New Insights into the Genomic Organization and Splicing of the *Doublesex* Gene, a Terminal Regulator of Sexual Differentiation in the Silkworm *Bombyx mori*

Jianping Duan1,2*, Hanfu Xu1*, Huizhen Guo1, David A. O’Brochta3, Feng Wang1, Sanyuan Ma1, Liying Zhang1, Xingfu Zha1, Ping Zhao1, Qingyou Xia1*

1 State Key Laboratory of Silkworm Genome Biology (Southwest University), Chongqing, PR China, 2 Henan Provincial Key Laboratory of Funiu Mountain Insect Biology, Nanyang Normal University, Nanyang, PR China, 3 Department of Entomology, University of Maryland, College Park, United States of America

Abstract

Sex-determination mechanisms differ among organisms. The primary mechanism is diverse, whereas the terminal regulator is relatively-conserved. We analyzed the transcripts of the *Bombyx mori* doublesex gene (*Bmdsx*), and reported novel results concerning the genomic organization and expression of *Bmdsx*. *Bmdsx* consists of nine exons and eight introns, of which two exons are novel and have not been reported previously. *Bmdsx* transcripts are spliced to generate seventeen alternatively-spliced forms and eleven putative trans-spliced variants. Thirteen of the alternatively-spliced forms and five of the putative trans-spliced forms are reported here for the first time. Sequence analysis predicts that ten female-specific, six male-specific splice forms and one splice form found in males and females will result in four female-specific, two male-specific Dsx proteins and one Dsx protein common to males and females. The Dsx proteins are expected to be functional and regulate downstream target genes. Some of the predicted Dsx proteins are described here for the first time. Therefore the expression of the *dsx* gene in *B. mori* results in a variety of cis- and trans-spliced transcripts and multiple Dsx proteins. These findings show that in *B. mori* there is a complicated pattern of *dsx* splicing, and that the regulation of splicing and sex-specific functions of lepidopteran *dsx* have evolved complexity.

Introduction

In insects sex is determined autonomously in somatic cells and the mechanisms that govern somatic sexual differentiation have been studied in many species. What has emerged from these studies is the realization that there are a diverse array of primary upstream signals in these various sex-determination hierarchies whereas the downstream genes in these hierarchies are relatively few and comparatively well-conserved [1]. The large number of diverse insects for which we have a molecular genetic understanding of the sex-determination hierarchy confronts us with the challenge of understanding how much divergence of the participating genes can be tolerated while maintaining functional conservation.

Sex-determination hierarchies have evolved in a retrograde manner from bottom to up as first proposed by Wilkins (1995). There are myriad primary mechanisms initiating the sex-determination cascade, such as the X:A ratio in *D. melanogaster* [2], the dominant maleness factor on Y chromosome in *C. capitata* [3], the haploid/diploidy and the number and allelic forms of the csd gene in *A. mellifera* [4], the activity of the zsd gene in *N. vitripennis* [5], and the dominant epistatic factor on W chromosome in *B. mori* [6,7] and so on. Unlike the diversity of genes involved in the upper part of the sex determination cascade, the genes and functions involved in the lower parts of the cascade are relatively well-conserved [8]. For example, tra (transformer)-mediated expression of *dsx* (*doublesex*) is prevalent in most dipterans and well-studied hymenopterans, however only in drosophilids is *Sex lethal* the master regulator of the *tra/dsx* mode of sex determination [9-11]. In these insects *Sex lethal* acts as the master gene responsible for initiating and maintaining the choice of sexual identity via an...
autoregulatory feedback loop [1,9-14]. In several dipterans and hymenopterans such as C. capitata, A. mellifera and N. vitripennis Sxl’s place upstream of dsx is occupied by tra homologues, so that tra initiates the choice of sexual identity and maintains that choice through autoregulation [5,10,15,16]. Both Sxl and tra genes are evolving rapidly, with tra being relatively more conserved than Sxl. So far, the homolog of tra in B. mori has not been identified, and whether the B. mori homolog of Sxl serves a sex-determining role remains unclear [17]. Sxl in D. melanogaster produces three early transcripts and seven late, sex-specific transcripts encoding multiple related polypeptides [18] whose functions are still unclear. Sxl in B. mori is not sex-specifically expressed, as is the case in other non-Drosophilid species [10,11], and therefore its role may also be similar in these species, although it appears not to act as the master regulator.

The dsx gene is the terminal regulator of sex differentiation at the bottom of the sex-determination cascade and its structure, expression and function are conserved among organisms [19]. The Drosophila dsx gene controls the development of sexually dimorphic features by producing two alternatively-spliced transcripts encoding two-related sex-specific Dsx proteins with common amino termini but sex-specific carboxy termini [20-24]. In addition, two transcripts are present in both sexes throughout the larval period [25]. This is the case in M. domestica and A. mellifera in which in addition to sex-specific transcripts there is a dsx transcript encoding a common Dsx protein in males and females [26,27]. In Aedes aegypti, A. assama, A. mylitta and B. mori a second female-specific Dsx protein is predicted [28-30]. Recently, we proposed that B. mori has a third female-specific Dsx protein predicted to arise from a novel trans-spliced transcript, Bmdsx-dsr2d [31]. The extent to which insects have evolved multiple sex-specific Dsx proteins warrants further study.

To date relatively few dsx splice forms and proteins have been described. The only exception to date has been in A. assama in which the dsx homologue results in six alternatively-spliced forms in females, resulting in two female-specific Dsx proteins [29]. Additionally, why ectopic expression of female Dsx proteins (DsxF) in male M. domestica [26] and B. mori [31,32] did not show any signs of morphological sex reversal has not been adequately explained. One possible reason might be a requirement for a female-specific cofactor for DsxF to promote female development and the existence of an antagonistic effect of the male form of Dsx, DsxM. So, while the general framework of dsx expression and Dsx function are fairly well established there are aspects that remain unclear. Here we present a comprehensive analysis of Bmdsx expression in conjunction with an analysis of the resulting transcripts that show a complex pattern of expression and splicing that has not been described before. The biological implications of this complex pattern of Bmdsx splicing remain to be determined, however it is certain to enlighten efforts to understand insect sex determination and its evolution.

