Evolutionary history of sexual differentiation mechanism in insects

Yasuhiko Chikami[1,2], Miki Okuno[3], Atsushi Toyoda[4,5], Takehiko Itoh[6], Teruyuki Niimi[1,2]*

[1]Division of Evolutionary Developmental Biology, National Institute for Basic Biology, 38 Nishigonaka, Myodaiji, Okazaki, Aichi, 444-8585, Japan

[2]Department of Basic Biology, School of Life Science, The Graduate University for Advanced Studies, SOKENDAI, 38 Nishigonaka, Myodaiji, Okazaki, Aichi, 444-8585, Japan

[3]Division of Microbiology, Department of Infectious Medicine, Kurume University School of Medicine, 67 Asahi-machi, Kurume, Fukuoka, 830-0011, Japan

[4]Comparative Genomics Laboratory, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

[5]Advanced Genomics Center, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

[6]School of Life Science and Technology, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo, 152-8550, Japan

***Corresponding author:** Teruyuki Niimi

E-mail: niimi@nibb.ac.jp

This PDF file includes:

Main Text

Figures 1 to 6

Abstract

Gain of alternative splicing gives rise to functional diversity in proteins and underlies the complexity and diversity of biological aspects. However, it is still not fully understood how alternatively spliced genes develop the functional novelty. To this end, we infer the evolutionary history of the *doublesex* gene, the key transcriptional factor in the sexual differentiation of arthropods. *doublesex* is controlled by sex-specific splicing and promotes both male and female differentiation in some holometabolan insects. In contrast, *doublesex* promotes only male differentiation in some hemimetabolan insects. Here, we investigate ancestral states of *doublesex* using *Thermobia domestica* belonging to Zygentoma, the sister group of winged insects. We find that *doublesex* of *T. domestica* expresses sex-specific isoforms but is only necessary for male differentiation of sexual morphology. This result ensures the hypothesis that *doublesex* was initially only used to promote male differentiation during insect evolution. However, *T. domestica doublesex* has a short female-specific region and upregulates the expression of *vitellogenin* homologs in females, suggesting that *doublesex* may have already controlled some aspects of feminization in the common ancestor of winged insects. Reconstruction of the ancestral sequence and prediction of the protein structure show that the female-specific isoform of *doublesex* has a long C-terminal disordered region in holometabolan insects, but not in non-holometabolan species. We propose that *doublesex* acquired a female-specific isoform and then underwent a change in the protein motif structure, which became essential for female differentiation in sexual dimorphisms.

Keywords: sexual differentiation, alternative splicing, *doublesex*, insect, Zygentoma

47

48 # Introduction

49 Sexual reproduction is widely used for transmitting genetic information from
50 one to the next generation in Metazoa. For reproductive success, animals evolved
51 diverse sex differences, i.e., sexual dimorphism, in morphology (Darwin 1871;
52 Geddes and Thomson 1889) which underlie eco-evolutionary dynamics such as
53 extinction rate and interspecific interaction (Fryxell et al. 2019). In the last three
54 decades, the genetic pathways that create sex and sexual dimorphism have been
55 elucidated in many animal species. Surprisingly, despite having a single origin
56 (Beukeboom and Perrin 2014), these pathways have undergone extensive changes
57 during animal evolution (Wilkins 1995; Bachtrog et al. 2014; Bopp et al. 2014;
58 Herpin and Schartl 2015).

59 The diversity has been attributed to differences in the composition of the
60 regulatory cascades. For example, in eutherians such as mice and humans, the master
61 regulator of sex is *Sex-determining region Y* (*Sry*), a member of the High Mobility
62 Group (HMG)-box transcriptional factor family (Gubbay et al. 1990; Sinclair et al.
63 1990; Koopman et al. 1991; Miyawaki et al. 2020), while *DM domain gene on the Y*
64 *chromosome* (*dmy*) of the *doublesex* and *mab-3* related transcriptional factor (DMRT)
65 family is the master sex-determining regulator in the medaka fish (Matsuda et al.
66 2002; Nanda et al. 2002). Diversification of the pathway governing sex
67 determination/differentiation are largely based on differences in their gene repertoires
68 (e.g., Hasselmann et al. 2008; Hattori et al. 2010; Sato et al. 2010; Takehana et al.
69 2014). In contrast, it has recently been discovered that the mechanisms of sexual

differentiation in Pterygota, i.e., winged insects, differ in outputs of the gene cascade, e.g., the promotion masculinization or feminization.

Sexually dimorphic morphology in Pterygota is formed during postembryonic development. *doublesex* (*dsx*), a member of the DMRT family, acts as a global regulator at the bottom of the cascade to govern over sex differentiation (Kopp, 2012; Verhulst and van de Zande, 2015). In many pterygote insects studied, *dsx* is controlled by sex-specific splicing. In Diptera, Coleoptera, and Lepidoptera, sex-specific Dsx protein variants are essential for promoting either male or female differentiation in sexual dimorphism (e.g., Hildreth 1965; Burtis and Baker 1989; Ohbayashi et al. 2001; Kijimoto et al. 2012; Ito et al. 2013; Shukla and Palli 2012; 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). However, in the sawfly *Athalia rosae* (Mine et al. 2017, 2021), the silverleaf whitefly *Bemisia tabaci* (Guo et al. 2018), the brown planthopper *Nilaparvata lugens* (Zhuo et al. 2018), the German cockroach *Blattella germanica* (Wexler et al. 2019), and the damselfly *Ischnura senegalensis* (Takahashi et al. 2019, 2021), *dsx* has sex-specific isoforms and is responsible for male differentiation of morphological traits during postembryonic development, but not needed for female differentiation. Thus, despite expressing sex-specific isoforms, *dsx*'s role in sexual differentiation in Pterygota is different, as it controls both male and female differentiation or only male differentiation for sexual morphogenesis.

dsx in crustaceans and arachnids is reported to be highly expressed in males without sex-specific isoforms (Kato et al. 2011; Pomerantz et al. 2015; Li et al. 2018; Panara et al. 2019). Accordingly, *dsx* is only required for male differentiation of morphological traits in the water flea *Daphnia magna*. Wexler et al. (2019) proposed a stepwise evolution in which *dsx* had acquired sex-specific isoforms and later had become essential for female differentiation. However, roles of *dsx* are more diverse than expected. In Hymenoptera, *dsx* is involved in female differentiation of reproductive organs in the honeybee *Apis mellifera* (Roth et al. 2019), while *dsx* is non-essential for female differentiation in the sawfly *At. rosae* (Mine et al. 2017, 2021). In the milkweed bug *Oncopeltus fasciatus*, *dsx* is involved in both female and male differentiation of the genital organs (Just et al. 2021). *dsx* in *Be. tabaci* positively regulates the expression of a yolk precursor gene *vitellogenin* in females while it is not essential for female morphology (Guo et al. 2018), implying that *dsx* has different functionality for morphogenesis and otherwise in females. Overall, estimating the evolutionary history of *dsx* in Pterygota is still a challenging task. Also, it is unclear what factors led to the feminizing roles of *dsx* (Hopkins and Kopp 2021).

The phylogenetic distance between crustaceans and Pterygota and the lack of information about outgroups more closely related to Pterygota may be the reason for the gap of understanding of how *dsx* evolved from a monofunctional to a bifunctional regulator in arthropods. In an attempt to close this gap, we decided to include the firebrat *Thermobia domestica* (Zygentoma) in our analysis of *dsx*. Zygentoma is the sister group of Pterygota (Misof et al. 2014), does not copulate, and displays simple sexual dimorphisms, i.e., non-aedeagus male penises and female ovipositors (Kristensen 1975; Matsuda 1976; Emeljanov 2014; Beutel et al. 2017; Boudinot

2018). Otherwise, there is little difference in morphology between females and males, as Darwin (1871: 348) noted, "The sexes do not differ." These features suggest that the level of sex differentiation in this species is very simple and likely to be ancestral. Thus, Zygentoma presents an ideal model for investigating the ancestral state of *dsx* in Pterygota. In this study, we investigated *dsx* in *T. domestica* and analyzed its functions in sexual differentiation. Also, we carried out the phylogenetic analysis, ancestral sequence reconstruction, and protein structure prediction to infer the evolutionary history of *dsx*.

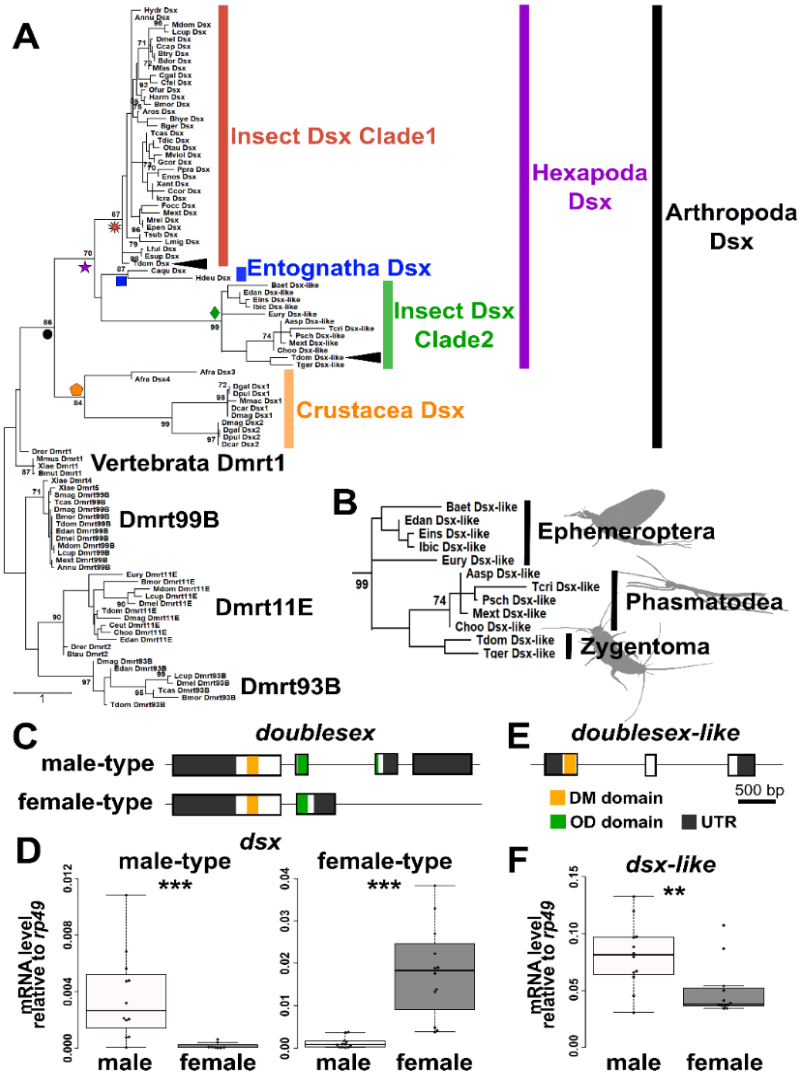
Results and Discussion

Molecular evolution of *dsx* homologs and gene duplication of *dsx* in insects

Five *doublesex* (*dsx*) homologs were found in the transcriptome database of the firebrat *Thermobia domestica*. To identify which of them corresponds to the *dsx* ortholog in *T. domestica*, we compared these to *dsx* homologs found in transcriptome/genome/protein databases of various arthropods and vertebrates (supplementary table 1) and performed molecular phylogenetic analyses based on the amino acid sequences of their DNA-binding domains. As a result, the pancrustacean *dsx* was grouped into a distinct clade from the other DMRT family genes (fig. 1A). Within this clade, four subclades were recognized: Insect Dsx Clade 1, Insect Dsx Clade 2, Entognatha Dsx Clade, and Crustacea Dsx Clade. Insect Dsx Clade 1 contained the *dsx* found in Pterygota including *Drosophila melanogaster*. This clade also contained one of the 5 *dsx* homologs of *T. domestica*. We consider it likely that this *dsx* homolog is the corresponding ortholog of *T. domestica*. This Entognatha Dsx

Clade also contained *dsx* from a springtail (Collembola) and a dipluran insect (Diplura). The Crustacea Dsx Clade contained *dsx* from branchiopods, including daphnids.

