

University of Groningen

## A conserved role of the duplicated Masculinizer gene in sex determination of the Mediterranean flour moth, *Ephestia kuehniella*

Visser, Sander; Voleníková, Anna; Nguyen, Petr; Verhulst, Eveline C.; Marec, František

*Published in:*  
 PLoS genetics

*DOI:*  
[10.1371/journal.pgen.1009420](https://doi.org/10.1371/journal.pgen.1009420)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
 Publisher's PDF, also known as Version of record

*Publication date:*  
 2021

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Visser, S., Voleníková, A., Nguyen, P., Verhulst, E. C., & Marec, F. (2021). A conserved role of the duplicated *Masculinizer* gene in sex determination of the Mediterranean flour moth, *Ephestia kuehniella*. *PLoS genetics*, 17(8), [e1009420]. <https://doi.org/10.1371/journal.pgen.1009420>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

## RESEARCH ARTICLE

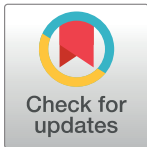
# A conserved role of the duplicated *Masculinizer* gene in sex determination of the Mediterranean flour moth, *Ephestia kuehniella*

Sander Visser<sup>1,2</sup>✉, Anna Voleníková<sup>1,2</sup>✉, Petr Nguyen<sup>1,2</sup>, Eveline C. Verhulst<sup>3</sup>, František Marec<sup>1\*</sup>✉

**1** Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic, **2** Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic, **3** Laboratory of Entomology, Wageningen University & Research, Wageningen, The Netherlands

✉ Current address: Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

\* [marec@entu.cas.cz](mailto:marec@entu.cas.cz)



## OPEN ACCESS

**Citation:** Visser S, Voleníková A, Nguyen P, Verhulst EC, Marec F (2021) A conserved role of the duplicated *Masculinizer* gene in sex determination of the Mediterranean flour moth, *Ephestia kuehniella*. PLoS Genet 17(8): e1009420. <https://doi.org/10.1371/journal.pgen.1009420>

**Editor:** Claude Desplan, New York University, UNITED STATES

**Received:** February 8, 2021

**Accepted:** June 22, 2021

**Published:** August 2, 2021

**Copyright:** © 2021 Visser et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data Availability Statement: The raw Nanopore PromethION long reads from an *Ephestia kuehniella* male generated in this study have been deposited in the Sequence Read Archive under accession number PRJNA683200. All splice variants of the *EkMasc* and *EkMascB* genes and all obtained sequences of the *PiMasc*, *CpMasc*, and *Ekdsx* genes are available at <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:200509> (doi: [10.17026/dans-zpj-ff5m](https://doi.org/10.17026/dans-zpj-ff5m)). All other sequences obtained in this study are available in GenBank under accession numbers

## Abstract

Sex determination in the silkworm, *Bombyx mori*, is based on *Feminizer* (*Fem*), a W-linked *Fem* piRNA that triggers female development in WZ individuals, and the Z-linked *Masculinizer* (*Masc*), which initiates male development and dosage compensation in ZZ individuals. While *Fem* piRNA is missing in a close relative of *B. mori*, *Masc* determines sex in several representatives of distant lepidopteran lineages. We studied the molecular mechanisms of sex determination in the Mediterranean flour moth, *Ephestia kuehniella* (Pyrilidae). We identified an *E. kuehniella* *Masc* ortholog, *EkMasc*, and its paralog resulting from a recent duplication, *EkMascB*. Both genes are located on the Z chromosome and encode a similar *Masc* protein that contains two conserved domains but has lost the conserved double zinc finger domain. We developed PCR-based genetic sexing and demonstrated a peak in the expression of *EkMasc* and *EkMascB* genes only in early male embryos. Simultaneous knock-down experiments of both *EkMasc* and *EkMascB* using RNAi during early embryogenesis led to a shift from male- to female-specific splicing of the *E. kuehniella* *doublesex* gene (*Ekdsx*), their downstream effector, in ZZ embryos and resulted in a strong female-biased sex-ratio. Our results thus confirmed the conserved role of *EkMasc* and/or *EkMascB* in masculinization. We suggest that the C-terminal proline-rich domain, we have identified in all functionally confirmed *Masc* proteins, in conjunction with the masculinizing domain, is important for transcriptional regulation of sex determination in Lepidoptera. The function of the *Masc* double zinc finger domain is still unknown, but appears to have been lost in *E. kuehniella*.

## Author summary

The sex-determining cascade in the silkworm, *Bombyx mori*, differs greatly from those of other insects. In *B. mori*, female development is initiated by *Fem* piRNA expressed from

MW505939–MW505946 and MW774596.

Numerical data including summary statistics for Fig 3, Fig 4, Fig 6, S1 Fig, and S11 Fig are available at <https://doi.org/10.17026/dans-zam-jj4k>.

**Funding:** This research was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 641456 awarded to F. M. and by grant 20-13784S of the Czech Science Foundation (CSF) awarded to F. M. A. V. and P. N. were supported by CSF grant 20-20650Y awarded to P. N. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

the W chromosome during early embryogenesis. *Fem* piRNA silences *Masculinizer* (*Masc*) thereby blocking the male pathway resulting in female development. It is currently unknown whether this cascade is conserved across Lepidoptera. In the Mediterranean flour moth, *Ephestia kuehniella*, we identified an ortholog of *Masc* and discovered its functional duplication on the Z chromosome, which has not yet been found in any other lepidopteran species. We provide two lines of evidence that the *EkMasc* and/or *EkMascB* genes play an essential role in masculinization: (i) they show a peak of expression during early embryogenesis in ZZ but not in WZ embryos and (ii) their simultaneous silencing by RNAi results in female-specific splicing of the *E. kuehniella doublesex* gene (*Ekdsx*) in ZZ embryos and in a female-biased sex ratio. Our results suggest a conserved role of the duplicated *Masc* gene in sex determination of *E. kuehniella*.

## Introduction

Sex determination in insects is under the control of a cascade of genes, each affecting the expression or splicing of the next gene in the pathway [1]. This cascade evolved from the bottom up with the most conserved gene, *doublesex* (*dsx*), at the bottom of the cascade [2] present in all insects studied to date [3–5]. The *dsx* gene is sex-specifically spliced by an upstream splicing factor, e.g. *transformer* (*tra*) in Hymenoptera, Coleoptera, and derived Brachycera (a sub-order of Diptera) [6,7]. In turn, the activity of *tra* is affected by the presence of a primary signal gene which can either activate or inactivate *tra* thereby acting as the initiator of sex determination. The upstream sex determination cascade seems to undergo rapid evolutionary change, as these genes are often replaced, duplicated, and reshuffled, as has been shown in multiple species. For example, in Hymenoptera, the feminizing gene *wasp overruler of masculinization* (*wom*) in the jewel wasp, *Nasonia vitripennis*, is a novel gene originating from a *p53* gene duplication, but has also been very recently duplicated, with both copies of the gene thought to be functional [8]. The gene initiating sexual differentiation in the honeybee *Apis mellifera*, *complementary sex determiner* (*csd*), originated via a duplication of the *feminizer* (*fem*) gene, an ortholog of *tra* [9,10]. In addition, duplications of *tra/fem* have been detected in many other hymenopteran species, though their role in sex determination is unknown [6,11,12]. Also in the housefly *Musca domestica* (Diptera), the masculinizing gene *male determiner* (*Mdmd*) originated through a gene duplication and is located in a locus containing multiple pseudocopies of the gene [13].

The insect order Lepidoptera (moths and butterflies) includes pollinators and several other beneficial species of high economic importance, e.g. the silkworm *Bombyx mori*, but also a large number of major pests of agricultural crops, such as the diamondback moth *Plutella xylostella* [14]. The lepidopteran sex determination cascade has diverged from other holometabolous insects as *tra* was presumably lost in this order [6,15]. The model species in lepidopteran sex determination research is *B. mori*, which has a WZ/ZZ sex chromosome constitution with a dominant feminizing W chromosome (Hasimoto 1933 as cited in [16]). In ZZ individuals, the Z-linked masculinizing gene *Masculinizer* (*Masc*) initiates male development and dosage compensation. In WZ individuals, the W-linked *Feminizer* (*Fem*) piRNA targets the *Masc* mRNA resulting in its degradation [17]. In the absence of *Masc* protein, the *B. mori doublesex* (*Bmdsx*) gene undergoes female-specific splicing resulting in female development. In the absence of the W chromosome, *Masc* is not suppressed by *Fem* piRNA and promotes male-specific splicing of *Bmdsx*, although the pathway through which *Masc* affects splicing of *Bmdsx* is currently unknown. The *Masc* protein contains two zinc finger motifs in the N-terminus

[17], a bipartite nuclear localization signal (bNLS) [18], and a masculinizing region containing two cysteine residues that are essential for the masculinizing activity [19]. An additional segment located in the C-terminus of the protein is essential for masculinization and dosage compensation but the exact functional motif has yet to be determined [19,20]. Masculinization through *Masc* seems to be a widely shared feature of the lepidopteran sex determination mechanism, and has been confirmed in *Trilochoa varians* (Bombycidae) [21], *Ostrinia furnacalis* (Crambidae) [22,23], *Agrotis ipsilon* (Noctuidae) [24], and *P. xylostella* (Plutellidae) [25]. In addition, regulation of dosage compensation has been confirmed in *O. furnacalis* [22] and suggested for *T. varians* [26] and *P. xylostella* [25]. Contrary to masculinization by *Masc*, feminization through *Fem* piRNA is not conserved in *T. varians*, a close relative of *B. mori* [21].