Materials and Methods

PCR with exon-specific primers

Total RNA was extracted from the head, integument, hemocytes, midgut, fat body, trachea, malpighian tubules, silk glands and gonads of both sexes using Trizol reagent (Invitrogen, USA) according to the manufacturer’s protocol. cDNA templates were synthesized by reverse transcription using M-MLV reverse transcriptase (Promega, USA) as described [31]. PCR amplifications were performed using primer set P1/P2 to assess the expression of Bmdsx transcripts in day-3 fifth-instar larvae of both sexes. The relative abundance of each band was quantified by densitometric measurement of the EB-stained gel using Quantity One® software (Bio-Rad, USA). P1 and P2 bound specifically to exons 1 and 6, respectively which are common to Bmdsx transcripts in both males and females. The following conditions were used: 94°C, 3 min followed by 30 cycles of 94°C, 30 s; 57°C, 40 s; and 72°C, 1 min; with a final 10 min extension at 72°C. The B. mori cytoplasmic actin 3 gene (Bmactin3) was used as the reference gene.

3’ RACE, sequencing validation and protein analysis

3’ RACE was conducted as described [31] for two successive rounds of PCR amplification, initially with P3/3P1 and then with P4/3P2 primer pairs. 3’ RACE was performed using GeneRacer™ according to the manufacture’s recommendations (Invitrogen, USA). GeneRacer™ Oligo-d(T) was used to initiate the reverse transcription reaction while primers P3 and P4 were specific to the 5’-terminal region of dsx transcripts. All amplified products were cloned and sequenced using routine methods. Sequencing data have been deposited in the GenBank under the accession numbers KF255805-KF255831. These sequencing results were then compared by BLAST analysis against the BGI B. mori genome assembly (http://silkworm.swu.edu.cn/silkdb/) [33,34] to confirm novel transcripts and genomic organization of Bmdsx. Transcripts were virtually translated and the predicted proteins were aligned using ClustalX [35] and GeneDoc [36] for analysis.

RT-PCR analysis

RT-PCR reactions were performed using cDNA templates synthesized as described above and the position-specific primers listed in Table 1. Primers P1, P3 and P4 were specific to exon 1. Primers P5, P9 and P2 were specific to exons 3n, 6n and 6, respectively. Primers P6 and P8 were specific to the 15-bp and 133-bp stretches in exons 3 (exon 3s) and 4 (exon 4s), respectively. Primer P7 was specific to exon 3/4, a splicing product between the 3’ end of unstretched exon 3 and the 5’ end of unstretched exon 4. Primers P1 was combined with primers P5, P6, P7, P8 and P9 to detect the alternative-splicing of exons 3n, 3s, 3/4, 4s and 6n, respectively. The PCR conditions were used as described above, using some modifications of the annealing temperatures according to Table 1.
Novel Exons and Splice Forms of dsx in B. mori

Table 1. Primers used for PCR and 3’RACE.

| Primer Sequence | Annealing temperature (°C) |
|-----------------|-----------------------------|
| P1 | TCCCTGTCGCCCTCTCCTAECT | 58 |
| P2 | GAAGTGGCGGCGGAATCCTGTG | 56 |
| P3 | TGCACGACAGCTCCACCG | 58 |
| P4 | CCACGTGATACCCCCCACTC | 60 |
| 3P1 | GCCTGACGATAGCAGTGAACG | 60 |
| 3P2 | CGCTACGTAACGGCATGACGATG | 60 |
| P5 | TGAAGGAAGAACCCAAAAC | 52 |
| P6 | CCAGATTTTCTATATGCTCC | 50 |
| P7 | TTTCAGCATTTTTCCTGCGC | 52 |
| P8 | TCACACGTGCCACATACT | 55 |
| P9 | TCCATTGACATAATCCAGG | 50 |

Results and Discussion

Unexpected alternatively-spliced and trans-spliced Bmdsx transcripts

Earlier studies reported that Bmdsx transcripts are spliced in a sex-specific manner to yield two female forms (Bmdsx84 and Bmdsx87) and one male form (Bmdsx44) [19,30]. In a recent study, we predicted the existence of a novel exon, 2n, and a third female-specific splice form Bmdsx31 [31]. To confirm this mode of splicing and expression, PCR was performed using the primer set P1/P2 binding specifically to the exons 1 and 8 that are common to all splice variants (Figure 1A). An unexpected expression pattern was observed showing extra amplification bands in both sexes, in addition to the bands predicted based on earlier studies (Figure 1B), indicating that some unidentified splice forms of Bmdsx remained to be described. Subsequently, all the comparatively high-abundance bands were cloned and sequenced. This analysis identified four female-specific (BmdsxE1, BmdsxE2, BmdsxE4, and BmdsxE5) and three male-specific (BmdsxM1, BmdsxM2, and BmdsxM4) splice forms (Figure 2A). BmdsxE1, BmdsxE2, and BmdsxE4 are well-known alternatively-spliced forms described in earlier studies, while transcripts BmdsxE4, BmdsxF5, BmdsxG2, and BmdsxH5 have not been described previously. BmdsxF5 and BmdsxG2 are identical to BmdsxE2 and BmdsxE4, respectively, except for the absence of exon 5. BmdsxE4 differs from BmdsxE1 by the presence and absence of a novel exon, 3n and a known exon 5, respectively, and the detection of splice form BmdsxE1 confirms our earlier report of the existence of novel exon, 2n [31]. In addition, bands for some isoforms in Figure 1B are overlapped, and the relative abundance of previously reported forms are high in each lane for tissues of both sexes. The new isoforms are evidently at a lower abundance than that for the major previously described isoforms (Figure 1C and 1D), indicating that some new isoforms might not functionally significant events. If the amount of an isoform is very low, it may be the by-product of splicing leakage, which can be tolerated due to its low frequency. We failed to find a clear band in Figure 1B for BmdsxE4, suggesting that BmdsxE4 may not be a functional event, and it may be a splicing error.