We found that several species of Zygentoma, Ephemeroptera (mayflies), and Phasmatodea (stick insects) contain a *dsx-like* homolog of the Insect Dsx Clade 2 (fig. 1A, B) as well as the dsx ortholog of the Insect Dsx Clade 1. These findings indicate that *dsx* was duplicated before the divergence of Zygentoma and that the two paralogs retained from the divergence of the pterygote insects until at least the divergence of Eumetabola (= Hemiptera + Thysanoptera + Psocodea + Holometabola). The molecular evolution of *dsx* has been inferred from *dsx* of some pterygote insects, mainly holometabolan insects (Wexler et al. 2014; Mawaribuchi et al. 2019), while the presence of a *dsx-like* gene may have been overlooked in their analyses. Here, we report that the genome of *T. domestica* also contains both *dsx* and *dsx-like*, reflecting the presumed ancestral state in Pterygota in terms of gene copy number of *dsx*. Gene duplication generally leads to neo-/sub-functionalization to allow functional diversification (c.f., Taylor and Raes 2004). In this study, we analyzed the expression profiles and functions of *dsx* as well as *dsx-like* in *T. domestica*.



158 **FIG. 1.** Molecular phylogeny and structural features of Doublesex in Arthropoda and
159 Vertebrata. (*A*) Molecular phylogeny of Doublesex and Mab-3 related transcriptional
160 factors (DMRT). The phylogenetic analysis was based on amino acid sequences of the
161 DNA binding domain (DM domain) of DMRT family and was performed by the
162 MEGA X after the multiple sequence alignment using the MAFFT software. The
163 maximum-likelihood method was applied. 97 operational taxonomic units (OTUs)
164 used for the phylogenetic analysis are listed in supplementary table 1. (B) Enlarged
165 view of insect Dsx Clade2 (*dsx-like* clade). The numerical value on each node is the
166 bootstrap supporting value. Bootstrap values < 70 are not shown. The node of each
167 clade is indicated by colored shapes: black circle, Arthropoda Dsx; orange pentagon,
168 Crustacea Dsx; purple star, Hexapoda Dsx; red sunburst, Insect Dsx Clade1; green
169 diamond, Insect Dsx Clade2; blue square, Entognatha Dsx. (C) Exon-intron structures
170 of *dsx* in *Thermobia domestica*. The upper and lower schematic images show the gene
171 structure of *dsx* male-type and female-type, respectively. (D) Expression level of *dsx*
172 in males and females of *T. domestica*. (E) Exon-intron structures of *dsx-like* of *T.*
173 *domestica*. (F) Expression level of *dsx-like* in males and females. The exon-intron
174 structure is determined by mapping the mRNA sequence of each gene to the genome

of *T. domestica*. The expression level (D and F) was measured by the RT-qPCR of *dsx* and *dsx-like* in the adult fat body and is indicated as the relative values to the expression of the reference gene, *ribosomal protein 49* (*rp49*). Each plot indicates 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 supplementary table 2.

Sex-specific splicing of *dsx* in *Thermobia domestica*

Splicing of *dsx* produces gives rise to sex-specific isoforms in all pterygote insects studied thus far, with the exception of the termite *Reticulitermes speratus* (Miyazaki et al. 2021), the silverleaf whitefly *Bemisia tabaci* (Guo et al. 2018), and the body louse *Pediculus humanus* (Wexler et al. 2019), suggesting that sex-specific splicing regulation of *dsx* was acquired before the divergence of Pterygota. To examine this hypothesis, we investigated the expression profile of *dsx* and *dsx-like* of *T. domestica*. Full-length mRNA sequences of *dsx-like* and *dsx* in *T. domestica* were determined by the RNA-seq and rapid amplification of cDNA ends (RACE) methods. Then, we investigated the gene structures to map the mRNA sequences to our genome database. *dsx* consists of five exons with two isoforms (fig. 1C): a long one (951 bp) and a short one (756 bp). RT-qPCR analysis showed that the long isoform and the short isoform were highly expressed in males and females, respectively (fig. 1D; Brunner-Munzel test, $P = 1.75\times10^{-6}$ and 2.20×10^{-16} in the long and the short isoforms). This fact indicates that *dsx* is controlled by sex-specific splicing. We refer to the male-biased isoform as *dsx* male-type and the female-biased isoform as *dsx* female-type. The *dsx-like* is expressed about two-fold higher in males than in females (Brunner-Munzel test, $P=0.00924$) and has three exons but no sex-specific isoform (fig. 1E, F), showing that *dsx-like* is not regulated by sex-specific splicing.

Our results give the further support that sex-specific splicing of *dsx* already existed in the common ancestor of Pterygota and Zygentoma (= Dicondylia), which

diverged ~421 million years ago (Ma). Misof et al. (2014) estimated that the common ancestor of *Daphnia* and hexapods occurred at ~508 Ma. Therefore, *dsx* sex-specific splicing regulation is an ancient feature of insects that was acquired between 508 and 421 Ma and has been conserved for ~400 million years in each taxon of Dicondylia.

Function of *dsx* for internal reproductive system and body size in *T. domestica*

Deciphering the role of *dsx* of *T. domestica* is essential for inferring the ancestral roles of *dsx* in pterygote insects. Hence, we conducted a functional analysis of not only *dsx* but also *dsx-like* since this paralog might also play a role in sexual differentiation.

To this end, we silenced *dsx* and *dsx-like* by RNA interference (RNAi). We quantified the expression of *dsx* and *dsx-like* in fat bodies of RNAi individuals by reverse transcriptional quantitative polymerase chain reaction (RT-qPCR). *dsx* silencing in females and *dsx-like* silencing in both sexes showed significantly decreased expression of each target genes compared to their expression in *enhanced green fluorescent protein* (*egfp*) RNAi controls (Brunner-Munzel test, $P = 0.0265$ in female *dsx*, 4.40×10^{-16} in male and female *dsx-like*; supplementary fig. 1A; supplementary table 2). The *dsx* RNAi males did not show a significant effect on *dsx* expression. Since it was suspected that outliers affected this result, we tested for outliers in *dsx* RNAi males and found one outlier (supplementary table 3). The reanalysis removing the outlier showed that *dsx* expression was significantly decreased in *dsx* RNAi males (Brunner-Munzel test, $P = 0.00545$; supplementary fig. 1B; supplementary table 2). Therefore, we concluded that *dsx* and *dsx-like* dsRNAs can knock down each target gene. Also, *dsx* RNAi had no effect on *dsx-like* expression and vice versa.

We also performed a double knockdown of *dsx* and *dsx-like*, to address the possibility that *dsx* and *dsx-like* are redundant. We specifically examined the effects of silencing on sexual dimorphism such as body size (fig. 2A) and reproductive systems. The body size, as measured by the pronotum width, was not affected in either knockdown group (fig. 2B; supplementary table 4, 5). In the gonads, the *dsx* RNAi, *dsx-like* RNAi, and double knockdown did not show any histological differences in testes, ovaries, and gametogenesis from the controls (fig. 2C; Supplementary Material online). On the other hand, 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 (fig. 2D). The number of sperm in the seminal vesicles of *dsx* RNAi males was lower than in the control group (fig. 2E; supplementary table 5; generalized linear model, $P = 0.00487$). Silencing of dsx or dsx-like or both did not affect normal differentiation of the female reproductive systems including the spermatheca (fig. 2F; supplementary fig. 2; Supplementary Material online). Also, there was no effect on the number of oocytes in any treatments (fig. 2G; supplementary table 4).

The lack of effect of RNAi on the gonads may be due to the timing of the 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, although, this experiment will be left to future studies, as our study focuses on the function of *dsx* 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). The effect on internal reproductive systems other

than the gonads are consistent with results in *Athalia rosae* (Mine et al. 2017, 2021) and *Blattella germanica* (Wexler et al. 2019).

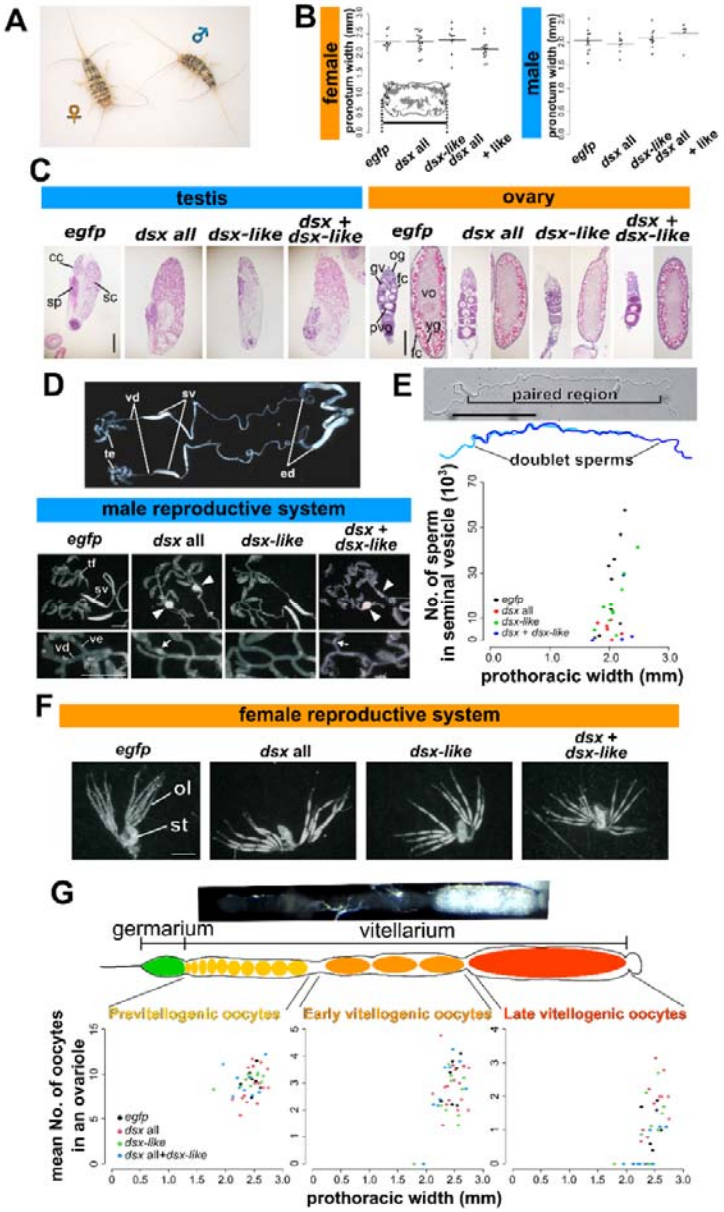


FIG. 2. Function of *doublesex* and *doublesex-like* for body size, internal reproductive system, and gametogenesis of *Thermobia domestica*. (*A*) A pair of *T. domestica*. The female looks much the same as the male. (*B*) Body size of RNAi treatment groups. The pronotum (prothoracic tergum) width was used for the index of the body size. The graph shows mean ± SE (standard error). The results of the generalized linear

model (GLM) analysis show in supplementary table 4 (female) and 5(male). Any significant effect can be detected in the RNAi treatments. Total $N = 49$ in females and 36 in males. (*C*) Histology of gonads in the RNAi groups. Paraffin. Hematoxylin-Eosin staining. In images of the ovary, the left and right panel in each treatment show germarium/previtellogenesis and vitellogenesis, respectively. (*D*) Effects of RNAi on male internal reproductive system. The upper photo shows the gross morphology of the reproductive systems in the non-treated male. The lower photos demonstrate the morphology of the RNAi males. The arrowheads show the rounded seminal vesicle. The lowest photos focused on the vas efferens. The arrows show the clogged sperm in the vas efferens. (*E*) Sperm of RNAi males. The upper photo and figure are sperm morphology in the non-treated male. The sperm forms doublet in the seminal vesicle. The lower figure shows the sperm number of the RNAi males. The results of the GLM analysis show in supplementary table 5. The significant effect was detected in the *dsx* RNAi treatment ($P = 0.00487$). Total $N = 29$. (*F*) Effects of RNAi on female internal reproductive system. (*G*) 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 the RNAi females along with the oogenetic stages. The results of the GLM analysis show in supplementary table 4. The number of the late vitellogenic oocytes was correlated with the pronotum width, although any significant effect can be detected in the RNAi treatments. Total $N = 42$ in each stage. 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. Each plot in (*B*), (*E*), and (*G*) indicates the value of each individual. cc, cystocyte; fc, follicle cell; gv, germinal vesicle; og, oogonia; ol, ovariole; pvo, previtellogenic oocyte; sc, spermatocyte; sp, sperm; st, spermatheca; sv, seminal vesicle; tf, testicular follicle; yg, yolk granule; ve, vas efferens; vd, vas deferens, vo, vitellogenic oocyte. Scales: 50 μm (*C*); 10 μm (*E*); 1000 μm (*D* and *F*).

