *vitellogenin* mRNA sequences. The picture is a screenshot of the integrative genome viewer (IGV). Vitellogenin genes are located on scaffold26052 of the assembled genome. *T. domestica* has three tandem-repeated *vitellogenin* homologs: *vitellogenin-1*, *vitellogenin-2*, and *vitellogenin-3*. (*B*) Expression level of *vitellogenin* homologs. The expression levels were calculated by the logCPM of transcriptome data in the fat body of males and females. All *vitellogenin* homologs show the high expression level in the fat body. (*C*) Difference in expression of *vitellogenin* homologs between sexes. The differential expression analysis was performed using the edgeR program. The expression difference is shown by the logFC value. When the values are more than 0, genes are expressed higher in females than males. Thus, all *vitellogenin* homologs shown here are expressed much higher in females than in males. Each value in (*B*) and (*C*) can be seen in Table 3.5.

Figure 3.9. Effect of RNAi treatment on the growth of the ovipositor. (*A*) The schematic images of the measured parts. (*B*) Effects of RNAi treatments on growth of ovipositor. Each plot indicates ovipositor length of each individual. The results of the generalized linear model analysis are shown in Table 3.4. Ovipositor length correlated with prothoracic width ($P = 2.00\times10^{-16}$), although any significant effects can be seen in RNAi treatments. Total $N = 38$. In each panel, the *egfp*, *dsx* all, *dsx-like*, and *dsx* + *dsx-like* indicates the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group, *dsx-like* dsRNA injected group, and both *dsx* sex-common region and *dsx-like* dsRNAs injected group, respectively.

Table 3.5. The expression level of *vitellogenin* genes in *Thermobia domestica*.

gene	logCPM	logFC
vitellogenin-1	15.93	15.28
vitellogenin-2	10.95	16.09
vitellogenin-3	16.13	13.66

In *dsx* RNAi males, all *vtg* mRNAs were expressed 45–1530-fold higher than controls (Figure 3.11A; Table 3.1: Brunner-Munzel test, $P = 2.87\times10^{-8}$, 6.60×10^{-16} and 2.80×10^{-4} in *vtg1*, *vtg2*, and *vtg3*). *vtg1* and *vtg3* mRNAs were significantly up-regulated in *dsx-like* RNAi males compared with controls (Figure 3.11A: Brunner-Munzel test, $P = 0.014$ and 0.0050 in *vtg1* and *vtg3*). In both *dsx* and *dsx-like* RNAi males, the effect was similar to that in *dsx* RNAi males (Figure 3.11A: Brunner-Munzel test, $P = 6.60\times10^{-16}$, 0.016, and 6.60×10^{-16} in *vtg1*, *vtg2*, and *vtg3*). I then found expression of all *vtg* genes was significantly reduced in *dsx* RNAi females (Figure 3.11B; Table 3.1: Brunner–Munzel test, $P = 0.043$, 0.0042, and 0.0062 in *vtg1*, *vtg2*, and *vtg3*). This reduction rate was



approximately 0.2−0.4-fold. Furthermore, *vtg* expression was significantly reduced in *dsx-like* RNAi females (Figure 3.11B; Brunner–Munzel test, $P = 0.0026$, 3.80×10^{-6}, and 1.49×10^{-5} in *vtg1*, *vtg2*, and *vtg3*) and both *dsx* and *dsx-like* RNAi females (Figure 3.11B; Brunner–Munzel test, $P = 0.031$, 0.00892, and 0.020 in *vtg1*, *vtg2*, and *vtg3*). These results show that *dsx* and *dsx-like* of *T. domestica* control *vtg* negatively in males and positively in females.

3.4. Discussion

3.4.1. Function of dsx *and* dsx-like *in* Thermobia domestica

In this chapter, I investigated the function of *dsx* and *dsx-like* genes in *T. domestica*. All previous studies on the roles of the *dsx* gene in insects focused on pterygote species, especially Holometabola. Therefore, information on *dsx* has been absent in apterygote insects. My attempt is the first study to provide data on the *dsx* gene in the apterygote as well as *T. domestica*.

3.4.1.1. Non-functionality in body size and gonadogenesis

I cannot detect any effects on the body size and gonad morphology by *dsx* and *dsx-like* RNAi. The lack of effect of RNAi on the gonads may be due to the timing of RNAi treatment, which was performed after gonadal differentiation. This is supported by a previous study (Klag 1977) suggesting that sex differences in gonads and germ cells are produced during embryogenesis. Embryonic RNAi is necessary to test this hypothesis,

Figure 3.11. Function of *doublesex* for *vitellogenin* expression in *Thermobia domestica*. (*A*) *vitellogenin* expression level in RNAi males. (*B*) *vitellogenin* expression level in RNAi females. The mRNA expression levels were measured by RT-qPCR analysis. The figures show the log-scale relative values of expression levels of three *vitellogenin* homologs to the reference gene, *ribosomal protein 49* (*rp49*). Each plot indicates mRNA expression levels of each individual. In each panel, the *egfp*, *dsx* all, *dsx-like*, and *dsx* + *dsx-like* indicates the *egfp* dsRNA injected group (control), *dsx* sex-common region dsRNA injected group, *dsx-like* dsRNA injected group, and both *dsx* sex-common region and *dsx-like* dsRNA injected group, respectively. The Brunner–Munzel test was performed to statistically analyze the difference in mRNA expression level between control and *dsx* or *dsx-like* RNAi groups. *P*-values were adjusted by the Holm's method. $*P < 0.05$, $**P < 0.001$, $***P < 0.0001$. $P \geq 0.05$ is not shown. Statistical results are described in Table 3.1. Total $N = 30$ (*vitellogenin-1*), 24 (*vitellogenin-2*) and 39 (*vitellogenin-3*) in males and 33 (*vitellogenin-1*), 33 (*vitellogenin-2*), and 34 (*vitellogenin-3*) in females.

although this experiment will be left to future studies, as my study focused on *dsx* function during postembryonic development. The lack of effect of *dsx* on the body size of *T. domestica* is consistent with studies in *D. melanogaster* (Hildreth 1965, Rideout et al. 2015).

3.4.1.2. Function in genital organ and fecundity

In this chapter, I found that the *dsx* RNAi in *T. domestica* affects male seminal vesicle morphology, storage number of sperm, and penial structure, while it failed to show effects on the female internal reproductive system, oocyte storage, and ovipositor morphology and growth. In addition, the *dsx-like* RNAi of the *T. domestica* showed no effects on sexually dimorphic morphology or gamete number in both sexes. These results indicate that *dsx* affects male differentiation of the sexually dimorphic morphology in *T. domestica*.

The lack of effect of the *dsx* RNAi in females may rely not only on the non-functionality of *dsx* on the female morphology in this species but also on the inefficiency of the RNAi. Indeed, the *dsx* RNAi efficiency is lower in females (~50%) than in males (~30%) based on the median (Figure 3.1). However, in this efficiency analysis, half of the females showed *dsx* expression levels below the lowest value in the control group. In addition, all *dsx* RNAi males in this analysis showed the effect on the morphology, while all *dsx* RNAi females did not (58 females in the *dsx* RNAi, 80 females including the *dsx* and *dsx-like* double RNAi). Given these results, it seems reasonable that the lack of effect of the *dsx* RNAi on female morphology in *T. domestica* is a consequence of non-functionality in *dsx* for female morphogenesis, rather than entirely depends on the RNAi inefficiency. Since all *dsx-like* RNAi individuals showed a lower level of *dsx-like* expression than that of the control group (Figure 3.1), the lack of effect of the *dsx-like* RNAi on male and female morphology suggests that *dsx-like* is not essential for male or female morphogenesis during post-embryonic development. The knockdown of both *dsx* and *dsx-like* showed only the same effect as the *dsx* RNAi alone. Therefore, it is unlikely that *dsx-like* functions redundantly with *dsx*. Thus, at this time, *dsx* and *dsx-like* in *T. domestica* are unlikely to be essential for female morphogenesis.

In summary, the analysis in this chapter suggests that *dsx* is required for male differentiation of sexual morphogenesis during post-embryonic development but not for female differentiation. Hence, it is suggested that *dsx* does not have a sex-antagonistic

function, i.e., promoting either male or female differentiation in the formation of sexually dimorphic morphology in *T. domestica*. Further, according to my results in this chapter, *dsx-like* does not seem to compensate for the function of *dsx* in the morphogenesis.

3.4.1.3. Significance of sex-specific splicing of *dsx* for sexual morphogenesis

The functional analysis in this chapter indicates that *dsx* is not essential for female morphogenesis during post-embryonic development in *T. domestica*. Thus, expression of *dsx* is not seem to be necessary for female morphogenesis. What is the significance of *dsx* in female morphogenesis? In general, alternative splicing tunes up the balance between functional and non-functional isoforms and results in precise physiological activity in cells and tissues (Keren et al. 2010). Similarly, it is possible that females of this species express the non-functional *dsx* isoforms for morphogenesis and thereby might prevent masculinization by expression of the functional *dsx*. This opinion is supported by the result that *dsx* in this species is essential for male differentiation of morphogenesis. Thus, although I cannot rule out other possibilities, this balancer hypothesis is a strong candidate for the significance of *dsx* splicing in females (see section 5.4).

3.4.2. Comparison of dsx *functionality for sexual morphogenesis in Insecta*

In this chapter, I conclude that, in *T. domestica* (Zygentoma), sexual morphology, e.g., reproductive systems and genital organs, formed during postembryonic development is controlled by *dsx* in males but is *dsx*-independent in females. The similar situation has been reported from non-holometabolan insects such as *Bl. germanica* (Dictyoptera: Wexler et al. 2019), and the brown planthopper *Nilaparvata lugens* (Hemiptera: Zhuo et al. 2018). Based on these data, I infer that *dsx* may not have been required for female differentiation of morphology at the common ancestor of Dicondylia (= Zygentoma + Pterygota) emerging ~421 Ma. This evolutionary model is supported by the recent studies on a sexual dimorphic body pigmentation in an Odonata species, *Ischnura senegarensis* (Takahashi et al., 2019, 2021). Moreover, my results and inference strongly support the hypothesis proposed by Wexler et al. (2019) that *dsx* had acquired the sex-specific splicing isoforms and later became essential for female differentiation.

To elucidate the timing of acquisition of roles of *dsx* in female morphogenesis during postembryonic development, I must interpret the role of *dsx* in Hymenoptera, the sister clade of the other Holometabola. Studies on the honeybee *Apis mellifera* showed through

genome editing that *dsx* controls female differentiation of the internal reproductive system under worker nutrition conditions (Roth et al. 2019). In the honeybee, sex differences in the gonads are established during embryogenesis (Lago et al. 2020). Therefore, the male-like reproductive organ in *dsx* mutant females in Roth et al. (2019) would show an effect during embryogenesis, not during postembryonic development. I cannot conclude whether *dsx* is non-essential for female morphogenesis in the honeybee, since information on the roles of *dsx* in sexual morphology is limited to gonads and heads of worker females. However, given that *dsx* does not affect heads in *Ap. mellifera* females (Roth et al. 2019), leg pigmentation, pheromone synthesis, and wing morphology in the parasitoid wasp *Nasonia vitripennis* females (Wang et al. 2022a, b), and sexual traits in *At. rosae* females (Mine et al. 2017, 2021), currently it is reasonable to infer that *dsx* is non-essential for female morphogenesis during postembryonic development in the common ancestor of Hymenoptera. This interpretation and the essential roles of *dsx* for female development in other holometabolan insects suggest that *dsx* became essential for feminization of morphology during postembryonic development at the common ancestor of holometabolan insects except for Hymenoptera (Aparaglossata) emerging ~327 Ma. In this hypothesis, *dsx* played the opposite role between sexes in sexual morphogenesis more than 100 million years after its female-specific isoform had appeared.

3.4.3. Functionality for female-specific gene expression and its comparison

In this chapter, I found that *dsx* performs opposite roles between sexes, i.e., repressive in males and promotive in females in *vtg* expression. *dsx-like* also shows sex-antagonistic functions for *vtg* expression. It is unlikely that this result is due to *dsx-like* regulating *dsx* transcription, as *dsx-like* did not affect *dsx* expression (Figure 3.2A). A possible hypothesis is that *dsx-like* may regulate *vtg* expression as a co-regulator that binds *dsx* or other transcription factors.

I do not know whether *dsx* of *T. domestica* oppositely controls genes other than *vtg* homologs between sexes since my analysis was limited to *vtg* homologs. However, results from these genes indicates that molecular function of *dsx* in this species includes the opposite function for some gene transcription in females and males. In *Be. tabaci*, *dsx* positively regulates *vtg* expression in females, even though it is non-essential for female differentiation of morphological traits (Guo et al. 2018). *dsx* of this species does not

negatively regulate *vtg* in males. Therefore, the functionality of *dsx* found in *T. domestica*, i.e., the opposing role in some genes' expression between sexes and functions that are non-essential for female morphogenesis, is a functionality that has not been reported in any insect or animal. These results show that *dsx* in females came into use separately among morphogenesis and other biological processes in evolution. (discussed in section 5.4.3).

Genes under *dsx* control in males are *dsx*-free in females of *I. senegalensis* (Takahashi et al. 2021), *Bl. germanica* (Wexler et al. 2019; Pei et al. 2021), and *Ni. lugens* (Zhuo et al. 2018). Feminizing roles of *dsx* in morphogenesis and other biological processes may have appeared in the common ancestor of Aparaglossata (or Holometabola) as an entirely novel function, i.e., neofunctionalization. In contrast, the contribution of *dsx* to some genes' expression in females of *T. domestica* (this study), *Be. tabaci* (Guo et al. 2018), *Ap. mellifera* (Velasque et al. 2018), and Aparaglossata raises the alternative hypothesis that the capacity for *dsx* contributing to female differentiation was already present in the common ancestor of Dicondylia. In this evolutionary model, the role of *dsx* in feminization of postembryonic morphogenesis in Aparaglossata could be due to extending its capability to control some genes in females. I currently cannot decide which of these hypotheses is appropriate. However, the latter hypothesis can reliably explain the presence of female-specific coding sequences of *dsx* and high expression of *dsx* female-type at the postembryonic stage, in non-aparaglossatan insects (discussed in section 5.4.3). The capacity to regulate some female genes may be a "minor function" of *dsx* in some non-holometabolan females, as predicted by Wexler et al. (2019).

3.4.4. Evolutionary history of dsx *functionality*

In this chapter, I investigated the function of *dsx* in *T. domestica* and inferred the status in the common ancestors of Dicondylia and Pterygota. This attempt would disambiguate the evolutionary history of *dsx* in Pterygota.

Currently, *dsx* is reported to be expressed in a male-specific manner and to be essential only for male differentiation in several species of crustaceans from Branchiopoda and Decapoda (Kato et al. 2011, Li et al. 2018). It is also known that *dsx* is expressed in a male-specific manner in other crustaceans and Chelicerata, although functional analysis has not been conducted in these species (Pomerantz et al. 2015, Gruzin et al. 2020). In

light of these facts and the discussion in this chapter, the basic function of *dsx* in arthropods is to promote male differentiation. This assumption is supported by the fact that Dmrt1 and Mab-3, the DMRT family members, are responsible for promoting male differentiation in vertebrates (Raymond et al. 2000, Kobayashi et al. 2004, Koopman 2009) and nematodes (Yi and Zarkower 1999), respectively. The role of *dsx* in morphogenesis may have transitioned from a monofunctional role in male differentiation, which it has long maintained during arthropod evolution, to an antagonistic function for male and female differentiation in the common ancestor of Aparaglossata.

The birth of the sex-antagonism of *dsx* in morphogenesis can be attributed to the acquisition of the female differentiation function, which is achieved by a female-specific isoform of *dsx* in Aparaglossata (e.g., Hildreth 1964, Burtis and Baker 1989, Kijimoto et al. 2012, Ohbayashi et al. 2001, Shukla and Palli 2012b, Ito et al. 2013, Gotoh et al. 2016, Xu et al. 2017). Based on the discussion in this chapter, the female-specific isoform was initially nonfunctional for morphogenesis and is assumed to have served as a balancer of isoforms that prevented the original masculinizing function of *dsx* in females. The balancer role of the female-specific splicing may have been co-opted for the function to promote female morphogenesis in the common ancestor of Aparaglossata. Therefore, the female-specific splicing could be regarded as a mechanism that resulted in a neo-functionalization to *dsx* without losing its original function of promoting male differentiation.

In this chapter, I propose possible evolutionary models for the functional evolution of *dsx*. Most importantly, female isoforms of *dsx* that were nonfunctional for morphogenesis were transformed to be responsible for female morphogenesis in the course of insect evolution. One remaining question is what changes in genomic space resulted in the feminizing function of *dsx* (Hopkins and Kopp 2021). In the next chapter, I will attempt to consider this question by investigating the molecular evolution of *dsx*.

Chapter 4

Evolution of C-terminal Motif in Dsx of Insecta

Abstract The *doublesex* plays crucial roles in both male and female differentiation in Aparaglossata. There has been mounting evidence that *dsx* obtained its feminizing role in sexual morphology more than 100 million years after the female-specific isoform appeared. However, what change in genomic space gave rise to the function of *doublesex* in female differentiation of sexual morphology remains unknown. To this end, I infer the ancestral sequences of the female-specific region and predict the structure of the ancestor proteins. Here, I uncover that the C-terminal segment of the *doublesex* occurred at the common ancestor of Aparaglossata. This result indicates that the C-terminal region of the female-specific isoform was extended after the Aparaglossata diverged. I also provide evidence that the Aparaglossata-specific region is disordered. Hence, the extension of the C-terminal disordered motif correlates with the appearance of the feminizing function of the *doublesex*. I propose that the C-terminal extension of the female-specific isoform is a candidate for the genomic change causing the neo-functionalization of the *doublesex*.

4.1. Introduction

Co-option of genes is achieved through various changes on genomic space, e.g., the acquisition of paralogs through gene duplication, the accumulation of mutations in cis-regulatory regions, the neo-functionalization of co-factors, and mutations in the coding (e.g., Ganfornina and Sánchez 1999, Mann and Carroll 2002, Carroll 2005). What changes link to the neo-functionalization of isoforms? This question has long been debated (e.g., Boue et al. 2003), and *dsx* would provide a case study for this question.