Some female isoforms are also found in some tissues of males, which are consistent with the results described [30,31]. Further verification of the presence of multiple splice forms of Bmdsx transcripts was obtained by performing 3’ RACE using primers P3 and P4 that are specific to the first common exon, exon 1 (Figure 2B). A total of twelve novel, alternatively-spliced and five novel, putatively trans-spliced forms were recovered, cloned and analyzed (Figure 3). The identification of splice forms BmdsxE1 and BmdsxM4 provided secondary confirmation for the existence and use of exon 2n [31]. There were three splice forms found in males and females that included exon 2n, Bmdsx1, Bmdsx2, and Bmdsx9 (Figure 3A). Splice forms Bmdsx2 and Bmdsx9 have 89-bp and 31-bp stretches at the 3’ end of exon 2n followed by a polyA tail, respectively. Bmdsx1 and Bmdsx9 are also found in both males and females. Bmdsx4 is also a nonsense splice form that includes only exon 2 while Bmdsx9 includes exon 3 to which 138 bp were added at the 3’ end. There were two splice forms in which exon 5, an exon common to all previously described transcripts in males and females, had been skipped, Bmdsx9 and Bmdsx9 which is also missing exons 2, 3 and 4. Splice form Bmdsx2 was identical to Bmdsx1 except exon 6 was replaced with novel exon 6n in Bmdsx9. Bmdsx6 was also missing exon 6 and was polyadenylated at the 5’ end of exon 5, a situation also found in BmdsxM4. Bmdsx5 is identical to Bmdsx7 except that splicing of exon 4 occurred 133-bp upstream of the usual exon 4 splice acceptor site. These extra nucleotides contained a potential binding site of Sex-essential (Sx), TTATTTTTTTATTTTTTTTTTTTTTTTTT. In Drosophila, Sx functions as the master regulatory gene regulating the alternative splicing of transformer (tra), which in turn promotes the splicing of exon 4 of Dmdsx transcripts in females [1,9,14]. It was reported that these negative splicing regulators BmPSI, BmHrp28 and BmIIMP specifically bound to the CE1 element at the 5’ end of exon 4 [37-39]. However, these three inhibitors might not be the master regulator upstream of Dmdsx in the sex-determining cascade in B. mori. The B. mori homolog of DmSxl, BmSxl, was identified previously [17] and while BmSxl may not serve the master regulator in B. mori as in other non-Drosophilid species. We cannot eliminate the possibility that BmSxl functions as a co-regulator of Bmdsx.

Trans-splicing is another method by which primary transcripts are processed in B. mori and D. melanogaster. Trans-splicing is involved in transcript processing of mod (mdg4) in B. mori and mod (mdg4) and Iola in D. melanogaster [40,41]. However, it is important to be aware of the possibility that some putative trans-splicing products may be experimental artifacts. Although eleven putative trans-spliced variants of Bmdsx have been cloned by 3’ RACE, including six trans-spliced variants reported previously (Figure 3B) [31], it is not clear at the moment whether these splice forms are the result of natural splicing events or are possible PCR artifacts arising from strand-switching.

Novel exons and revised genomic organization of Bmdsx

Previous analyses of Bmdsx showed that it is composed of six exons and five introns, including an alternative 5’ splice site
in intron 5 resulting in an additional 15-bp at the 3’ end of exon 3. In our current understanding of the genomic organization of \textit{Bmdsx}, the 3’ end of exon 6 is undetermined [19,30]. This classic understanding of the structure of \textit{Bmdsx} can now be substantially revised in three ways in light of our findings (Figure 4A). First, the presence of novel exons 2n, 3n, and 6n has been confirmed by the successful identification of new splice forms containing these novel exons. Second, there is a

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Unexpected expression of \textit{Bmdsx} in tissues from males and females. (A) Previously-reported structure of \textit{Bmdsx}. Horizontal arrow indicates the position of the primers used for expression analysis of \textit{Bmdsx}. Vertical arrows show the positions of stop codons. The black box is the stretch of 15 bp in exon 3. Exons 1, 2, 5 and 6 are common to both sexes, and exons 3 and 4 are female-specific. (B) Expression analysis of \textit{Bmdsx} in tissues from males and females. Arrowheads shows the positions of the splice forms previously-reported. \textit{Bmactin3} was used as the internal control for message abundance. (C) The relative abundance (%) of each amplification band corresponding to the tissues of females. (D) The relative abundance (%) of each amplification band corresponding to the tissues of males. doi: 10.1371/journal.pone.0079703.g001}
\end{figure}
138-bp stretch at the 3' end of exon 2 and an alternative splice site that add 133 bp to the 5' end of exon 4. Third, exon 6 has been confirmed to be 525bp. Based on these observations we now know that Bmdsx consists of nine exons with eight introns that vary markedly in length from 287 to 55569 bp (Table 2). All of the splice sites in Bmdsx abided by the GT-AG rule [42] and an analysis of the 5' splicing sites of Bmdsx introns revealed that only those associated with introns 1 and 6 conformed to the consensus sequence GTRAGY [43]. In addition to minor deviations at the 5' splice sites of introns 4, 5 and 7, major deviations from the consensus GTRAGY were observed at the 5' splice sites of introns 2 and 3. Analysis of the 3' splice sites revealed that the length of the pyrimidine stretches associated with the 3' splice sites did not deviate significantly from the consensus number of 8.49±1.57 [19], and most of the 3' splice sites were canonical (Figure 4C, Table 2). However, the pyrimidine stretch preceding novel exon 6n deviated markedly from the consensus number (6/12), indicating that the 3' end of intron 7 contained a weak splice site, and its splicing might require a positive splicing regulator binding to exon 6n (Figure 4C, Table 2).