Function of *dsx* for morphology in genital organs of *T. domestica* and evolution of the function of *dsx* for sexual morphogenesis in insects

The sexually dimorphic morphology can be seen in the external genital organs, i.e., male penis and female ovipositor (fig. 3A). Males of *T. domestica* have unpaired small external genitalia on the abdominal segment IX. Females have an ovipositor consisted of two paired appendage-like structure on the abdominal segment VIII and IX. Males of the *dsx* knockdown groups (*dsx* only and both *dsx* and *dsx-like* RNAi) was transformed into two pairs of appendage-like structures resembling the female ovipositor (fig. 3B, C; Supplementary Material online). This effect was not observed

in *dsx-like* RNAi males. Our results indicate that *dsx* is essential for male differentiation of morphological traits in *T. domestica*. In contrast to males, our analysis showed that none of the RNAi treatments affected female ovipositors at external morphological, tissue, or cellular levels (fig. 3D, E; supplementary fig. 3; Supplementary Material online). We then measured the length of the female ovipositor in the RNAi-treated groups to examine the involvement of *dsx* and *dsx-like* in the growth of female morphology. The results showed that *dsx* and *dsx-like* RNAi had no significant effect on ovipositor length (fig. 3F, G; supplementary table 4).

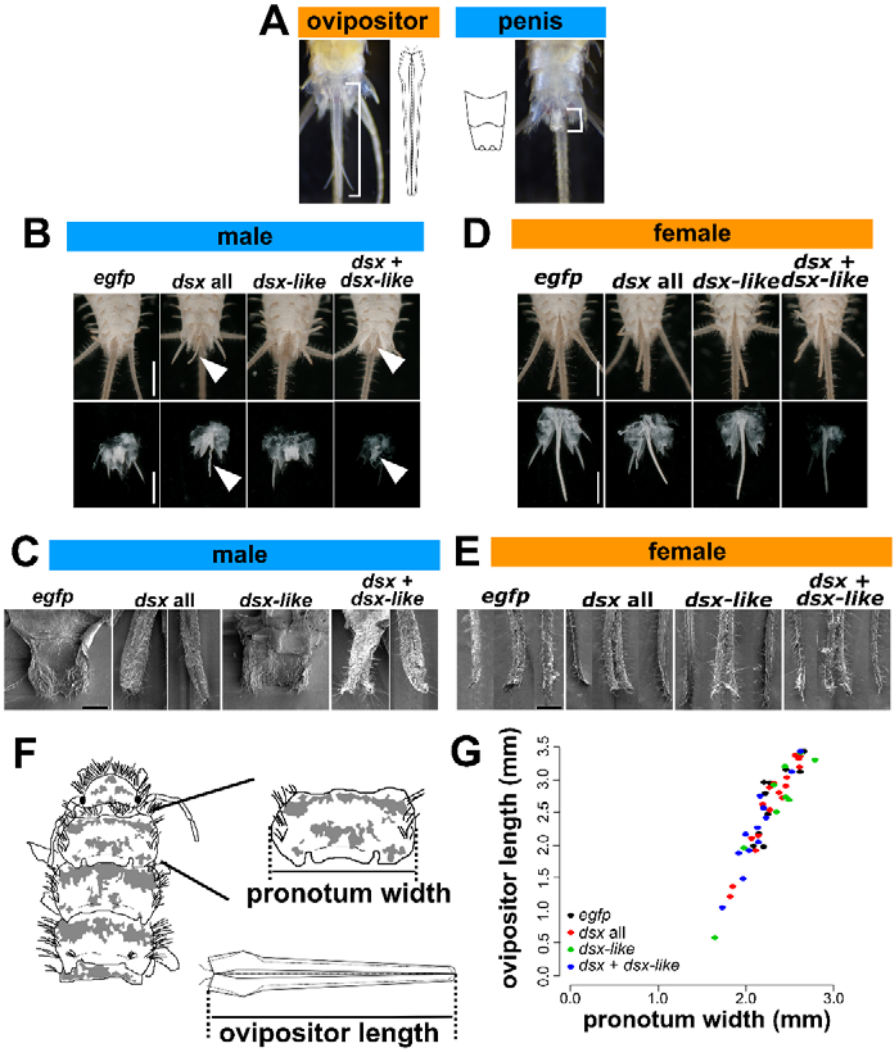
The lack of effect of *dsx* RNAi in females is due to that *dsx* is not essential for female differentiation of morphology during postembryonic development or that *dsx* knockdown is inefficient in the females. Compared to the knockdown efficiency of *dsx* in the males (~30% at median), that in the females is ~50% (supplementary table 2). However, given that half of the female individuals in the RT-qPCR analysis in the fat body had lower *dsx* expression than the minimum value of the control ones (supplementary fig. 1A), it can be assumed that *dsx* expression is suppressed in a certain number of females used in each analysis. In addition, *dsx* RNAi showed no effect on morphology in all 58 females (80 females including *dsx* and *dsx-like* double RNAi) analyzed in this study. Thus, it is reasonable to conclude that *dsx* is not essential for female differentiation of morphology during postembryonic development in *T. domestica*. Also, our results indicate that *dsx-like* is not essential for the sexual differentiation of morphology during postembryonic development in *T. domestica*. The knockdown of both *dsx* and *dsx-like* showed only the same effect as *dsx* RNAi alone. Thus, it is unlikely that *dsx-like* functions redundantly with *dsx*.

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 of non-holometabolan insects such as *T. domestica* (Zygentoma: this study), *Bl. germanica* (Dictyoptera: Wexler et al. 2019) and the brown planthopper *Nilaparvata lugens* (Hemiptera: Zhuo et al. 2018). There could be the possibility of tissue-specific effects of *dsx* found in some holometabolan females (e.g., Ledón-Rettig et al. 2017). However, this possibility would be unlikely at least in these species since *dsx* was knocked down by systematic RNAi and was not reported to affect female morphology at this time. Based on these facts, we estimate that *dsx* may not be essential for female differentiation of morphology at the common ancestor of Dicondylia, ensuring the hypothesis proposed by Wexler et al. (2019).

To elucidate the timing of the acquisition of the role of *dsx* in female morphogenesis during postembryonic development, we must interpret the role of *dsx* in Hymenoptera, the basal clade of Holometabola. Studies in 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). Thus, the male-like reproductive organ in *dsx* mutant females in Roth et al. (2019) would show an effect during embryogenesis, not during postembryonic development. We cannot conclude whether *dsx* is not essential for female morphogenesis in the honeybee, since the information on the roles of *dsx* in sexual morphology is limited to gonads and heads of the worker females. However, given that *dsx* does not affect heads in *Ap. mellifera* females (Roth et al. 2019), wings in the parasitoid wasp *Nasonia vitripennis* females (Wang et al. 2020), and sexual traits in *At. rosae* females (Mine et al. 2017, 2021), at this time, it is reasonable to infer that *dsx* was not essential for female morphogenesis during postembryonic development in the common ancestor of Hymenoptera. This interpretation and the

349 essential roles of *dsx* for female development in the other holometabolan insects
350 suggest that *dsx* became essential for feminization of morphology during
351 postembryonic development at the common ancestor of holometabolan insects except
352 for Hymenoptera (=Aparaglossata) emerging ~327 Ma.



353

354 **FIG. 3.** Function of *doublesex* and *doublesex-like* for genital organs in *Thermobia*
355 *domestica*. (*A*) Sexually dimorphic traits of *T. domestica*. Females possess an
356 ovipositor and males have a penis. (*B*) Effects of RNAi treatments on male penial
357 structure. The upper images show the ventral side of the male abdomen. The lower
358 images focus on the male penis. The arrowheads indicate the ovipositor-like structure
359 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 the one image. In these images, the left panels show the ovipositor valvula II (inner sheath) -like structure. The right panels exhibit the ovipositor valvula I (outer sheath)-like structure. The detail description can be referred in Supplementary Material online. (*D*) Effects of RNAi treatments on female ovipositor. The upper images show the ventral side of the female abdomen. The 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 one exhibits the valvula I. The results of the histological observation are in supplementary fig. 3. The detail description can be referred in Supplementary material online. (*F*) The schematic images of the measured parts. (*G*) Effects of RNAi treatments on growth of ovipositor. Each plot indicates the ovipositor length of each individual. The results of the generalized linear model analysis show in supplementary table 4. The ovipositor length was correlated with the prothoracic width ($P = 2.00\times10^{-16}$), although any significant effects can be seen in the 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. Scales: 1 cm (*B* and *D*); 50 µm (*C* and *E*).

Cryptic role of *doublesex* for female-specific transcripts in *T. domestica* and its opposite role between sexes

dsx in *T. domestica* does not seem to have conflicting functions between sexes in postembryonic morphogenesis. On the other hand, other biological processes remain to be considered. We tested whether *dsx* contributes to the 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 have shown that *vtg* in pterygote insects is controlled by *dsx* (e.g., Suzuki et al. 2003; Shukla and Palli 2012; Thongsaiklaing et al. 2018). Our RNA-seq analysis showed that three *vtg* homologs, i.e., *vtg1*, *vtg2*, and *vtg3*, were expressed female-specifically in the fat body in *T. domestica* (supplementary fig. 4; supplementary table 6). We analyzed the expression of *vtg* in the fat bodies of *dsx*, *dsx-like*, or both genes RNAi groups by RT-qPCR.

In *dsx* RNAi males, all *vtg* mRNAs were expressed 45–1530-fold higher than the controls (fig. 4A; supplementary table 2: Brunner-Munzel test, $P = 2.87\times10^{-8}$,

395 6.60×10^{-16} and 2.80×10^{-4} in *vtg1*, *vtg2*, and *vtg3*). *vtg1* and *vtg3* mRNAs were
396 significantly up-regulated in *dsx-like* RNAi males compared to the controls (fig. 4A:
397 Brunner-Munzel test, $P = 0.0139$ and 0.00497 in *vtg1* and *vtg3*). In both *dsx* and *dsx-*
398 *like* RNAi males, the effect was similar to that in *dsx* RNAi males (fig. 4A: Brunner-
399 Munzel test, $P = 6.60\times10^{-16}$, 0.0162, and 6.60×10^{-16} in *vtg1*, *vtg2*, and *vtg3*). Then, we
400 found that the expression of all *vtg* genes was significantly reduced in *dsx* RNAi
401 females (fig. 4B; supplementary table 2: Brunner-Munzel test, $P = 0.0433$, 0.00422,
402 and 0.00623 in *vtg1*, *vtg2*, and *vtg3*). This reduction rate was approximately 0.2–0.4-
403 fold. Furthermore, *vtg* expression was significantly reduced in *dsx-like* RNAi females
404 (Brunner-Munzel test, $P = 0.00256$, 3.80×10^{-6}, and 1.49×10^{-5} in *vtg1*, *vtg2*, and *vtg3*)
405 and both *dsx* and *dsx-like* RNAi females (Brunner-Munzel test, $P = 0.0305$, 0.00892,
406 and 0.0197 in *vtg1*, *vtg2*, and *vtg3*) (fig. 4B). These results show that *dsx* and *dsx-like*
407 of *T. domestica* control *vtg* negatively in males and positively in females.