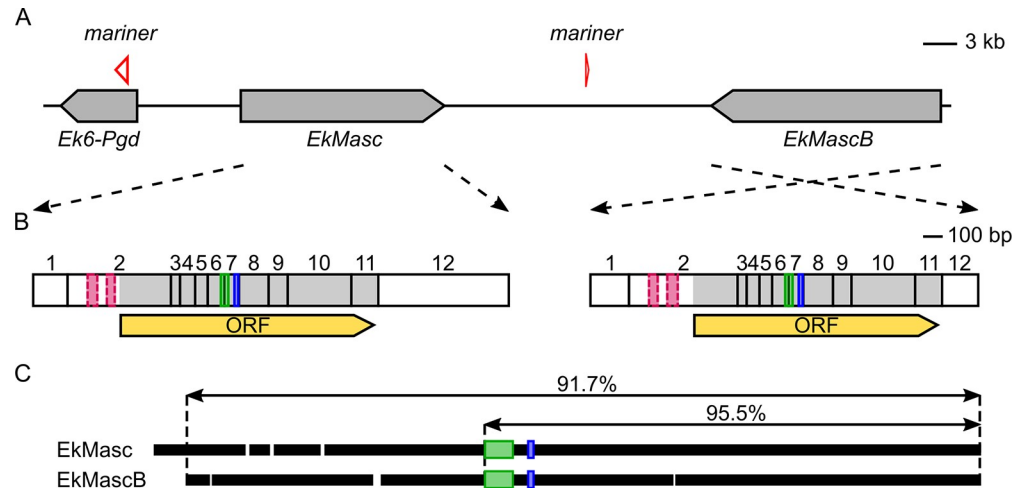
Considering the economic importance of Lepidoptera, extensive research into sex determination mechanisms is still lacking in this large insect order. Therefore, we set out to identify the sex determination mechanism in the Mediterranean flour moth, *Ephestia kuehniella*. This moth was one of early models of genetics in Lepidoptera (see *Anagasta kuehniella* in [27]) and later a model for sex chromosome research [28]. It used to be a serious pest of dried food products, in particular flour, but due to improved hygiene regulations, it is no longer considered a major pest [29]. Nowadays, *E. kuehniella* is widely accepted as a factitious host that is easy to rear in large quantities, and thus its eggs and larvae are used as a food source for a broad range of beneficial arthropods used in biological control [30–32]. Abiotic and biotic factors affecting the growth and development of *E. kuehniella* have been studied extensively [33,34], but egg production could be considerably increased by alteration of the sex ratio of the population. Males are able to sire the complete offspring of at least 9 females [35], and therefore the number of males in the population can be reduced while retaining complete fertilization of the population. Therefore, identification of the sex determining mechanism in *E. kuehniella* may potentially provide candidate target genes to alter the sex ratio, thereby increasing egg production.

Here, we identified a *Masc* ortholog in *E. kuehniella*, *EkMasc*, and its paralog resulting from a recent duplication, *EkMascB*. We developed a PCR-based genetic sexing method in *E. kuehniella* to be able to analyze *EkMasc* and *EkMascB* function using RNAi during early embryogenesis. Our results show that despite the loss of the zinc finger motifs, their role in masculinization is intact. In addition, we characterized a male-specific splice variant of *EkMasc* and *EkMascB*, skipping the exon containing the masculinizing region. Finally, we identified conserved regions in the C-terminus of all functionally confirmed lepidopteran *Masc* proteins that are likely involved in masculinization. Overall, our results substantially add to the understanding of the core genes of the lepidopteran sex determination cascade and its evolution, but also provide target genes for sex ratio optimization for mass rearing purposes.

## Results

### Identification of *EkMasc* and its paralog *EkMascB*

Initial sequences of *Masculinizer* (*EkMasc*) in *E. kuehniella* were identified using reverse transcription PCR (RT-PCR). Full cDNA sequences were subsequently obtained using 5' and 3' rapid amplification of cDNA ends (RACE) followed by sequencing. Two variants of *EkMasc* were consistently found, but we were not sure if they were allelic variants or duplication of the *Masc* gene. A BLASTn search of the two complete cDNA sequences against the draft genome assembly of *E. kuehniella* revealed a single scaffold with both copies of the gene, strongly suggesting that the two sequences obtained are not allelic variants but signatures of a real *EkMasc* duplication, which we denoted *EkMascB*. The open reading frames of *EkMasc* and *EkMascB* are in opposite orientation (tail-tail) with ~23 kb distance between the two genes (Fig 1A). In



**Fig 1. Genomic organization of *EkMasc* and *EkMascB* genes in *Ephestia kuehniella*.** (A) Graphical representation of *EkMasc* and *EkMascB* on the assembled scaffold with an ortholog of the 6-*Pgd* gene (*Ek6-Pgd*) upstream of *EkMasc* and two inactive mariner transposase genes. (B) Representation of complete transcripts of *EkMasc* and *EkMascB* with the exons, open reading frame (ORF; yellow), bipartite nuclear localization signal (bNLS; green), masculinizing domain (blue), and degenerated zinc finger domains in exon 2 (dashed pink) indicated. (C) Alignment of the *EkMasc* and *EkMascB* protein sequences showing the amount of homology between the two proteins, white spaces are indels, green and blue boxes represent the bNLS and masculinizing domains, respectively.

<https://doi.org/10.1371/journal.pgen.1009420.g001>

addition, an ortholog of the 6-phosphogluconate dehydrogenase (*Ek6-Pgd*) gene was identified in the same scaffold in close proximity to *EkMasc*. Within the intron of *Ek6-Pgd* between exons 7 and 8, a complete mariner transposon lacking an intact open reading frame was identified and a similar but partial mariner transposon was identified between the *EkMasc* and *EkMascB* genes, containing only the 3' region of the transposon. The two mariner transposon sequences share 93% sequence identity.

### Localization of *EkMasc* and *EkMascB* on the Z chromosome

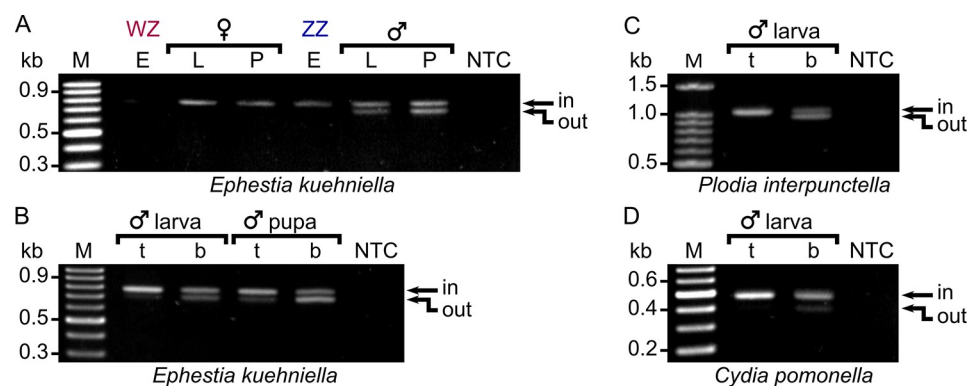
In *B. mori* and *Plutella xylostella*, *Masc* is located on the Z chromosome [17,25]. As with most lepidopterans, the Z chromosome in *E. kuehniella* is present in one copy in females but in two copies in males [36]. Gene content of the Z chromosome between lepidopteran species is generally highly conserved [37–39], but no information regarding the content of the Z chromosome of *E. kuehniella* is available. Therefore, we verified the location of *EkMasc* on the Z chromosome in *E. kuehniella* using quantitative real-time PCR (qPCR) on genomic DNA. We tested two hypotheses regarding the location of the two genes, (i) *EkMasc* and *EkMascB* are Z-linked resulting in a female to male ratio of 0.5 and (ii) both genes are autosomal resulting in a ratio of 1 [37]. Data from qPCR with genomic DNA of *E. kuehniella* females and males targeting both *EkMasc* and *EkMascB* genes at the same time were normalized against the autosomal reference gene, *Acetylcholinesterase 2* (*Ace-2*) (S1 Fig). An unpaired two-tailed *t*-test for unequal variances showed a statistically significant difference in the female and male normalized quantities ( $P = 0.0049$ ), thus ruling out the autosomal hypothesis. In addition, a comparison of the male normalized quantities against the female normalized quantities multiplied by 2 showed no significant difference between ratios ( $P = 0.8415$ ), therefore strongly suggesting that both *EkMasc* and *EkMascB* are located on the Z chromosome. These normalized *EkMasc* + *EkMascB* quantities were compared between females and males showing that the normalized female to male ratio of *EkMasc* + *EkMascB* in *E. kuehniella* was  $0.505 \pm 0.032$ .

Our Southern hybridization data consistently showed that the two signals present in both sexes are stronger in males than in females (S2 Fig), which is in line with the qPCR results and confirms the Z-linkage of *EkMasc* and *EkMascB*. This further corroborates our conclusion that there are two *EkMasc* copies, because the single Z chromosome in females shows two bands in Southern hybridization, ruling out allelic variation.

### Nucleotide and protein comparison of *EkMasc* and *EkMascB*

*EkMasc* and *EkMascB* complete cDNA sequences were used in a BLASTn search against the draft genome to identify and annotate the exons of both genes. Next, we aligned *EkMasc* to *EkMascB* to compare the structure of both genes. All exons of *EkMasc* were also present in *EkMascB*, and both genes contained an open reading frame (S3 Fig). Divergence between the two genes, i.e. single nucleotide polymorphisms (SNPs), insertions and deletions (indels), was strongest in the 5'- and 3'- UTRs and in the intron sequences. We obtained five splice variants for *EkMasc* and seven splice variants for *EkMascB* from sequencing cloned RT-PCR and RACE-PCR products. Six of these splice variants (two for *EkMasc* and four for *EkMascB*) appeared to be present at a very low frequency as they were not visibly amplified and contained early stop codons as a consequence of exon-skipping, intron-retention, or premature polyadenylation. Therefore, we considered these six variants biologically non-relevant.