In Chapters 2 and 3, I inferred that the female-specific isoforms of dsx were initially non-functional for morphogenesis, followed by obtaining the roles in female differentiation in the common ancestor of Aparaglossata. Since the acquiring timing differs much between the female-specific splicing and function for female morphogenesis,

the feminizing role in morphogenesis of *dsx* cannot be explained only by the emergence of the female-specific new exon. One of the remaining questions is what changes led to the feminizing function of *dsx* for morphogenesis (Hopkins and Kopp 2021). Currently, there is no hypothesis to answer this question. It is unlikely that the neo-functionalization in *dsx* is explained by gene duplication, accumulation of mutations in cis-regulatory regions, or gain of co-factor function. First, the feminizing role is unlikely to be derived from the paralog because of my finding that *dsx-like* does not affect morphological feminization (Chapter 3). Second, it may be hard to explain the neo-functionalization by the emergence of novel targets of *dsx* since the male- and female-specific isoforms share the DNA binding domain (DM domain) (Chapter 2). Third, the novel roles of co-factors do not seem to lead to the feminizing function of *dsx* since the *intersex*, the essential partner of *dsx* for female differentiation (Yang et al. 2008, Gotoh et al. 2016, Morita et al. 2019, Xu et al. 2019), is involved in female morphogenesis even in *Ni. lugens* (Zhang et al. 2021), where *dsx* does not contribute to female differentiation in morphology (Zhuo et al. 2018). The remaining candidate is a change in the coding regions. Thus, I examined the following possibilities in this chapter: the accumulation of mutations in the amino acid sequence or the acquisition of new exons.

In this chapter, to examine the change in the coding region, I focus on the female-specific region of *dsx*. In *D. melanogaster*, the female-specific region is essential for female differentiation via physically binding to itself, other transcription factors, and transcript cofactors (An and Wensink 1995, Erdman 1996, Ghosh et al. 2019, Romero-Pozuelo et al. 2019). In addition, the comprehensive study analyzing the dN/dS ratio of *dsx* in Holometabola shows that the female-specific region is conserved in Diptera, Lepidoptera, and Coleoptera, whereas they are diversified in Hymenoptera (Baral et al. 2019). This result suggests that, in the former orders, the region is constrained by strong purifying selection. Thus, the female-specific region may be functionally important in these orders of Aparaglossata, whereas the functional significance may be lower in Hymenoptera. The difference in the selection also implies that roles of the female-specific region may be changed in the course of evolution. Thus, the changes in the female-specific region that occurred in the common ancestor of Aparaglossata may be a candidate factor linked to the neo-functionalization of *dsx*. To this end, I compared the sequence of the female-specific region between *D. melanogaster* and *T. domestica*, reconstructed

ancestral sequences of the region, predicted *dsx* protein structures, and examined exon-intron structures in insects.

4.2. Methods

4.2.1. Ancestral Sequence Reconstruction

To infer the sequence evolution of *dsx*, I conducted an ancestral sequence reconstruction (ASR) of the C-terminal sequences of the *dsx* female-type homologous sequence. First, I searched homologous sequences to *dsx* female-type from NCBI protein/transcript shotgun assembly databases and previous studies. The searches in the NCBI databases were performed by BLAST search. I closely examined the alignment results of the BLAST and selected sequences with at least 10 amino acids aligned with the female-specific region of each query sequence. I do not know whether some of these sequences are expressed in females and contribute to female morphogenesis, as these sequences are not necessarily to have investigated expression and function in the species. I decided that it was not problem to use these sequences since I focused on the evolution of sequences homologous to *dsx* female-type in each insect taxa. In Diptera, I set *dsx* female-type of *D. melanogaster* (Accession #: NP_001287220) as a query and obtained 9 sequences. In Lepidoptera, I used *dsx* female-type of *Bo. mori* (NP_001036871) as a query and get 10 sequences. In Coleoptera, *dsx* female-type of *Tribolium castaneum* (AFQ62106) was set in a query and then 10 sequences were obtained. I used *dsx* female-type of *Ap. mellifera* (NP_001128407) and *At. rosae* (XP_012262256) as queries to search hymenopteran sequences. I also searched some hymenopteran sequences from the NCBI databases based on a previous study (Baral et al. 2019). 10 hymenopteran sequences were obtained. In Psocodea and Hymenoptera, I searched the databases to set the sequences of *Pediculus humanus* (QGB21102) and *Rhodonius prolixus* (QGB21099) as queries. Wexler et al. (2019) showed that *dsx* of *Pediculus humanus* (Psocodea) has isoforms without sex-specificity. In this study, based on the blast search and exon structure, I regarded that the PhDsx1 in Wexler et al. (2019) is homologous to the *dsx* female-type. The sequences of *Ni. lugens* (AWJ25056) and *Bl. germanica* (QGB21105 and QGB21106) were obtained from the database based on previous studies (Zhuo et al. 2018, Wexler et al. 2019). I selected two sequences from *Bl. germanica*, as this species

has two female-specific *dsx* isoforms (Wexler et al. 2019). The sequences of *Cryptocercus punctulatus* and *I. senegalensis* were obtained from previous studies (Miyazaki et al. 2021, Takahashi et al. 2021). In *T. domestica*, the sequence identified in this study was used. The sequence names are listed in Table 4.1. I then manually extracted the OD domain and performed multiple sequence alignments (MSA) using the MAFFT version 7 (Katoh et al. 2013) with the -linsi option to use an accuracy option, L-INS-i. I reconstructed ancestral sequences (AS) from the MSA using MEGA X software. The maximum-likelihood method was applied to the ASR. The JTT + G model was chosen as a substitution model by AIC-based model selection. The guide tree was reconstructed based on previously reported phylogenetic relationships (Wiegmann et al. 2011, Misof et al. 2014, Li et al. 2017, Peters et al. 2017, Zhang et al. 2018, Kawahara et al. 2019, McKenna et al. 2019, Gustafson et al. 2020) (Figure 4.1). I selected the most probable sequences for the following analyses. In Aparaglossata (Node 77) and Holometabola (Node 87) AS, almost all probabilities of sites were more than 0.9. The except sites were sites 83 and 98 in Node 77 and sites 77–79 and 83 in Node 87. These sites other than sites 77 had probabilities > 0.5. Thus, I concluded that the AS in Aparaglossata and Holometabola, which I considered the most critical, was reconstructed with sufficient reliability. Any residues had the probabilities = 0 in the Aparaglossata-specific region of Holometabola AS. In contrast, in non-holometabolan insects, since my taxon sampling is limited to several species (Eumetabola in Node 92, Neoptera in Node95, Pterygota in Node 96), the probabilities of some sites are lower than 0.5. These low probable sites are not necessarily confident. To conclude with reliability, it is no doubt that analyses based on a larger number of species will be essential. However, all sites of the Aparaglossata-specific region in these AS were gaps with the probabilities > 0.9. The result of the sites of the Aparaglossata-specific region seems to be relatively reliable in my analysis. Thus, my conclusion that the Aparaglossata-specific region occurred in the common ancestor of Aparaglossata would be confident. To compare the sequences, I then performed MSA of the most probable reconstructed ancestral sequences and the sequence of *D. melanogaster* using MAFFT version 7.

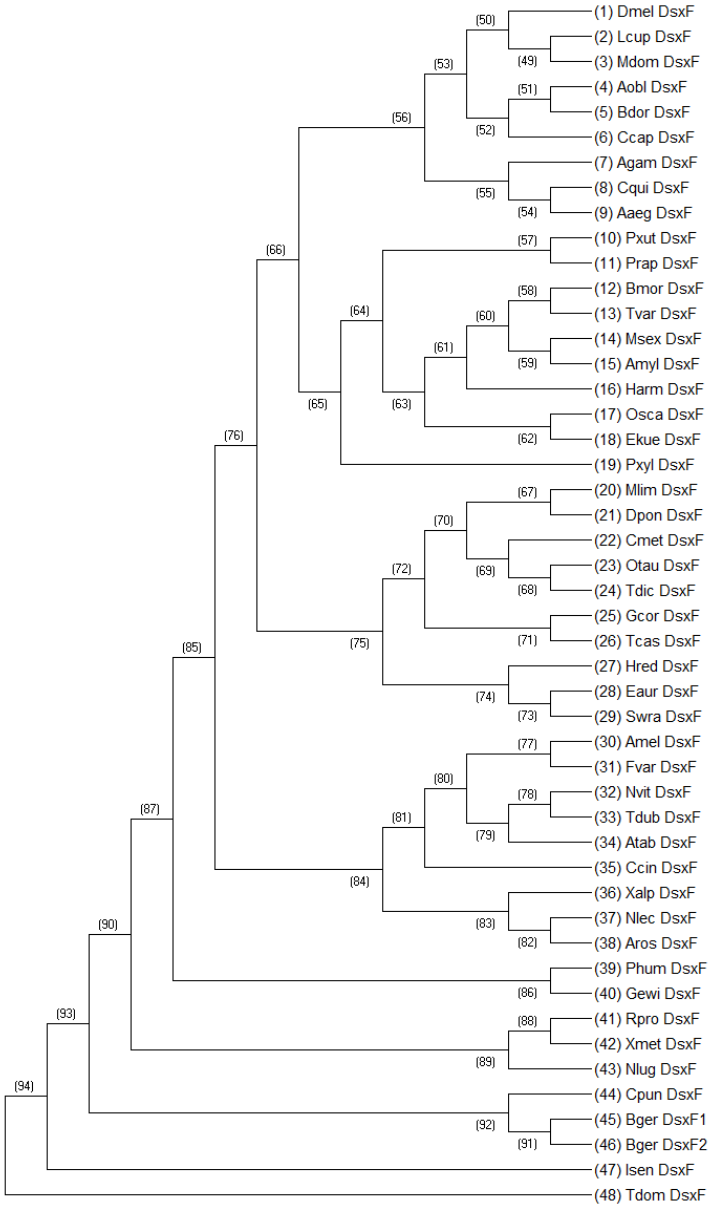


Figure 4.1. The guide tree used for the ancestral sequence reconstruction. The tree topology was reconstructed based on previous phylogenetic studies (Wiegmann et al. 2011, Misof et al. 2014, Li et al. 2017, Peters et al. 2017, Zhang et al. 2018, Kawahara et al. 2019, McKenna et al. 2019, Gustafson et al. 2020). The topology is here: "((((((((((Dmel_DsxF,(Lcup_DsxF,Mdom_DsxF)),((Aobl_DsxF,Bdor_DsxF),Ccap_DsxF)),(Agam_DsxF,(Cqui_DsxF,Aaeg_DsxF))),(((Pxut_DsxF,Prap_DsxF),((((Bmor_DsxF,Tvar_DsxF),(Msex_DsxF,Amyl_DsxF)),Harm_DsxF),(Osca_DsxF,Ekue_DsxF))),Pxyl_DsxF)),((((Mlim_DsxF,Dpon_DsxF),(Cmet_DsxF,(Otau_DsxF,Tdic_DsxF))),(Gcor_DsxF,Tcas_DsxF)),(Hred_DsxF,(Eaur_DsxF,(Swra_DsxF,Ains_DsxF))))),((((Amel_DsxF,Fvar_DsxF),((Nvit_DsxF,Tdub_DsxF),Atab_DsxF)),Ccin_DsxF),(Xalp_DsxF,(Nlec_DsxF,Aros_DsxF)))),(Phum_DsxF,Gewi_DsxF)),((Rpro_DsxF,Xmet_DsxF),Nlug_DsxF)),(Cpun_DsxF,(Bger_DsxF1,Bger_DsxF2))),Isen_DsxF),Tdom_DsxF);". The OTU names can be referred to in Table 4.1.

Table 4.1. The taxa list used for the ancestral sequence reconstruction of *dsx*.

ID	Order	Family	species	Accession No.	source	database2
Agam_DsxF	Diptera	Culicidae	*Anopheles gambiae*	AAZ78363	NCBI	Protein
Cqui_DsxF	Diptera	Culicidae	*Culex quinquefasciatus*	AJB28478	NCBI	Protein
Aaeg_DsxF	Diptera	Culicidae	*Aedes aegypti*	ABD96571	NCBI	Protein
Aobl_DsxF	Diptera	Tephritidae	*Anastrepha obliqua*	AAY25166	NCBI	Protein
Bdor_DsxF	Diptera	Tephritidae	*Bactrocera dorsalis*	ACN24617	NCBI	Protein
Lcup_DsxF	Diptera	Calliphoridae	*Lucilia cuprina*	ADG37649	NCBI	Protein
Dmel_DsxF	Diptera	Drosophilidae	*Drosophila melanogaster*	NP_001287220	NCBI	Protein
Mdom_DsxF	Diptera	Muscidae	*Musca domestica*	AAR23812	NCBI	Protein
Ccap_DsxF	Diptera	Tephritidae	*Ceratitis capitata*	AAN63598	NCBI	Protein
Bmor_DsxF	Lepidoptera	Bombycidae	*Bombyx mori*	NP_001036871	NCBI	Protein
Tvar_DsxF	Lepidoptera	Bombycidae	*Trilocha varians*	BAS02078	NCBI	Protein
Amyl_DsxF	Lepidoptera	Saturniidae	*Antheraea mylitta*	ADL40853	NCBI	Protein
Msex_DsxF	Lepidoptera	Sphingidae	*Manduca sexta*	XP_037293923	NCBI	Protein
Harm_DsxF	Lepidoptera	Noctuidae	*Helicoverpa armigera*	AHF81656	NCBI	Protein
Osca_DsxF	Lepidoptera	Crambidae	*Ostrinia scapulalis*	BAJ25851	NCBI	Protein
Ekue_DsxF	Lepidoptera	Pyralidae	*Ephestia kuehniella*	CAG7465060	NCBI	Protein
Pxut_DsxF	Lepidoptera	Papilionidae	*Papilio xuthus*	XP_013171086	NCBI	Protein
Prap_DsxF	Lepidoptera	Pieridae	*Pieris rapae*	BBA83992	NCBI	Protein
Pxyl_DsxF	Lepidoptera	Plutellidae	*Plutella xylostella*	XP_037963447	NCBI	Protein
Hred_DsxF	Coleoptera	Hydroscaphidae	*Hydroscapha redfordi*	GDMJ01014513	NCBI	TSA
Swra_DsxF	Coleoptera	Aspidytidae	*Sinaspidytes wrasei*	GDNH01030794	NCBI	TSA
Eaur_DsxF	Coleoptera	Carabidae	*Elaphrus aureus*	GDPI01009550	NCBI	TSA
Tcas_DsxF	Coleoptera	Tenebrionidae	*Tribolium castaneum*	AFQ62106	NCBI	Protein
Gcor_DsxF	Coleoptera	Tenebrionidae	*Gnatocerus cornutus*	BAW32685	NCBI	Protein
Tdic_DsxF	Coleoptera	Scarabaeidae	*Trypoxylus dichotomus*	BAM93343	NCBI	Protein

Otau_DsxF	Coleoptera	Scarabaeidae	*Onthophagus taurus*	AEX92940	NCBI	Protein
Cmet_DsxF	Coleoptera	Lucanidae	*Cyclommatus metallifer finae*	BAO23811	NCBI	Protein
Dpon_DsxF	Coleoptera	Curculionidae	*Dendroctonus ponderosae*	XP_019767419	NCBI	Protein
Mlim_DsxF	Coleoptera	Ripiphoridae	*Macrosiagon limbatum*	GDPU01025064	NCBI	TSA
Aros_DsxF	Hymenoptera	Tenthredinidae	*Athalia rosae*	XP_012262256	NCBI	Protein
Nlece_DsxF	Hymenoptera	Diprionidae	*Neodiprion lecontei*	GEES01058869	NCBI	TSA
Xalp_DsxF	Hymenoptera	Xyelidae	*Xyela alpigena*	GBVH01020704	NCBI	TSA
Ccin_DsxF	Hymenoptera	Cephidae	*Cephus cinctus*	GEFG01014249	NCBI	TSA
Dall_DsxF	Hymenoptera	Braconidae	*Diachasma alloeum*	THK32978	NCBI	Protein
Atab_DsxF	Hymenoptera	Braconidae	*Asobara tabida*	ATE86739	NCBI	Protein
Nvit_DsxF	Hymenoptera	Pteromalidae	*Nasonia vitripennis*	NP_001155990	NCBI	Protein
Tdub_DsxF	Hymenoptera	Pteromalidae	*Trichomalopsis dubius*	ACJ65505	NCBI	Protein
Fvar_DsxF	Hymenoptera	Apidae	*Frieseomelitta varia*	QEK21874	NCBI	Protein
Amel_DsxF	Hymenoptera	Apidae	*Apis mellifera*	NP_001128407	NCBI	Protein
Phum_DsxF	Psocodea	Pediculidae	*Pediculus humanus*	QGB21102	NCBI	Protein
Gewi_DsxF	Psocodea	Trichodectidae	*Geomydoecus ewingi*	GCXD01024249	NCBI	TSA
Rpro_DsxF	Hemiptera	Reduviidae	*Rhodnius prolixus*	QGB21099	NCBI	Protein
Xmet_DsxF	Hemiptera	Peloridiidae	*Xenophyes metoponcus*	GDEM01064407	NCBI	TSA
Nlug_DsxF	Hemiptera	Delphacidae	*Nilaparvata lugens*	AWJ25056	NCBI	Protein
Cpun_DsxF	Dictyoptera	Cryptocercidae	*Cryptocercus punctulatus*		Miyazaki et al. (2021)	
Bger_DsxF1	Dictyoptera	Ectobiidae	*Blattella germanica*	QGB21105	NCBI	Protein
Bger_DsxF2	Dictyoptera	Ectobiidae	*Blattella germanica*	QGB21106	NCBI	Protein
Isen_DsxF	Odonata	Coenagrionidae	*Ischnura elegans*	-	Takahashi et al. (2019)	-
Tdom_DsxF	Zygentoma	Lepistimatidae	*Thermobia domestica*	-	this study	-

4.2.2. Protein Structure Prediction

To infer the evolution of protein structures of *dsx*, I conducted the protein structure prediction. The ancestral sequences reconstructed by the above section were used for the protein structure prediction. The protein structure prediction was performed using the

Alphafold2-based algorism (ColabFold: Mirdita et al. 2021) with the default option. The accuracy of predictions was evaluated based on the predicted Local distance difference test (plDDT) score that was automatically calculated on the ColabFold. I selected a model with the highest average plDDT score in each prediction. The average plDDT scores were 81.824 (Aparaglossata), 89.165 (Holometabola), 87.376 (Eumetabola), 90.721 (Neoptera), and 90.720 (Pterygota). The plDDT scores were more than 70 in the helical structure predicted as the α-helix loop of the female-specific *dsx* region. Generally, predicted structures of plDDT>70 are regarded to be a confident prediction (cf., Tunyasuvunakool et al. 2021). Therefore, I assessed the α-helix loop of the female-specific region of *dsx* as the confidently predicted structure. The graph of the plDDT score of each model is shown in Figure 4.2. The 3D models of predicted structures were visualized with the PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC.). On the viewer, I colored the female-specific region and the Aparaglossata-specific region with red color and the green color, respectively.

4.2.3. Comparisons of exon-intron structure

To compare the exon-intron structure among insects, I took information of the gene structure of *dsx* from the previous studies: Bruce and Baker (1989) in *Drosophila melanogaster*, Mine et al. (2017) in *Athalia rosae*, Wexler et al. (2019) in *Pediculus humanus*, *Rhodonius proxius*, and *Blattella germanica*, Takahashi et al. (2019) in *Ischnura elegans*, this study in *T. domestica*. Then, I mapped these structures on the phylogenetic relationship of Insecta (Misof et al. 2014). In addition, to investigate the conservation of the exon-intron junction of *dsx* among insects, I compared the sequences of the Dsx isoforms among some insect species. First, the sex-specific isoforms within species were aligned and identified the junction. Then, the junction of the female-specific isoform among species were compared by multiple sequence alignments. The multiple sequence alignments were conducted by the MAFFT software.