408 Our results indicate that *dsx* has opposite roles between sexes, i.e., repressive
409 in males and promotive in females, in *vtg* expression. *dsx-like* also has the opposite
410 functions for *vtg* expression in males and females. It is unlikely that this result is due
411 to *dsx-like* regulating *dsx* transcription, as *dsx-like* did not affect *dsx* expression
412 (supplementary fig. 1A). A possible hypothesis is that *dsx-like* might regulate *vtg*
413 expression as one of the co-regulators that bind *dsx* or other transcription factors.

414 We do not know whether *dsx* of *T. domestica* oppositely controls genes other
415 than the *vtg* homologs between sexes since our analysis was limited to *vtg* homologs.
416 However, the results from these genes indicates that the molecular function of *dsx* in
417 this species includes the opposite function for some genes' transcription in females
418 and males. In *Be. tabaci*, *dsx* positively regulates *vtg* expression in females, even
419 though it is not 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 the function that are not essential for female morphogenesis, is a functionality that has not been reported in any insect or animal. This functionality indicates that even if *dsx* can oppositely function for some genes' expression between sexes, it does not necessarily have opposite functions in morphogenesis between sexes. This difference in the functionality might be due to differences in genes under *dsx* control between morphogenesis and other aspects such as the yolk synthesis in females.

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). It was thought that 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 ability of *dsx* to be involved in female differentiation was already present in the common ancestor of Dicondylia and later became essential for the female morphogenesis in the common ancestor of Aparaglossata. In this evolutionary scenario, the role of *dsx* in the feminization of postembryonic morphogenesis in Aparaglossata could be due to extending its capability to control some genes in females, i.e., functional expansion. We cannot decide which of these hypotheses is appropriate, at this time. However, the latter scenario can well explain the presence of female-specific coding sequences of *dsx* and high expression of *dsx* female-type

during postembryonic development, in non-aparaglossatan insects. The capability to regulate some female genes might be a "minor function" of *dsx* in non-holometabolan females predicted by Wexler et al. (2019).

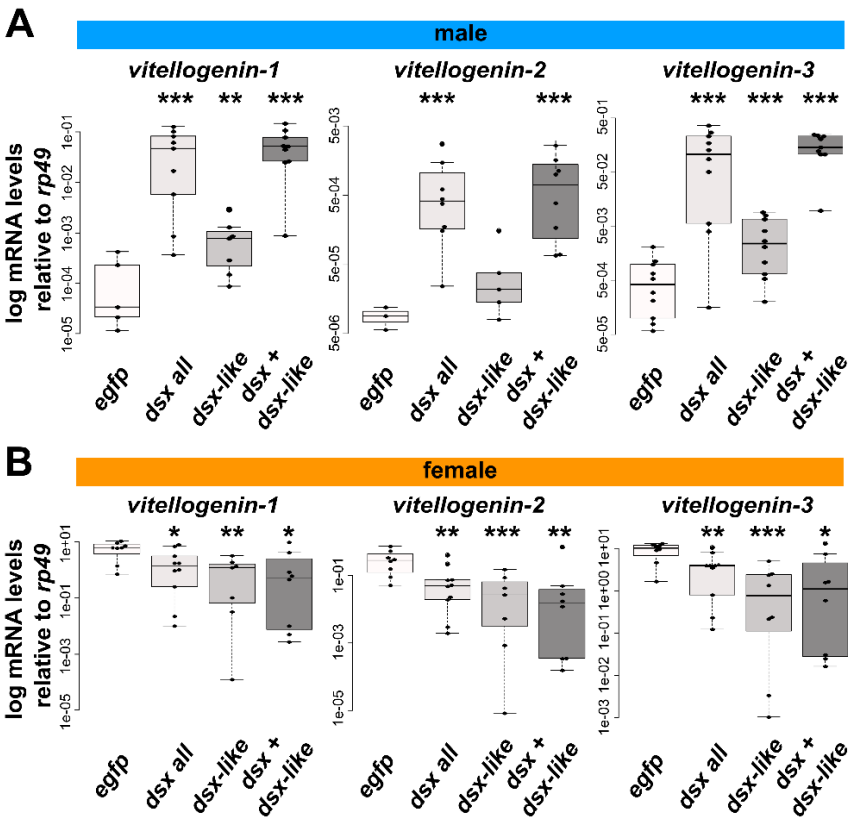


FIG. 4. Function of *doublesex* for *vitellogenin* expression in *Thermobia domestica*. (*A*) *vitellogenin* expression level in RNAi males. (*B*) *vitellogenin* expression level in RNAi females. The mRNA expression levels were measured by the RT-qPCR analysis. The figures show the log-scale relative values of the expression level of three *vitellogenin* homologs to the reference gene, *ribosomal protein 49* (*rp49*). Each plot indicates the mRNA expression level 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* dsRNAs injected group, respectively. The Brunner-Munzel test method were performed to statistically analyze the difference in mRNA expression level between the control and the *dsx* or *dsx-like* RNAi groups. The *P*-values were adjusted by the Holm's method. **P*<0.05, ***P*<0.001, ****P*<0.0001. *P* ≥ 0.05 is not shown. The statistical results were described in supplementary table 2.

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.

Evolution of C-terminus disordered region of *dsx* female-type

One of the puzzling problems is how *dsx* became recruited for female differentiation of morphological traits (cf., Hopkins and Kopp 2021). Here, we 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 that in *D. melanogaster* (53 aa) (supplementary fig. 5). The OD domain is essential for female differentiation in *D. melanogaster*, as it physically binds to *dsx* itself, transcription factors, and co-activators (An and Wensink 1995; Erdman 1996; Ghosh et al. 2019; Romero-Pozuelo et al. 2019). Therefore, we 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, we obtained sequences of *dsx* female-type from 48 insect species based on the National Center for Biotechnology Information (NCBI) protein/transcriptome shotgun assembly database and previous studies (supplementary table 7) and reconstructed ancestral sequences of *dsx* female-type. Our 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 the common ancestor of the other taxon (fig. 5A; supplementary fig. 6; supplementary table 8). This motif is conserved within Aparaglossata in our dataset although moderate sequence diversification was observed (supplementary fig. 6). In our dataset, almost all sequences of this motif were not found in species in which *dsx* is not essential for the female differentiation of morphological traits during postembryonic development. Exceptionally, *dsx* of *At. rosae* had an amino acid sequence in the region corresponding to this motif, but our

results of ancestral sequence reconstruction showed that this sequence was acquired in parallel with Aparaglossata.

The Aparaglossata-specific region is located in the distal (C-terminal) part of the female-specific region in *D. melanogaster*. This distal region is a disordered region, i.e., a mobile region that lacks a fixed structure, following an α-helix loop in the proximal (N-terminal side) region (Yang et al. 2008). To investigate whether the acquisition of the disordered region occurred in Aparaglossata, we predicted the protein structure of *dsx* female-type ancestral sequences of Pterygota, Neoptera, Eumetabola, Holometabola, and Aparaglossata. According to the Alphafold2 algorism-based structure prediction, the female-specific region of *dsx* in the common ancestor of Aparaglossata had a proximal α-helix loop structure, and a distal random coil indicating a disordered region (fig. 5B). 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 was also predicted in the common ancestors of taxon other than Aparaglossata. The random coil following the α-helix structure was predicted in all common ancestors, but its length was shorter than that of the common ancestor of Aparaglossata. It is essential to determine the structure via nuclear magnetic resonance or cryo-electron microscopy methods to conclude the details at the structural level, although, our theoretical predictions suggest that the disordered region following the α-helix structure in the female-specific region may have been extended in the common ancestor of Aparaglossata.

Our results suggest that both the extension of the disordered region following the α-helix loop in the female-specific region of *dsx* and the feminizing function of *dsx* for morphology occurred in the common ancestor of Aparaglossata. At present, the causality between these two events is uncertain, as we do not know which of the

events appeared earlier. In general, disordered regions in transcription factors play essential roles in transcriptional activity through post-translational modifications and binding to co-activators and nucleic acids (Liu et al. 2006; Darling and Uversly 2018). Furthermore, 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 the extension of the C-terminal region of *dsx* female-type may have been a key event associated with the acquisition of the female-differentiating roles of *dsx* in morphology during postembryonic development. This functional evolution of the "non-functional" isoform by the coding mutation is also consistent with an evolutionary process of alternative splicing isoforms theoretically predicted lacking empirical evidence (Keren et al. 2010).

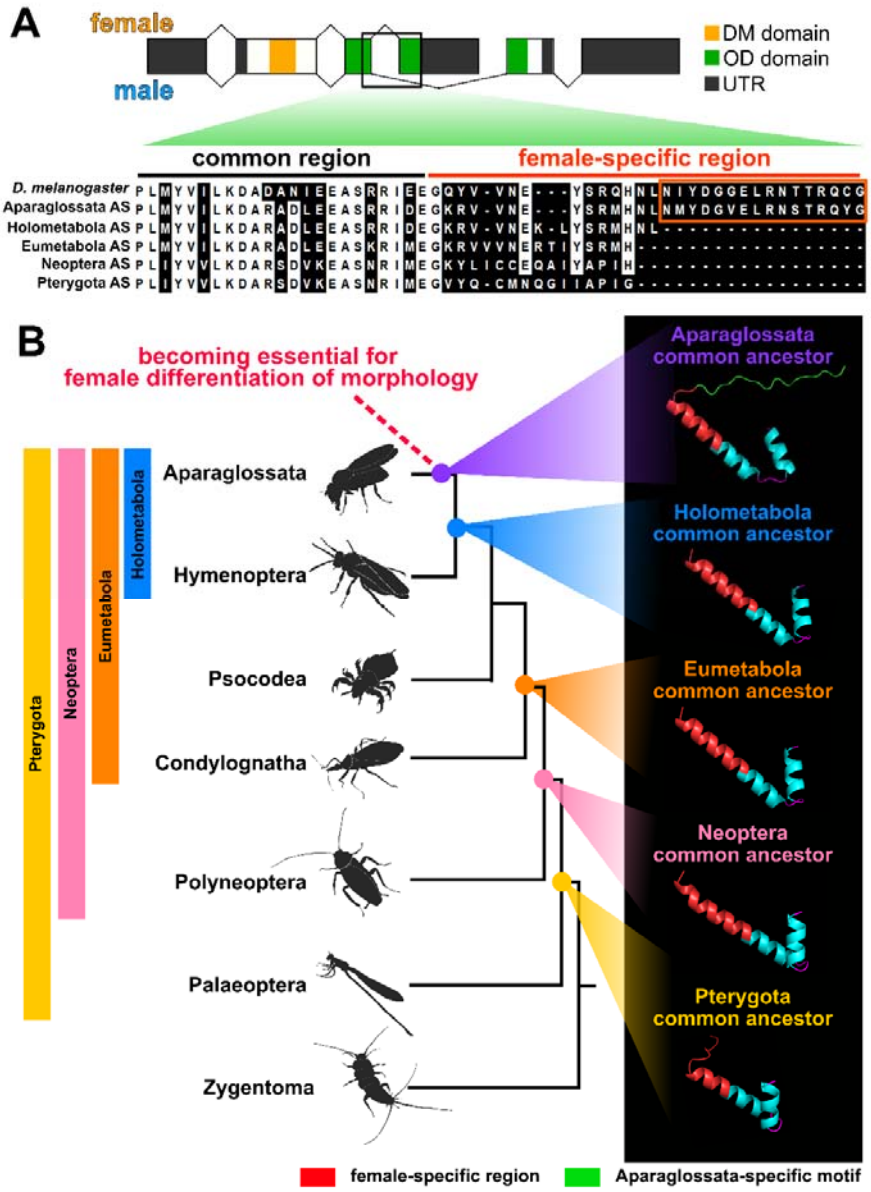


FIG. 5. Evolution of C-terminal sequence of *doublesex* in insects. (*A*) Ancestral sequences (AS) of *dsx* in insects. The AS were reconstructed from 49 *dsx* proteins of insects by the maximum likelihood methods of the MEGA X. The information on the species and proteins used for the AS reconstruction is listed in supplementary table 7. The most probable sequences were applied. The results of the AS reconstruction are described in supplementary table 8. 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 the residues in the 80% taxa. The Aparaglossata-specific motif is indicated by the orange frame. (*B*) Predicted protein structures of *dsx* female-type in common ancestors of insect taxa. The phylogenetic

539 relationship is based on the topology of Misof et al. (2014). The 3D images in the
540 right panel indicate the predicted structures of the OD domain including the female-
541 specific region of *dsx*. The protein structures were predicted by the AlphaFold2-based
542 algorism (ColabFold: Mirdita et al. 2021). The red region of the 3D image indicates
543 the female-specific region. The green region shows the Aparaglossata-specific motif.
544 The information on the evaluated values (predicted local distance difference test:
545 plDDT) of the prediction is shown in the Material and Methods section and
546 supplementary fig. 10.