The remaining six splice variants (three for each gene) were abundantly present (S3 Fig). In both genes, the first and second splice variant, *EkMasc-1* and *EkMasc-2* (and *EkMascB-1* and *EkMascB-2*), contained the bipartite nuclear localization signal (bNLS) [18] and two cysteine residues that are presumably essential for masculinization [19] and differed only in splicing of the terminal (non-coding) exon (Figs 1B and S4). These two splice variants are the only variants detected during early embryogenesis (Fig 2A). The third splice variant, termed masculinizing domain skipping *Masc* (*EkMasc<sup>ms</sup>* and *EkMascB<sup>ms</sup>*) skips the exon containing the masculinizing domain (exon VII), resulting in an early stop codon (S3 and S4 Figs). This splice variant appears later during development and exclusively in males (Fig 2A). Splice forms containing the masculinizing domain appear to predominate in the testes, while *EkMasc<sup>ms</sup>* and *EkMascB<sup>ms</sup>* appear to predominate in somatic tissues (Fig 2B). Due to the importance of the masculinizing domain for *Masc* function, we investigated whether the *Masc<sup>ms</sup>* splice form is



**Fig 2. Alternative splicing of *Masc* in different life stages of *Ephestia kuehniella* and different tissues in *E. kuehniella*, *Plodia interpunctella*, and *Cydia pomonella*.** (A) Alternative splicing of *EkMasc* and *EkMascB* in WZ and ZZ samples of different life stages, embryo 16 hpo (E), larva (L), and pupa (P), where the exon containing the masculinization domain (exon VII) is either spliced in or spliced out. (B) Male testis (t) and whole body samples minus the testis (b) show different splicing patterns in *E. kuehniella* with the testis sample showing predominantly splicing-in of exon VII in both larva and pupa stage, and the body samples showing both splice types. This splice pattern is conserved in males of the closely related *P. interpunctella* (C) and the more distant *C. pomonella* (D).

<https://doi.org/10.1371/journal.pgen.1009420.g002>

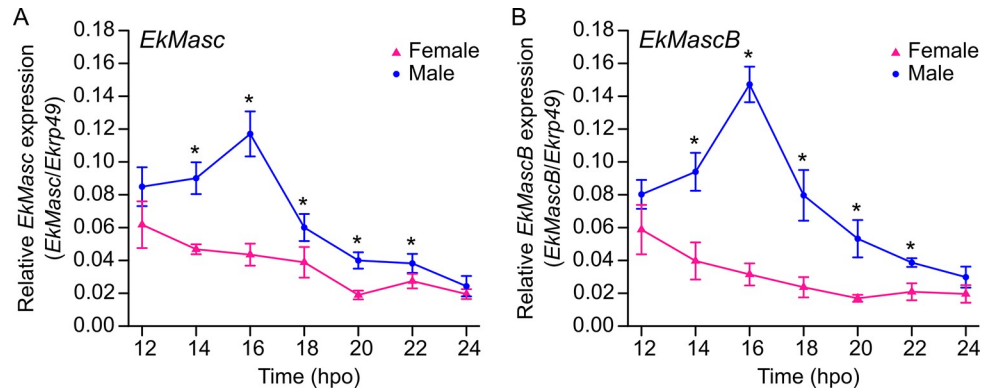
present in other lepidopterans or if it is specific to *E. kuehniella*. The presence and distribution of the *Masc<sup>ms</sup>* splice form was tested in the closely related Indian meal moth, *Plodia interpunctella*, and the distant codling moth, *Cydia pomonella*. In both species, *Masc<sup>ms</sup>* is found in male somatic tissues but not in the testes, and results in a premature stop codon, similar to *E. kuehniella* (Figs 2C, 2D and S4). In *P. interpunctella*, we also found that the exon containing the masculinizing region is shorter than the annotated *PiMasc* sequence, resulting in an exon of similar size to the homologous exon in *EkMasc*.

The complete *EkMasc* mRNA sequence containing the full open reading frame is 2425 bp (*EkMasc-1*) or 2667 bp long (*EkMasc-2*) depending on the splice variation in the 3' terminal exon of the gene and encodes a 435 amino acid protein. Similarly, the complete mRNA sequence of *EkMascB* is 2346 bp (*EkMascB-1*) or 2644 bp (*EkMascB-2*) long and encodes a slightly shorter protein of 418 amino acids. Comparison of *EkMasc* with *EkMascB* shows that they share 91.7% identity over the entire length of the proteins and 95.5% identity in the region likely to be important for masculinization [19], as seen in Fig 1C. *EkMasc* and *EkMascB* differ from each other by 6 indels and 23 substitutions (S5 Fig). Both the male determining domain [19] and the bNLS domain [18] are present in both *EkMasc* and *EkMascB*. In addition, two substitutions are present within the spacer sequence of the bNLS domain, but the domain itself and the masculinizing domain share complete homology (i.e. complete amino acid sequence identity). Interestingly, the two zinc-finger motifs present in the N-terminus of *BmMasc* [17] and most other lepidopteran *Masc* proteins [21,22,24,25] were lost in both *EkMasc* and *EkMascB*. Remnants of the nucleotide sequence of the double zinc finger motif can still be recognized at the nucleotide level based on homology to *PiMasc* and remain part of the transcripts of both genes. However, multiple mutations have occurred in these regions of both genes, and their open reading frames start downstream of the degenerated motifs (Fig 1B). The nucleotide composition of the degenerated zinc finger motifs suggests that the sequences degenerated independently in the two genes.

The *EkMasc<sup>ms</sup>* and *EkMascB<sup>ms</sup>* transcripts have a premature stop codon, but we identified two potential open reading frames, each consisting of approximately half of the *EkMasc-1* and *EkMasc-2* (and *EkMascB-1* and *EkMascB-2*) open reading frames (S4 Fig). These potential proteins consist of 182 amino acids and/or 231 amino acids for *EkMasc<sup>ms</sup>* protein, and 166 amino acids and/or 230 amino acids for *EkMascB<sup>ms</sup>* protein. Even though splicing of *EkMasc<sup>ms</sup>* (and *EkMascB<sup>ms</sup>*) disrupts the standard bNLS domain located at the boundaries of exons 6 and 7, the *EkMasc<sup>ms</sup>* and *EkMascB<sup>ms</sup>* proteins translated from the first segment of the genes do contain a bNLS domain which is now comprised of exon 6 and 8 regions (S4 Fig).

## Functional analysis of *EkMasc* and *EkMascB*

Expression levels of all *EkMasc* and *EkMascB* splice forms were measured during embryogenesis using quantitative reverse transcription PCR (qRT-PCR) to determine sex-specific expression patterns for both genes separately. Embryos were sexed using an in-house developed PCR-based sexing system that shows an additional band in female samples (S6 Fig). Both *EkMasc* and *EkMascB* were expressed throughout embryogenesis and both genes followed the same pattern of expression level differences between the sexes (Fig 3). At 12 hours post oviposition (hpo), no significant difference in expression level of both genes was observed between the sexes. After 12 hpo, expression levels of *EkMasc* and *EkMascB* increased in males until they reached a maximum around 16 hpo, while the expression levels decreased in females simultaneously. After 16 hpo, expression levels decreased in males and reached levels comparable to females at approximately 24 hpo. We performed an unpaired two-tailed *t*-test with unequal

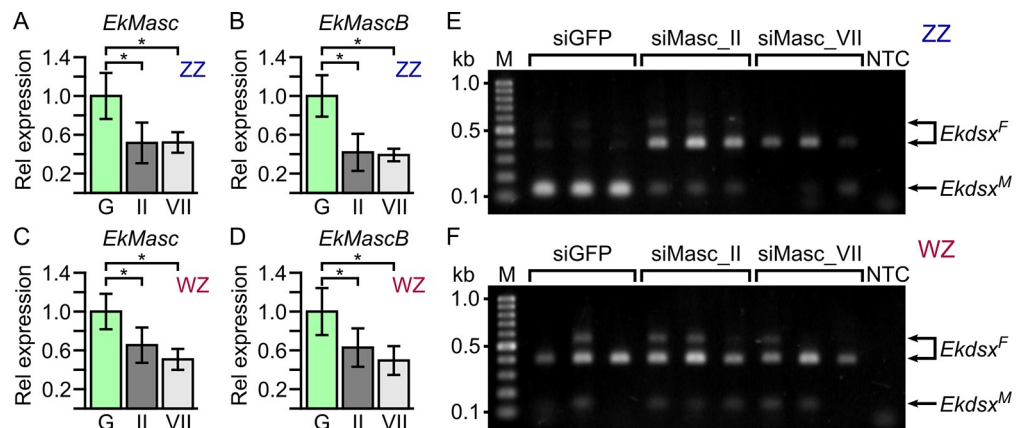


**Fig 3. Expression of *EkMasc* (A) and *EkMascB* (B) during early embryogenesis in female (pink) and male (blue) samples of *Ephestia kuehniella*.** Each time point has at least 3 biological replicates. Significant differences between sexes are indicated (\*; unpaired two-tailed *t*-test;  $P < 0.05$ ) and error bars indicate standard deviation.

<https://doi.org/10.1371/journal.pgen.1009420.g003>

variances to test for differences between the sexes, showing statistically significant differences at 14–22 hpo for both genes (S1 Table).