4.3. Results

Here, I found that the C-terminal sequences, including the oligomerization (OD) domain of the *dsx* female-type, is much shorter in *T. domestica* (38 aa) than in *D.*

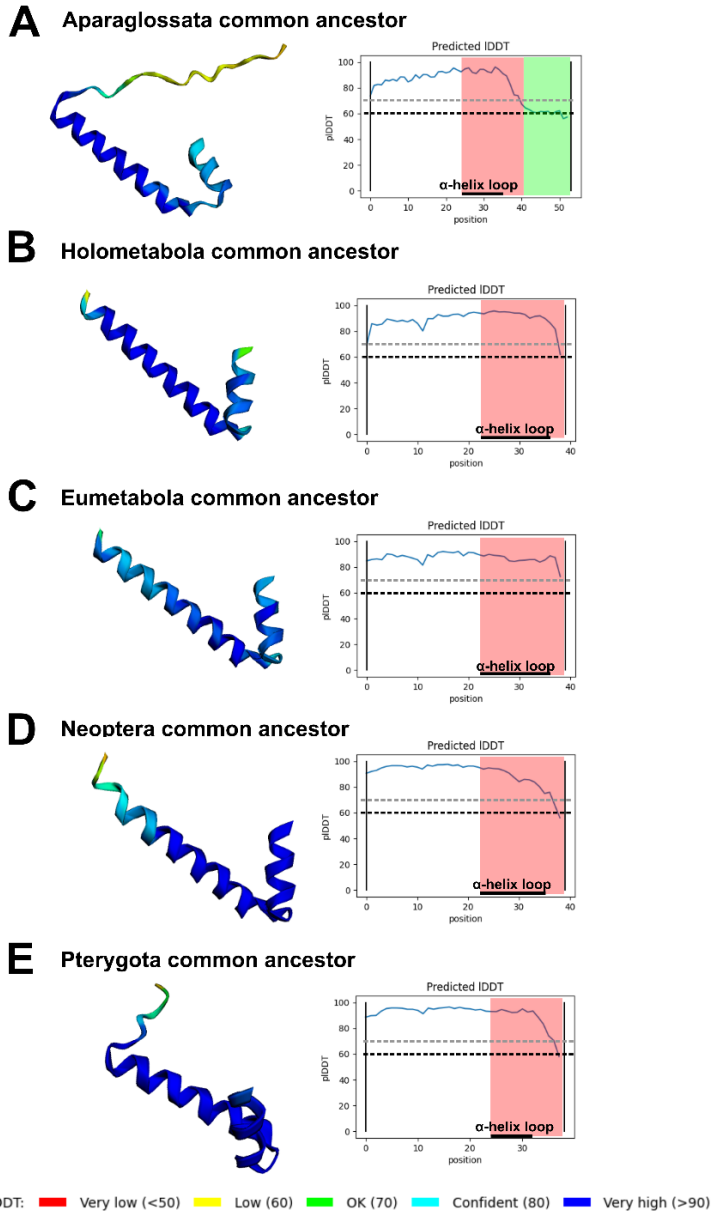


Figure 4.2. Accuracy of structure predictions of *dsx* female-type. Results of prediction of *dsx* female-type structure in the common ancestor of Aparaglossata (*A*), the common ancestor of Holometabola (*B*), the common ancestor of Eumetabola (*C*), the common ancestor of Neoptera (*D*), and the common ancestor of Aparaglossata (*D*). In each panel, the right 3D model shows the predicted structure of *dsx* female-type colored by its predicted local distance difference test (plDDT) score. The legend of color in the 3D model is shown at the bottom of the figure. The left graph indicates the plDDT score in each residue. The female-specific region is shown by red background. The Aparaglossata-specific region is colored by green. The black bar at the bottom of each graph shows the region predicted as an α-helix loop in the female-specific region. The black and gray dotted lines indicate plDDT = 60 and 70.

melanogaster (53 aa) (Figure 4.3). Therefore, I hypothesized that the additive region

found in *D. melanogaster* occurred at the common ancestor of Aparaglossata in which *dsx* became essential for female morphogenesis. To test this hypothesis, I obtained sequences of *dsx* female-type from 48 insect species based on the National Center for Biotechnology Information protein/transcriptome shotgun assembly database and previous studies (Table 4.1) and reconstructed ancestral sequences of *dsx* female-type.

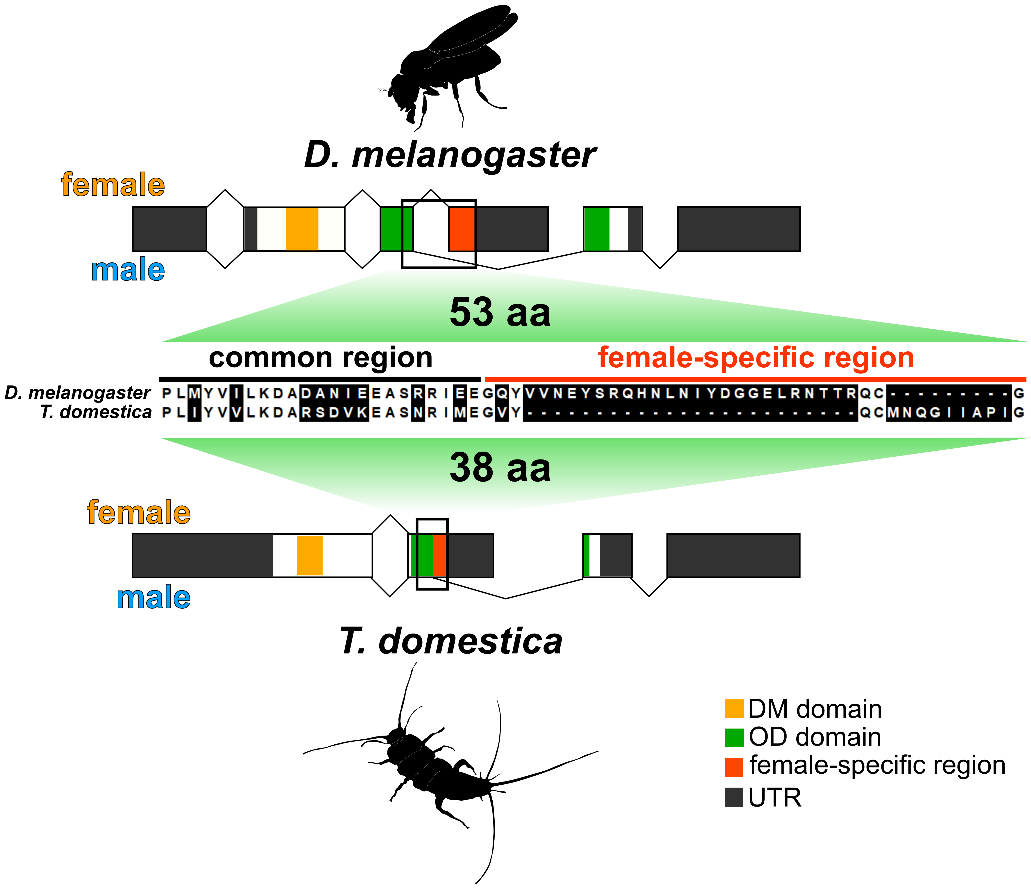
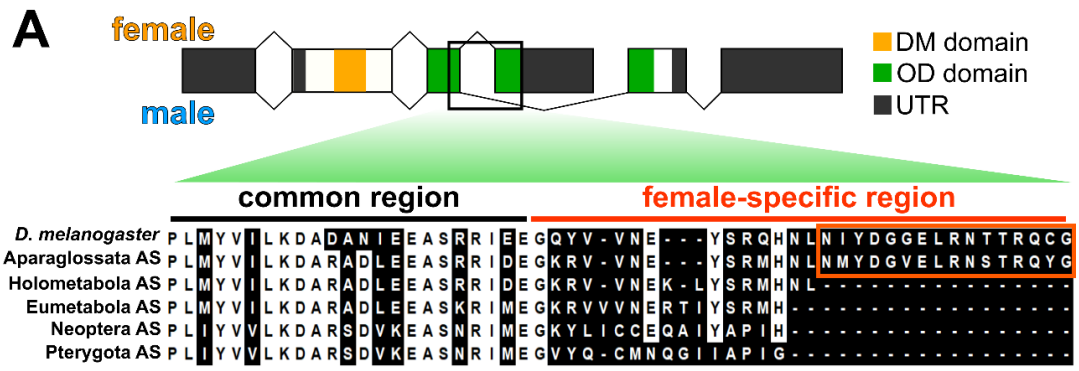
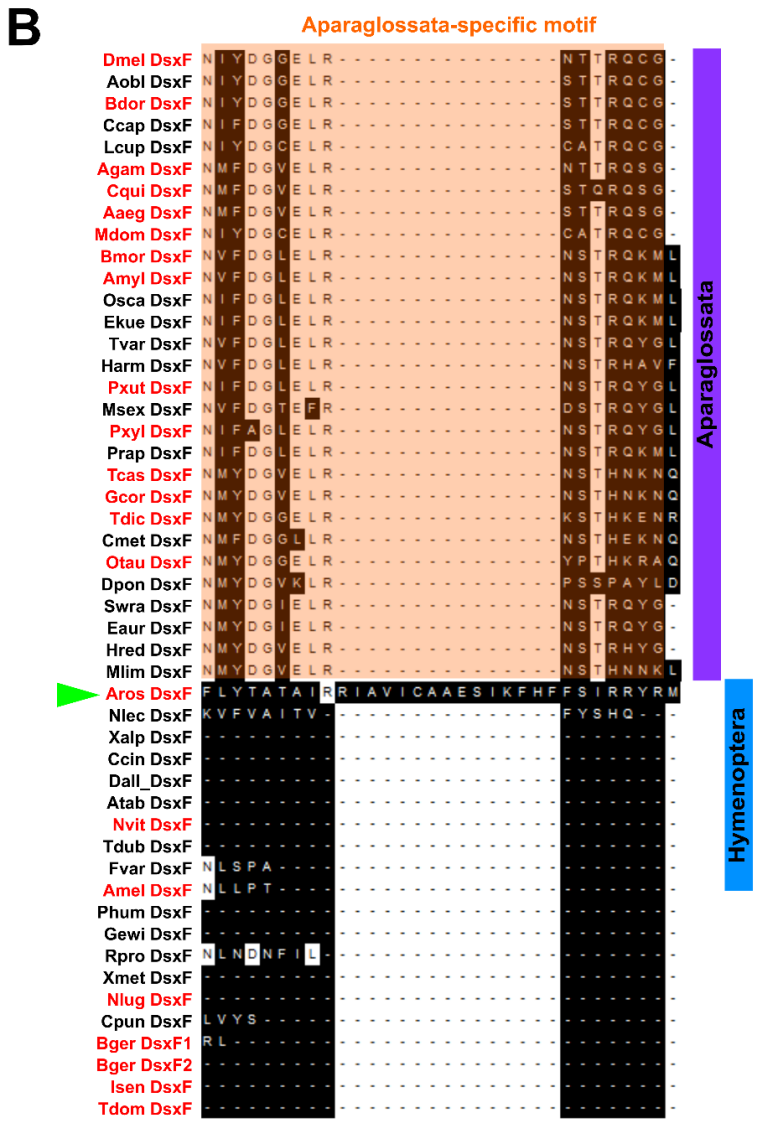


Figure 4.3. Comparison of C-terminal sequences of *dsx* female-type between *Drosophila melanogaster* and *Thermobia domestica*. The upper schematic figure shows the gene structure of *dsx* in *D. melanogaster*. The lower schematic figure indicates the gene structure of *dsx* in *T. domestica*. The female-specific region is shown by the orange color. The middle image is the result of the multiple sequence alignment (MSA) of C-terminal region between two species. The MSA was performed using the MAFFT software. The white background indicates the matched residues between the species in the MSA. The female-specific region is much shorter in *T. domestica* (38 aa) than in *D. melanogaster* (53 aa).

My ancestral sequence reconstruction revealed that the C-terminal 16-amino acid region of *dsx* female-type found in the common ancestor of Aparaglossata was absent in common ancestors of other taxa (Figure 4.4A). This motif is conserved within

Aparaglossata in my dataset although moderate sequence diversification was observed (Figure 4.4B). In my dataset, almost all sequences of this motif were not found in species in which *dsx* is non-essential for female differentiation of morphological traits during





postembryonic development. Exceptionally, *dsx* of *At. rosae* showed an amino acid sequence in the region corresponding to this motif, but my results of ancestral sequence reconstruction showed that this sequence was acquired in parallel with Aparaglossata.

Sequence diversification was also observed in female-specific regions other than the Aparaglossata-specific motif. This portion shows an α-helix loop structure in *D. melanogaster* and binds to co-factors such as the Intersex, which is responsible for the role of *dsx* in female differentiation (Bayrer et al. 2005, Yang et al. 2008). Hence, I investigated whether differences in this portion result in structural changes by predicting the structure of *dsx* female-type ancestral sequences of Pterygota, Neoptera, Eumetabola, Holometabola, and Aparaglossata using Alfafold2 algorism-based structure prediction. According to the Alphafold2 the female-specific region of *dsx* in the common ancestor of Aparaglossata had a proximal α-helix loop structure (Figure 4.5). This structure was similar to that of *D. melanogaster* determined by a crystal structural analysis (Yang et al. 2008). The proximal α-helix loop structure also predicted the common ancestors of taxon other than Aparaglossata. This result supports that the proximal α-helix loop structure in the Dsx female-type is shared among insect ancestors regardless of its sequence diversity.

Finally, I compared the exon-intron structure of *dsx* to investigate whether the C-terminal extension in the *dsx* female-type is due to the appearance of a new exon. To this end, I mapped the *dsx* mRNA sequences of some insects to the genome (Figure 4.6) and were ordered along with the insect phylogeny (Misof et al. 2014). The number of exons

Figure 4.4. Evolution of C-terminal sequence of *doublesex* in insects. (*A*) Ancestral sequences (AS) of *dsx* in insects. The AS were reconstructed from 49 *dsx* proteins of insects by the maximum likelihood methods of MEGA X. Information on the species and proteins used for the AS reconstruction is listed in Table 4.1. The most probable sequences were applied. The upper scheme indicates the *dsx* gene structure of *D. melanogaster*. The lower image shows the result of the multiple sequence alignments (MSA) of *dsx* sequences by MAFFT. The oligomerization domain sequences at C-terminal side were used for the MSA. The white background in the MSA result indicates the conserved sites that share residues in the 80% of taxa. The Aparaglossata-specific motif is indicated by the orange frame. (*B*) Multiple sequence alignments of the Aparaglossata-specific motif. The multiple sequence alignment of the C-terminal sequences was performed by the MAFFT software. This image shows the region around the Aparaglossata-specific motif indicated by the orange color. The species in the sequence name can be obtained from the Table 4.1. The names indicated by the red color show species in which the functional analysis of *dsx* has been performed. The green arrowhead exhibits the sequence of *Athalia rosae* that has the amino acid sequences corresponding to the motif.

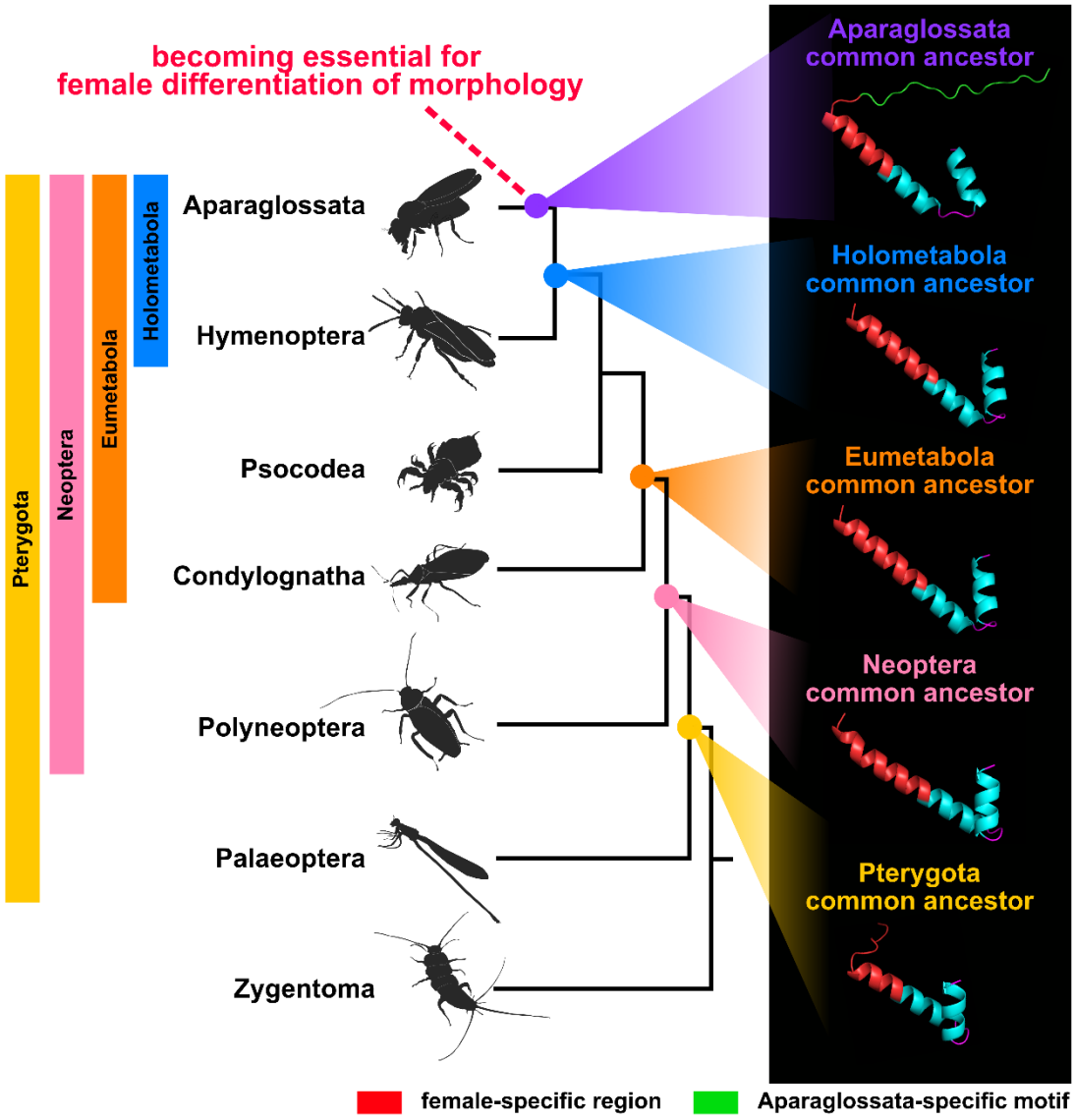


Figure 4.5. Predicted protein structures of *dsx* female-type in common ancestors of insect taxa. The phylogenetic relationship is based on topology from Misof et al. (2014). The 3D images in the right panel indicate the predicted structures of the OD domain including the female-specific region of *dsx*. The protein structures were predicted by the AlphaFold2-based algorism (ColabFold: Mirdita et al. 2021). The red region of the 3D image indicates the female-specific region. The green region shows the Aparaglossata-specific motif. Information on the evaluated values (predicted local distance difference test: plDDT) of the prediction is shown in the Methods section and Figure 4.2.

in *dsx* differed among species. These differences were often observed in exons common to both sexes and in male-specific exons. For example, *At. rosae* had two male-specific exons, and *D. melanogaster* had one male-specific exon. However, this difference in exon number was not found in the female-specific exon. The only exception was *Bl. germanica*.