547

548 **On the origin of outputs of the sexual differentiation mechanism**

549 Recent findings in insects (e.g., Mine et al. 2017; Guo et al. 2018; Zhuo et al.
550 2018; Wexler et al. 2019; Takahashi et al. 2021), including this study, have shown
551 that sexual differentiation mechanisms are diverse in their outputs as well as their
552 gene repertoires. The diversity in the output is attributed to the functional diversity of
553 a single gene, *dsx*, for sexual differentiation of morphogenesis during postembryonic
554 development. The evolutionary origin of the diversity in the output is one of the
555 enigmatic problems in sexual development. Information on the roles of *dsx* is limited
556 to some traits in some species and cannot be available in many non-aparaglossatan
557 species although functional analyses of *dsx* have been rapidly progressing using
558 emerging model species. Unquestionably, comprehensive information on functions of
559 *dsx* for sexually dimorphic morphology from wider taxa is essential for fully tracing
560 the evolution of *dsx*. We propose, albeit premature, as one of the possibilities to be
561 considered, the hypothesis by which *dsx* might have become essential for female
562 differentiation in sexual morphology by expanding its cryptic feminizing role, i.e.,
563 functions for some female genes' expression, in association with mutations in female-
564 specific motifs (fig. 6). This scenario can explain how single genes acquire novel
565 outputs of sexual development although our hypothesis does not prevent any other
566 alternative hypothesis from being proposed.

The diversity of mechanisms that produce animal sex is a model case of developmental systemic drift (True and Haag 2001; Haag and True 2021). The functional diversity of a single gene and its evolutionary process has not been focused on in the context of developmental system drift to date due to its poor examples. Our evolutionary scenario may be one hypothesis explaining the origin of the system drift in the function of single genes. In this study, we have mainly discussed the functionality of *dsx* for sexual differentiation of morphology during postembryonic development. Therefore, it is unclear whether our conclusions and evolutionary scenarios apply to sexual behavior, including sex pheromone secretion and courtship, as well as sexual determination and gonadal differentiation during embryogenesis. Detailed studies of sex differences at various levels across insect taxa will test our evolutionary scenario and will fully reconstruct the evolutionary history of *dsx* and sexual differentiation mechanisms.

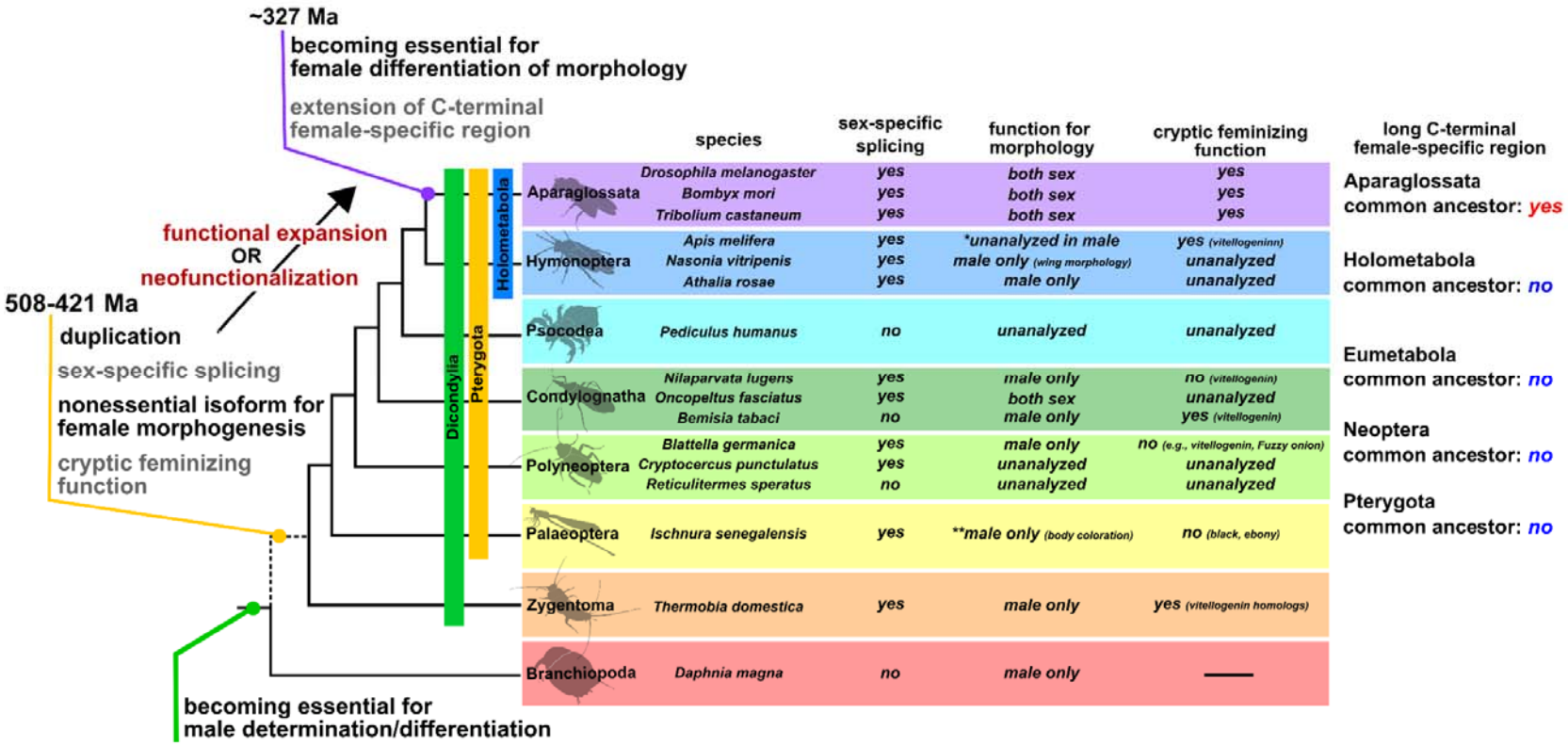


FIG. 6. Schematic image of the evolutionary scenario of *doublesex* proposed in this study and the feature of *dsx* in insects. *dsx* in arthropods may have been initially involved in only male determination/differentiation based on findings in crustaceans and chelicerates (Kato et al. 2011; Pomerantz et al. 2015; Li et al. 2018; Panara et al. 2019). In our hypothesis, the female isoform of *dsx* may have been not essential for female differentiation of morphological traits at least from the common ancestor of Dicondylia to the common ancestor of Aparaglossata but might have contributed to the expression of some genes in females ("cryptic feminizing function" in the figure). This "seemingly non-functional" isoform might have become essential for female differentiation of morphological traits at the common ancestor of Aparaglossata emerging at ~327

million years ago (Ma) through extending its cryptic feminizing function ("functional expansion" in the figure) or, alternatively, through acquiring entirely novel function ("neofunctionalization" in the figure). The extension of the C-terminus amino acid sequences in the female-specific region might be involved in the functional expansion/neofunctionalization of *dsx*. The common ancestor between Branchiopoda and Hexapoda may have had the male-specific expressed *dsx*. It is estimated that the sex-specific splicing control and the gene duplication of *dsx* occurred from the common ancestor between Branchiopoda and Hexapoda emerging at ~508 Ma to the common ancestor of Dicondylia at ~421 Ma. The phylogenetic relationship and the divergence time refer to Misof et al. (2014). The dotted line in the phylogenetic relationship indicates that the taxa that occurred from the common ancestor between Branchiopoda and Dicondylia to the common ancestor of Dicondylia are omitted. Here, we also show information on the current knowledge of *dsx* features in insects and a branchiopod. In Aparaglossata, since there are many studies, we show only three representative species. The information was based on: 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 (2012) in *Tribolium castaneum* (Coleoptera), Roth et al. (2019) and Velasque et al. (2018) in *Apis mellifera* (Hymenoptera), Wang et al. (2020) 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. 2014). The "unanalyzed" means the functional analyses of *dsx* have not been performed in the relevant species. Information on the roles of *dsx* of some species in female morphogenesis is limited to some body parts: e.g., body coloration in *I. senegalensis* (Takahashi et al. 2021), wing morphology in *Na. vitripenis* (Wang et al. 2020), and worker morphology in *Ap. mellifera* (Roth et al. 2019). The asterisk (*) in *Ap. mellifera* indicates that the functional analysis of *dsx* in males was not conducted although the 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 the color formation of the males and andromorphic females but not gynomorphic females (see Takahashi et al. 2021), suggesting that *dsx* is not essential for the female color development. In our hypothesis, the essential roles of *dsx* for female development in *O. fasciatus* (Just et al. 2021) may have occurred in parallel with Aparaglossata.

Materials and Methods

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 our 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. For examining the roles of *dsx* and *dsx-like* in the postembryonic morphogenesis, colonies of hatched nymphs were reared up to the fourth instar in a six-well plate and then transferred into 24-well plates to be kept individually. 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, we 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.

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. The timing of Hemi- or holometabolan insects can be estimated 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

637 and length of its styli during the fourth to ninth instar under our breeding conditions.
638 These features can be used to estimate molt timing, but it is difficult to apply these
639 criteria to experiments using adults or a large number of nymphs. To resolve this
640 problem, we used leg regeneration after autotomy and time-lapse imaging to estimate
641 the molt timing of *T. domestica*. Autotomy occurs at the joint between the trochanter
642 and femur in *T. domestica*. An autotomized leg regenerates after one molt (*Buck and*
643 *Edwards, 1990*). For the RNAi analysis during postembryonic development, we
644 amputated a right hindleg at the autotomic rift, using tweezers, and observed whether
645 the leg had regenerated. This test enabled us to rapidly estimate the molt timing. For
646 the RNA-seq and the RT-qPCR analysis, the time-lapse imaging was used to
647 determine the precise time of molt. We build a time-lapse imaging system with a
648 network camera system (SANYO, Tokyo, Japan) set in an incubator at 37°C
649 (supplementary fig. 7A). Photos of insects in the 24-well plate were taken every five
650 minutes. We created a time-lapse movie from the photos every 12 hours using ImageJ
651 1.52a (https://imagej.nih.gov/ij/) and observed whether the insects molted
652 (supplementary fig. 7B).