To verify the function of *EkMasc* and *EkMascB*, we performed a simultaneous knock-down of both genes by injecting short interfering RNAs (siRNAs) targeting a conserved region into eggs 1–2 hpo. We performed two separate knock-down experiments using siRNA targeting either exon II (siMasc\_II) or exon VII (siMasc\_VII), which contains the masculinizing domain (S7 Fig). For the third treatment, we injected a negative control siRNA designed against the *green fluorescent protein* (*GFP*) gene, as previously done [17]. We assessed knock-down efficiency by qRT-PCR at 16 hpo (Fig 4 and S2 Table), the peak of *EkMasc* and *EkMascB* expression levels, as this is presumably the time point at which the expression of *EkMasc* and *EkMascB* is essential for male development. The average expression level of *EkMasc* in ZZ embryos after injection with siMasc\_II and siMasc\_VII was reduced to 51.6% ( $P = 0.0311$ ) and



**Fig 4. RNAi knock-down effects in ZZ and WZ *Ephestia kuehniella* individuals 16 hours post oviposition.** The fold-change in expression levels of *EkMasc* (A and C) and *EkMascB* (B and D) in ZZ (A and B) and WZ (C and D) individuals injected with siMasc\_II (II) or siMasc\_VII (VII) are shown relative to the control individuals injected with siGFP (G). Statistically significant differences are indicated by \* (unpaired one-tailed *t*-test;  $P < 0.05$ ; S2 Table), error bars indicate standard deviation. In E and F, effects of sex-specific splicing of *Ekdsex* 48 hpo after injection with control, siMasc\_II, and siMasc\_VII are shown in ZZ and WZ individuals, respectively. Note that splicing of *Ekdsex* shifts from predominantly male-specific in the control to mostly female-specific in siMasc\_II- and siMasc\_VII-treated ZZ individuals (E), while splicing does not differ between treatments in WZ individuals (F).

<https://doi.org/10.1371/journal.pgen.1009420.g004>



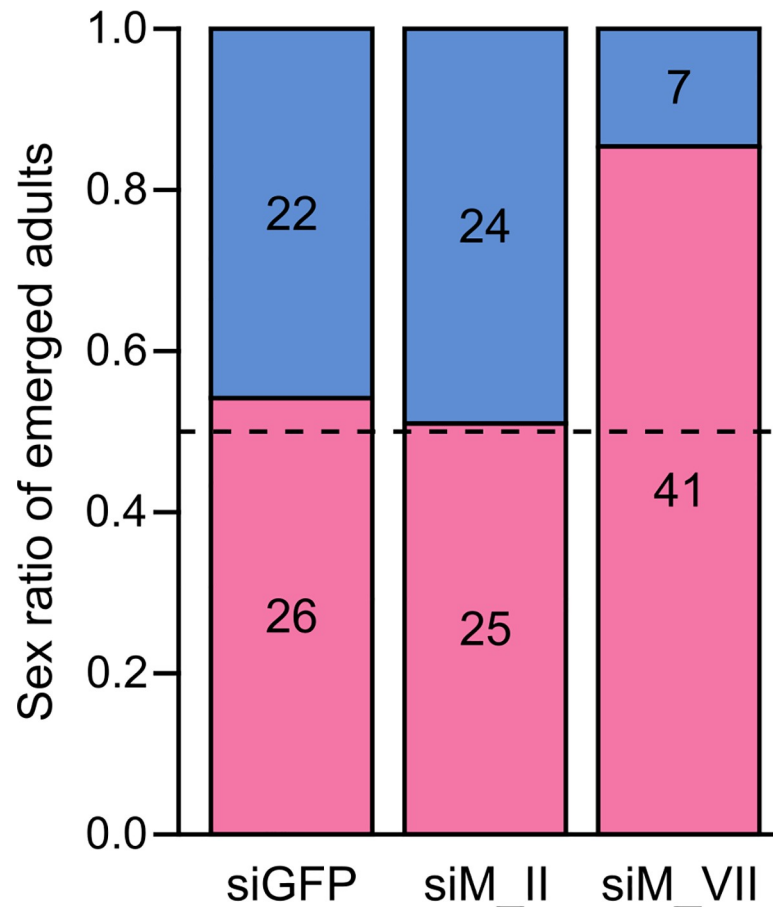
52.0% ( $P = 0.0300$ ), respectively (Fig 4A), and *EkMascB* expression in ZZ embryos was reduced to 41.8% ( $P = 0.0137$ ) and 39.1% ( $P = 0.0164$ ), respectively (Fig 4B). In WZ embryos, expression of *EkMasc* was reduced to 65.4% ( $P = 0.0335$ ) and 50.7% ( $P = 0.0034$ ) after injection with siMasc\_II and siMasc\_VII, respectively, while *EkMascB* showed reduced expression levels of 62.8% ( $P = 0.0416$ ) and 49.5% ( $P = 0.0090$ ), respectively (Fig 4C and 4D).

The *doublesex* (*dsx*) gene is often used as indicator of sexual development in knock-down experiments involving sex determination genes [21,25]. We identified an ortholog of this gene in *E. kuehniella*, *Ekdsx*, and confirmed its sex-specific splicing using RT-PCR in pupae (S8 Fig). A single male-specific splice form was identified in males and two dominant female-specific splice forms were identified in females (S8A and S8B Fig). This is similar to the splicing structure identified in *B. mori* [40,41]. In addition, a third female-specific splice form was observed in females (S8C Fig), but we failed to sequence this third female-specific splice form, and thus its splicing structure is currently unknown. Splicing of *Ekdsx* was female-specific in all early embryos, but transitions to male-specific splicing were observed only in ZZ individuals, 16–18 hpo at 21–22°C (S8D Fig). After injection of siRNA, sex-specific splicing of *Ekdsx* was assessed at 48 hpo to ensure that *Ekdsx* splicing stabilized in either the female- or male-specific isoform(s). WZ individuals injected with any of the three siRNAs (control siGFP, siMasc\_II, and siMasc\_VII) showed predominantly female-specific splicing of *Ekdsx* as expected (Fig 4F). ZZ individuals injected with control siGFP showed predominantly male-specific splicing of *Ekdsx* while those injected with siMasc\_II or siMasc\_VII showed predominantly female-specific splicing of *Ekdsx* (Fig 4E) comparable to WZ individuals. Approximately 400 eggs injected with siGFP were left to develop of which 12% (48/400) hatched and survived to adulthood. Similarly, approximately 10% (49/500) of the eggs injected with siMasc\_II and approximately 8% (48/620) of the eggs injected with siMasc\_VII hatched and survived to adulthood. Adults were phenotypically sexed based on external genitalia, revealing sex-ratios not significantly differing from 0.5 in those injected with siGFP ( $P = 0.5637$ ) and siMasc\_II ( $P = 0.8864$ ) (Fig 5). Adults developing from embryos injected with siMasc\_VII showed a strong female-biased sex-ratio of 0.85, which differed significantly from 0.5 ( $P = 9.226e-07$ ). The heads of females that developed from siMasc\_VII-injected embryos were used to test the genetic sex of the individual, which showed that all phenotypic females have WZ sex chromosomes, suggesting a male-killing effect of RNAi with siMasc\_VII. Adult males that developed from siMasc\_II- and siMasc\_VII-injected embryos showed normal morphology of the external genitalia. The testes of specimens examined cytologically also showed normal morphology and a typical purple-red coloration and contained large numbers of mature sperm (S9 Fig). We found no evidence of ovarian tissue or cells, suggesting that the RNAi-treated male survivors are true, probably fertile males.

## Masculinizer comparison in Lepidoptera

We compared all published and functionally confirmed Masc protein sequences to identify any additional widely shared domains and/or features of the proteins. Previously, three conserved domains were identified in BmMasc: (1) the double zinc finger domain [17], (2) the bNLS domain [18], and (3) the masculinizing domain [19]. The two cysteine residues that define the masculinizing domain and the bNLS domain are present in all species. The double zinc finger motif is conserved in most species with the exceptions of *E. kuehniella*, which lost both motifs in both copies of the gene, and potentially in *Trilocha varians* that has a deletion in the second zinc finger [21].

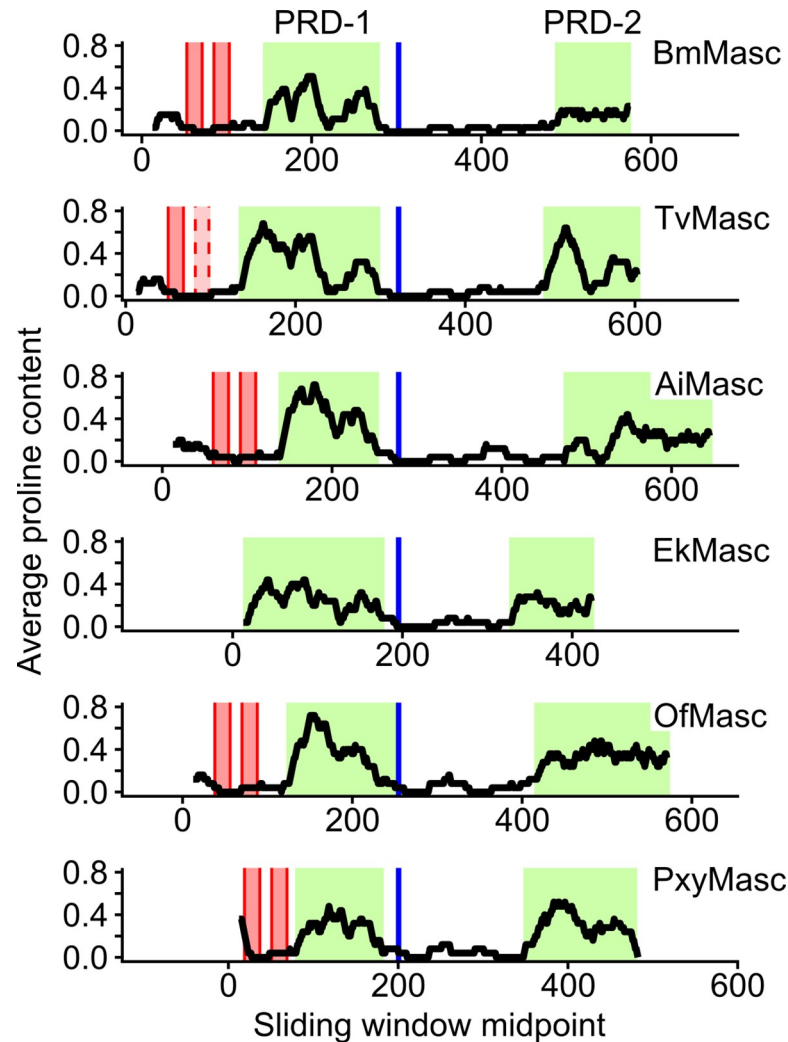
After aligning all functionally confirmed lepidopteran Masc protein sequences, we observed a relatively high level of proline residues in all proteins (S10 Fig). The average proline content



**Fig 5. Sex ratio of emerged *Ephesia kuehniella* adults injected as 1–2 hpo embryos with siGFP, siMasc\_II (siM\_II), and siMasc\_VII (siM\_VII).** The number of adult females (pink) and males (blue) emerged in each treatment are indicated in the figure. The sex ratio (F:M) of emerged adults injected with siGFP or siMasc\_II is approximately 0.5 (indicated by a dashed line), while the sex ratio of siMasc\_VII-injected individuals is strongly female-biased ( $> 0.8$ ).