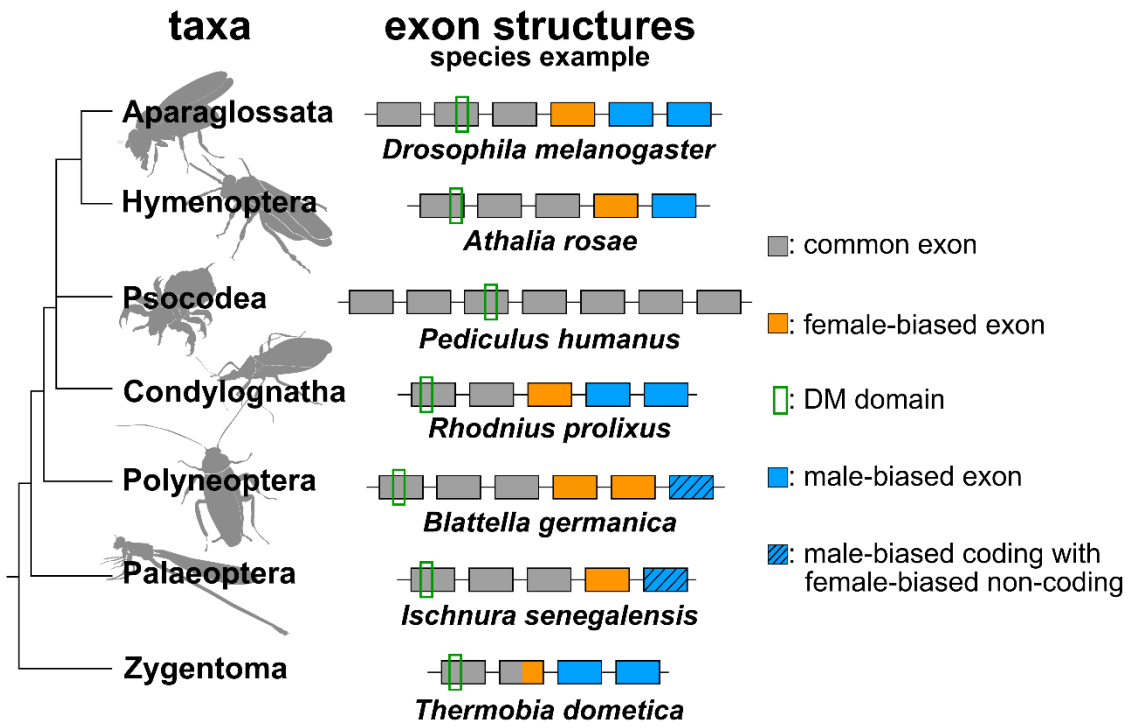


Figure 4.6. Exon structure of *dsx* among insect taxa. The exon-intron structure is based on the previous studies: Bruce and Baker (1989) in *Drosophila melanogaster*, Mine et al. (2017) in *Athalia rosae*, Wexler et al. (2019) in *Pediculus humanus*, *Rhodonius proxius*, and *Blattella germanica*, Takahashi et al. (2019) in *Ischnura elegans*, this study in *Th. domestica*. The coding region of *dsx* is shown. The phylogenetic relationship is that of Misof et al. (2014).

This cockroach had two female-specific exons, while the other species analyzed had only one female-specific exon. In addition, I found that the sex-specific exon-intron junction is maintained among insects (Figure 4.7). These results suggest that the C-terminal extension in the *dsx* female-type is unlikely to have originated through a neo-exonization.

4.4. Discussion

In this chapter, I compared the amino acid sequences of Dsx female-type among insect species and showed that its most C-terminal motif is added in the common ancestor of Aparaglossata, when the feminizing function of *dsx* for morphology occurred. In my analysis, there was no evidence that a novel exon gave rise to the extension of the C-terminal motif of the Dsx female-type. These results support that the appearance of the

feminizing role of *dsx* in morphogenesis may be involved to the mutation in the coding

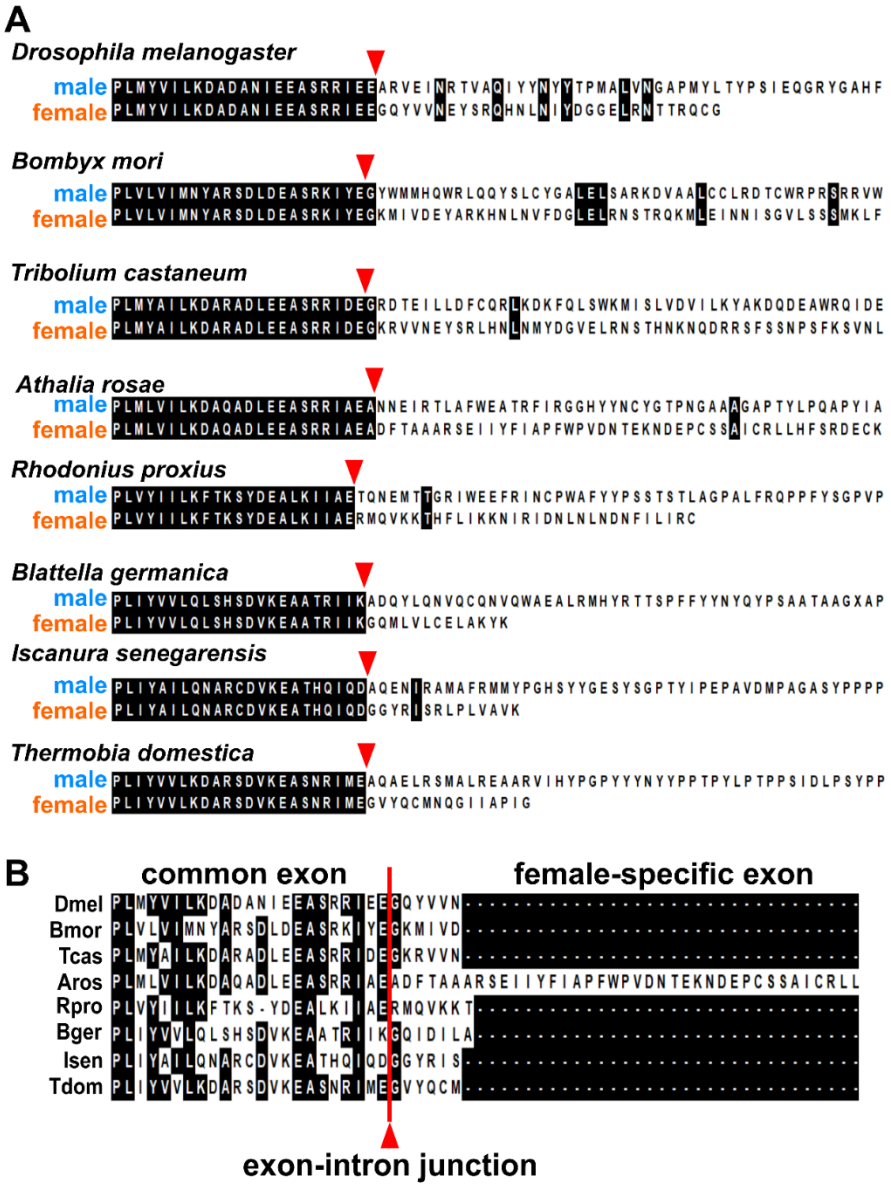


Figure 4.7. Exon-intron junction of *dsx* among insect species. (*A*) Alignments of the sex-specific isoforms within species. The arrowheads indicate the exon-intron junction. The locations of the junction are based on the previous studies such as Burtis and Baker (1988) in *D. melanogaster*, Duan et al. (2013) in *Bo. mori*, Shukla and Palli (2012b) in *Tr. castaneum*, Mine et al. (2017) in *A. rosae*, Wexler et al. (2019) in *R. proxius* and *Bl. germanica*, Takahashi et al. (2019) in *I. senegarensis*, and this study (see Chapter 2) in *Th. domestica*. (*B*) Alignments of the female-specific isoforms among insect species. The red bar and arrowhead show the exon-intron junction of *dsx*. Dmel, *D. melanogaster*: Bmor, *Bo. mori*: Tcas, *Tr. castaneum*: Aros, *At. rosae*: Rpro, *R. proxius*: Bger, *Bl. germanica*: Isen, *I. senegarensis*: Tdom, *Th. domestica*. All multiple sequence alignments (MSA) were conducted by the MAFFT software. The white background in the MSA result indicates the conserved sites that share residues in the 100% (*A*) or 60% (*B*) of taxa

region of *dsx*, i.e., the extension of the C-terminal motif.

Currently, the causality between the extension of the C-terminal motif and the appearance of the feminizing role of *dsx* is uncertain, as I do not know which event appeared first. Wang et al. (2019) showed that in the diamondback moth *Plutella xylostella*, when the Aparaglossata-specific motif is specifically broken by deletion or frameshift mutations using the CRISPR/Cas9 method, the female morphology is transformed into the intersexual phenotype. This result indicates that the Aparaglossata-specific motif is essential for female differentiation of morphology in *P. xylostella*. These facts suggest that extension of the C-terminal region of *dsx* female-type may have been a key event associated with acquiring the female-differentiating roles of *dsx* in morphology. This hypothesis will be tested by CRISPR-mediated Knock-in experiments to add the Aparaglossata-specific motif to *dsx* in non-Aparaglossata and to assess their effects on biological processes such as gene expression and cell fate.

In *D. melanogaster*, the Aparaglossata-specific motif is the disordered region following the α-helix loop in the female-specific region of *dsx* (Yang et al. 2008). Hence, the extension of such disordered portion might be linked to the neo-functionalization of *dsx*. In *D. melanogaster*, the C-terminal disordered region does not bind to the Ix (Yang et al. 2008). Thus, it is unknown how the disordered region contributes to transcriptional regulation in females of *D. melanogaster*. Yang et al. (2008) propose that this region might interact with co-factors other than the Ix. Indeed, some proteins physically bind to *dsx* (An et al. 1995, Erdman et al. 1996, Ghosh et al. 2019, Romero-Pozuelo et al. 2019). However, the prediction by Yang et al. has not been examined yet. Furthermore, in general, disordered regions in transcription factors play essential roles in transcriptional activity through post-translational modifications or binding to co-activators and nucleic acids (Liu et al. 2006, Darling and Uversly 2018). It is a future task whether the *dsx* female-type acts via these mechanisms. CRISPR-mediated genome editing in the disordered region will uncover relationships between the region and female differentiation of morphology in Aparaglossata.

Since *dsx* in the non-Aparaglossata species is not essential for female differentiation of morphology (Chapter 3) and does not possess the C-terminal motif conserved between the aparaglossatan orders (this chapter), the C-terminal region of *dsx* might have been less subject to a purifying selection. Alternatively, Baral et al. (2019) proposed that there may

be a link between the higher number of nonsynonymous mutations in the female-specific region of Hymenopteran species and the sociality. If latter is true, the sequence diversification in the female-specific region of non-Aparaglossata might have resulted from the positive selection in each lineage. Here, it is not clear which of these two hypotheses is correct. To test these hypotheses, a class or subphyla-level re-analysis should be necessary. This "taxonome" (= taxon + ome) approach would be helpful for a better understanding of the evolution of *dsx*.

Chapter 5

General Discussion and Conclusion Remarks

Abstract In this thesis, I demonstrate the function of *dsx* for sexual morphology in the apterygote insect, *Thermobia domestica*, and infer the evolutionary process of *dsx* in Insecta. In this final chapter, I overview my results and discuss the implication of the evolutionary history of *dsx* for arthropod systematics and diversity in the sex-determining cascade in Metazoa. Finally, I attempt to extend the evolutionary model in this thesis to the exaptive process of alternative splicing isoforms and propose some opinions on the neo-functionalization of splicing isoforms.

5.1. Overview of evolutionary history of *dsx* in Arthropoda

In this thesis, to understand the evolutionary history of *dsx*, I investigated the molecular evolution of *dsx* and the function of *dsx* in the apterygote insect, *Thermobia domestica*. In the Chapter 2, my phylogenetic analysis provided evidence of the *dsx* duplication before the appearance of the last common ancestor of Dicondylia. In addition, I revealed the sex-specific splicing manner in *dsx* of *T. domestica* and inferred that the splicing control of *dsx* emerged before the divergence of Dicondylia. In the Chapter 3, the nymphal RNAi assays showed that *dsx* in *T. domestica* contributes to male differentiation of morphology but is not required for female differentiation in morphogenesis. In contrast, I found that *dsx* in *T. domestica* plays antagonistic roles between sexes in controlling the expression of the female-specific genes. Thus, I inferred that the non-functional isoform became utilized for the female morphogenesis from the common ancestor of Holometabola to the appearance of Aparaglossata. The "non-functional" isoform is not essential for female morphogenesis during the post-embryonic development but may already have the opposite functionality between sexes for some gene expression, which might be a exaptive stage of *dsx* in females. In the Chapter 4, I

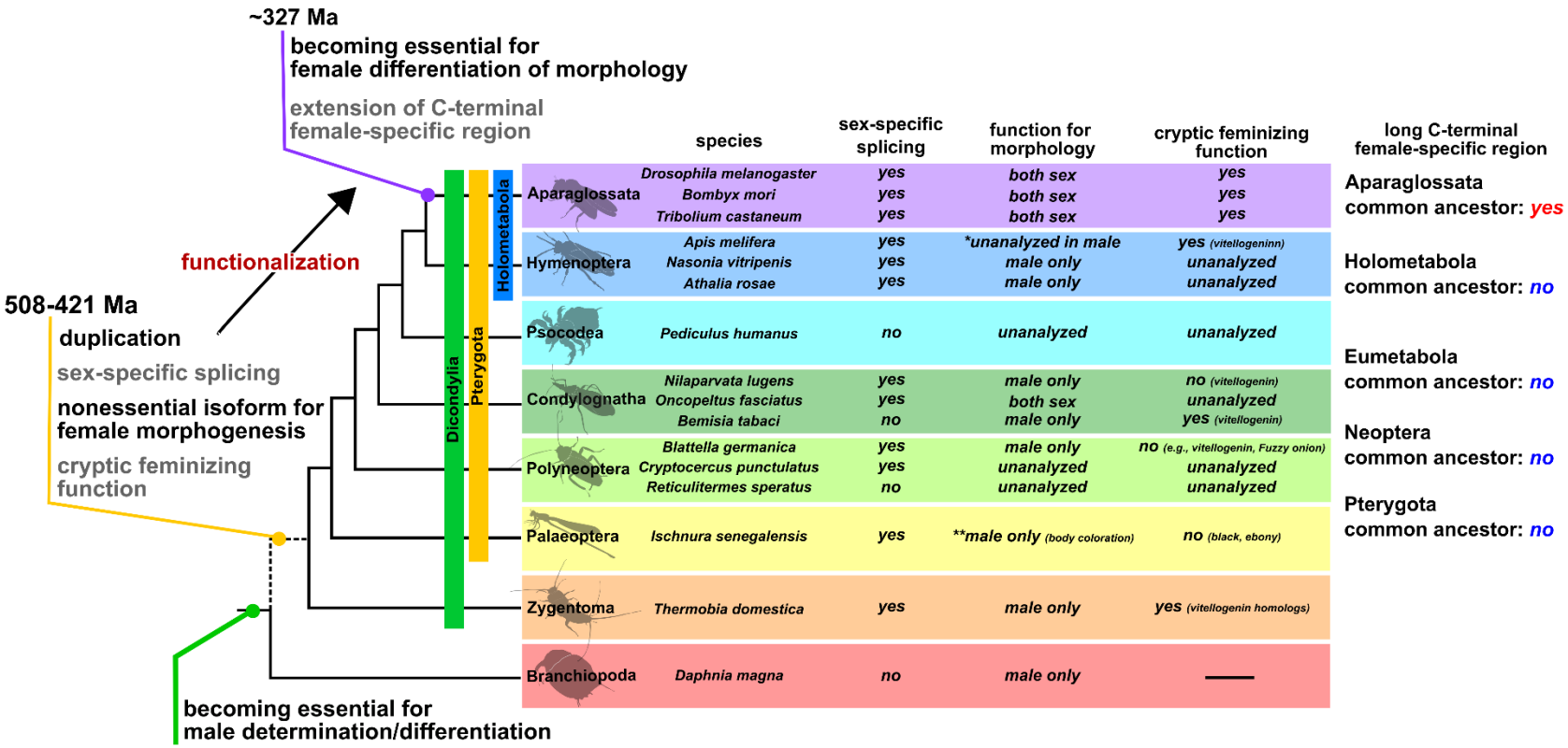
uncovered that *dsx* in the non-Aparaglossata species does not possess the C-terminal motif conserved among the Aparaglossata orders. Thus, the acquisition of this Aparaglossata-specific motif is presumed to be one of the key events linked to the emergence of the feminizing roles of *dsx* in morphogenesis.

In this thesis, I demonstrate the first insight on the sex-specific splicing and the function of *dsx* in the hexapod species except for Pterygota. Data on the roles of *dsx* are limited to some traits in some species and are unavailable in many non-aparaglossatan species, although functional analyses of *dsx* have been rapidly progressing using emerging model species (e.g., Mine et al. 2017, Zhuo et al. 2018, Wexler et al. 2019, Takahashi et al. 2019). Unquestionably, comprehensive information on *dsx* functions for sexually dimorphic morphology from wider taxa is essential for fully tracing the evolution of *dsx* (Wexler et al. 2019). I propose, albeit prematurely, the hypothesis by which *dsx* may become essential for female differentiation in sexual morphology by expanding its cryptic feminizing role, i.e., the function for some female genes' expression, in association with the extension of the female-specific motif rather than acquisition of the female-specific exon (Figure 5.1). On the basis of my results and evolutionary model, I strongly support some hypotheses on the evolution of *dsx* proposed in the last decade. In addition, my results provide the novel hypothesis on the evolutionary history of *dsx*.