653 **De novo genome assembly**

654 A whole genome of *T. domestica* was sequenced to analyze the exon-intron
655 structure of *dsx*. We selected an adult female of *T. domestica* from our stock colony
656 and removed its alimentary canal. Genomic DNA was extracted from the sample
657 using DNeasy Blood and Tissue Kit (QIAGEN K.K., Tokyo, Japan). A paired-end
658 library was constructed from 1 µg of the DNA using TruSeq DNA PCR-Free LT
659 Sample Prep kits (Illumina K.K., Tokyo, Japan) following the manufacturer's
660 instructions. The library was run on a sequencer (HiSeq 2500; Illumina K.K., Tokyo,
661 Japan). We obtained 417 Gb of raw reads and assembled them using Platanus v1.2.4

assembler (*Kajitani et al. 2014*) after removal of the adapter sequences. The genome sequence can be obtained from the DNA Data Bank in Japan (Accession number: DRA005797; Bioproject: PRJDB5781).

Transcriptome analysis

To search for *doublesex* (*dsx*) and *vitellogenin* (*vtg*) homologs, we 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. We 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*). Information about the samples can be obtained from the National Center for Biotechnology Information (NCBI) BioSample database (Accession number: SAMN18175012–SAMN18175021).

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 Dsx homologs was performed using the amino acid sequences of the DM domain. We used the Dsx sequences of *D. melanogaster* as a query and obtained 97 metazoan DMRT family proteins from the NCBI and the i5k databases (https://i5k.nal.usda.gov/) and our genome data of *T. domestica* by the BLAST analysis (listed in supplementary table 1). We then aligned the sequences using MAFFT version 7 (Katoh et al. 2013) with the -linsi option (to use an accuracy option, L-INS-i) and manually extracted the DM domain, which consisted of 61 amino acids (supplementary fig. 8). The result of the multiple sequence alignment can be obtained from supplementary sequence file 1. Molecular phylogenetic analysis of the aligned sequences was performed using a maximum likelihood method after selecting a substitution model (JTT matrix-based model) with MEGA X (Kumar et al. 2018). Bootstrap values were calculated after 1000 replications.

Full-length cDNA and exon-intron structures

To elucidate the exon-intron structures of Dsx and Dsx-like, we determined the full-length cDNA sequences using a Rapid Amplification of cDNA Ends (RACE) method and performed a BLAST analysis for our genome database of *T. domestica*. We 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

712 extraction and ethanol precipitation. For 5□ -RACE analysis, mRNAs were purified
713 from 75 µg of the total RNAs using Dynabeads mRNA Purification kit (Thermo
714 Fisher Scientific K.K., Tokyo, Japan) following the manufacturer's instruction. We
715 then ligated an RNA oligo at the 5'-end of the mRNA using GeneRacer Advanced
716 RACE kits (Thermo Fisher Scientific K.K., Tokyo, Japan). For 3□-RACE analysis,
717 we ligated an RNA oligo of the SMART RACE cDNA Amplification Kit (Takara Bio
718 Inc., Shiga, Japan) at 3□-end of the total RNA during reverse transcription. First
719 stranded (fs-) cDNA was generated from the RNAs using SuperScript III Reverse
720 Transcriptase (Thermo Fisher Scientific K.K., Tokyo, Japan). We used primers
721 specific to the RNA oligos and performed RACE analysis by nested RT-PCR using
722 Q5 High-Fidelity DNA polymerase (New England BioLabs Japan Inc., Tokyo,
723 Japan). The primers specific to *dsx* and *dsx-like* were made from sequences of the
724 relevant genomic regions and are listed in supplementary table 9. The amplicons were
725 separated using the agarose gel-electrophoresis and cloned using TOPO TA Cloning
726 Kit for Sequencing (Thermo Fisher Scientific K.K., Tokyo, Japan) following the
727 manufacture's protocol. We used a DH5α *Escherichia coli* strain (TOYOBO CO.,
728 LTD., Osaka, Japan) as the host cell. Plasmids were extracted using the alkaline lysis
729 and purified by phenol-chloroform and ethanol precipitation. The nucleotide
730 sequences of the cloned amplicons were determined from the purified plasmids by the
731 Sanger Sequencing service of FASMAC Co. Ltd. (Kanagawa, Japan). We then
732 searched the genomic region of the full-length cDNA sequences of *dsx* and *dsx-like*
733 via local blastn analysis.

734 **Reverse transcription-quantitative PCR (RT-qPCR)**

735 To quantitative mRNA expression levels, we performed RT-qPCR analysis.
736 For investigating the sex-specific expression profile of *dsx* and *dsx-like*, we used the

737 fat body of adults of *T. domestica* since the sexes can be distinguishable by the
738 external morphology at this stage. Fat bodies also exhibit sex-specific physiological
739 functions in adults. Thirteenth instar individuals and adults after molting were
740 sampled for investigating roles of the genes in the sexually dimorphic morphology
741 and the *vitellogenin* expression, respectively. The sample sizes are reported in the
742 figure legends and supplementary table 2. We dissected the individuals in PBS and
743 collected their fat body in 2 ml tubes containing TRI Reagent (Molecular Research
744 Center Inc., Ohio, USA). The fat bodies then were disrupted using a TissueLyser LT
745 small beads mill (QIAGEN K.K., Tokyo, Japan). These disrupted samples were
746 preserved at −80°C until used. Total RNA was extracted from the samples according
747 to the manufacture's protocol for the TRI Reagent. Extracted RNA was treated with
748 2% RNase-free DNase I (New England BioLabs Japan Inc., Tokyo, Japan) at 37°C for
749 40 minutes and purified by phenol/chloroform extraction and ethanol precipitation.
750 We measured the concentration of the total RNA using a spectrophotometer (DS-11+,
751 Denovix Inc., Wilmington, USA). fs-cDNA was synthesized from 350 ng of the total
752 RNA using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific K.K.,
753 Tokyo, Japan). We diluted the fs-cDNA to 1:2 with MilliQ water and preserved it at
754 −30°C until it was used in RT-qPCR assay. The RT-qPCR assays were performed
755 using a LightCycler 96 instrument (Roche, Basel, Switzerland) according to the
756 manufacture's protocol with the THUNDERBIRD SYBR qPCR Mix (TOYOBO Co.
757 Ltd., Osaka, Japan). The reaction volume was 10 µl. We used 1 µl of the fs-cDNA as
758 templates. The preparation of the RT-qPCR solution proceeded on ice. The protocol
759 of the RT-qPCR was as follows: preincubation at 95°C for 600 seconds and 45 cycles
760 of three-step reactions, such as denaturation at 95°C for 15 seconds, annealing at 60°C
761 for 15 seconds and extension at 72°C for 45 seconds. We used *ribosomal protein 49*

(*rp49*) as a reference gene, as described by Ohde et al. (2011). We designed primer sets of the target genes by the Primer3Web version 4.1.0 (Untergasser et al. 2012) following the manufacture's recommended condition of the THUNDERBIRD SYBR qPCR Mix. We confirmed the primers' specificity using melting curves ranging from 65°C to 95°C. We selected primer sets exhibiting a single peak. The primers are listed in supplementary table 9. 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. In these removed data, a technical error was suspected. We 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. 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/web/packages/brunnermunzel/index.html). Holm's method was used for multiple comparison analyses between the control and treatments. The data are listed in supplementary table 2. In the *dsx* expression of the RNAi male, we performed the Smirnov-Grubbs (SG) test for ΔCt value using the *grubbs.test* function of the "outliers" package in R (https://cran.r-project.org/web/packages/outliers/index.html) (supplementary table 3). An outlier was detected in the *dsx* RNAi male. We repeatedly performed the SG test using the data excluding the outlier. No further outliers were detected. Lastly, we re-analyzed the data, excluding the outlier, using the BM test (supplementary table 2).

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

787 differentiation of insects is generally assumed to be a cell-autonomous mechanism
788 that is independent of systemic hormonal- control (Verhulst and van de Zande 2015)
789 as discussed in De Loof and Huybrechts (1998) and Bear and Monteiro (2013) and
790 progresses during postembryonic development. Therefore, nymphal RNAi is the most
791 effective tool to investigate the roles of genes on sexual trait formation during
792 postembryonic development. To reduce the risk of off-target effects, the dsRNA was
793 designed to avoid the region of the DM domain. We also confirmed that the dsRNA
794 had no contiguous matches of more than 20 bases with other genes on the genome by
795 BLAST (blastn option). To produce templates for the dsRNA, we cloned the regions
796 of *dsx* and *dsx-like* from the fs-cDNA using the same method as the RACE analysis.
797 We amplified the template DNAs from purified plasmids with PCR using Q5 High-
798 Fidelity DNA Polymerase and purified the amplified DNA with the
799 phenol/chloroform extraction and the ethanol precipitation. dsRNA was synthesized
800 from the purified DNA using Ampliscribe T7-Flash Transcription kits (Epicentre
801 Technologies, Co., Wisconsin, USA). We designed the PCR primers using the
802 Primer3Web version 4.1.0 (Untergasser et al. 2012). The PCR primers are listed in
803 supplementary table 9. In nymphal RNAi analysis, we injected the dsRNAs repeatedly
804 into the abdomen of the nymphs of *T. domestica* with each molt from the fourth or
805 fifth instar to thirteenth instar to sustain the RNAi effect during postembryonic
806 development. The initial stage was the same within a single experiment. This repeated
807 RNAi treatment was effective in some insects such as *Blattella germanica* (Wexler et
808 al. 2019). We sampled the individuals one, three, and five days after molting, using
809 phenotypic observations, analysis of *dsx* knockdown effects, and the oocyte number.
810 To determine the sex of individuals, we initially observed the gonads: testis and
811 ovary. In our RNAi analysis, the gonads completely formed and there was no

812 difference between the control and *dsx* RNAi individuals in external morphology (fig.
813 2C). Therefore, individuals with testis were males and those with ovaries were
814 females. *T. domestica* molts throughout its life, even after sexual maturation, and
815 produces *vtg* during each adult instar (Rousset and Bitsch 1993). To analyze the *vtg*
816 mRNA levels, we also injected the dsRNAs of *dsx* and *dsx-like* repeatedly into the
817 females and males every three days from 12 hours after molting. We sampled the
818 females and males at 720±20 minutes after subsequently molts.

819 **Phenotype observation**

820 We dissected thirteenth instar individuals in PBS using tweezers and removed
821 the thoraxes, reproductive systems, and external genital organs. We took images using
822 the digital microscope system (VHX-5000, KEYENCE, Tokyo, Japan). The thoraxes
823 and external genital organs were fixed with FAA fixative (formaldehyde: ethanol:
824 acetic acid = 15:5:1) at 25°C overnight and then preserved in 90% ethanol. We used
825 the length of the prothorax as an indicator of body size. To measure the prothoracic
826 width, the prothoracic notum was removed from the fixed thorax after treatment with
827 10% NaOH solution at 60°C for 30 minutes to dissolve the soft tissues. The notum
828 was mounted in Lemosol on a microscope slide. The prepared specimens were imaged
829 using a KEYENCE VHX-5000. With the microscope at 50×, the length of the notum
830 was measured. The ovipositor length was also measured using the microscope at 20×
831 and 50×. To count the sperm number, sperm was collected from seminal vesicles and
832 diluted with 5 ml MilliQ water. 50 µl of the diluted sperm was spotted on a
833 microscope slide and dried overnight. We technically replicated the measurement
834 three times for ovipositor length and six times in sperm number and calculated these
835 means. Measurement was performed by blinding the treatment. We counted the
836 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. We 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 supplementary table 4 (female) and 5 (male).

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).

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 (Li_2CO_3). 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

862 microscope slides coated with egg white-glycerin and stained using Delafield's
863 Hematoxylin and Eosin staining. After staining with the hematoxylin, the slides were
864 washed with 1% hydrochloric acid-ethanol for 40 seconds. The stained slides were
865 enclosed with Canada balsam. We observed the slides on an optical microscope
866 (Olympus, Tokyo, Japan) and took photos using a digital single-lens reflex camera
867 (Nikon, Tokyo, Japan).