<https://doi.org/10.1371/journal.pgen.1009420.g005>

of lepidopteran proteins is 5.52% (*B. mori*), 5.86% (*P. xylostella*), and 5.35% (*P. interpunctella*), while the proline content in their respective Masc proteins is 11.56%, 17.37%, and 13.73%. Given the less than 0.4% variation in the proline contribution across the total protein databases between the species, we assume that *E. kuehniella* has similar proline levels across all proteins, whereas EkMasc and EkMascB consist of 16.09% and 17.22% proline residues, respectively. To determine the distribution of the proline residues across the protein sequence, we performed a sliding window analysis with window size 25 on the Masc proteins previously confirmed to have a masculinizing function [17,21,22,24,25] and EkMasc (Fig 6). Due to the high level of identity between the two Masc sequences in *E. kuehniella*, we excluded EkMascB from the analysis, however, a comparison of EkMasc and EkMascB is shown in S11 Fig. The analysis identified two regions with an increased proline content in all species. The first proline-rich domain (PRD-1) is localized between the second zinc finger domain and the masculinizing domain, and the second proline-rich domain (PRD-2) at the C-terminus of the proteins (Fig 6). The *BmMasc* mutant of *B. mori*, which lacked the first 294 amino acids of the protein, was able to masculinize BmN-4 cells, while the same mutant additionally missing the final 75 amino acids failed to do so [19]. As indicated in Fig 7A, this region of the protein corresponds



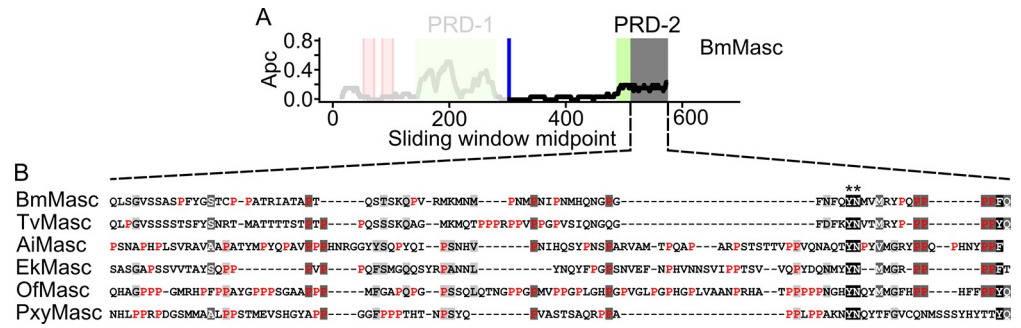
**Fig 6. Sliding window analysis of proline distribution in functionally confirmed Masc proteins of six lepidopteran species.** Sliding window analysis revealed two regions in all Masc proteins with increased proline content, proline-rich domain 1 (PRD-1) and proline-rich domain 2 (PRD-2) (shaded green). Graphs are aligned by the masculinizing domain in TvMasc (blue). Also indicated are the two zinc finger domains (shaded red with solid outline), the second zinc finger domain in TvMasc has a dashed outline due to a deletion potentially resulting in the loss of the domain. Species used for this analysis: *Bombyx mori* (Bm), *Trilocha varians* (Tv), *Agrotis ipsilon* (Ai), *Ephestia kuehniella* (Ek), *Ostrinia furnacalis* (Of), and *Plutella xylostella* (Pxy).

<https://doi.org/10.1371/journal.pgen.1009420.g006>

to PRD-2. In addition to PRD-2, we also identified two consecutive tyrosine-asparagine (YN) amino acid residues in this region that are conserved in all six species (Fig 7B). Only in *P. interpunctella* Masc, this tyrosine-asparagine motif is missing, but it should be noted that this region of the *PiMasc* transcript has not yet been confirmed by RT-PCR and the *PiMasc* protein has not been functionally confirmed.

## Discussion

Duplications of sex determining genes have been described in several insect species. In particular, duplications of the *transformer* (*tra*) gene, the main splicing regulator of *dsx*, are common in the order Hymenoptera [6,10–12]. In Lepidoptera, only a few species have been investigated, but apart from the primary signal gene in the silkworm, i.e. *Feminizer* (*Fem*), which



**Fig 7. Analysis of the terminal region essential for BmMasc function as identified in Katsuma et al. (2015) [19].** In addition to the masculinizing domain (blue), the shaded grey region in (A) shows a region essential for Masc protein function, coinciding with the proline-rich domain 2 (PRD-2) (green). This region of BmMasc was aligned to the functionally confirmed Masc protein sequences in other Lepidoptera (B). Apart from a tyrosine-asparagine motif (\*\*), which is conserved in all species, the only conserved feature of this region is the relatively high proline (red) content. Apc, average proline content. Species used for this analysis: *Bombyx mori* (Bm), *Trilocho varians* (Tv), *Agrotis ipsilon* (Ai), *Ephestia kuehniella* (Ek), *Ostrinia furnacalis* (Of), and *Plutella xylostella* (Pxy).

<https://doi.org/10.1371/journal.pgen.1009420.g007>

consists of a high copy tandem repeat generating *Fem* piRNA [17], no duplications of lepidopteran sex determination genes have been reported. Here, we report a duplication of the male-determining gene *Masculinizer* (*Masc*) in *E. kuehniella*. Both the *EkMasc* and *EkMascB* genes lack the tandem zinc finger motif, but do contain the masculinizing domain. We developed a PCR-based genetic sexing method in *E. kuehniella* and analyzed the effects of simultaneous knock-down of these genes on sexual development. Functional analysis showed that the simultaneous knock-down of *EkMasc* and *EkMascB* results in female-specific splicing of *Ekdsx* and lethality in males.

The masculinizing function of *Masc* was initially identified in *B. mori* [17] and is conserved in *Trilocho varians*, *Ostrinia furnacalis*, *Agrotis ipsilon*, and *Plutella xylostella* [21,22,24,25]. Similar to the other species, knock-down of *EkMasc* and *EkMascB* in *E. kuehniella* results in female-specific splicing of *Ekdsx*. The similarity in expression patterns and the high level of protein similarity between *EkMasc* and *EkMascB* show no clear indications that the two proteins have diverged in their function in male determination. In particular, the high level of conservation between the proteins in the male-determining domain and the C-terminus suggests a conserved function in sex determination as these protein segments are indicated to be essential for *Masc* function [19]. This is further supported by the ability of *O. furnacalis* and *T. varians* *Masc* proteins to induce both male-specific splicing of *dsx* and dosage compensation in *B. mori* cells, even though they share low levels of amino acid identity [26].

Simultaneous knock-down of *EkMasc/ EkMascB* showed female-specific *Ekdsx* splicing in genetic males, but no feminized males were found among adults. However, when using the siRNA targeting the exon containing the masculinizing domain (siMasc\_VII), we observed a highly female-biased sex ratio that is likely due to the male-killing effect caused by the loss of dosage compensation, similar to what was found in *B. mori* [17,20] and hypothesized in *O. furnacalis* [22] and *P. xylostella* [25]. Surprisingly, this male-killing effect was not observed when using the siRNA targeting exon 2, even though both siRNAs showed similar effects on the *EkMasc/ EkMascB* expression levels and *Ekdsx* splicing pattern. We observed additional splice variants during embryogenesis present at 40–64 hpo that could be targeted by siMasc\_II but not siMasc\_VII (S12 Fig). This could theoretically result in different knock-down effects of siMasc\_II compared to siMasc\_VII at later stages of development, and thus in the reduced effect observed for this siRNA, i.e. the lack of male lethality. However, whether the difference

in male-killing effects is caused by the different exons being targeted by the two siRNAs, or other differences between these siRNAs, is currently uncertain.

The presence of female-specific *Ekdsx* in genetic males injected with siMasc\_II, but the absence of a male-killing effect and lack of persistent phenotypic feminization provide insights into the sex determination mechanism in *E. kuehniella*. Under normal conditions, splicing of *Ekdsx* in ZZ embryos transitions from female- to male-specific at 16–18 hpo, indicating that sex is determined at this point of development. After injections with siMasc\_II, splicing of *Ekdsx* was assessed at 48 hpo, showing predominantly female-specific splicing in genetic males 30–32 hours after the sex of the individual is normally established. However, RNAi is transient and the expression levels of *EkMasc/EkMascB* likely increased after a number of hours, resulting in a reversal of female-specific *Ekdsx* splicing to male-specific splicing of *Ekdsx* >48 hpo and the development of a male phenotype. This suggests a strong plasticity in timing of sexual development in *E. kuehniella*.