My phylogenetic analysis provides the first phylogenetical evidence of the duplication of *dsx* before the emergence of the common ancestor of Pterygota. This inference ensures the hypothesis of the previous study on the sequence comparison (Price et al. 2015). Further, based on my results and previous phylogenetic analysis (Wexler et al. 2019), I infer that *dsx* experienced duplication events independent of multiple lineages, e.g., a predatory tick, branchiopods, insects. Generally, gene duplication can lead to the neo-functionalization of genes (True and Carroll 2002). Indeed, a lineage-specific paralog of *dsx* in a dogface butterfly contributes to forming structural coloration in wings (Rodriguez-Caro et al. 2021). However, the roles of *dsx* paralogs in other arthropods have remained unknown: e.g., *Daphnia magna* (Branchiopoda, Kato et al. 2011) and *Nilaparvata lugens* (Hemiptera, Zhuo et al. 2018). I uncovered that *dsx-like* plays sex-antagonistic roles in the *vitellogenin* expression (Chapter 3), suggesting that *dsx* and *dsx-like* may act cooperatively. In contrast, I cannot provide evidence that the feminizing function of the *dsx* isoform appeared with the replacement of the role of

its paralog as *dsx-like* does not contribute to female morphogenesis during the post-embryonic development. In this study, it is still unclear what biological processes *dsx-like* is involved in since I did not conduct the functional analysis of *dsx-like* during embryogenesis. Investigating the roles of *dsx-like* and *dsx* in embryogenesis will shed light on the evolutionary history of the paralogs and the isoforms.

Wexler et al. (2019) proposed that *dsx* had acquired the sex-specific isoforms before the occurrence of Neoptera and later became essential for female development. My functional and comparative analyses are consistent with Wexler et al.'s hypothesis and allow more precise identification of the period of when the isoforms of *dsx* were gained. The major difference from Wexler et al.'s hypothesis is that *dsx* serves the antagonistic function between sexes for some gene expression. Previous studies (e.g., Wexler et al. 2019, Mine et al. 2021) was forced to make roles of *dsx* integrated into a single hierarchy in sexual differentiation since species used by them shows no evidence on the sex-antagonistic roles of *dsx* in any aspects of sexual differences. Alternatively, I infer that the female isoform of *dsx* had initially the sex-antagonistic roles in some female-specific gene expression and later became essential for female differentiation of morphogenesis in the post-embryonic development. According to my model, the function of *dsx* during post-embryonic development is not standardized across aspects of sexual differentiation but is viewed as distinct between morphogenesis and other physiological processes in sexual development. This means that the roles of *dsx* in sexual morphogenesis have different evolutionary origins from those in other sexually dimorphic processes. This opinion coincides with the results of several studies using pterygote species (Guo et al. 2018, Velasque et al. 2018), where *dsx* contributes to only masculinization of morphology and upregulates female-specific genes in females.


~327 Ma
becoming essential for
female differentiation of morphology
extension of C-terminal
female-specific region
functionalization
508-421 Ma
duplication
sex-specific splicing
nonessential isoform for
female morphogenesis
cryptic feminizing
function
becoming essential for
male determination/differentiation
Holometabola
Dicondylia
Pterygota
species
sex-specific splicing
function for morphology
cryptic feminizing function
Aparaglossata
Drosophila melanogaster
Bombyx mori
Tribolium castaneum
yes
yes
yes
both sex
both sex
both sex
yes
yes
yes
Hymenoptera
Apis melifera
Nasonia vitripenis
Athalia rosae
yes
yes
yes
*unanalyzed in male
male only
male only
yes (vitellogeninn)
unanalyzed
unanalyzed
Psocodea
Pediculus humanus
no
unanalyzed
unanalyzed
Condylognatha
Nilaparvata lugens
Oncopeltus fasciatus
Bemisia tabaci
yes
yes
no
male only
both sex
male only
no (vitellogenin)
unanalyzed
yes (vitellogenin)
Polyneoptera
Blattella germanica
Cryptocercus punctulatus
Reticulitermes speratus
yes
yes
no
male only
unanalyzed
unanalyzed
no (e.g., vitellogenin, Fuzzy onion)
unanalyzed
unanalyzed
Palaeoptera
Ischnura senegalensis
yes
**male only (body coloration)
no (black, ebony)
Zygentoma
Thermobia domestica
yes
male only
yes (vitellogenin homologs)
Branchiopoda
Daphnia magna
no
male only
—
long C-terminal female-specific region
Aparaglossata
common ancestor: yes
Holometabola
common ancestor: no
Eumetabola
common ancestor: no
Neoptera
common ancestor: no
Pterygota
common ancestor: no

5.2. Implication for arthropod systematics

My results and evolutionary model have significance on arthropod systematics. This implication is apart from the subject of this thesis, but I would like to mention it here considering the current status of arthropod systematics. In this thesis, I discovered that the roles of *dsx* in female morphogenesis and its C-terminal region are present among Aparaglossata orders studied so far (Chapters 3 and 4). Furthermore, I inferred that these features occurred at the common ancestor of Aparaglossata. Therefore, these features of *dsx* are autapomorphies, i.e., derived characters, of Aparaglossata and strong candidates of synapomorphies, i.e., shared derived characters, of aparaglossatan species. Aparaglossata is the recently established clade supported by the phylogenomics (Misof

Figure. 5.1. Schematic diagram of the evolutionary history of *doublesex* proposed in this study and the feature of *dsx* in insects. The phylogenetic relationship and divergence time are referenced in Misof et al. (2014). The dotted line in the phylogenetic relationship indicates that the taxa occurring from the common ancestor between Branchiopoda and Dicondylia to the common ancestor of Dicondylia are omitted. This information was based on the following studies: Hildreth (1965), Bruce and Baker (1989) and Cloudh et al. (2014) in *Drosophila melanogaster* (Diptera), Ohbayashi et al. (2001), Suzuki et al. (2003), and Xu et al. (2017) in *Bombyx mori* (Lepidoptera), Shukla and Palli (2012b) in *Tribolium castaneum* (Coleoptera), Roth et al. (2019) and Velasque et al. (2018) in *Apis mellifera* (Hymenoptera), Wang et al. (2022a, b) in *Nasonia vitripenis* (Hymenoptera), Mine et al. (2017, 2021) in *Athalia rosae* (Hymenoptera), Wexler et al. (2019) in *Pediculus humanus* (Psocodea) and *Blattella germanica* (Dictyoptera), Zhuo et al. (2018) in *Nilaparvata lugens* (Hemiptera), Just et al. (2021) in *Oncopeltus fasciatus* (Hemiptera), Guo et al. (2018) in *Bemisia tabaci* (Hemiptera), Miyazaki et al. (2021) in the wood roach *Cryptocercus punctulatus* and *Reticulitermes speratus* (Dictyoptera), Takahashi et al. (2019, 2021) in *Ischnura senegalensis* (Odonata), this study in *Thermobia domestica* (Zygentoma), and Kato et al. (2011) in *Daphnia magna* (Branchiopoda). In Condylognatha, information on *dsx* in the blood-sucking bug *Rhodnius prolixus* is omitted. *R. prolixus* has sex-specific isoforms of *dsx* whose function has not been investigated (Wexler et al. 2019). The term "unanalyzed" means the functional analyses of *dsx* have not been performed in the relevant species. The asterisk (*) in *Ap. mellifera* indicates that the functional analysis of *dsx* in males was not conducted although gonad differentiation of female workers was affected by *dsx* knockouts (Roth et al. 2019; see main text). The double-asterisk (**) in *I. senegalensis* shows that this species has polymorphic coloration in females, i.e., gynomorph (normal female color) and andromorph (male-like color) and that *dsx* is involved in color formation of males and andromorphic females but not gynomorphic females (see Takahashi et al. 2021), suggesting that *dsx* is non-essential for the female color development.

et al. 2014) and shares only one trait, i.e., the absence of a paraglossa that is an appendage segment at the tip of the labium (Peters et al. 2014, Beutel et al. 2017). Since this clade has been supported only by the "lost" trait, the feminizing function and the C-terminal motif may be the first "gained" synapomorphies of Aparaglossata. Future taxonome studies on *dsx* of Holometabola will examine this opinion.

According to Sober (2000), traits do not define biological groups but are evidence of monophyletic groups' existence. Also, examining synapomorphies allows us to verify phylogenetic hypotheses (Hennig 1966, Wiley 1981). Systematics has proposed many synapomorphic traits in arthropod clades and has provided amounts of evidence on the being of monophyletic groups, especially confirmed by the molecular phylogeny (Beutel et al. 2017, Giribet and Edgecombe 2019). The examples can be seen in the oogenetic mode of Myriapoda (Miyachi et al. 2012, Yahata et al. 2018), the embryonic formation and the blastokinesis in Polyneoptera (Mashimo et al. 2014), and the mouthparts in Labiocarida (= Remipedia + Hexapoda) (Schwentner et al. 2017). These studies mainly focus on morphological traits. Moreover, in the last decade, some studies have used gene expression patterns as synapomorphies of arthropod clades: e.g., Hox genes in Myriapoda (Janssen and Budd 2010) and segmentation gene *cap-n-collar* in Mandibulata (Sharma et al. 2014a). In contrast to these accumulations of information, no or few synapomorphies have been found in many arthropod clades reconstructed by transcriptome-based phylogenetic studies such as Multicrustacea (Lozano-Fernandez et al. 2019) and the Chelicerata-clade consisted of Xiphosura + Ricinulei (Sharma et al. 2014b, Ballesterons and Sharma 2019, Ballesterons et al. 2022). Approaches other than the traditional focus may provide support for these clades. To my knowledge, my argument that the feminizing role of *dsx* would be a putative synapomorphy of Aparaglossata is the first attempt to utilize the gene function as synapomorphic support of an arthropod clade proposed by phylogenomic data. Similarly, functional information of genes may become one of the effective means of systematics and provide a new guideline in the phylogenomic era. Currently, gene analysis tools, especially genome editing tools, have developed remarkably in Arthropoda (e.g., Bassett and Liu 2014, Nakanishi et al. 2014, Gilles et al. 2015, Martin et al. 2016, Watanabe et al. 2017, Ohde et al. 2018, Xue et al. 2018, Dermauw et al. 2020, Matsuoka et al. 2021, Li et al. 2022, Sharma et al. 2022, Shirai et

al. 2022). Hence, these recent advances would facilitate the accumulation of systematic support for arthropod clades in the phylogenomic era.

5.3. Implication for diversity in sex-determining cascade in Metazoa

The results in this thesis provide a novel aspect of diversity in the sex-determining cascade. In the last three decades, many studies have elucidated the genetic pathways that create sex and sexual dimorphism in various animal species. Despite a single origin of the animal sex (Beukeboom and Perrin 2014), these pathways have undergone extensive changes throughout evolution (Wilkins 1995, Bachtrog et al. 2014, Bopp et al. 2014, Herpin and Schartl 2015).

The evolution of sex-determining genes is mainly studied in vertebrates (reviewed in Nagahama et al. 2021). The studies on vertebrates support that the upstream genes in the sex-determining cascade turn over more frequently than the downstream components. For example, in eutherians such as mice and humans, the master regulator of sex is the *Sex-determining region Y* (*Sry*), a member of the High Mobility Group-box transcriptional factor family (Gubbay et al. 1990, Sinclair et al. 1990, Koopman et al. 1991, Miyawaki et al. 2020), while *DM domain gene on the Y chromosome* (*dmy*) of the DMRT family is the master sex-determining regulator in the medaka fish (Matsuda et al. 2002, Nanda et al. 2002). Wilkins (1995) initially postulated this evolutionary tendency and later called it a bottom-up model. In this model, diversity in gene components among species or taxa is higher in the initial factor than in the bottom factor of sex-determining cascades. The initial factors are variable among species or taxa and sometimes distinct among the population within species. The bottom factor is Dmrt1, a member of the DMRT family (e.g., Mawaribuchi et al. 2019) and conserved among almost all vertebrates studied. Hence, the degree of diversity of sex-determining cascades can be represented as a funnel-shaped scheme.

The bottom-up model is also applied to sex-determining cascades in insects. As in the vertebrates, the initial factors are diverse in insects. For instance, the primary factors of dipteran species are *Sex-lethal* in *D. melanogaster* (Bell et al. 1988), *Musca domestica male determiner* in *Musca domestica* (Sharma et al. 2017), *Yob* in *Anopheles gambiae* (Krzywinska et al. 2016), *Nix* in *Aedes aegypti* (Hall et al. 2015), and *Maleness-*

on-the-Y in *Ceratitis capitata* (Meccariello et al. 2019). Further, the initial trigger is more diversified in *Bombyx mori*. In this species, a piRNA-mediated system underlies the sex-determining cascade. The piRNA, *Fem*, is expressed in females and silences the male determiner, *Masc* mRNA (Kiuchi et al. 2014, Sakai et al. 2016).

The intermediate and bottom factors tend to be conserved among taxa. Generally, the intermediate factor of insects is the RNA binding protein, *transformer* (*tra*). The bottom factor is *dsx* (Kopp et al. 2012) and is retained among almost all arthropods studied. *tra* controls the female-specific splicing of *dsx* together with the RNA binding protein, *transformer-2* (Hoshijima et al. 1991). *tra* is absent and does not contribute to the sex determination in some insects such as Lepidoptera, Culicomorpha, Bibionomorpha, and Hemiptera (Geuverink and Beukeboom 2014). In *Bo. mori* (Lepidoptera), *insulin-like growth factor II mRNA binding protein* is involved in the sex-specific splicing of *dsx* (Suzuki et al. 2010). In *Ni. lugens*, the serine/arginine-rich protein-encoding gene, *female determiner* controls the female-specific splicing of *dsx* (Zhuo et al. 2021). In contrast to these exceptions, the *tra-dsx* system is broadly conserved in Holometabola: e.g., *D. melanogaster* (Hoshijima et al. 1991), *Tribolium. castaneum* (Shukla and Palli 2012a), *Apis mellifera* (Hasselmann et al. 2008), and *Nasonia vitripenis* (Verhulst et al. 2010). Also, *tra* contributes to the female differentiation in *Blattella germanica* (Wexler et al. 2019). Hence, the *tra-dsx* module would be the pervasive core in regulating sex determination in pterygote insects. The conserved bottom factor, *dsx*, integrates the sexual signal and underlies various aspects of sexual traits such as morphogenesis, behavior, and physiology (Kopp et al. 2012).

On the basis of insights into insects, the bottom-up model was extended to the hourglass model (Bopp et al. 2014). In the hourglass model of sex-determining cascades, the sex-determining pathway is divided into three parts, i.e., instruction, transduction, and execution. The instruction is the primary signal or triggers of the sex-determining cascade. The transduction indicates the *tra-dsx* module. The execution includes various aspects of sexual traits. Therefore, the instruction and execution are variable among species. In contrast, the middle part, i.e., transduction, is strongly conserved among insect taxa.

Diversity in the bottom-up and hourglass models is primarily due to differences in the composition of the regulatory cascade. In addition, recent studies on insects (e.g., Mine et al. 2017, Guo et al. 2018, Zhuo et al. 2018, Wexler et al. 2019, Takahashi et al. 2021),

including this thesis, suggest that the sex-determining cascade is diverse independently of its gene repertoires. This diversity is attributed to the function of a single gene, *dsx*, for sex differentiation of morphogenesis during postembryonic development. In Insecta, *dsx* can be divided into two types based on its roles in morphogenesis, i.e., promoting only male differentiation in Zygentoma, Odonata, Dictyoptera, Hemiptera, and Hymenoptera (see the Chapter 3) and differentiating both male and female morphologies in Aparaglossata. Dsx is the bottom factor of the insect sex-determining cascade (Kopp 2012) and plays a role in the output of the cascade. The difference in the *dsx* functionality leads to distinct outputs in the sex-determining pathway regulating morphogenesis, i.e., only masculinization (single-output) or masculinization/feminization (double-output). This diversity in the outputs is ascribed to the function of *dsx* and is independent of the gene repertoire diversity. Indeed, the pathway with the single-output shares the genes with the pathway with the double-output. For example, the pathway with the single output in several species contains the *transformer* gene, which controls *dsx* in the double output pathway (Hasselmann et al. 2008, Wexler et al. 2019). Since diversity in the sex-determining cascade has been equivalent primarily to differences in the gene repertoires, diversity in the output, independent of the gene composition, would represent a new aspect for describing diversity in the sex-determining cascade.

Diversity in the *tra-dsx* module can be seen in the function of *dsx* and the exon-intron structure of the *tra* gene. *tra* has a poison exon, i.e., stop codon-including exon, and has the non-functional isoform in males of many holometabolan insects studied (Geuverink and Beukeboom 2014). As a result, *dsx* pre-mRNA is spliced via *tra*-independent constitutive splicing in males. In contrast, *tra* does not retain such a poison exon and produces a seemingly functional isoform with an extended ORF in males of *Bl. germanica* (Wexler et al. 2019). Hence, *tra* might initially have produced functional isoforms with a complete ORF and later obtained the non-functional isoform with the poison exon in males. Investigating this speculation requires identifying the isoform sequences of *tra* in earlier branched groups such as Zygentoma and inferring the evolutionary history of *tra* isoforms. Further, future works focusing on the roles of *tra* in males of species with the functional *tra* isoform will shed light on the evolutionary process of *tra*.

5.4. Exaptive process of alternative isoforms

In this thesis, I stated that *dsx* has differences in function between insect taxa, i.e., single- or double-output. *dsx* in the current insect species may well document the evolutionary process in the role of alternatively spliced genes. Here, I discuss the exaptive process of alternatively spliced genes based on my results and evolutionary model. I compare the evolutionary history of *dsx* with previous hypotheses on the evolution of alternative isoforms. Through this comparison, I attempt to generalize the model of this thesis, albeit incompletely, to the context of the exaptation in isoforms.

5.4.1. Recruitment of new function in alternative isoforms

The most important conclusion of this thesis is the co-option of non-functional isoforms of *dsx* to female morphogenesis (Chapter 3). I then raised the possibility that the neo-functionalization could be linked to changes in the coding region (Chapter 4). Given that splicing isoforms are often non-functional or have no visible function (Nielsen and Graveley 2010, Baralle and Giudice 2017), new isoforms and exons would not necessarily be at high adaptive status. The general predictions are that such isoforms will become more adaptive through later mutation accumulation (Boue et al. 2003, Xing and Lee 2006, Keren et al. 2010). In these predictions, non-functional isoforms come into use through mutations in coding elements. Therefore, the functionalization of such isoforms fits well with my model of the *dsx* function for sexual morphology.