868 **Ancestral Sequence Reconstruction**

869 To infer the sequence evolution of the *dsx*, we conducted an ancestral
870 sequence reconstruction (ASR) of the C-terminal sequences of the *dsx* female-type
871 homologous sequence. First, we searched homologous sequences to *dsx* female-type
872 from NCBI protein/transcript shotgun assembly databases and previous studies. The
873 searches in the NCBI databases were performed by BLAST search. We closely
874 examined the alignment results of the BLAST and selected sequences with at least 10
875 amino acids aligned with the female-specific region of each query sequence. We do
876 not know whether some of these sequences are expressed in females and contribute to
877 female morphogenesis, as these sequences are not necessarily to have investigated
878 expression and function in the species. We decided that it was not problem to use
879 these sequences since we focused on the evolution of sequences homologous to *dsx*
880 female-type in each insect taxa. In Diptera, we set *dsx* female-type of *D. melanogaster*
881 (Accession #: NP_001287220) as a query and obtained 9 sequences. In Lepidoptera,
882 we used *dsx* female-type of *B. mori* (NP_001036871) as a query and get 10 sequences.
883 In Coleoptera, *dsx* female-type of *Tribolium castaneum* (AFQ62106) was set in a
884 query and then 10 sequences were obtained. We used *dsx* female-type of *Ap. mellifera*
885 (NP_001128407) and *At. rosae* (XP_012262256) as queries to search hymenopteran
886 sequences. We 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, we 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, we 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). We 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 supplementary table 7. We 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) (supplementary sequence file 2). We 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) (supplementary fig. 9). We selected the most probable sequences for the following analyses. The results of ASR can be seen in supplementary table 8. The probabilities of sites of AS that we focused on are listed in supplementary table 10. 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, we concluded that the AS in Aparaglossata and Holometabola, which we 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 our 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 our analysis. Thus, our conclusion that the Aparaglossata-specific region occurred in the common ancestor of Aparaglossata would be confident. To compare the sequences, we then performed MSA of the most probable reconstructed ancestral sequences and the sequence of *D. melanogaster* using MAFFT version 7 (fig. 5A).

Protein Structure Prediction

To infer the evolution of protein structures of *dsx*, we conducted the protein structure prediction. The ancestral sequences reconstructed by the above section were used for the protein structure prediction. The sequences were obtained from supplementary sequence file 3. 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. We 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, we 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 supplementary fig. 10. The 3D models of predicted structures were visualized with the PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC.). On the viewer, we colored the female-specific region and the Aparaglossata-specific region with red color and the green color, respectively.

Data availability

The draft genome data was deposited in the DNA Data Bank of Japan (Accession number: DRA005797; Bioproject: PRJDB5781). The raw read data of the transcriptome was in the NCBI Sequence Read Archive (Accession numbers: SRR13870115–SRR13870124; Bioproject: PRJNA707122). The sequences of *dsx* male-type, *dsx* female-type, and *dsx-like* are also in GenBank (Accession numbers: MW711323, MW711324, and MW711325, respectively).

Acknowledgments

We express gratitude to Dr. Daniel Bopp (University of Zürich) for his comments and encouragement for this manuscript. We would like to thank Dr. Takahiro Ohde (National Institute for Basic Biology; Kyoto University) for his help to extract genomic DNA of *T. domestica* and for a great advice to the discussion. We are also

962 grateful to Dr. Toshiya Ando, Dr. Taro Nakamura, Dr. Shinichi Morita, Dr. Hiroki
963 Sakai, and Dr. Tatsuro Konagaya (National Institute for Basic Biology) for technical
964 advice and discussion on this manuscript. We also express our gratitude to Dr. Satoshi
965 Miyazaki (Tamagawa university) for providing sequences of *dsx* in *Cryptocercus*
966 *punctulatus*. Computations were performed on the NIG supercomputer at ROIS,
967 National Institute of Genetics and the Data Integration and Analysis Facility, National
968 Institute for Basic Biology. We thank the Model Plant Research Facility, NIBB
969 Bioresource Center for providing the network camera system. This work was
970 supported by the JSPS KAKENHI Grant numbers JP25660265, JP16H02596, and
971 JP16H06279 (PAGS) for TN and the Sasakawa Scientific Research Grant from The
972 Japan Science Society for YC.

973 **Author Contributions:**

974 YC and TN conceived this study. YC performed all experiments, observations, and
975 analyses other than the genome sequence and assembly. AT sequenced the genome.
976 MO and TI performed the de novo genome assembly. YC and TI wrote the
977 manuscript. All authors commented on the manuscript.

978 **Competing Interest Statement**

979 The authors declare that have no competing interests.

980

981 # References

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

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

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

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

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

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

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

999 Boudinot BE. 2018. A general theory of genital homologies for the Hexapoda
1000 (Pancrustacea) derived from skeletomuscular correspondences, with emphasis on
1001 the Endopterygota. *Arthropod Struct Dev.* **47:** 563–613.

1002 Buck C, Edwards JS. 1990. The effect of appendage and scale loss on instar duration
1003 in adult firebrats, *Thermobia domestica* (Thysanura). *J Exp Biol.* **151:** 341–347.

1004 Burtis KC, Baker BS. 1989. *Drosophila doublesex* gene controls somatic sexual
1005 differentiation by producing alternatively spliced mRNAs encoding related sex–
1006 specific polypeptides. *Cell* **56:** 997–1010.

1007 Byrne BM, Gruber MABG, Ab G. 1989. The evolution of egg yolk proteins. *Prog*
1008 *Biophys Mol Biol*. **53:** 33–69.

1009 Clough E, Jimenez E, Kim YA, Whitworth C, Neville MC, Hempel LU, Pavlou HJ,
1010 Chen ZX, Sturgill D, Dale RK. 2014. Sex- and tissue-specific functions of
1011 *Drosophila* Doublesex transcription factor target genes. *Dev Cell* **31:** 761–773.

1012 Darling AL, Uversky VN. 2018. Intrinsic disorder and posttranslational
1013 modifications: the darker side of the biological dark matter. *Front Genet*. **9:** 158.

1014 Darwin C. 1871. The Descent of Man, and Selection in Relation to Sex vol. 1.
1015 London: John Murray.

1016 De Loof A, Huybrechts R. 1998. “Insects do not have sex hormones”: a myth? *Gen*
1017 *Comp Endocrinol*. 111: 245–260.

1018 Emeljanov AF. 2014. The evolutionary role and fate of the primary ovipositor in
1019 insects. *Entomol Rev*. **94:** 367–396.

1020 Erdman SE, Chen HJ, Burtis KC. 1996. Functional and genetic characterization of the
1021 oligomerization and DNA binding properties of the *Drosophila doublesex* proteins.
1022 *Genetics* **144:** 1639–1652.

1023 Fryxell DC, Weiler DE, Kinnison MT, Palkovacs EP. 2019. Eco–evolutionary
1024 dynamics of sexual dimorphism. *Trends Ecol Evol.* **34:** 591–594.

1025 Geddes P, Thomson JA. 1889. *The evolution of sex*. London: Walter Scott.

1026 Ghosh N, Bakshi A, Khandelwal R, Rajan SG, Joshi R. 2019. The Hox gene
1027 *Abdominal-B* uses Doublesex^{F} as a cofactor to promote neuroblast apoptosis in the
1028 *Drosophila* central nervous system. *Development* **146:** dev175158.

1029 Gotoh H, Zinna RA, Warren I, DeNieu M, Niimi T, Dworkin I, Emlen DJ, Miura T,
1030 Lavine LC. 2016. Identification and functional analyses of sex determination genes
1031 in the sexually dimorphic stag beetle *Cyclommatus metallifer*. *BMC Genomics* **17:**
1032 250.

1033 Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Münsterberg A, Vivian N,
1034 Goodfellow P, Lovell-Badge R. 1990. A gene mapping to the sex-determining
1035 region of the mouse Y chromosome is a member of a novel family of
1036 embryonically expressed genes. *Nature* **346:** 245–250.

1037 Guo L, Xie W, Liu Y, Yang Z, Yang X, Xia J, Wang S, Wu Q, Zhang Y. 2018.
1038 Identification and characterization of *doublesex* in *Bemisia tabaci*. *Insect Mol Biol.*
1039 **27:** 602–632.

1040 Gustafson GT, Baca SM, Alexander AM, Short AEZ. 2020. Phylogenomic analysis of
1041 the beetle suborder Adephaga with comparison of tailored and generalized
1042 ultraconserved element probe performance. *Syst Entomol.* **45:** 552–570.

1043 Haag ES, True JR. 2021. Developmental system drift. In: Nuño de la Rosa L, Müller
1044 GB, editors. Evolutionary developmental biology: a reference guide. Switzerland:
1045 Springer Nature Switzerland AG. P. 99–110.

1046 Hasselmann M, Gempe T, Schiøtt M, Nunes-Silva CG, Otte M, Beye M. 2008.
1047 Evidence for the evolutionary nascence of a novel sex determination pathway in
1048 honeybees. *Nature* **454:** 519–522.

1049 Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, Fernandino JI,
1050 Somoza GM, Yokota M, Strüssmann CA. 2012. A Y-linked anti-Müllerian
1051 hormone duplication takes over a critical role in sex determination. *Proc Natl Acad*
1052 *Sci USA.* **109:** 2955–2959.

1053 Hayward A, Takahashi T, Bendena WG, Tobe SS, Hui JH. 2010. Comparative
1054 genomic and phylogenetic analysis of vitellogenin and other large lipid transfer
1055 proteins in metazoans. *FEBS Letters* **584:** 1273–1278.

1056 Herpin A, Schartl M. 2015. Plasticity of gene-regulatory networks controlling sex
1057 determination: of masters, slaves, usual suspects, newcomers, and usurpators.
1058 *EMBO Reports* **16:** 1260–1274.

1059 Hildreth PE. 1965. Doublesex, a recessive gene that transforms both males and
1060 females of *Drosophila* into intersexes. *Genetics* **51:** 659–678.

1061 Hildreth PE, Lucchesi JC. 1963. A gene which transforms males and females into
1062 intersexes. *Proc 11th Intern Congr Genet.* **1:** 171.

1063 Hopkins BR, Kopp A. 2021. Evolution of sexual development and sexual dimorphism
1064 in insects. *Curr Opin Genet Dev.* **69:** 129–139.

1065 Ito Y, Harigai A, Nakata M, Hosoya T, Araya K, Oba Y, Ito A, Ohde T, Yaginuma T,
1066 Niimi T. 2013. The role of *doublesex* in the evolution of exaggerated horns in the
1067 Japanese rhinoceros beetle. *EMBO Reports* **14:** 561–567.

1068 Just J, Laslo M, Lee YJ, Yarnell M, Zhang Z, Angelini DR. 2021. Distinct
1069 developmental mechanisms influence sexual dimorphisms in the milkweed bug
1070 *Oncopeltus fasciatus*. *bioRxiv*. doi: 10.1101/2021.05.12.443917.

1071 Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M,
1072 Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T.
1073 2014. Efficient de novo assembly of highly heterozygous genomes from whole-
1074 genome shotgun short reads. *Genome Res.* **24:**1384–95.

1075 Kato Y, Kobayashi K, Watanabe H, Iguchi T. 2011. Environmental sex determination
1076 in the branchiopod crustacean *Daphnia magna*: deep conservation of a *Doublesex*
1077 gene in the sex-determining pathway. *PLOS Genet.* **7:** e1001345.

1078 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
1079 7: improvements in performance and usability. *Mol Biol Evol.* **30:** 772–780.

1080 Kawahara AY, Plotkin D, Espeland M, Meusemann K, Toussaint EFA, Donath A,
1081 Gimnich F, Frandsen PB, Zwick A, dos Reis M, et al. 2019. Phylogenomics
1082 reveals the evolutionary timing and pattern of butterflies and moths. *Proc Natl*
1083 *Acad Sci USA.* **116:** 22657–22663.