The revised structure of Bmdsx and its comparison to Aadsx, the dsx gene from Antheraea assama basal member of the superfamily Bombycoidea, suggests that the ancestral dsx gene gained and lost exons as well as gaining an intron. The Aadsx gene contains seven exons, including a 15-bp stretch at the 3' end of exon 2, and undergoes alternative splicing to produce seven transcripts (Figure 4B) [29]. The first exon of Aadsx is homologous to Bmdsx exons 1 and 2, suggesting that the ancestral form of dsx had an A. assama-type first exon that acquired an intron and exon 2n in the lineage leading to B. mori. The corresponding homologous relationship of the second, fourth and sixth exons of Aadsx to Bmdsx exons 3, 4 and 5 indicated that Aadsx third and fifth exons were lost, whereas the novel exons 3n and 6n were acquired in B. mori lineage.

**Tissue and sex-specific expression**

Exon-specific primers were used to test for the presence of transcripts that included specific exons in the head, integument, hemocytes, midgut, fat body, Malpighian tubules, trachea, silk glands and gonads of both sexes (Figure 5A). Exon 2n was known to be expressed in both males and females [31] and it was not included in this analysis. Exon 3n was alternatively-spliced in a female-specific manner and transcripts containing this exon were readily detected. A low level of transcripts containing exon 3n was also detected in male gonads (Figure 5B). An identical pattern of alternative splicing and expression were seen for transcripts containing exons 3 and 4 as described below (Figure 5B), suggesting that the sex-specific splicing of exon 3n is controlled by an unknown regulatory factor upstream to Bmdsx in the sex-determination cascade. In D. melanogaster, exon 4 of Dmdsx contains a TRA/TRA2 binding site (dsxRE/PRE); no such sequences have been detected in exon 3n of Bmdsx. Because exon 3n contains a canonical 5' splice site, it suggests that a putative negative regulator was recruited to control splicing of exon 3n similar to exons 3 and 4. Exon 3 with the 15 bp 3' stretch (exon 3s), exon...
4 with the 133 bp 5' stretch (exon 4s) and exon 3/4 arising from the splicing of exons 3 and 4 were all expressed in the female soma and gonads, as well as at low levels in the male gonads (Figure 5B) [19,30]. It is interesting to speculate that the gain or loss of the 133-bp stretch at the 5' end of exon 4 might be related to the recruitment of a Sxl-like factor since a Sxl-like binding site was found here. Although Sxl might not serve a master regulator in B. mori [10,11,17], considering the presence of a poly(U) tract, the RNA binding and splicing regulation of tra by Sxl protein and the autoregulation of Sxl

Figure 3. Novel alternatively-spliced and trans-spliced forms of Bmdsx. (A) Novel alternatively-spliced forms of Bmdsx. Boxes are Bmdsx exons, and the “V” lines indicate introns. Black and brown boxes are a stretch of 15 bp in exon 3 and novel exons, respectively. Grey and red triangles show the position of previously reported and newly-discovered stop codons, respectively. Hollow triangle shows the position of the newly-discovered stop codons in novel splice forms expected to be degraded by NMD. Black asterisk indicates the nonsense splice forms expected to be degraded by nonstop-mediated decay (NSD). (B) Novel putative trans-spliced variants of Bmdsx. Underlined splice forms indicate putative trans-spliced variants previously reported. Dashed “V” lines represent the potential trans-splicing events between Bmdsx transcripts and transcripts of other genes. The boxes downstream of the dashed “V” lines are the exons of Bmdsx-related genes, Bmdsr. Grey triangles indicate the position of reported stop codons. Green asterisks indicate the putative trans-spliced events, which occurs upstream of the termination codons.

doi: 10.1371/journal.pone.0079703.g003
itself [44], we suspect that the choice of the 3’ splice site of intron 5 also involves a Sxl-like factor.

Exon 6n, as described earlier, is part of the 3’ UTR of Bmdsx(Figure 3A). Our expression analysis of transcripts containing exon 6n showed that they were found almost exclusively in male tissue (Figure 5B). Exon 6n-containing transcripts were also detected in the gonads of females but the size of the 6n-containing transcripts were significantly larger than those detected in males (Figure 5B). Exon 6n-containing transcripts were isolated, cloned and sequenced, resulting in the discovery of three novel splice forms including one female (Bmdsx_F10) and two male (Bmdsx_M6 and Bmdsx_M7) splice forms (Figure 5C). Bmdsx_F9 was not detected under those conditions and we suspect that the sensitivity of our RT-PCR was such that if this transcript was present in only a few cells at a low level, we would have failed to detect it. In Drosophila, alternative splicing of msl-2 5’UTR is affected by the master regulator Sxl. Although alternative splicing does not affect the msl-2 ORF, it represses the translation of the mRNA leading to the absence of MSL-2 in males [45]. The protein coding capacity of transcripts containing exon 6n is not changed and consequently we speculate that its use may be not only regulated by some key upstream gene in the sex-determining cascade, but also its presence likely modulates subsequent downstream information transfer in the sex-determination cascade.