The pattern of sex-specific expression of *Masc* was previously shown only in *B. mori* [17]. In early male embryos, expression of *Masc* increases and subsequently decreases rapidly, while expression in female embryos gradually decreases after oviposition [17]. Here we show that differential expression of *EkMasc* and *EkMascB* also occurs during early embryogenesis in *E. kuehniella*. As found in *B. mori*, the window of differential expression is narrow in *E. kuehniella*, starting at approximately 12–14 hpo, with a maximum around 16 hpo, and ending before 24 hpo. The peak of *EkMasc* and *EkMascB* expression in males corresponds to the first occurrence of the male-specific splice form of *Ekdsx*, suggesting a close interaction between these genes. However, previous studies have shown that dosage maintenance of duplicated genes occurs at both the transcriptional and translational levels in eukaryotes and that repression of translation occurs rapidly rather than gradually [42]. Therefore, it should be noted that even though both *EkMasc* and *EkMascB* show sex-specific differential expression, translational repression might prevent one of the genes to be truly functional. Future research should therefore focus on the separate knock-down/-out of *EkMasc* and *EkMascB* to determine whether one or both genes are involved in sex determination and/or dosage compensation in *E. kuehniella*.

We identified a male-specific splice variant of *EkMasc* and *EkMascB* lacking the exon essential for masculinization in *E. kuehniella*, i.e. *EkMasc<sup>ms</sup>* and *EkMascB<sup>ms</sup>*, which is not detected at 16 hpo, but is detected from 40–64 hpo onwards. We further confirmed this splice variant in the closely related *P. interpunctella* (Pyralidae) and the distant codling moth, *C. pomonella* (Tortricidae). In all three species, *Masc<sup>ms</sup>* results from an early stop codon and therefore leads to a truncated protein. Interestingly, an alternative splice variant of *BmMasc*, *Masc-S*, has been identified [43] that lacks the exon containing the *Fem* piRNA target sequence and is therefore not sensitive to the down-regulation by *Fem* piRNA. The regulation of alternative splicing of *Masc* in *B. mori* is largely unknown, but the gene *exuperantia* (*Bmexu*) seems to play a role in this process, as it binds to *Masc* but not *Masc-S* RNA [43]. In *B. mori*, this splice variant is present in both sexes and is essential for the development of the external female genitalia. In contrast, the *Masc<sup>ms</sup>* splice variant is found predominantly in male somatic tissues, but its function is unclear. However, the pattern of splicing and its apparent conservation across Lepidoptera is remarkable. In *P. interpunctella* and *C. pomonella*, but not *E. kuehniella*, the *Masc<sup>ms</sup>* protein does contain the two zinc finger domains, and therefore in these species *Masc* might have a tertiary function aside from masculinization and dosage compensation.

The flour moth *E. kuehniella* is the first species described in which the *Masc* zinc finger motifs were lost naturally. Initially, the dosage compensation function of *Masc* in *B. mori* was predicted to involve the zinc finger domains in the N-terminus of the protein due to their known ability to bind DNA and RNA [17]. However, recent data revealed that the

masculinizing domain of BmMasc, rather than the zinc finger domains, regulates dosage compensation [20,26]. Male-specific lethality of *EkMasc*/*EkMascB*-silenced individuals suggests that dosage compensation also occurs during embryogenesis in *E. kuehniella* and that *EkMasc* and/or *EkMascB* regulate(s) this process either directly or indirectly. This supports the dispensable role of the zinc finger domains in both sex determination and dosage compensation as found in *B. mori* [19,20,26]. High conservation in other lepidopteran species, however, suggests an important and conserved function of the zinc finger domains for the Masc protein function. What this function is, remains currently unknown, but one proposed hypothesis states that the domains increase the binding efficiency of Masc to target RNA or DNA [20]. In addition, it was hypothesized that the absence of the zinc finger domain in a mutant strain of *B. mori* might be compensated by another zinc finger-containing gene, *Bmzmf-2*. Overexpression of *Bmzmf-2* in BmN cells, derived from female ovarian cells, resulted in partial male-specific splicing of *Bmdsx*, but only if both zinc finger domains were intact [44]. In *E. kuehniella*, we were able to identify a potential ortholog, *Ekzmf-2*, containing both zinc finger domains (S13 Fig). Therefore, it is possible that the loss of the zinc finger domains in *EkMasc* and *EkMascB* is compensated by the presence of *Ekzmf-2* or another zinc finger domain-containing protein. However, why this loss of the zinc finger domains occurred only in *E. kuehniella*, even though *zmf-2* is likely to be conserved in Lepidoptera, is unclear. Alternatively, it is possible that the combined expression of *EkMasc* and *EkMascB* can compensate for the potentially reduced efficiency of Masc protein in *E. kuehniella*, thereby negating the loss of the zinc finger domains. Individual knock-out experiments of *EkMasc* and *EkMascB* could provide more clarity regarding the role of the zinc finger domains in *E. kuehniella* and Lepidoptera in general.

Similar to the *transformer* gene in Diptera, Coleoptera, and Hymenoptera, the protein sequence of Masc is highly variable in Lepidoptera. Apart from the previously identified masculinizing domain, the two zinc finger motifs and the bNLS domain, a shared feature of the Masc proteins is the relatively high level of proline content concentrated in two proline-rich domains, PRD-1 and PRD-2. Proline-rich repeats are known to have protein-protein binding capabilities and are often involved in binding of transcription factors [45,46]. The presence of a proline-rich domain at the C-terminus of the protein is another similarity between Masc and TRA [3], the latter being a confirmed splicing regulator of *dsx* in three insect orders. In addition, it has been shown that not only the masculinizing domain but also a segment of the C-terminus (within the final 75 amino acids) of the protein is essential for masculinization [19]. The C-terminus of Masc coincides with PRD-2, not only in *B. mori* but in all lepidopteran Masc proteins identified. Therefore, we propose that PRD-2, rather than the zinc fingers, of the Masc proteins in conjunction with the masculinizing domain are important for transcriptional regulation and sex determination in Lepidoptera, either through providing protein stability as previously suggested [19] or through binding of transcription factors, or both.

In conclusion, we identified the first duplication of *Masc* in Lepidoptera, *EkMasc* and its paralog *EkMascB*, on the Z chromosome of *E. kuehniella*. Knock-down of *EkMasc* and *EkMascB* revealed male lethality similar to that previously found in other lepidopteran species, confirming that *Masc* is a good target gene to eliminate males, a trait that might be exploited in the future to shift the sex ratio of populations of *E. kuehniella* to increase egg production. In addition, we have identified another, seemingly conserved splice form of *Masc*, *Masc<sup>ms</sup>*, which reveals an additional, previously unknown level of complexity to the lepidopteran sex determination pathway. Overall, our data contribute to the understanding of *Masc* function in *E. kuehniella* and Lepidoptera in general and provide further information regarding *Masc* as a potential target for lepidopteran pest management.

## Materials and methods

### Insects

A laboratory strain (WT-C02) of the Mediterranean flour moth, *E. kuehniella*, was used for all experiments. This wild-type strain was established in 2002 from individuals collected in Boršov nad Vltavou, Czech Republic, and has been maintained on artificial diet [47] under 12:12 (L:D) conditions at 20–22°C. The codling moth, *Cydia pomonella*, samples used were obtained from the laboratory strain Krym-61; its origin and rearing were described earlier [48]. Larvae of the Indian meal moth, *Plodia interpunctella*, were obtained from a local population collected in České Budějovice, Czech Republic.

### Identification and isolation of *EkMasc* and *EkMascB* sequences

To identify *Masculinizer* in *E. kuehniella*, reverse transcription PCR (RT-PCR) was done using RNA from approximately 50 pooled eggs 24 hours post oviposition (hpo) and single male pupa samples, both in duplicate. Total RNA was isolated using TRI Reagent (Sigma-Aldrich, St. Louis, MO) according to the manufacturer's protocol, using chloroform for phase separation. RNAs were dissolved in 40 µL nuclease-free water and their concentrations were measured on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Approximately 1 µg RNA per sample was converted to cDNA using the ImProm-II Reverse Transcription System kit (Promega, Madison, WI) using the Oligo(dT)<sub>15</sub> primer supplied with the kit and with a final concentration of 3 mM MgCl<sub>2</sub> according to the manufacturer's instructions. Primers were designed based on the putative *Masc* transcript sequence of *P. interpunctella* (annotated as maker-scaffold92-augustus-gene-0.127-mRNA-1) obtained through LepBase (lepbase.org; [49]) using Geneious 9.1.6 (<https://www.geneious.com>; [50]) with default settings. All primers used in the article can be found in S3 Table. PCR was carried out in a final volume of 10 µL containing 0.2 µM of each primer (Masc\_F1 and Masc\_R1), 0.2 mM of each dNTP, 1× Ex *Taq* PCR buffer, 0.025 units of Ex *Taq* DNA polymerase (TaKaRa, Otsu, Japan), and 1 µL of cDNA. PCR amplification was performed using a standard thermocycling program of 94°C for 3 min initial denaturation; 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 1 min; with a final extension at 72°C for 3 min. PCR amplification was confirmed by gel electrophoresis using a 1.5% agarose gel in 1× TAE buffer and visualized using ethidium bromide (EtBr). The remaining PCR products were pooled, purified using the Wizard SV Gel and PCR Clean-Up System (Promega), cloned into the pGEM-T Easy vector (Promega), and finally sequenced (SEQme, Dobříš, Czech Republic) to obtain the initial *EkMasc* and *EkMascB* sequences (see S1 Methods for details).