The remaining question is what changes in the coding region could result in the functionalization of the non-functional isoforms. In this thesis, I identified the synapomorphic nature of the C-terminal disordered region of *dsx* in Aparaglossata (Chapter 4). In addition, it is generally known that many differences in coding among splicing isoforms do not significantly disrupt structural motifs in protein domains (Lareu et al. 2004, Blujan et al. 2013, Reixachs-Solé and Eyras 2021). Frequently, differences among isoforms are found in intrinsically disordered regions responsible for protein-protein interactions (Ule and Blencowe 2019). Thus, one of the roles of alternative splicing is to underpin the pleiotropy of genes via the re-wiring of diverse networks of protein-protein interactions (Ule and Blencowe 2019). Since this statement concerns the differences in isoforms among cells or tissues within individuals, its evolutionary significance has been unknown. Integrating this argument with my model, I assume that

changes in terminal disordered regions of alternative isoforms may be one of the factors leading to the functionalization of non-functional isoforms.

In addition, I proposed that female-specific splicing control of *dsx* serves as a balancer of isoforms in female morphogenesis (Chapter 3). Alternative splicing as a balancer is known with examples (Kelemen et al. 2013), and an evolutionary hypothesis has been proposed that isoforms arising as a byproduct of balance adjusting later acquire new functions (Keren et al. 2010). On the basis of these results and prediction, I speculate that balancer roles may be one of the exaptive states of the non-functional isoform before its neo-functionalization.

5.4.2. Expression level before neo-functionalization in alternative isoforms

An assumption of the prediction on the functionalization of isoforms is the neutrality of alternative isoforms (Boue et al. 2003, Xing and Lee 2006, Keren et al. 2010). If an isoform is neutral or slightly deleterious to cellular activity, such isoforms are released from purifying selection and are allowed to accumulate mutations. This functional neutrality is attained by low expression of alternative isoform (detailed in section 1.2.). However, the neutral model requiring the low-expression nature does not seem to apply to the *dsx* evolution since *dsx* is expressed to the same level in males and females or higher in females than in males (Chapter 2, Wexler et al. 2019).

What is the interpretation of the higher expression of *dsx* isoforms in female cells? One of the answers may be to tune up non-functional or functional isoforms between sexes. *dsx* is regulated by little sex-biased transcription (Chapter 2). If *dsx* is not controlled by sex-specific splicing, female cells would be masculinized by the original function of *dsx*. In other words, the female-specific splicing produces the non-functional isoform and turns off the activity of *dsx* in cells constituting female-specific traits. Thus, to prevent the masculinizing effects in female cells, it is assumed that almost all of its transcripts are produced as female-specific isoforms.

5.4.3. Exaptive states before neo-functionalization in alternative isoforms

Based on the balancer model discussed in this thesis, the female-specific isoform of *dsx* is a byproduct of balancing isoforms. Indeed, the non-functionality of *dsx* in the female differentiation of morphogenesis supports that the isoform is the byproduct at least in the cells constituting female-specific morphology (Chapter 3).

In insects studied, the isoform in females retains a long ORF. In addition, based on the positions and sequences of female-specific exons, the non-functional isoform has been conserved among taxa during insect evolution (Chapter 4). On the other hand, neutral or non-functional isoforms are rapidly replaced across species and sometimes lost or became non-sense ones with very short ORFs, suggesting that isoforms need to have some function to be conserved across species (Graveley 2001, Xing and Lee 2006, Kelemen et al. 2013). Indeed, the insect sex-determining gene, *transformer*, came to produce a non-sense isoform in the course of evolution (Wexler et al. 2019). The resolution of this seeming contradiction may be clued in the sex-antagonistic function of *dsx* against the female-specific expression of *vitellogenin* homologs.

As discussed above, the sex-antagonistic role of *dsx* in *vitellogenin* expression of *T. domestica* indicates that *dsx* serves different functions between female morphogenesis and other aspects of female specificity. Simply, such "cryptic" functionality of *dsx* explains the conservation of the female-specific isoform with the long ORF that is not essential for the female-specific morphology. Due to the "cryptic" function, the splicing mechanism and the isoform of *dsx* may be under purifying selection and conserved among insect females. More generally, the "cryptic" function of alternative isoforms is an exaptive role, leading to long-term maintenance of the "byproduct" isoform in other cells. This opinion also means that the same splicing mechanism is responsible both for tuning up isoform balance and for the production of isoforms with "cryptic" functions.

In this thesis, I found a sex-antagonistic function of *dsx* only in the regulation of *vitellogenin* transcription. This result does not mean that the role of *dsx* in females is only to promote *vitellogenin* expression. Since this thesis focused on sex differences represented during post-embryonic development, it remains unknown whether *dsx* contributes to female determination during embryogenesis. Therefore, there might be other "cryptic" functions. The search for this "cryptic" function will be the subject of future work.

The opinions in this section are not conclusive, as they currently require several assumptions. They do not rule out other possibilities and are only speculations. Moreover, testing these opinions is not easy since we can only know the isoform status of the current organism and cannot detect isoforms that have been lost in the past. However, a comprehensive comparison via a large-scale taxon sampling would provide a more

complete inference of the evolutionary history of isoform function and allow us to examine my hypotheses at the correlation level. In addition, the balancer model could be further explored using constitutive approaches that alter the splicing balance using genome editing methods such as altering splice sites by CRISPR-mediated knock-in.

5.5. Conclusion

In this thesis, I infer the evolutionary history of the function of *dsx* in insects. My results strongly support some hypotheses on the evolution of *dsx* (Chapters 2 and 3). In addition, I proposed the new model based on the sex-antagonistic function of *dsx* for *vitellogenin* expression and the change in the C-terminal motif (Chapters 3 and 4).

The primary question in this thesis is whether coding mutations cause neo-functionalization of alternative isoforms. To this end, I attempt to generalize my model for *dsx* to the exaptive process of alternative isoforms. This attempt shows that the evolutionary process of *dsx* shows the recruitment of isoforms to new roles and authenticates several hypotheses such as the balancer model. Overall, I provide possible speculation that neo-functionalization of genes can be attained by the co-option of alternative isoforms as a byproduct of balancing isoforms via accumulation of mutations in coding motifs. This exaptive process is consistent with an earlier statement (Lareau et al. 2004, Papasaikas and Valcárcel 2012, Ule and Blencowe 2019) that alternative isoforms are material for Jacob (1977) 's "evolutionary tinkering."

In this thesis, one of the subjects is to investigate what exaptive functions exist before neofunctionalization. One of such roles is the balancer of isoforms, as mentioned above. Also, I proposed cryptic roles such as the upregulation of *vitellogenin* of the female-specific *dsx* isoform as putative exaptive functions before the neo-functionalization (Chapter 3). Such an exaptive state may lead to maintaining the isoforms with the complete coding region and may resolve the seeming contradiction that the frequent loss or turnover vs. the conservation of "low-adaptive" isoforms among species.

I postulated at least two exaptive states before the neo-functionalization of isoforms. These exaptive states might have different significance for the evolution of *dsx* isoforms: cryptic functions such as upregulation of *vitellogenin* may have resulted in the

maintenance of isoforms with complete ORFs in females. Alternatively, a balancing function of female-specific isoforms could result in the high expression of isoforms in female cells involved in morphogenesis. On the basis of the general prediction about the functional evolution of alternative isoforms (Graveley 2001, Xing and Lee 2006, Kim et al. 2008, Keren et al. 2010, Kelemen et al. 2013) and my discussion, more generally, I speculate that the step before acquiring a new function of alternative isoforms is the high conservation of the isoforms due to their cryptic function and the high expression of the isoforms due to their balancer role.

The evolutionary process of antagonistic roles in a single gene has been overlooked in splicing evolution. To infer the evolution of such antagonism through splicing isoforms, I investigated dsx functionality, which shows the sex-antagonistic roles in the morphogenesis of Aparaglossata. Then, my results and hypothesis suggest that dsx initially acquired sex-antagonistic roles in some female-specific gene expression and subsequently obtained the function for female morphogenesis. Therefore, the antagonistic nature of functionality in a single gene may evolve not at once but gradually. This speculation could be applied to the evolution of functional antagonism of pleiotropic genes, which play roles in many aspects of biological processes.

The question of what coding mutation leads to new functions of isoforms is also the issue of this paper. Here, I proposed that the extension of the C-terminal region is linked to the neo-functionalization of isoforms (Chapter 4). In *Drosophila melanogaster*, the relevant motif of *dsx* is the disordered region (Yang et al. 2008). The C-terminal disordered region less disrupts protein structure and function (Lareu et al. 2004, Blujan et al. 2013, Reixachs-Solé and Eyras 2021). Therefore, mutation into these low invasive sites might bring on escaping from purifying selection and be a factor underpinning the neo-functionalization of isoforms. This opinion should be tested in the future, as it does not predate any alternative hypothesis at this time.

Finally, the evolutionary history of *dsx* in this thesis recalls the genomenclature by Brosius and Gould (1992). The authors coined four terms to describe exaptation at the molecular genetic level: "nuon," which is any structure encoded in nucleic acids; "potonuon," meaning a nuon that could potentially recruit new roles during evolution; "naptonuon, " which is a potonuon that dis not acquire a new function; "xaptonuon, " meaning a potonuon that was co-opted for a new role. Graur (1993) harshly criticized the

authors for their unhelpful and unthoughtful terminology. These terms are then rarely used and have been forgotten from the front pages of evolutionary biology. However, as in the rebuttal by Brosius and Gould (1993), these terms could successfully integrate the exaptation process. Using these terms, the exaptive process can be organized in the course of evolution as nuon → potonuon → naptonuon or xaptonuon (Brosius and Gould 1993). These terms can well be applied to the evolutionary process of *dsx*: nuon can be referred to as the male-specific expressed *dsx* of the common ancestor of Branchiopoda and Dicondylia; potonuon as the female-specific isoform that was "non-functional" in female morphogenesis before the appearance of the common ancestor of Aparaglossta; naptonuon as the female-specific isoform in non-Aparaglossata; xaptonuon as the female-specific isoforms that function in female morphogenesis in Aparaglossata. Then, for example, we can propose that the shift from potonuon to xaptonuon in splicing isoforms may correlate with mutations in C-terminal motifs. I expect that the use of these terms can be helpful to generalize the cases of neo-functionalization via alternative isoforms.

References

An W, Wensink PC. 1995. Integrating sex-and tissue-specific regulation within a single *Drosophila* enhancer. *Genes Dev.* **9:** 256–266.

Alexander DE. 2018. A century and a half of research on the evolution of insect flight. *Arthropod Struct Dev.* **47:** 322–327.

Altschmied J, Delfgaauw J, Wilde B, Duschl J, Bouneau L, Volff JN, Schartl M. 2002. Subfunctionalization of duplicate *mitf* genes associated with differential degeneration of alternative exons in fish. *Genetics* **161:** 259–267.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex determination: why so many ways of doing it? *PLoS Biol.* **12:** e1001899.

Ballesteros JA, Santibáñez-López CE, Baker CM, Benavides LR, Cunha TJ, Gainett G, Ontano AZ, Setton EVW, Arango CP, et al. 2022. Comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida. *Mol Biol Evol.* **39**: msac021.

Ballesteros JA, Sharma PP. 2019. A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst Biol.* **68**: 896-917.

Baral S, Arumugam G, Deshmukh R, Kunte K. 2019. Genetic architecture and sex–specific selection govern modular, male–biased evolution of *doublesex*. *Sci Adv.* **5:** eaau3753.

Baralle FE, Giudice J. 2017. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol.* **18:** 437–451.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misguitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ, Blencowe BJ. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338:** 1587–1593.

Bassett AR, Liu JL. 2014. CRISPR/Cas9 and genome editing in *Drosophila*. *J Genet Genom*. **41:** 7–19.

Bayrer JR, Zhang W, Weiss MA. 2005. Dimerization of *doublesex* is mediated by a cryptic ubiquitin-associated domain fold: implications for sex-specific gene regulation. *J Biol Chem*. **280:** 32989–32996.

Bear A, Monteiro A. 2013. Both cell-autonomous mechanisms and hormones contribute to sexual development in vertebrates and insects. *Bioessays* **35:** 725–732.

Bell LR, Maine EM, Schedl P, Cline TW. 1988. *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55:** 1037–1046.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA*. **74:** 3171–3175.

Beukeboom LW, Perrin N. 2014. *The evolution of sex determination*. Oxford: Oxford University Press.

Beutel RG, Yavorskaya MI, Mashimo Y, Fukui M, Meusemann K. 2017. The phylogeny of Hexapoda (Arthropoda) and the evolution of megadiversity. *Proc Arthropod Embryol Soc Jap.* **51:** 1–15.

Boise LH, González-García M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nunez G., Thompson CB. 1993. *bcl-x*, a *bcl-2*-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **74:** 597–608.

Blencowe BJ. 2017. The relationship between alternative splicing and proteomic complexity. *Trends Biochem Sci.* **42:** 407–408.

Buljan M, Chalancon G, Dunker AK, Bateman A, Balaji S, Fuxreiter M, Babu MM. 2013. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr Opin Struct Biol.* **23:** 443–450.

Bopp D, Saccone G, Beye M. 2014. Sex determination in insects: variations on a common theme. *Sex Dev.* **8:** 20–28.

Boudinot BE. 2018. A general theory of genital homologies for the Hexapoda (Pancrustacea) derived from skeletomuscular correspondences, with emphasis on the Endopterygota. *Arthropod Struct Dev.* **47:** 563–613.

Boue S, Letunic I, Bork P. 2003. Alternative splicing and evolution. *Bioessays* **25:** 1031–1034.

Brosius J. 1991. Retroposons—seeds of evolution. *Science* **251:** 753–753.

Brosius J. 2019. Exaptation at the molecular genetic level. *Sci China Life Sci.* **62:** 437–452.

Brosius J, Gould SJ. 1992. On" genomenclature": a comprehensive (and respectful)

taxonomy for pseudogenes and other" junk DNA". *Proc Natl Acad Sci USA*. **89:** 10706–10710.

Brosius J, Gould SJ. 1993. Molecular constructivity. *Nature* **365:** 102–102.

Buck C, Edwards JS. 1990. The effect of appendage and scale loss on instar duration in adult firebrats, *Thermobia domestica* (Thysanura). *J Exp Biol.* **151:** 341–347.

Burtis KC, Baker BS. 1989. *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex–specific polypeptides. *Cell* **56:** 997–1010.

Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO. 2017. Alternative splicing and the evolution of phenotypic novelty. *Philos Trans R Soc B Biol Sci.* **372:** 20150474.

Byrne BM, Gruber MABG, Ab G. 1989. The evolution of egg yolk proteins. *Prog Biophys Mol Biol.* **53:** 33–69.

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol.* **3:** e245.

Casinos A. 2017. From Cuénot's préadaptation to Gould and Vrba's exaptation: a review. *Biol J Linn Soc.* **121:** 239–247.

Chen L, Tovar-Corona JM, Urrutia AO. 2012. Alternative splicing: a potential source of functional innovation in the eukaryotic genome. *Int J Evol Biol.* **2012:** 596274.

Chipman AD. 2010. Parallel evolution of segmentation by co-option of ancestral gene regulatory networks. *Bioessays* **32:** 60–70.

Chipman AD. 2021. Developmental exaptation. In: de la Rosa LN, Müller GB, editors. Evolutionary developmental biology: a reference guide. Switzerland (Cham): Springer Nature. p. 29–38.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12:** 1–8.

Darling AL, Uversky VN. 2018. Intrinsic disorder and posttranslational modifications: the darker side of the biological dark matter. *Front Genet.* **9:** 158.

Darwin C. 1859. *On the origin of species by means of natural Selection, or the preservation of favoured races in the struggle for life 1 ed.* London: John Murray.

Darwin C. 1871. *The descent of man, and selection in relation to sex vol. 1*. London: John Murray.

Darwin C. 1872. *On the origin of species by means of natural Selection, or the preservation of favoured races in the struggle for life 6 ed.* London: John Murray.

De Loof A, Huybrechts R. 1998. "Insects do not have sex hormones": a myth? *Gen Comp Endocrinol.* **111:** 245–260.

Dennett DC. 1995. *Darwin's dangerous idea: evolution and the meanings of life.* New York: Simon and Schuster.

Dermauw W, Jonckheere W, Riga M, Livadaras I, Vontas J, Van Leeuwen T. 2020. Targeted mutagenesis using CRISPR-Cas9 in the chelicerate herbivore *Tetranychus urticae*. *Insect Biochem Mol Biol.* **120:** 103347.

Duan J, Xu H, Guo H, O'Brochta DA, Wang F, Ma S, Zhang L, Zha X, Zhao P, Xia Q. 2013. New insights into the genomic organization and splicing of the *doublesex* gene, a terminal regulator of sexual differentiation in the silkworm *Bombyx mori*. *PLoS ONE* **8:** e79703.

Emeljanov AF. 2014. The evolutionary role and fate of the primary ovipositor in insects. *Entomol Rev.* **94:** 367–396.

Endo Y, Liu Y, Kanno K, Takahashi M, Matsushita M, Fujita T. 2004. Identification of the mouse *H-ficolin* gene as a pseudogene and orthology between mouse *ficolins A/B* and human *L-/M-ficolins*. *Genomics* **84:** 737–744.

Erdman SE, Chen HJ, Burtis KC. 1996. Functional and genetic characterization of the oligomerization and DNA binding properties of the *Drosophila doublesex* proteins. *Genetics* **144:** 1639–1652.

Futuyma DJ, Kirkpatrick M. 2017. *Evolution 4th ed*. New York: Oxford Press.

Ganfornina MD, Sánchez D. 1999. Generation of evolutionary novelty by functional shift. *Bioessays* **21**: 432–439.

Geuverink E, Beukeboom LW. 2014. Phylogenetic distribution and evolutionary dynamics of the sex determination genes *doublesex* and *transformer* in insects. *Sex Dev.* **8:** 38–49.

Ghosh N, Bakshi A, Khandelwal R, Rajan SG, Joshi R. 2019. The Hox gene *Abdominal-B* uses Doublesex^{F} as a cofactor to promote neuroblast apoptosis in the *Drosophila* central nervous system. *Development* **146:** dev175158.

Gilbert W. 1978. Why genes in pieces? *Nature* **271:** 501.

Gilles AF, Schinko JB, Averof M. 2015. Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle *Tribolium castaneum*. *Development* **142:** 2832–2839.

Giribet G, Edgecombe GD. 2019. The phylogeny and evolutionary history of arthropods. *Curr Biol.* **29:** 592–602.

Glassford WJ, Johnson WC, Dall NR, Smith SJ, Liu Y, Boll W, Noll M, Rebeiz N. 2015. Co-option of an ancestral hox-regulated network underlies a recently evolved morphological novelty. *Dev Cell* **34:** 520–531.

Gómez-Redondo I, Planells B, Navarrete P, Gutiérrez-Adán A. 2021. Role of alternative splicing in sex determination in vertebrates. *Sex Dev.* **15:** 381–391.

Gotoh H, Zinna RA, Warren I, DeNieu M, Niimi T, Dworkin I, Emlen DJ, Miura T, Lavine LC. 2016. Identification and functional analyses of sex determination genes in the sexually dimorphic stag beetle *Cyclommatus metallifer*. *BMC Genom*. **17:** 250.