Therefore, in summary, with the exception of Bmdsx_F10, all Bmdsx transcripts are primarily female-specific in the soma and gonads, with low levels of expression in the male gonads, which does not affect testis development. Bmdsx_F10 is specifically expressed in the female gonads. Additionally, Bmdsx_M6 is specifically expressed in male trachea, while the remaining six Bmdsx_Ms are specifically expressed in males. These expression patterns may be related to the sex-specific and tissue-specific functions of these different transcripts.

**BmDsx protein**

When compared to dsx reported in other species, the situation in *B. mori* is complicated. Our observations beg the question as to the function of these various alternatively-spliced

---

**Figure 4.** Classic and revised gene models and expression patterns of Bmdsx. (A) Classic and second revised gene model of Bmdsx. Boxes represent Bmdsx exons. Black boxes strand for astretch of 15 bp in exon 3, and brown boxes represent novel exons or extended regions of exons. 6n indicates a novel exon upstream of exon 6. Dashed vertical arrows show the position of stop codons. “V” lines indicate splicing events between two exons. The vertical lines within exons indicate the boundaries between different splice junctions leading to different splice forms. (B) Classic gene model of the dsx gene of A. assama, Aadsx. The black box shows the position of a 15-bp stretch of nucleotides in exon 2. Vertical arrows indicate the position of stop codons. (C) Comparative genomic structure of the A. assama and B. mori dsx genes. Percentages above and below represent the identities of the corresponding nucleic and amino acid sequences between Aadsx and Bmdsx, respectively. Asterisks and dot indicate the canonical and weak splicing acceptor sites, respectively. Brackets indicate the female-specific exons.

doi: 10.1371/journal.pone.0079703.g004
transcripts. To begin to address that question we have considered the potential of the transcripts isolated and analyzed here to produce proteins that might function in sexual development. Generally, a Dsx protein with regulatory function on downstream targets must contain several characteristics such as a common N-terminal region in both sexes, a sex-specific C-terminal region, a conserved DNA binding domain (OD1 domain) containing two overlapping binding sites with conserved cysteine and histidine residues, a conserved oligomerization domain (OD2 domain), the N-terminal region of the OD2 domain common to both sexes and containing some important nonpolar amino acids, the C-terminal region of the OD2 domain located in the female-specific C-terminal region, and a glycine located in the first position of the female-specific C-terminal region.

Of the ten female-specific Bmdsx transcripts, BmdsxF4 is predicted to result in a truncated nonfunctional protein, BmDsxF4, because of a premature termination codon (PTC) present in exon 3n (Figure 2A and 6, Table 3). Because PTC-containing mRNA can be targeted for degradation by nonsense-mediated mRNA decay (NMD) to avoid producing truncated proteins with potentially deleterious functions [46], we suspect that BmdsxF4 will be degraded by NMD during the first few rounds of translation. The acquisition of a PTC by exon 3n in females may be a method of regulating BmDsxF translation. The remaining nine female-specific Bmdsx transcripts are expected to result in two functional (BmDsxF1 and BmDsxF3) [31,32] and two potentially functional (BmDsxF2 and BmDsxF5) Dsx proteins, which have three different C-termini (Figure 6, Table 3). The C-terminus of Dsx protein generally serves a regulatory domain, and therefore different C-termini possibly cause BmDsxF proteins to have different effects on the control of their downstream target genes involved in female sexual development. BmDsxF1 is encoded by four transcripts, which use a common 5’ end but differ at their 3’ ends by the use of alternative polyadenylation sites. BmdsxF5 contains the 133-bp stretch at the 5’ end of exon 4 which contains a termination codon resulting in the production of a novel female-specific C-terminus of 38 aa (Figure 3A and 6), which possibly causes BmDsxF5 to have a different effect on the control of downstream target genes related to female sexual development.

Of the seven male-specific Bmdsx transcripts, two, BmdsxM2 and BmdsxM5, might result in truncated non-functional proteins. In the transcript of BmdsxM2, exon 5 is skipped (Figure 2A), resulting in a putative protein (BmDsxFM2) without the male-specific C-terminus (Figure 6, Table 3). Translation of BmdsxM5 could result in a protein (BmDsxFM4) without the male-specific C-terminus as well as the OD2 domain (Figure 6, Table 3). Consequently, we speculate based on the structure of these transcripts that BmdsxF2 and BmdsxM5 will undergo nonsense-mediated degradation, avoiding the production of truncated proteins. Transcripts BmdsxM7, BmdsxM4 and BmdsxM7 only differ at their 3’ ends (Figure 2A, 3A and 5C). We suspect that BmdsxF4 will be degraded shortly after splicing because of the very short distance between the stop codon (TAG) and the

---

**Table 2.** Exon/intron boundaries of the Bmdsx gene.

| Exon | Size (bp) | 5’ splice site | Intron | Size (bp) | 3’ splice site | N° of Y |
|------|-----------|----------------|--------|-----------|---------------|---------|
| 1    | 67(SUTR)+602 | CG/GT AAGT | 1      | 31300     | TTTTATGATTG A TAG/A | 8       |
| 2n+a | 81        | AG/GT ATTA | 2      | 45831     | TAATTTATGTT A CAG/C | 8       |
| 112  |           |             |        |           |               |         |
| 170  |           |             |        |           |               |         |
| 2n   | 144       | AG/GT ACCG | 3      | 2761      | TAATCTGGATT T CAG/G | 7       |
| 136  |           |             |        |           |               |         |
| 282  |           |             |        |           |               |         |
| 3n   | 108       | AG/GT AAGA | 4      | 287       | AGATCTTGTCC G CAG/G | 7       |
| 3n   | 82        | CA/GT ACCG | 5b     | 4077      | AATTCCACTTT A CAG/T | 9       |
| 3n   |           |             | 2     | 4204      | ATTTTCTTTTT G TAG/G | 11      |
| 4n   | 97        | TA/GT AATA | 7      | 4169      |               |         |
| 169  |           |             | 4d    | 55589     | GTCCGACCCTT G CAG/G | 7       |
| 296  |           |             | 5b    | 218       | TTATGCTGAAAAC A AAC/G | 6       |
| 199  |           |             | 6     | 17        |               |         |
| 6n   | 112       | AG/GT AAAC | 7c    | 27810     | TTTTGCTGAAAC A AAC/G | 6       |
| 5   | 525       |               |       |           |               |         |
| consensus | YYYYYYYYYYYY NY AG/G | 8.49±1.57 |       |           |               |         |