Full-length cDNA sequences of *EkMasc* and *EkMascB* were obtained by rapid amplification of cDNA ends PCR (RACE-PCR) for both 3'- and 5'- ends (see S1 Methods for details). In short, to obtain the 3' region of the cDNA, first strand cDNA synthesis was prepared using 1 µg of the RNA sample from pooled 24 hpo eggs, the adapter primer (AP), and the ImProm-II Reverse Transcription System (Promega) as described above. Two rounds of PCR amplification followed. In the first round, the Masc\_F1 primer and the abridged universal amplification primer (AUAP) were used. In the second round, the primers were substituted by Masc\_VII\_F1 and AUAP and the template DNA consisted of a 100× diluted PCR product of the first round. For the isolation of the 5'-UTR of *EkMasc* and *EkMascB*, first strand cDNA was synthesized using 1 µg of the same RNA sample with the ImProm-II Reverse Transcription System as described above, but using the Masc\_V\_R1 primer. The 5'-RACE-PCR was performed as described previously [51]. Full-length cDNA sequences of *EkMasc* and *EkMascB*, two splice variants each, are available (GenBank accession numbers MW505939-MW505942). In

addition, all splice variants of *EkMasc* and *EkMascB* were aligned and can be accessed at <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:200509> (doi: [10.17026/dans-zpj-ff5m](https://doi.org/10.17026/dans-zpj-ff5m)).

## Genome assembly

To assess the splicing structure and genome organization of *EkMasc* and *EkMascB*, we sequenced the male genome of *E. kuehniella* using Oxford Nanopore technology. DNA was extracted from 5 male larvae using cetyltrimethylammonium bromide (CTAB) DNA extraction as described elsewhere [52] and sequenced on the Nanopore PromethION by Novogene (HK) Co., Ltd. (Hong Kong, China). The sequencing yielded 16.2 Gb of data equivalent to ca. 36.8× genome coverage (assuming a genome size of 440.1 Mb, as determined earlier [53]) with N50 of 24.8 kb. The long reads were deposited in the Sequence Read Archive under accession number PRJNA683200. The reads were assembled using Canu version 1.8 [54] with the following parameters, genomeSize = 440.1m and minReadLength = 10000. The complete cDNA sequences of *EkMasc* and *EkMascB* obtained through RACE PCRs were used in a BLASTn search against the assembled genome and exons were annotated using Geneious 9.1.6. A fragment of the scaffold containing *EkMasc*, *EkMascB*, and an additional 20 kb flanking sequence on either side (GenBank accession number MW774596) was further analyzed using BLASTn in LepBase and NCBI to identify genes and transposable elements. For LepBase, BLASTn searches were only restricted to the CDS database of *P. interpunctella*, while the search using NCBI was restricted to the nucleotide collection of Lepidoptera.

## Assessment of copy number and localization of *Masc* using Southern hybridization

DNA was extracted from single female and male pupae using CTAB DNA extraction as described elsewhere [52]. Concentrations were measured on a Qubit 3.0 Fluorometer using the dsDNA BR Assay Kit (Invitrogen, Carlsbad, CA), and approximately 4 µg DNA per sample was used for DNA digestion reactions. DNA was double digested using *NdeI* × *NotI*, *DraI* × *NheI* (all Fermentas, Vilnius, Lithuania), or *AgeI* × *BspHI* (New England Biolabs, Ipswich, MA) enzymes (see S1 Methods for details). All digested DNA was separated by electrophoresis on a 1% TBE agarose gel and subsequently transferred by capillary transfer to an Amersham Hybond-N+ nylon membrane (GE Healthcare, Milwaukee, WI).

A probe specific to *EkMasc* with high homology to *EkMascB* (96%) was made by PCR-labeling (see S1 Methods for details). The probe was labeled with digoxigenin-11-dUTPs (Roche Diagnostics, Mannheim, Germany) using primers *Masc\_Sb\_F* and *Masc\_Sb\_R*. For Southern hybridization, 100 ng of the probe was used. The Southern blot assay was performed as described previously [55] with some modifications [56].

## Z-linkage of *EkMasc* and *EkMascB* by quantitative real-time PCR

To assess the localization of *EkMasc* and *EkMascB* on the Z chromosome, we performed quantitative real-time PCR (qPCR) using genomic DNA as a template and *Acetylcholinesterase 2* (*Ace-2*) as an autosomal reference gene as described previously [39] (see S1 Methods for details). This method relies on the comparison of female and male samples as the ratio between Z-linked genes and autosomal genes is 1:2 in females and 2:2 in males. DNA was isolated from single female and male *E. kuehniella* larvae in triplicate using the NucleoSpin DNA Insect kit (Macherey-Nagel, Düren, Germany) as described elsewhere [57]. The 10 µL qPCR mixture consisted of 1× Xceed SG qPCR Mix Lo ROX (Institute of Applied Biotechnologies, Prague, Czech Republic), 400 nM of each forward and reverse primer (for *EkMasc/EkMascB* primers *Masc\_F\_VIa* and *Masc\_R\_VI*; for *EkAce-2* primers *Ek\_Ace2\_F* and *Ek\_Ace2\_R*), and 10 ng of



genomic DNA. Data was analyzed as described previously [39]. Briefly, the ratios between the target gene and the autosomal reference gene were calculated using the formula  $R = [(1 + E_{\text{Reference}})^{\text{Ct}_{\text{Reference}}}] / [(1 + E_{\text{Target}})^{\text{Ct}_{\text{Target}}}]$ , where  $E$  is the primer efficiency and  $\text{Ct}$  the cycle threshold value. Two hypotheses were tested statistically by an unpaired two-tailed  $t$ -test for unequal variances: (i) the *EkMasc* and *EkMascB* genes are located on the Z chromosome (the expected female to male ratio is 0.5) and (ii) both genes are autosomal (the female to male ratio is 1). The male and female  $R$  values were either compared directly, or female values were multiplied by 2 to test both hypotheses. The average  $R$  value for females and males was calculated and plotted using R version 3.5.2 [58].

### Tissue-specific splicing of *Masc*

After identification of a splice variant encoding a truncated *Masc* protein, which lacks the two cysteine residues essential for masculinization, i.e. skipping exon VII, we tested the presence of this variant during different developmental stages in *E. kuehniella* females and males. This splice variant is referred to as masculinizing “domain” skipping *Masc* (*Masc<sup>ms</sup>*). For the embryonic stage, we used sexed embryos 16 hpo isolated for the *EkMasc* and *EkMascB* expression analysis (see below). In addition, RNA was isolated from single final instar larvae and single 2- or 3-day-old pupae using TRI Reagent. RNA was subsequently DNase-treated with the Invitrogen TURBO DNA-free Kit (Thermo Fisher) according to the manufacturer’s instructions, and 1  $\mu\text{g}$  RNA was converted to cDNA using the ImProm-II Reverse Transcription System as described above. To test splice variation, we performed PCR using primers *Masc\_bmd\_qF1* and *Masc\_R\_X*, and 1  $\mu\text{l}$  of cDNA. The same PCR mix and profile were used as for the initial isolation of *EkMasc* and *EkMascB*. PCR products were visualized on a 1.5% TAE agarose gel stained with EtBr.

In addition, we tested whether the *Masc<sup>ms</sup>* splice variant was present in the testis and/or the rest of the body of *E. kuehniella*, *P. interpunctella*, and *C. pomonella* (see [S1 Methods](#) for identification of *Masc* in *C. pomonella*, primer design, and testes dissection). RNA was extracted and processed as described in the previous paragraph. The obtained cDNA samples were tested by PCR using primers *qPiMasc\_F1* x *qPiMasc\_R2* and *CpMasc\_F4* x *CpMasc\_R2* for *P. interpunctella* and *C. pomonella*, respectively. Samples of *E. kuehniella* were tested by PCR using primers *Masc\_bmd\_qF1* x *Masc\_R\_X*. The PCR mix and thermocycling settings were the same as for the initial identification of *EkMasc* and *EkMascB*, and the PCR products were separated and visualized as described above. The non-sex-specific splice variant of *PiMasc* was uploaded to GenBank (accession number MW505946). In addition, all obtained sequences of *PiMasc* and *CpMasc* were aligned and can be accessed at <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:200509> (doi: [10.17026/dans-zpj-ff5m](https://doi.org/10.17026/dans-zpj-ff5m)).

### PCR-based genetic sexing of *E. kuehniella*

Sexing of *E. kuehniella* can be done phenotypically in fifth instar larvae, pupae, and adults, however, sexing during embryogenesis and early larval stages is not possible. Therefore, we developed a PCR-based genetic sexing method for *E. kuehniella*. For this, DNA from three female and three male fifth instar larvae was isolated individually using the NucleoSpin DNA Insect kit (Macherey-Nagel). We observed that the primers *Masc\_F1* and *Masc\_R1* targeting *EkMasc* and *EkMascB* consistently showed off-target amplification in female samples, resulting in two bands in female samples and a single, expected band in male samples. Further research revealed that there are many partial copies of *EkMasc* on the W chromosome and that the second band of smaller size on the gel in female samples comes from the amplification of these degenerate *EkMasc* fragments derived from the W chromosome ([S6 Fig](#)). We tested the

primers Masc\_F1 and Masc\_R1 at different annealing temperatures and found consistent results at temperatures ranging from 52–60°C with optimal results at 55°C. The 10 µL PCR mix contained 0.2 µM of each primer, 0.2 mM dNTPs, 1× OneTaq Quick-Load reaction buffer, 0.025 units of OneTaq DNA polymerase (New England Biolabs), and 5 ng of DNA. The thermocycling profile consisted of initial denaturation at 94°C for 3 min; 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s and extension at 68°C for 45 s; with a final extension at 68°C for 3 min. This protocol was additionally confirmed using DNA samples from at least 20 adults, 10 pupae and 10 larvae of the WT-C02 strain in multiple independent experiments.