Gould SJ. 2002. *The structure of evolutionary theory*. Harvard University Press.

Gould SJ, Vrba ES. 1982. Exaptation—a missing term in the science of form. *Paleobiology.* **8:** 4–15.

Grabherr, MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q. 2011. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol.* **29:** 644.

Grantham ME, Brisson JA. 2018. Extensive differential splicing underlies phenotypically plastic aphid morphs. *Mol Biol Evol.* **35:** 1934–1946.

Graur D. 1993. Molecular deconstructivism. *Nature* **363:** 490–490.

Graveley BR. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17:** 100–107.

Graveley BR. 2008. The haplo-spliceo-transcriptome: common variations in alternative splicing in the human population. *Trends Genet.* **24**: 5–7.

Gruzin M, Mekheal M, Ruhlman K, Winkowski M, Petko J. 2020. Developmental expression of *doublesex*-related transcripts in the common house spider, *Parasteatoda tepidariorum*. *Gene Expr Patterns* **35:** 119101.

Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Münsterberg A, Vivian N, Goodfellow P, Lovell-Badge R. 1990. A gene mapping to the sex-determining region

of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346:** 245–250.

Gueroussov S, Gonatopoulos-Pournatzis T, Irimia M, Raj B, Lin ZY, Gingras AC, Blencowe BJ. 2015. An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **349:** 868–873.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* **59:** 307–321.

Guo L, Xie W, Liu Y, Yang Z, Yang X, Xia J, Wang S, Wu Q, Zhang Y. 2018. Identification and characterization of *doublesex* in *Bemisia tabaci*. *Insect Mol Biol.* **27:** 602–632.

Gustafson GT, Baca SM, Alexander AM, Short AEZ. 2020. Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance. *Syst Entomol.* **45:** 552–570.

Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler JK, Sharakhova MV, Elahi R, Anderson MAE, Chen XG, et al. 2015. A male-determining factor in the mosquito *Aedes aegypti*. *Science* **348:** 1268–1270.

Hasselmann M, Gempe T, Schiøtt M, Nunes-Silva CG, Otte M, Beye M. 2008. Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* **454:** 519–522.

Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, Fernandino JI, Somoza GM, Yokota M, Strüssmann CA. 2012. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc Natl Acad Sci USA.* **109:** 2955–2959.

Hayward A, Takahashi T, Bendena WG, Tobe SS, Hui JH. 2010. Comparative genomic and phylogenetic analysis of *vitellogenin* and other large lipid transfer proteins in metazoans. *FEBS Lett.* **584:** 1273–1278.

Hennig W. 1966. *Phylogenetic systematics*. London: University of Illinois Press.

Herpin A, Schartl M. 2015. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Rep.* **16:** 1260–1274.

Hildreth PE. 1965. Doublesex, a recessive gene that transforms both males and females of *Drosophila* into intersexes. *Genetics* **51:** 659–678.

Hildreth PE, Lucchesi JC. 1963. A gene which transforms males and females into intersexes. *Proc 11th Intern Congr Genet.* **1:** 171.

Hopkins BR, Kopp A. 2021. Evolution of sexual development and sexual dimorphism in insects. *Curr Opin Genet Dev.* **69:** 129–139.

Hoshijima K, Inoue K, Higuchi I, Sakamoto H, Shimura Y. 1991. Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. *Science* **252:** 833–836.

Howes TR, Summers BR, Kingsley DM. 2017. Dorsal spine evolution in threespine sticklebacks via a splicing change in *MSX2A*. *BMC Biol.* **15:** 1–16.

Ito Y, Harigai A, Nakata M, Hosoya T, Araya K, Oba Y, Ito A, Ohde T, Yaginuma T, Niimi T. 2013. The role of *doublesex* in the evolution of exaggerated horns in the Japanese rhinoceros beetle. *EMBO Rep.* **14:** 561–567.

Jacob F. 1977. Evolution and tinkering. *Science* **196**: 1161–1166.

Janssen R, Budd, GE. 2010. Gene expression suggests conserved aspects of *Hox* gene regulation in arthropods and provides additional support for monophyletic Myriapoda. *EvoDevo* **1:** 1–11.

Just J, Laslo M, Lee YJ, Yarnell M, Zhang Z, Angelini DR. 2021. Distinct developmental mechanisms influence sexual dimorphisms in the milkweed bug *Oncopeltus fasciatus*. *bioRxiv*. doi: 10.1101/2021.05.12.443917.

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 Genet.* **7:** e1001345.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30:** 772–780.

Kawahara AY, Plotkin D, Espeland M, Meusemann K, Toussaint EFA, Donath A, Gimnich F, Frandsen PB, Zwick A, dos Reis M, et al. 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc Natl Acad Sci USA.* **116:** 22657–22663.

Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. 2013. Function of alternative splicing. *Gene* **514:** 1–30.

Keren H, Lev-Maor G, Ast G. 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet.* **11:** 345–355.

Kijimoto T, Moczek AP, Andrews J. 2012. Diversification of *doublesex* function underlies morph-, sex-, and species-specific development of beetle horns. *Proc Natl Acad Sci USA*. **109:** 20526–20531.

Kim E, Goren A, Ast G. 2008. Alternative splicing: current perspectives. *Bioessays* **30:** 38–47.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT–genotype. *Nat Biotechnol.* **37:** 907–915.

Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, Ishihara G, Kawaoka S, Sugano S, Shimada T, et al. 2014. A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* **509:** 633–636.

Klag J. 1977. Differentiation of primordial germ cells in the embryonic development of *Thermobia domestica*, Pack. (*Thysanura*): an ultrastructural study. *J Embryol exp Morph.* **38:** 93–114.

Kobayashi T, Matsuda M, Kajiura-Kobayashi H, Suzuki A, Saito N, Nakamoto M, Shibata N, Nagahama Y. 2004. Two DM domain genes, DMY and *DMRT1*, involved in testicular differentiation and development in the medaka, *Oryzias la*tipes. *Dev Dyn.* **231:** 518–526.

Koopman P. 2009. Sex determination: the power of *DMRT1*. *Trends Genet.* **25:** 479–481.

Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. 1991. Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351:** 117–121.

Kopelman NM, Lancet D, Yanai I. 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat Genet.* **37:** 588–589.

Kopp A. 2012. *Dmrt* genes in the development and evolution of sexual dimorphism. *Trends Genet.* **28:** 175–184.

Kristensen NP. 1975. The phylogeny of hexapod “orders”. A critical review of recent accounts. *J Zool Syst Evol Res.* **13:** 1–44.

Krzywinska E, Dennison NJ, Lycett GJ, Krzywinski J. 2016. A maleness gene in the malaria mosquito *Anopheles gambiae*. *Science* **353:** 67–69.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* **35:** 1547–1549.

Kyrou K, Hammond AM, Gaizi R, Franjc N, Burt A, Beaghton AK, Nolan T, Crisanti A.

2018. A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat Biotechnol.* **36:** 1062–1066

Lago DC, Martins JR, Dallacqua RP, Santos DE, Bitondi MM, Hartfelder K. 2020. Testis development and spermatogenesis in drones of the honey bee, *Apis mellifera* L. *Apidologie* **51:** 935–955.

Li H, Leavengood JM, Chapman EG, Burkhardt D, Song F, Jiang P, Liu J, Zhou X, Cai W. 2017. Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *Proc R Soc B.* **284:** 20171223.

Li S, Li F, Yu K, Xiang J. 2018. Identification and characterization of a *doublesex* gene which regulates the expression of insulin-like androgenic gland hormone in *Fenneropenaeus chinensis*. *Gene* **649:** 1–7.

Li R, Meng Q, Qi J, Hu L, Huang J, Zhang Y, Yang J, Sun J. 2022. Microinjection-based CRISPR/Cas9 mutagenesis in the decapoda crustaceans, *Neocaridina heteropoda* and *Eriocheir sinensis*. *J Exp Biol.* **225:** jeb.243702.

Li Z, Tiley GP, Galuska SR, Reardon CR, Kidder TI, Rundell RJ, Barker MS. 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc Natl Acad Sci USA.* **115:** 4713–4718.

Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. 2006. Intrinsic disorder in transcription factors. *Biochemistry* **45:** 6873–6888.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real–time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25:** 402–408.

Lozano-Fernandez J, Giacomelli M, Fleming JF, Chen A, Vinther J, Thomsen PF, Glenner H, Palero F, Legg DA, Iliffe TM, et al. 2019. Pancrustacean evolution illuminated by taxon-rich genomic-scale sata sets with an expanded Remipede sampling. *Genome Biol Evol.* **11:** 2055–2070.

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* **8:** e1000506.

McLennan DA. 2008. The concept of co-option: why evolution often looks miraculous. *Evol: Educ Outreach* **1:** 247–258.

MacLean DW, Meedel TH, Hastings KEM. 1997. Tissue-specific alternative splicing of ascidian *troponin I* isoforms - redesign of a protein isoform generating mechanism during chordate evolution. *J Biol Chem.* **272:** 32115–32120.

Mann RS, Carroll SB. 2002. Molecular mechanisms of selector gene function and evolution. *Curr Opin Genet Dev.* **12:** 592–600.

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

Martin A, Serano JM, Jarvis E, Bruce HS, Wang J, Ray S, Barker CA, O'Connell LC, Patel NH. 2016. CRISPR/Cas9 Mutagenesis reveals versatile roles of *hox* genes in crustacean limb specification and evolution. *Curr Biol.* **26:** 14–26.

Mashimo Y, Beutel RG, Dallai R, Lee CY, Machida R. 2014. Embryonic development of Zoraptera with special reference to external morphology, and its phylogenetic implications (Insecta). *J Morphol.* **275:** 295–312.

Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N, Hori H, Hamaguchi S, Sakaizumi M. 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417:** 559–563.

Matsuda R.1976. *Morphology and Evolution of the Insect Abdomen: With Special Reference to Developmental Patterns and their Bearings upon Systematics.* Oxford: Pergamon Press.

Matsuoka Y, Nakamura T, Watanabe T, Barnett AA, Noji S, Mito T, Extavour CG. 2021. Establishment of CRISPR/Cas9-based knock-in in a hemimetabolous insect: Targeted gene tagging in the cricket *Gryllus bimaculatus*. *bioRxiv*. doi.org/10.1101/2021.05.10.441399.

Mawaribuchi S, Ito Y, Ito M. 2019. Independent evolution for sex determination and differentiation in the DMRT family in animals. *Biol Open* **8:** bio041962.

McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, Donath A, Escalona HE, Friedrich F, Letsch H, et al. 2019. The evolution and genomic basis of beetle diversity. *Proc Natl Acad Sci USA.* **116:** 24729–24737.

Meccariello A, Salvemini M, Primo P, Hall B, Koskinioti P, Dalíková M, Gravina A, Gucciardino MA, Forlenza F, Gregoriou ME, et al. 2019. *Maleness-on-the-Y (MoY)* orchestrates male sex determination in major agricultural fruit fly pests. *Science* **365:** 1457–1460.

Mine S, Sumitani M, Aoki F, Hatakeyama M, Suzuki MG. 2017. Identification and functional characterization of the sex-determining gene *doublesex* in the sawfly, *Athalia rosae* (Hymenoptera: Tenthredinidae). *Appl Entomol Zool.* **52:** 497–509.

Mine S, Sumitani M, Aoki F, Hatakeyama M, Suzuki MG. 2021. Effects of functional depletion of *doublesex* on male development in the sawfly, *Athalia rosae*. *Insects* **12:** 849.

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* **30:** 1188–1195.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* **37:** 1530–1534.

Mirdita M, Ovchinnikov S, Steinegger M. 2021. ColabFold-Making protein folding accessible to all. *bioRxiv*. doi: 10.1101/2021.08.15.456425.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346:** 763–767.

Miyachi Y, Yahata K. 2012. Morphological study of ovarian structures in *Scolopendromorph centipedes* (Myriapoda: Chilopoda) with special reference to the position of oocyte growth. *Proc Arthropod Embryol Soc Jap.* **47:** 21–28.

Miyawaki S, Kuroki S, Maeda R, Okashita N, Koopman P, Tachibana M. 2020. The mouse *Sry* locus harbors a cryptic exon that is essential for male sex determination. *Science* **370:** 121–124.

Miyazaki S, Fujiwara K, Kai K, Masuoka Y, Gotoh H, Niimi T, Hayashi Y, Shigenobu S, Maekawa K. 2021. Evolutionary transition of *doublesex* regulation in termites and cockroaches: from sex-specific splicing to male-specific transcription. *Sci Rep.* **11:** 15992.

Modrek B, Lee CJ. 2003. Alternative splicing in the human, mouse, and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet.* **34:** 177–180.

Morita S, Ando T, Maeno A, Mizutani T, Mase M, Shigenobu S, Niimi T. 2019. Precise staging of beetle horn formation in *Trypoxylus dichotomus* reveals the pleiotropic roles of *doublesex* depending on the spatiotemporal developmental contexts. *PLOS Genet.* **15:** e1008063.

Naftaly AS, Pau S, White MA. 2021. Long-read RNA sequencing reveals widespread sex-specific alternative splicing in threespine stickleback fish. *Genome Res.* **31:** 1486–1497.

Nagahama Y, Chakraborty T, Paul-Prasanth B, Ohta K, Nakamura M. 2021. Sex determination, gonadal sex differentiation, and plasticity in vertebrate species. *Physiol Rev.* **101:** 1237–1308.

Nakanishi T, Kato Y, Matsuura T, Watanabe H. 2014. CRISPR/Cas-mediated targeted mutagenesis in *Daphnia magna*. *PLoS One* **9:** e98363.

Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, Shan Z, Haaf T, Shimizu N, Shima A, Schmid M, Schartl M. 2002. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc Natl Acad Sci USA*. **99:** 11778–11783.

Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463:** 457–463.

Ohbayashi F, Suzuki MG, Mita K, Okano K, Shimada T. 2001. A homologue of the *Drosophila doublesex* gene is transcribed into sex-specific mRNA isoforms in the silkworm, *Bombyx mori*. *Comp Biochem Physiol B Biochem Mol Biol*. **128:** 145–158.

Ohde T, Takehana Y, Shiotsuki T, Niimi T. 2018. CRISPR/Cas9-based heritable targeted mutagenesis in *Thermobia domestica*: a genetic tool in an apterygote development model of wing evolution. *Arthropod Struct Dev.* **47:** 362–369.

Ohde T, Yaginuma T, Niimi T. 2011. Nymphal RNAi analysis reveals novel function of scalloped in antenna, cercus and caudal filament formation in the firebrat, *Thermobia domestica*. *J Insect Biotechnol Sericol.* **80:** 101–108.

Panara V, Budd GE, Janssen R. 2019. Phylogenetic analysis and embryonic expression of panarthropod Dmrt genes. *Front Zool.* **16:** 1–18.

Papasaikas P, Valcárcel J. 2012. Splicing in 4D. *Science* **338:** 1547–1548.

Pei XJ, Fan YL, Bai Y, Bai TT, Schal C,Zhang ZF, Chen N, Li S, Liu TX. 2021. Modulation of fatty acid elongation in cockroaches sustains sexually dimorphic hydrocarbons and female attractiveness. *PLOS Biol.* **19:** e3001330.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* **33:** 290–295.

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, et al. 2017. Evolutionary History of the Hymenoptera. *Curr Biol.* **27:** 1013–1018.

Peters RS, Meusemann K, Petersen M, Mayer C, Wilbrandt J, Ziesmann T, Donath A, Kjer KM, Aspöck U, Aspöck, H, et al. 2014. The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. *BMC Evol Biol.* **14:** 52.

Pickrell JK, Pai AA, Gilad Y, Pritchard JK. 2010. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* **6:** e1001236.

Pomerantz AF, Hoy MA, Kawahara AY. 2015. Molecular characterization and evolutionary insights into potential sex-determination genes in the western orchard predatory mite *Metaseiulus occidentalis* (Chelicerata: Arachnida: Acari: Phytoseiidae). *J Biomol Struct Dyn.* **33:** 1239–1253.

Price DC, Egizi A, Fonseca DM. 2015. The ubiquity and ancestry of insect *doublesex*. *Sci Rep.* **5:** 1–9.

Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, Zarkower D. 2000. *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* **14:** 2587–2595.

R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reixachs-Solé M, Eyras E. 2022. Uncovering the impacts of alternative splicing on the proteome with current omics techniques. *Wiley Interdiscip Rev RNA*. e1707.

Rideout EJ, Narsaiya MS, Grewal SS. 2015. The sex determination gene *transformer* regulates male-female differences in *Drosophila* body size. *PLoS Genet.* **11:** e1005683.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140.

Rodriguez-Caro F, Fenner J, Bhardwaj S, Cole J, Benson C, Colombara AM, Papa R, Brown MW, Martin A, Range RC, et al. 2021. Novel *doublesex* duplication associated with sexually dimorphic development of dogface butterfly wings. *Mol Biol Evol.* **38:** 5021–5033.

Rogers TF, Palmer DH, Wright AE. 2021. Sex-specific selection drives the evolution of alternative splicing in birds. *Mol Biol Evol.* **38:** 519–530.

Romero-Pozuelo J, Foronda D, Martín P, Hudry B, Merabet S, Graba Y, Sánchez-Herrero E. 2019. Cooperation of axial and sex specific information controls *Drosophila* female genitalia growth by regulating the *Decapentaplegic* pathway. *Dev Biol.* **454:** 145–155.

Roth A, Vleurinck C, Netschitailo O, Bauer V, Otte M, Kaftanoglu O, Page RE, Beye M. 2019. A genetic switch for worker nutrition-mediated traits in honeybees. *PLOS Biol.* **17:** e3000171.

Rousset A., Bitsch C. 1993. Comparison between endogenous and exogenous yolk proteins along an ovarian cycle in the firebat *Thermobia domestica* (Insecta, Thysanura). *Comp Biochem Physiol B Biochem Mol Biol.* **104:** 33–44.

Sato Y, Shinka T, Sakamoto K, Ewis AA, Nakahori Y. 2010. The male-determining gene *SRY* is a hybrid of *DGCR8* and *SOX3*, and is regulated by the transcription factor *CP2*. *Molecular and Cellular Biochemistry* **337:** 267–275.

Sakai H, Sumitani M, Chikami Y, Yahata K, Uchino K, Kiuchi T, Katsuma S, Aoki F, Sezutsu H, Suzuki MG. 2016. Transgenic expression of the piRNA-resistant *Masculinizer* gene induces female-specific lethality and partial female-to-male sex reversal in the silkworm, *Bombyx mori*. *PLoS Genet.* **12:** e1006203.

Schaefke B, Sun W, Li YS, Fang L, Chen W. 2018. The evolution of posttranscriptional regulation. *Wiley Interdiscip Rev RNA* **9:** e1485.

Schwentner M, Combosch DJ, Nelson JP, Giribet G. 2017. A phylogenomic solution to the origin of insects by resolving crustacean-hexapod relationships. *Curr Biol.* **27:** 1818–1824.