/a Diagonal line indicates the boundary between exon and intron. Left sequence, exon; Right sequence, intron.
/b Exons alternatively-spliced at different 5’ splice sites.
/c Introns alternatively-spliced at different 3’ splice or 5’ splice sites.
/d Intron containing weak 3’ splice site.

N° of Y: Average number of pyrimidines from 5 to 17 nucleotides upstream of 3’ splice site in B. mori [19].
poly(A) tail as well as a lack of a polyadenylation signal (AAUAAA). The origin of transcript BmdsxM4 is not clear though it appears not to have arisen from an internal priming event, due to the absence of a short-A stretch following the polyadenylation site, which eliminates nonspecific hybridization of oligo-d(T) to the short-A stretch as playing a role in the genesis of this transcript. The remaining male-specific splice forms are predicted to result in functional (BmDsxM1) [47] and potentially functional (BmDsxM3) Dsx proteins (Figure 6, Table 3). BmDsxM3 is different from BmDsxM1 by a 27-aa insertion between OD1 and OD2 domains, which may affect the binding affinity of BmDsxMs to downstream target sites.

Five splice forms were cloned from males and females, however, only one, BmdsxM5, results in a potentially functional protein (BmDsx3). Despite the presence of a short-A stretch following the polyadenylation site, a consensus AAUAAA at 22 nucleotides upstream of the poly(A) tail indicates that BmdsxM5 probably results from the genuine use of a polyadenylation site. Its validity needs to be tested further. Two of the remaining splice forms, BmdsxM1 and BmdsxM4, are expected to be degraded by nonstop-mediated mRNA decay (NSD) due to the absence of a translation termination codon [48]. These transcripts may have arisen as a result of internal priming because of the presence of a short-A stretch following the
right polyadenylation site, providing a nonspecific annealing site for oligo-d(T). Putative protein BmDsx2 encoded by splice forms

\[ Bmdsx_2 \]
and

\[ Bmdsx_3 \]
lacks an OD2 domain and C-terminus (Figure 6, Table 3), and these splice forms may result from internal priming due to the presence of short-A stretches following incorrect polyadenylation sites. Consequently, we speculate that the protein BmDsx2 will not be synthesized.

Therefore, in summary, besides the confirmed proteins BmDsxF1, BmDsxF3 and BmDsxM1 [31,32,47], four candidate Dsx proteins (Female-specific BmDsxF2 and BmDsxF5; Male-specific BmDsxM3; Male- and female-expressed BmDsx3) are expected from Bmdsx that have the potential to regulate the expression of downstream sexual development-related targets. In some cases, multiple splice forms differed only in their 3' UTR and there is the potential that these differences, while having no effect on the protein produced, could affect translation efficiency.

Each of the six sex-specific BmDsx proteins (BmDsxF1, BmDsxF2, BmDsxF3, BmDsxF5, BmDsxM1, BmDsxM3) could be divided into three regions (Figure 6). Two were found in proteins from both males and females and the third was sex-specific. The amino terminus was common to all proteins and

Figure 6. An alignment using ClustalX of the putative BmDsx proteins encoded by all the alternatively-spliced forms of Bmdsx. The sequence is divided into (A) the region common to both sexes, (B) the female-specific region, (C) the male-specific region and (D) the C terminal region common to females and males. Overline indicates the positions of the DNA-binding domain (OD1 domain) and the oligomerization domain (OD2 domain). Asterisks and triangles indicate the position of key residues in the OD1 and OD2 domains, respectively. The red triangle indicates the position of a conserved glycine, the first amino acid of female-specific C-terminal region. Site1(Red) and site2(Blue) indicate the positon of the two overlapping binding sites aligned with cysteines and histidines that coordinate Zn²⁺. Black arrow indicate the candidate Dsx proteins that have regulatory function on downstream targets, and grey arrow show the Dsx proteins without regulatory function.

doi: 10.1371/journal.pone.0079703.g006
was either 215 or 242 amino acids in length containing the DNA binding domains OD1 and OD2. The sex-specific regions included the C-terminus which in females could be 31 aa, 38 aa or 49 aa in length while the male-specific C-terminal region was 64 aa. Comparison of the BmDsx proteins with the reportedly sex-specific Dsx proteins of \textit{D. melanogaster}, \textit{A. assama} and \textit{A. mylitta} showed that OD1 and OD2 domains and the sex-specific C-termini were very conserved in the three Lepidopterans but some what diverged from \textit{Drosophila} (Figure 7). The 27 aa arising from exon 2n was unique to \textit{B. mori}. It was common for the three Lepidopterans to have the two female-specific C-termini of 31 aa and 49 aa while the female-specific C-terminus of 38 aa derived from the 133-bp stretch at the 5' end of exon 4 was unique to \textit{B. mori}.