1084 Keren H, Lev-Maor G, Ast G. 2010. Alternative splicing and evolution:
1085 diversification, exon definition and function. *Nat Rev Genet.* **11:** 345–355.

1086 Kijimoto T, Moczek AP, Andrews J. 2012. Diversification of *doublesex* function
1087 underlies morph-, sex-, and species-specific development of beetle horns. *Proc*
1088 *Natl Acad Sci USA.* **109:** 20526–20531.

1089 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome
1090 alignment and genotyping with HISAT2 and HISAT–genotype. *Nat Biotechnol.*
1091 **37:** 907–915.

1092 Klag J. 1977. Differentiation of primordial germ cells in the embryonic development
1093 of *Thermobia domestica*, Pack. (*Thysanura*): an ultrastructural study. *J Embryol*
1094 *exp Morph.* **38:** 93–114.

1095 Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. 1991. Male
1096 development of chromosomally female mice transgenic for *Sry*. *Nature* **351:** 117–
1097 121.

1098 Kopp A. 2012. *Dmrt* genes in the development and evolution of sexual dimorphism.
1099 *Trends Genet.* **28**: 175–184.

1100 Kristensen NP. 1975. The phylogeny of hexapod “orders”. A critical review of recent
1101 accounts. *J Zool Syst Evol Res.* **13:** 1–44.

1102 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular
1103 evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* **35:**
1104 1547–1549.

1105 Lago DC, Martins JR, Dallacqua RP, Santos DE, Bitondi MM, Hartfelder K. 2020.
1106 Testis development and spermatogenesis in drones of the honey bee, *Apis mellifera*
1107 L. *Apidologie* **51:** 935–955.

1108 Li H, Leavengood JM, Chapman EG, Burkhardt D, Song F, Jiang P, Liu J, Zhou X,
1109 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.

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.

Ledón-Rettig C, Zattara E, Moczek A. 2017. Asymmetric interactions between *doublesex* and tissue- and sex-specific target genes mediate sexual dimorphism in beetles. *Nat Commun.* **8:** 14593.

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.

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.

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.

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.

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.

1152 Morita S, Ando T, Maeno A, Mizutani T, Mase M, Shigenobu S, Niimi T. 2019.
1153 Precise staging of beetle horn formation in *Trypoxylus dichotomus* reveals the
1154 pleiotropic roles of *doublesex* depending on the spatiotemporal developmental
1155 contexts. *PLOS Genet.* **15:** e1008063.

1156 Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, Shan Z, Haaf T,
1157 Shimizu N, Shima A, Schmid M, Schartl M. 2002. A duplicated copy of DMRT1
1158 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*.
1159 *Proc Natl Acad Sci USA.* **99:** 11778–11783.

1160 Ohbayashi F, Suzuki MG, Mita K, Okano K, Shimada T. 2001. A homologue of the
1161 *Drosophila doublesex* gene is transcribed into sex-specific mRNA isoforms in the
1162 silkworm, *Bombyx mori*. *Comp Biochem Physiol B Biochem Mol Biol*. **128:** 145–
1163 158.

1164 Ohde T, Yaginuma T, Niimi T. 2011. Nymphal RNAi analysis reveals novel function
1165 of scalloped in antenna, cercus and caudal filament formation in the firebrat,
1166 *Thermobia domestica*. *J Insect Biotechnol Sericol.* **80:** 101–108.

1167 Panara V, Budd GE, Janssen R. 2019. Phylogenetic analysis and embryonic
1168 expression of panarthropod Dmrt genes. *Front Zool.* **16:** 1–18.

1169 Pei XJ, Fan YL, Bai Y, Bai TT, Schal C,Zhang ZF, Chen N, Li S, Liu TX. 2021.
1170 Modulation of fatty acid elongation in cockroaches sustains sexually dimorphic
1171 hydrocarbons and female attractiveness. *PLOS Biol.* **19:** e3001330.

1172 Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015.
1173 StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.
1174 *Nat Biotechnol.* **33:** 290–295.

1175 Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A,
1176 Podsiadlowski L, Petersen M, Lanfear R, et al. 2017. Evolutionary History of the
1177 Hymenoptera. *Curr Biol.* **27:** 1013–1018.

1178 Pomerantz AF, Hoy MA, Kawahara AY. 2015. Molecular characterization and
1179 evolutionary insights into potential sex-determination genes in the western orchard
1180 predatory mite *Metaseiulus occidentalis* (Chelicerata: Arachnida: Acari:
1181 Phytoseiidae). *J Biomol Struct Dyn.* **33:** 1239–1253.

1182 R Core Team. 2020. R: A language and environment for statistical computing. R
1183 Foundation for Statistical Computing.

1184 Rideout EJ, Narsaiya MS, Grewal SS. 2015. The sex determination gene transformer
1185 regulates male-female differences in *Drosophila* body size. *PLoS Genet.* **11:**
1186 e1005683.

1187 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for
1188 differential expression analysis of digital gene expression data. *Bioinformatics* **26**:
1189 139–140.

1190 Romero-Pozuelo J, Foronda D, Martín P, Hudry B, Merabet S, Graba Y, Sánchez-
1191 Herrero E. 2019. Cooperation of axial and sex specific information controls
1192 Drosophila female genitalia growth by regulating the Decapentaplegic pathway.
1193 *Dev Biol.* **454:** 145–155.

1194 Roth A, Vleurinck C, Netschitailo O, Bauer V, Otte M, Kaftanoglu O, Page RE, Beye
1195 M. 2019. A genetic switch for worker nutrition-mediated traits in honeybees. *PLOS*
1196 *Biol.* **17:** e3000171.

1197 Rousset A., Bitsch C. 1993. Comparison between endogenous and exogenous yolk
1198 proteins along an ovarian cycle in the firebrat *Thermobia domestica* (Insecta,
1199 *Thysanura*). *Comp Biochem Physiol B Biochem Mol Biol.* **104:** 33–44.

1200 Sato Y, Shinka T, Sakamoto K, Ewis AA, Nakahori Y. 2010. The male-determining
1201 gene *SRY* is a hybrid of *DGCR8* and *SOX3*, and is regulated by the transcription
1202 factor *CP2*. *Mol Cell Biochem.* **337:** 267–275.

1203 Shukla JN, Palli SR. 2012. Doublesex target genes in the red flour beetle, *Tribolium*
1204 *castaneum*. *Sci Rep.* **2:** 948.

1205 Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW,
1206 Frischauf AM, Lovell-Badge R, Goodfellow PN. 1990. A gene from the human
1207 sex-determining region encodes a protein with homology to a conserved DNA-
1208 binding motif. *Nature* **346:** 240–244.

1209 Suzuki MG, Funaguma S, Kanda T, Tamura T, Shimada T. 2003. Analysis of the
1210 biological functions of a *doublesex* homologue in *Bombyx mori*. *Dev Gene Evol.*
1211 **213:** 345–354.

1212 Takaku Y, Suzuki H, Ohta I, Ishii D, Muranaka Y, Shimomura M, Hariyama T. 2013.
1213 A thin polymer membrane, nano-suit, enhancing survival across the continuum
1214 between air and high vacuum. *Proc Natl Acad Sci USA.* **110:**7631–7635.

1215 Takahashi M, Okude G, Futahashi R, Takahashi Y, Kawata M. 2021. The effect of
1216 the *doublesex* gene in body colour masculinization of the damselfly *Ischnura*
1217 *senegalensis*. *Biol Lett.* **17:** 20200761.

1218 Takahashi M, Takahashi Y, Kawata M. 2019. Candidate genes associated with color
1219 morphs of female-limited polymorphisms of the damselfly *Ischnura senegalensis*.
1220 *Heredity* **122:** 81–92.

1221 Takehana Y, Matsuda M, Myosho T, Suster ML, Kawakami K, Shin-I T, Kohara Y,
1222 Kuroki Y, Toyoda A, Fujiyama A, et al. 2014. Co-option of *Sox3* as the male-
1223 determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nat*
1224 *Commun.* **5:** 4157.

1225 Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and
1226 old ideas. *Annu. Rev. Genet.* **38:** 615–643.

1227 Thongsaiklaing T, Passara H, Nipitwathanaphon M, Ngernsiri L. 2018. Identification
1228 and characterization of *doublesex* from the pumpkin fruit fly, *Bactrocera tau*
1229 (Diptera: Tephritidae). *European Journal of Entomology* **115:** 602–613.

1230 True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary
1231 trajectories. *Evol. Dev.* **3:** 109–119.

1232 Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A,
1233 Cowie A, Meyer C, Laydon A, et al. 2021. Highly accurate protein structure
1234 prediction for the human proteome. *Nature* **596:** 590–596.

1235 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG.
1236 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40:** e115.

1237 Velasque M, Qiu L. Mikheyev AS. 2018. The D*oublesex* sex determination pathway
1238 regulates reproductive division of labor in honey bees. *bioRxiv* doi:
1239 10.1101/314492.

1240 Verhulst EC, van de Zande L. 2015. Double nexus—Doublesex is the connecting
1241 element in sex determination. *Brief Funct Genom.* **14:** 396–406.

1242 Wang Y, Chen X, Liu Z, Xu J, Li X, Bi H, Andongma AA, Niu C, Huang Y. 2019.
1243 Mutation of *doublesex* induces sex-specific sterility of the diamondback moth
1244 *Plutella xylostella*. *Insect Biochem. Mol. Biol.* 112: 1138.

1245 Wang Y, Rensink A, Fricke U, Riddle MC, Trent C, van de Zande L, Verhulst EC.
1246 2020. Sexually dimorphic traits and male-specific differentiation are actively
1247 regulated by Doublesex during specific developmental windows in *Nasonia*
1248 *vitripennis*. *bioRxiv* doi: 10.1101/2020.04.19.048553.

1249 Wexler J, Delaney EK, Belles X, Schal C, Wada-Katsumata A, Amicucci MJ, Kopp
1250 A. 2019. Hemimetabolous insects elucidate the origin of sexual development via
1251 alternative splicing. *eLife* **8:** e47490.

1252 Wexler JR, Plachetzki DC, Kopp A. 2014. Pan-metazoan phylogeny of the DMRT
1253 gene family: a framework for functional studies. *Dev Gene Evol.* **224:** 175–181.

1254 Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone
1255 MA, Cassel BL, Bayless KM, Heimberg AM, et al. 2011. Episodic radiations in
1256 the fly tree of life. *Proc Natl Acad Sci USA.* **108:** 5690–5695.

1257 Wilkins AS. 1995. Moving up the hierarchy: a hypothesis on the evolution of a
1258 genetic sex determination pathway. *Bioessays* **17:** 71–77.

1259 Xu J, Zhan S, Chen S, Zeng B, Li Z, James AA, Tan A, Huang Y. Sexually dimorphic
1260 traits in the silkworm, *Bombyx mori*, are regulated by *doublesex*. 2017. *Insect*
1261 *Biochem Mol Biol*. **80:** 42–51.

1262 Yang Y, Zhang W, Bayrer JR, Weiss MA. 2008. Doublesex and the regulation of
1263 sexual dimorphism in *Drosophila melanogaster*: structure, function, and
1264 mutagenesis of a female-specific domain. *J Biol Chem.* **283:** 7280–7292.

1265 Zhang SQ, Che LH, Li Y, Liang D, Pang H, Ślipiński A, Zhang P. 2018. Evolutionary
1266 history of Coleoptera revealed by extensive sampling of genes and species. *Nat*
1267 *Commun.* **9:** 1–11.

1268 Zhuo JC, Hu QL, Zhang HH, Zhang MQ, Jo SB, Zhang CX. 2018. Identification and
1269 functional analysis of the *doublesex* gene in the sexual development of a
1270 hemimetabolous insect, the brown planthopper. *Insect Biochem Mol Biol*. **102:** 31–
1271 42.