Schwentner M, Richter S, Rogers DC, Giribet G. 2018. Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics. *Proc R Soc Lond B Biol Sci.* **285:** 20181524.

Sharma PP, Gupta T, Schwager EE, Wheeler WC, Extavour CG. 2014a. Subdivision of arthropod *cap-n-collar* expression domains is restricted to Mandibulata. *EvoDevo* **5:** 1–11.

Sharma PP, Kaluziak ST, Pérez-Porro AR, González VL, Hormiga G, Wheeler WC, Giribet G. 2014b. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol Biol Evol.* **31:** 2963–2984.

Sharma A, Heinze SD, Wu Y, Kohlbrenner T, Morilla I, Brunner C, Wimmer EA, van de Zande L, Robinson MD, Beukeboom LW, Bopp D. 2017. Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*. *Science* **356:** 642–645.

Sharma A, Pham MN, Reyes JB, Chana R, Yim WC, Heu CC, Kim D, Chaverra-Rodriguez D, Rasgon JL, Harrell RA, et al. 2022. Cas9-mediated gene-editing in the black-legged tick, *Ixodes scapularis*, by embryo injection and ReMOT control.

iScience 103781. doi.org/10.1016/j.isci.2022.103781.

Shirai Y, Piulachs MD, Belles X, Daimon T. 2022. DIPA-CRISPR is a simple and accessible method for insect gene editing. *Cell Rep Method.* **2:** 100215.

Shubin N. 2020. *Some Assembly Required: Decoding Four Billion Years of Life, from Ancient Fossils to DNA* New York: Pantheon Books.

Shukla JN, Palli SR. 2012a. Sex determination in beetles: production of all male progeny by parental RNAi knockdown of *transformer*. *Sci Rep.* **2:** 1–9.

Shukla JN, Palli SR. 2012b. Doublesex target genes in the red flour beetle, *Tribolium castaneum*. *Sci Rep.* **2:** 948.

Simpson GG. 1944. *Tempo and mode in Evolution* London: Oxford University Press.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346:** 240–244.

Singh P, Ahi EP. 2022. The importance of alternative splicing in adaptive evolution. *Mol Ecol.* online. doi.org/10.1111/mec.16377.

Singh P, Börger C, More H, Sturmbauer C. 2017. The role of alternative splicing and differential gene expression in cichlid adaptive radiation. *Genome Biol Evol.* **9:** 2764–2781.

Snodgrass RE. 1957. A revised interpretation of the external reproductive organs of male insects. *Smithson Misc Collect.* **135:** 1–6.

Smith C, Barber I, Wootton RJ, Chittka L. 2004. A receiver bias in the origin of three–spined stickleback mate choice. *Proc Royal Soc B.* **271:** 949–955.

Smith CC, Rieseberg LH, Hulke BS, Kane NC. 2021. Aberrant RNA splicing due to genetic incompatibilities in sunflower hybrids. *Evolution* **75:** 2747–2758.

Smith CC, Tittes S, Mendieta JP, Collier-Zans E, Rowe HC, Rieseberg LH, Kane NC. 2018. Genetics of alternative splicing evolution during sunflower domestication. *Proc Natl Acad Sci USA.* **115:** 6768–6773.

Sober E. 2000. *Philosophy of Biology 2nd ed.* Colorado: Westview Press.

Sorek R, Shamir R, Ast G. 2004. How prevalent is functional alternative splicing in the

human genome? *Trends Genet.* **20:** 68–71.

Steward RA, de Jong MA, Oostra V, Wheat CW. 2022. Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nat Commun.* **13:** 1–12.

Stevens M, Oltean S. 2019. Modulation of the Apoptosis Gene Bcl-x Function Through Alternative Splicing. *Front Genet.* **10:** 804.

Su Z, Wang J, Yu J, Huang X, Gu X. 2006. Evolution of alternative splicing after gene duplication. *Genome Res.* **16:** 182–189.

Suzuki MG, Imanishi S, Dohmae N, Asanuma M, Matsumoto S. 2010. Identification of a male-specific RNA binding protein that regulates sex-specific splicing of *Bmdsx* by increasing RNA binding activity of *BmPSI*. *Mol Cell Biol.* **30:** 5776–5786.

Suzuki MG, Funaguma S, Kanda T, Tamura T, Shimada T. 2003. Analysis of the biological functions of a *doublesex* homologue in *Bombyx mori*. *Dev Gene Evol.* **213:** 345–354.

Takaku Y, Suzuki H, Ohta I, Ishii D, Muranaka Y, Shimomura M, Hariyama T. 2013. A thin polymer membrane, nano-suit, enhancing survival across the continuum between air and high vacuum. *Proc Natl Acad Sci USA.* **110:** 7631–7635.

Takahashi M, Okude G, Futahashi R, Takahashi Y, Kawata M. 2021. The effect of the *doublesex* gene in body colour masculinization of the damselfly *Ischnura senegalensis*. *Biol Lett.* **17:** 20200761.

Takahashi M, Takahashi Y, Kawata M. 2019. Candidate genes associated with color morphs of female-limited polymorphisms of the damselfly *Ischnura senegalensis*. *Heredity* **122:** 81–92.

Takehana Y, Matsuda M, Myosho T, Suster ML, Kawakami K, Shin-I T, Kohara Y, Kuroki Y, Toyoda A, Fujiyama A, et al. 2014. Co-option of *Sox3* as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nat Commun.* **5:** 4157.

Talavera D, Vogel C, Orozco M, Teichmann SA, de la Cruz X. 2007. The (in)dependence of alternative splicing and gene duplication. *PLoS Comput Biol.* **3:** e33.

Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* **38:** 615–643.

Telonis-Scott M, Kopp A, Wayne ML, Nuzhdin SV, McIntyre LM. 2009. Sex-specific splicing in *Drosophila*: widespread occurrence, tissue specificity and evolutionary

conservation. *Genetics* **181:** 421–434.

Thongsaiklaing T, Passara H, Nipitwathanaphon M, Ngernsiri L. 2018. Identification and characterization of *doublesex* from the pumpkin fruit fly, *Bactrocera tau* (Diptera: Tephritidae). *Europ J Entomol.* **115:** 602–613.

Tovar-Corona JM, Castillo-Morales A, Chen LU, Olds BP, Clark JM, Reynolds SE, Pittendrigh BR, Feil EJ, Urrutia AO. 2015. Alternative splice in alternative lice. *Mol Biol Evol.* **32:** 2749–2759.

Toyota K, Kato Y, Sato M, Sugiura N, Miyagawa S, Miyakawa H, Watanabe H, Oda S, Ogino Y, Hiruta C, et al. 2013. Molecular cloning of *doublesex* genes of four cladocera (water flea) species. *BMC Genom*. **14:** 239.

Tress ML, Abascal F, Valencia A. 2017. Alternative splicing may not be the key to proteome complexity. *Trends Biochem Sci.* **42:** 98–110.

True JR, Carroll SB. 2002. Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol.* **18:** 53–80.

Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon A, et al. 2021. Highly accurate protein structure prediction for the human proteome. *Nature* **596:** 590–596.

Ule J, Blencowe BJ. 2019. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol Cell* **76**: 329–345.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40:** e115.

Velasque M, Qiu L. Mikheyev AS. 2018. The D*oublesex* sex determination pathway regulates reproductive division of labor in honey bees. *bioRxiv* doi: 10.1101/314492.

Verhulst EC, Beukeboom LW, van de Zande L. 2010. Maternal control of haplodiploid sex determination in the wasp *Nasonia*. *Science* **328:** 620–623.

Verhulst EC, van de Zande L. 2015. Double nexus—Doublesex is the connecting element in sex determination. *Brief Funct Genom*. **14:** 396–406.

Verta JP, Jacobs A. 2021. The role of alternative splicing in adaptation and evolution. *Trends Ecol Evol.* **37:** 299–308.

Volff JN, Zarkower D, Bardwell VJ, Schartl M. 2003. Evolutionary dynamics of the DM domain gene family in metazoans. *J Mol Evol.* **57:** S241–S249.

Wang Y, Chen X, Liu Z, Xu J, Li X, Bi H, Andongma AA, Niu C, Huang Y. 2019. Mutation of *doublesex* induces sex-specific sterility of the diamondback moth *Plutella xylostella. Insect Biochem. Mol Biol.* **112:** 1138.

Wang L, Miura M, Bergeron L, Zhu H, Yuan J. 1994. *Ich-1*, an *Ice/ced-3*-related gene, encodes both positive and negative regulators of programmed cell death. *Cell* **78:** 739–750.

Wang Y, Sun W, Fleischmann S, Millar JG, Ruther J, Verhulst EC. 2022a. Silencing Doublesex expression triggers three-level pheromonal feminization in *Nasonia vitripennis* males. *Proc R Soc B*. 289:20212002.

Wang Y, Rensink A, Fricke U, Riddle MC, Trent C, van de Zande L, Verhulst EC. 2022b. Doublesex regulates male-specific differentiation during distinct developmental time windows in a parasitoid wasp. *Insect Biochem Mol Biol.* 142:103724.

Watanabe T, Noji S, Mito T. 2017. Genome editing in the cricket, *Gryllus bimaculatus*. *Methods Mol Biol.* **1630:** 219–233.

Wexler J, Delaney EK, Belles X, Schal C, Wada-Katsumata A, Amicucci MJ, Kopp A. 2019. Hemimetabolous insects elucidate the origin of sexual development via alternative splicing. *eLife* **8:** e47490.

Wexler JR, Plachetzki DC, Kopp A. 2014. Pan-metazoan phylogeny of the DMRT gene family: a framework for functional studies. *Dev Gene Evol.* **224:** 175–181.

Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone MA, Cassel BL, Bayless KM, Heimberg AM, et al. 2011. Episodic radiations in the fly tree of life. *Proc Natl Acad Sci USA.* **108:** 5690–5695.

Wiley EO. 1981. *Phylogenetics: the theory and practice of phylogenetic systematics.* New York: John Wiley and Sons.

Wilkins AS. 1995. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* **17:** 71–77.

Xiao J, Sekhwal MK, Li P, Ragupathy R, Cloutier S, Wang X, You FM. 2016. Pseudogenes and their genome-wide prediction in plants. *Int J Mol Sci.* **17:** 1991.

Xing Y, Lee C. 2006. Alternative splicing and RNA selection pressure — evolutionary consequences for eukaryotic genomes. *Nat Rev Genet.* **7:** 499–509.

Xu J, Yu Y, Chen K, Huang Y. 2019. *Intersex* regulates female external genital and imaginal disc development in the silkworm. *Insect Biochem Mol Biol.* **108**: 1–8.

Xu J, Zhan S, Chen S, Zeng B, Li Z, James AA, Tan A, Huang Y. 2017. Sexually dimorphic traits in the silkworm, *Bombyx mori*, are regulated by *doublesex*. *Insect Biochem Mol Biol.* **80:** 42–51.

Xue WH, Xu N, Yuan XB, Chen HH, Zhang JL, Fu SJ, Zhang CX, Xu HJ. 2017. CRISPR/Cas9-mediated knockout of two eye pigmentation genes in the brown planthopper, *Nilaparvata lugens* (Hemiptera: Delphacidae). *Insect Biochem Mol Biol*. **93:** 19–26.

Yahata K, Chikami Y, Umetani E. 2018. Morphological study of the ovary in *Hanseniella caldaria* (Myriapoda; Symphyla): The position of oocyte-growth and evolution of ovarian structure in Arthropoda. *Arthropod Struct Dev.* **47:** 655–661.

Yang Y, Zhang W, Bayrer JR, Weiss MA. 2008. Doublesex and the regulation of sexual dimorphism in *Drosophila melanogaster*: structure, function, and mutagenesis of a female-specific domain. *J Biol Chem*. **283:** 7280–7292.

Yi W, Zarkower D. 1999. Similarity of DNA binding and transcriptional regulation by *Caenorhabditis elegans* MAB-3 and *Drosophila melanogaster* DSX suggests conservation of sex determining mechanisms. *Development* **126:** 873–881.

Zhang SQ, Che LH, Li Y, Liang D, Pang H, Ślipiński A, Zhang P. 2018. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nat Commun.* **9:** 1–11.

Zhang HH, Xie YC, Li HJ, Zhuo JC, Zhang CX. 2021. Pleiotropic roles of the orthologue of the *Drosophila melanogaster intersex* gene in the brown planthopper. *Genes* **12:** 379.

Zhu L, Wilken J, Phillips NB, Narendra U, Chan G, Stratton SM, Kent SB, Weiss MA. 2000. Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes Dev.* **14:** 1750–1764.

Zhuo JC, Hu QL, Zhang HH, Zhang MQ, Jo SB, Zhang CX. 2018. Identification and functional analysis of the *doublesex* gene in the sexual development of a hemimetabolous insect, the brown planthopper. *Insect Biochem Mol Biol.* **102:** 31–42.

Zhuo JC, Zhang HH, Hu QL, Zhang JL, Lu JB, Li HJ, Xie YC, Wang WW, Zhang Y, Wang HQ. 2021. A feminizing switch in a hemimetabolous insect. *Sci Adv.* **7:** eabf9237.

Zou X, Schaefke B, Li Y, Jia F, Sun W, Li G, Liang W, Reif T, Heyd F, Gao Q, Tian S, Li Y, Tang Y, Fang L, Hu Y, Chen W. 2021. Mammalian splicing divergence is shaped by drift, buffering in trans, and a scaling law. *Life Sci Alliance* **5:** e202101333.

Appendix

Appendix 1. Primer sequences used in this study.

No.	expeiment	target	primer name	primer sequence (5' to 3')	length	note
1	5' RACE	*doublesex* (*dsx*)	Tdom_*dsx*_RACE_01	TCGCGTGACAAGGAAGAAGGCCCCGG	26	gene specific primer
2	5' RACE	*dsx*	Tdom_*dsx*_RACE_02	AGGCCCCGGAATTGAAGAAGCACCT	25	nested gene specific primer
3	5' RACE	*doublesex-like* (*dsx-like*)	Tdom_*dsx-like*_RACE_01	CACTTTGAAAACGCAGGGCTGGATG	25	gene specific primer
4	5' RACE	*dsx-like*	Tdom_*dsx-like*_RACE_02	GGGCTGGATGTTCGCTGTAGTTGAA	25	nested gene specific primer
5	3' RACE	*dsx*	Tdom_*dsx*_RACE_03	GCTTCTTCAATTCCGGGGCCTTCTTCC	27	gene specific primer
6	3' RACE	*dsx*	Tdom_*dsx*_RACE_04	TCAATTCCGGGGCCTTCTTCCTTGTCA	27	nested gene specific primer
7	3' RACE	*dsx-like*	Tdom_*dsx-like*_RACE_03	AGACAGCAGCCAAATGACGTCAAGA	25	gene specific primer
8	3' RACE	*dsx-like*	Tdom_*dsx-like*_RACE_04	ACAGCAGCCAAATGACGTCAAGAGA	25	nested gene specific primer
9	RT-qPCR	*ribosomal protein 49* (*rp49*)	Tdom_*rp49*_RT-qPCR_F	ACCCACCATAGTCAAGAAGCGGA	23	reference gene
10	RT-qPCR	*ribosomal protein 49* (*rp49*)	Tdom_*rp49*_RT-qPCR_R	AACTGTCCCTTAAACCGCCTTCG	23	reference gene
11	RT-qPCR	*dsx* male-specific region	Tdom_*dsx*_male_RT-qPCR_F	GCTACCGCTTGAAACATTGCCTT	23	
12	RT-qPCR	*dsx* male-specific region	Tdom_*dsx*_male_RT-qPCR_R	AGTGCCATGGATCGTAATTCTGCT	24	
13	RT-qPCR	*dsx* female-specific region	Tdom_*dsx*_female_RT-qPCR_F	CTACCGCTTGAAACATTGCCTTT	23	
14	RT-qPCR	*dsx* female-specific region	Tdom_*dsx*_female_RT-qPCR_R	TGCCCTGATTCATGCATTGA	20	
15	RT-qPCR	*dsx* common region	Tdom_*dsx*_RT-qPCR_F	ACCCAGCCATGCCTCCTAATGTA	23	
16	RT-qPCR	*dsx* common region	Tdom_*dsx*_RT-qPCR_R	CTTCGAGCGTCCTTCAGAACGAC	23	
17	RT-qPCR	*dsx-like*	Tdom_*dsx-like*_RT-qPCR_F	ACGGGTTGTTGCTTTACATCTGT	23	

18	RT-qPCR	*dsx-like*	Tdom_*dsx-like*_RT-qPCR_R	TCTCTTGACGTCATTTGGCTGCT	23	
19	RT-qPCR	*vitellogenin-1*	Tdom_*vitellogenin-1*_RT-qPCR_F	TGCTCCATTCAACAACCAGC	20	
20	RT-qPCR	*vitellogenin-1*	Tdom_*vitellogenin-1*_RT-qPCR_R	AGCCCAGATGAACTTGACGA	20	
21	RT-qPCR	*vitellogenin-2*	Tdom_*vitellogenin-2*_RT-qPCR_F	CCAGTGATGGTGGCAATTCAGGA	23	
22	RT-qPCR	*vitellogenin-2*	Tdom_*vitellogenin-2*_RT-qPCR_R	TGTGGCTGTGACTGTCGTTTTGT	23	
23	RT-qPCR	*vitellogenin-3*	Tdom_*vitellogenin-3*_RT-qPCR_F	CACCAGCGATGTTGACGAGAAGA	23	
24	RT-qPCR	*vitellogenin-3*	Tdom_*vitellogenin-3*_RT-qPCR_R	GCTCAAACTCAGGCTCAAGTGGA	23	
25	dsRNA	*egfp*	*egfp*_dsRNA_F	ATCATGGCCGACAAGCAGAA	20	control of RNAi assay
26	dsRNA	*egfp*	*egfp*_dsRNA_R	AACTCCAGCAGGACCATGTG	20	control of RNAi assay
29	dsRNA	*dsx* common region	Tdom_*dsx*_dsRNA_F	CCAAGCCCAAGACGAAGC	18	
30	dsRNA	*dsx* common region	Tdom_*dsx*_dsRNA_R	CCGACTGTTACATTAGGAGGC	21	
31	dsRNA	*dsx-like*	Tdom_*dsx-like*_dsRNA_F	CTCATGTGCAGTGATGTGGC	20	
32	dsRNA	*dsx-like*	Tdom_*dsx-like*_dsRNA_R	TGCGTCAATGAACAGCGAAA	20	
33	dsRNA	pCR4-TOPO vector	T7-pCR4-TOPO_F	taatacgactcactatagggAGACCACGTCCTGCAGGTTTAAACG	45	T7 flanked
34	dsRNA	pCR4-TOPO vector	T7-pCR4-TOPO_R	taatacgactcactatagggAGACCACCGAATTGAATTTAGCGGC	45	T7 flanked