In conclusion, the \textit{dsx} gene in the domesticated silkworm displays characteristics that distinguish it from the other insects in which it has been studied. Seventeen different alternatively-spliced transcripts and eleven putative trans-spliced transcripts, potentially encoding for four female-specific, two male-specific and one shared functional Dsx protein. Understanding the roles and functions of these proteins in the sexual differentiation of \textit{B. mori} is expected to enlighten our understanding of sexual differentiation in other organisms and the data reported here suggest that there is much more to be learned, reminiscent of \textit{Sxl} and \textit{dsx} in the model insect \textit{D. melanogaster}.

| Candidate protein | Splice form |
|-------------------|-------------|
| BmDsxF1\textsuperscript{a} | Bmdsx\textsuperscript{F1}, Bmdsx\textsuperscript{F6}, Bmdsx\textsuperscript{F8}, Bmdsx\textsuperscript{F10} |
| BmDsxF2\textsuperscript{a} | Bmdsx\textsuperscript{F2}, Bmdsx\textsuperscript{F5} |
| BmDsxF3\textsuperscript{a} | Bmdsx\textsuperscript{F3}, Bmdsx\textsuperscript{F9} |
| BmDsxF4 | Bmdsx\textsuperscript{F4} |
| BmDsxF5\textsuperscript{a} | Bmdsx\textsuperscript{F7} |
| BmDsxF6\textsuperscript{b} | Bmdsx\textsuperscript{M1}, Bmdsx\textsuperscript{M4}, Bmdsx\textsuperscript{M7} |
| BmDsxF7 | Bmdsx\textsuperscript{M2} |
| BmDsxF8\textsuperscript{b} | Bmdsx\textsuperscript{M3}, Bmdsx\textsuperscript{M5} |
| BmDsxF9 | Bmdsx\textsuperscript{M6} |
| BmDsx1 | Bmdsx\textsuperscript{1} |
| BmDsx2 | Bmdsx\textsuperscript{2}, Bmdsx\textsuperscript{3} |
| BmDsx3\textsuperscript{c} | Bmdsx\textsuperscript{5} |
| BmDsx4 | Bmdsx\textsuperscript{4} |

\textsuperscript{a} Female Dsx protein having regulatory function on downstream targets as described [31,32].
\textsuperscript{b} Female Dsx protein having potentially regulatory function on downstream targets.
\textsuperscript{c} Male Dsx protein having regulatory function on downstream targets as described [47].
\textsuperscript{d} Male Dsx protein having potentially regulatory function on downstream targets.
\textsuperscript{e} Common Dsx protein having potentially regulatory function on downstream targets.
\textsuperscript{f} Transcripts possibly resulting from internal priming or some unknown error.

doi: 10.1371/journal.pone.0079703.t003
Figure 7. An alignment using ClustalX of Dsx proteins of *B. mori* (*Bm*), *A. mylitta* (*Amy*), *A. assama* (*Aa*) and *D. melanogaster* (*Dm*). Dashed line above the sequences indicates the positions of the conserved DNA-binding domain (OD1 domain) and oligomerization domain (OD2 domain). The rectangular box shows the position of 27 aa arising from exon 2n. The brackets indicate proteins with the same sex-specific C-termini.

doi: 10.1371/journal.pone.0079703.g007
Acknowledgements

The authors wish to acknowledge Dr. D.J. Cheng, Dr. T.C. Cheng, Dr. G.H. Wang, Dr. S.P. Liu, Dr. H.W. He, Dr. Y. Ling, Dr. C. Liu and Dr. Fei Wang for their comments on the manuscript. They also want to thank the editor and reviewers for their constructive suggestions and comments.

References

1. Sánchez L (2008) Sex-determining mechanisms in insects. Int J Dev Biol 52: 837-856. doi:10.1387/ijdb.072396sl. PubMed: 18956315.
2. Cline TW (1993) The Drosophila sex determination signal: how do flies count to two? Trends Genet 9: 385-390. doi: 10.1016/0168-9525(93)90138-8. PubMed: 8310555.
3. Wilkhoft U, Franz G (1996) Identification of the sex-determining region of the Ceratitis capitata Y chromosome by deletion mapping. Genetics 144: 737-745. PubMed: 888534.
4. Beye M, Hasselmann F, Fendri MK, Page RE, Omholt SW (2003) The gene card is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell 114: 419-429. doi:10.1016/S0092-8674(03)00606-8. PubMed: 12941271.
5. Verhulst EC, Beukeboom LW, van de Zande L (2010) Maternal control of haplodiploid sex determination in the wasp Nasonia. Science 328: 620-623. doi:10.1126/science.1185805. PubMed: 20431014.
6. Fuji T, Shimada T (2007) Sex determination in the silkworm, Bombyx mori: a female determinant on the W chromosome and the sex-determining gene cascade. Sem Cell Dev Biol 18: 379-388. doi:10.1016/j.scdb.2007.02.008. PubMed: 17440905.
7. Suzuki MG (2010) Sex determination: insights from the silkworm. J Genet 89: 357-363. doi:10.1007/s12041-010-0047-5. PubMed: 20877002.
8. Raymond CS, Shamu CE, Shen MM, Selfert KJ, Hirsch B et al. (1998) Evidence for evolutionary conservation of sex-determining genes. Nature 391: 691-695. doi:10.1038/35618. PubMed: 9490411.
9. Shukla JN, Nagaraju J (2010) Doublesex: a conserved downstream gene controlled by diverse upstream regulators. J Genet 89: 341-356. doi:10.1016/j.jgenet.2010.05.001. PubMed: 20570131.
10. Bopp D, Calhoun G, Horabin JI, Samuels M, Schedl P (1996) Sex-specific control of Sex- lethal is a conserved mechanism for sex determination in the genus Drosophila. Development 122: 971-982. PubMed: 8631274.
11. Penalva LO, Sakamoto H, Navarro-Sabaté A, Sakashita E, Granadino MC, Verhulst EC, van de Zande L, Beukeboom LW (2010) Insect sex determination in the genus Drosophila reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. Genetics 177: 1733-1741. doi:10.1534/genetics.107.078980. PubMed: 17947149.
12. Salvemini M, Mauro U, Lombardo F, Milano A, Zazzaro V et al. (2011) Genomic organization and splicing evolution of the doublesex gene, B. mori regulator of sexual differentiation, in the dengue and yellow fever mosquito Aedes aegypti. BMC Evol Biol 11: 41. doi:10.1186/1471-2148-11-41. PubMed: 21310052.
13. Shukla JN, Nagaraju J (2011) Two female-specific DSX proteins are encoded by the sex-specific transcripts of dsx, and are required for female sexual differentiation in two wild silkworm species, Antheraea assama and Antheraea mylitta. BMC Evol Biol 11: 41. doi:10.1186/1471-2148-11-41. PubMed: 21310052.
14. Suzuki MG (2010) Sex determination in insects: a binary decision based on alternative splicing. Curr Opin Genet Dev 21: 395-400. doi:10.1016/j.gde.2010.05.001. PubMed: 20633649.
15. Salz HK (2011) Sex determination in insects: a binary decision based on alternative splicing. Curr Opin Genet Dev 21: 395-400. doi:10.1016/j.gde.2010.05.001. PubMed: 20633649.
16. O’Neill MT, Belote JM (1992) Interspecific comparison of the transformer gene of Drosophila reveals an unusually high degree of evolutionary divergence. Genetics 131: 113-128. PubMed: 1592233.
17. Gerpke T, Hasselmann M, Schiett M, Hause G, Otte M et al. (2009) Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. PLOS Biol 7: e1000222. PubMed: 19841734.
18. Niimi T, Sahara K, Oshima H, Yasukochi Y, Ikeo K et al. (2006) Molecular cloning and chromosomal localization of the Bombyx Sex-lethal gene. Genome 49: 263-268. doi:10.1139/G05-108. PubMed: 16604109.
19. Samuels ME, Schedl P, Cline TW (1991) The complex set of late transcripts from the Sex-lethal gene encodes multiple related polypeptides. Mol Cell Biol 11: 3584-3602. PubMed: 1710769.
20. Suzuki MG, Ohbayashi F, Mitika K, Shimada T (2001) The mechanism of sex-specific splicing of the doublesex gene is different between Drosophila melanogaster and Bombyx mori. Insect Biochem Mol Biol 31: 1201-1211. doi:10.1016/S0965-1748(01)00067-4. PubMed: 11563953.

Author Contributions

Conceived and designed the experiments: JD QX. Performed the experiments: JD. Analyzed the data: JD QX. Contributed reagents/materials/analysis tools: HG SM LZ BW. Wrote the manuscript: JD QX DO. Provided suggestions about designing the experiments: DO XZ PZ.

Cite this article: Novo...
37. Suzuki MG, Imanishi S, Dohmae N, Nishimura T, Shimada T et al. (2008) Establishment of a novel in vivo sex-specific splicing assay system to identify a trans-acting factor that negatively regulates splicing of Bombyx mori dsx female exons. Mol Cell Biol 28: 333-343. doi: 10.1128/MCB.01528-07. PubMed: 17967886.

38. Wang Z, Zhao M, Li D, Zha X, Xia Q et al. (2009) BmHrp28 is a RNA-binding protein that binds to the female-specific exon 4 of Bombyx mori dsx pre-mRNA. Insect Mol Biol 18: 795-803. doi:10.1111/j.1365-2583.2009.00943.x. PubMed: 19853667.

39. Suzuki MG, Imanishi S, Dohmae N, Asanuma M, Matsumoto S (2010) Identification of a male-specific RNA binding protein that regulates sex-specific splicing of Bmdsx by increasing RNA binding activity of BmPSI. Mol Cell Biol 30: 5776-5786. doi:10.1128/MCB.00444-10. PubMed: 20986652.

40. Shao W, Zhao QY, Wang XY, Xu XY, Tang Q et al. (2012) Alternative splicing and trans-splicing events revealed by analysis of the Bombyx mori transcriptome. RNA 18: 1395-1407. doi:10.1261/rna.029751.111. PubMed: 22627775.

41. McManus CJ, Duff MO, Elliper-Mains J, Graveley BR (2010) Global analysis of trans-splicing in Drosophila. Proc Natl Acad Sci U S A 107: 12975-12979. doi:10.1073/pnas.1007586107. PubMed: 20615941.

42. Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 50: 349-383. doi:10.1146/annurev.bi.50.070181.002025. PubMed: 6791577.

43. Stephens RM, Schneider TD (1992) Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. J Mol Biol 226: 1124-1136. doi: 10.1016/0022-2836(92)900320-J. PubMed: 1474582.

44. Samuels ME, Bopp D, Colvin RA, Rosigno RF, Garcia-Blanco MA et al. (1994) RNA binding by Sxl proteins in vitro and in vivo. Mol Cell Biol 14: 4975-4990. PubMed: 7516475.

45. Penalva LO, Sánchez L (2003) RNA binding protein sex- lethal (Sxl) and control of Drosophila sex determination and dosage compensation. Microbiol Mol Biol Rev 67: 343-359. doi:10.1128/MMBR.67.3.343-359.2003. PubMed: 12966139.

46. Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76: 51-74. doi: 10.1146/annurev.biochem.76.050106.093909. PubMed: 17352659.

47. Suzuki MG, Funaguma S, Kanda T, Tamura T, Shimada T (2005) Role of the male BmDSX protein in the sexual differentiation of Bombyx mori. Evol Dev 7: 58-68. doi:10.1111/j.1525-142X.2005.00707.x. PubMed: 15642090.

48. Fasken MB, Corbett AH (2005) Process or perish: quality control in mRNA biogenesis. Nat Struct Mol Biol 12: 482-488. doi:10.1038/nsmb845. PubMed: 15933735.