### Expression analysis of *EkMasc* and *EkMascB* during embryogenesis

To measure the expression patterns of *EkMasc* and *EkMascB*, DNA and RNA was simultaneously isolated from single embryos at two-hour intervals, starting at 12 hpo and ending at 24 hpo. RNA was isolated using TRI Reagent according to the manufacturer's protocol and pellets were stored in ethanol at –80°C until further use (see [S1 Methods](#)). DNA was simultaneously isolated from the organic phase using the back extraction protocol as described by Thermo Fisher Scientific (see [S1 Methods](#) for details). This DNA was used to determine the sex of the individual by PCR using sex-specific markers as described above. After sex determination by PCR, ethanol was removed from the corresponding RNA samples of three to six females and males, and RNA pellets were dissolved in 10 µL of diethyl pyrocarbonate (DEPC)-treated water. The complete RNA samples were DNase-treated with the Invitrogen TURBO DNA-free Kit according to the manufacturer's instructions, and all RNA was used directly for conversion to cDNA using the ImProm-II Reverse Transcription System according to the manufacturer's instructions. The cDNA was prepared using a combination of Random Primers and Oligo (dT)<sub>15</sub> Primer (1:1) supplied with the kit, and with a concentration of 3 mM MgCl<sub>2</sub>. All cDNA samples were diluted 3× with nuclease-free water before use in qRT-PCR experiments.

qRT-PCR experiments were done on cDNA from the embryo samples using the gene *Ekrp49* as a reference to calculate relative expression levels (see [S1 Methods](#) for identification of *Ekrp49* and primer design for qRT-PCR). The experiments were performed as described above for qPCR, but 2 µL of the 3× diluted cDNA was used as a template. A minimum of three biological replicates per sex and time point were used and three technical replicates were used for each sample. Since the number of samples exceeded a single plate experiment, at least five samples (in triplicate) were repeated between plates to correct for between-plate variation, and different genes were measured on separate plates. Relative copy numbers were calculated according to the formula described above in Z-linkage of *EkMasc* and *EkMascB* using qPCR, for *EkMasc* and *EkMascB* separately. An unpaired two-tailed *t*-test with unequal variances was used to test for significant differences in expression levels between sexes of the same age. Statistical analysis and visualization of the data were done using R version 3.5.2 [58].

### Functional analysis of *EkMasc* and *EkMascB*

To assess the function of *EkMasc* and *EkMascB* in sex determination, two short interfering RNAs (siRNAs; see [S3 Table](#)) were designed to target *EkMasc* and *EkMascB* simultaneously, based on published recommendations [59,60]. We tested two siRNAs, one targeting the start of the open reading frame (exon II) and one targeting the exon coding for the two cysteine residues essential for masculinization (exon VII) (see [S1 Methods](#) for details on siRNA design). Custom synthetic siRNA duplexes siMasc\_II and siMasc\_VII (see [S3 Table](#) for sequence details) were obtained from Sigma-Aldrich. The siRNA duplexes were dissolved in 100 µL nuclease-free water, then the solution buffer contained 100 mM potassium acetate, 30 mM

HEPES, and 2 mM magnesium acetate. In addition, one siRNA published earlier [17], designed against GFP (siGFP), was used as a negative control. All siRNAs were injected individually, no combinations were tested.

Eggs were collected within the first hour after oviposition and injected using a FemtoJet Microinjector (Eppendorf, Hamburg, Germany) (see [S1 Methods](#) for details). DNA and RNA were isolated simultaneously from a subset of injected single embryos at 16 hpo (to test knock-down efficiency) or 48 hpo (to test *Ekdsx* splicing, see below) using TRI Reagent as described above in the expression analysis. DNA was used to sex the individuals, and subsequently RNA from a minimum of three confirmed WZ and three confirmed ZZ samples was converted to cDNA. Relative expression levels of *EkMasc* and *EkMascB* were measured and calculated using qRT-PCR as described in the expression analysis. Differences in *EkMasc* and *EkMascB* expression levels between control and knock-down treatments were statistically tested by an unpaired one-tailed *t*-test for unequal variances, for both sexes separately.

Remaining eggs were left to develop and hatched approximately 7–10 days post injection. Larvae were transferred to artificial medium and left to develop until adulthood. The adults were phenotypically sexed and additionally assessed for potential abnormalities in reproductive organs, by dissection using a stereo microscope. In selected males, the testes were dissected in physiological solution, fixed in a Carnoy's fixative for 15 min, and spread on slides as described previously [48]. The slides were stained with either 1.5% lactic acetic orcein and examined using a phase contrast microscope or 0.5 µg/mL DAPI in antifade based on DABCO and examined using a fluorescence microscope [48]. In addition, heads of adults were used for CTAB DNA isolation (see above) and genetically sexed as described in section "Expression analysis of *EkMasc* and *EkMascB* during embryogenesis". The observed adult sex ratios were tested for deviations from the expected sex ratio using the Pearson's chi-squared test for goodness-of-fit with the expected frequency of 0.5 for both sexes.

To assess the effects of *Masc* knock-down on sex determination, splicing patterns of *doublesex* (*dsx*) are often used [17,25]. Therefore, we identified a *dsx* ortholog in *E. kuehniella* (see [S1 Methods](#) for details). Primers were designed using Geneious 9.1.6 in exons II and V, flanking the female-specific exons. PCR was done by using the mixture and thermocycling program described above, but with an annealing temperature of 60°C using primers *dsx\_dR\_F2* and *dsx\_dR\_R2* and cDNA from either female or male pupa samples. PCR products, run on a 1.5% TAE agarose gel stained with EtBr, showed alternative splicing between sexes. PCR products of both sexes were purified, cloned, and sequenced as described above. After confirmation of *Ekdsx* by sequencing, we used the same PCR conditions with cDNA from the embryo time series (see "Expression analysis of *EkMasc* and *EkMascB* during embryogenesis") to identify the timing of sex-specific splicing of *Ekdsx* in *E. kuehniella*. We also used the same PCR on cDNA of a minimum of three WZ and ZZ embryos (of 48 hpo) injected with any of the three siRNAs to determine differential *Ekdsx* splicing. The products were run on a 1.5% TAE agarose gel and stained with EtBr. All obtained *Ekdsx* sequences are included as an alignment and can be accessed at <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:200509> (doi: [10.17026/dans-zpj-ff5m](https://doi.org/10.17026/dans-zpj-ff5m)).

### Sliding window analysis of proline content in lepidopteran Masc proteins

We compared all functionally confirmed lepidopteran Masc proteins to identify any other potentially conserved features/domains and we noticed a relatively high proline content in all lepidopteran Masc proteins. To identify patterns in the distribution of these proline residues, we used a sliding window approach. Therefore, we calculated the average proline content across a window of 25 amino acids (aa) and shifted the window by steps of 1 aa until the end of

the protein using Microsoft Excel 2013. This process was performed for all functionally confirmed Masc protein sequences, i.e. *Bombyx mori* [17], *Trilocha varians* [21], *Agrotis ipsilon* [24], *Ostrinia furnacalis* [22], and *Plutella xylostella* [25], including EkMasc and EkMascB. An initial trial using window sizes of 10, 20, 25, and 50 aa showed that window size 25 aa was optimal for visualization of the data. The data were visualized using R version 3.5.2 [58] with the ggplot2 package [61], and domains were indicated using Inkscape 0.92 (<https://inkscape.org/>).

## Supporting information

**S1 Table. Significance test of *EkMasc* and *EkMascB* expression level comparison between sexes during embryogenesis in *Ephestia kuehniella*.** An unpaired two-tailed *t*-test for unequal variances was used. Expression levels significantly differing from each other between sexes ( $P < 0.05$ ) are indicated in bold.

(PDF)

**S2 Table. Significance test of knock-down effects of siRNA treatments in *Ephestia kuehniella* using an unpaired one-tailed *t*-test for unequal variances.** Expression levels significantly differing from each other between treatments ( $P < 0.05$ ) are indicated in bold.

(PDF)

**S3 Table. Overview of primers and siRNAs used in this study.**

(PDF)

**S1 Fig. *EkMasc* and *EkMascB* Z-linkage assessed by qPCR using genomic DNA of *Ephestia kuehniella*.** Relative copy numbers were assessed for female and male samples ( $n = 3$  for both sexes). Indicated are hypothetical *EkMasc* and *EkMascB* female to male ratios relative to male copy numbers corresponding to the autosomal hypothesis (both *EkMasc* and *EkMascB* located on autosomes; F:M ratio = 1.00), and the Z chromosomal hypothesis (F:M ratio = 0.50). Error bars indicate standard deviation.

(TIF)

**S2 Fig. Southern blot assay using *EkMasc* probe in *Ephestia kuehniella*.** Two signals can be identified in the genomic DNA of both female and male samples double digested with (1) *NdeI* x *NotI*, (2) *DraI* x *NheI*, and (3) *AgeI* x *BspHI*. Arrows indicate highly diffused bands. Note that female signals are weaker than male signals. Indicated are the marker (M) in bp, female (♀) and male (♂) samples.

(TIF)

**S3 Fig. Exon-intron map of the *EkMasc* and *EkMascB* genes in *Ephestia kuehniella*.** Two poly-adenylation sites are indicated for *EkMasc*, i.e. *EkMasc-1* and *EkMasc-2*, as well as the open reading frame (grey) including the start (ATG) and stop codon (\*). Additionally shown is the splice variant *EkMasc<sup>ms</sup>* skipping exon VII (dashed grey line) and its corresponding premature stop codon (grey \*). Similar splice patterns were observed for the *EkMascB* gene.

(TIF)

**S4 Fig. Schematic representation of the effects of splicing on the translation of the *EkMasc* and *EkMascB* genes.** (A) Schematic drawings of the main alternative splice forms of *EkMasc* and *EkMascB*. Indicated are the exons (white boxes), the degenerated zinc finger motifs (dashed pink boxes), the bipartite nuclear localization signal (green box), the masculinizing domain (blue box), and the open reading frame (ORF). Note that the masculinizing domain is absent in *EkMasc<sup>ms</sup>* and *EkMascB<sup>ms</sup>* splice variants and that there are two potential ORFs for these splice forms. (B) Schematic representation of the proteins translated from each of the

transcripts (in black or grey) with the same domains indicated as in **A**. The first of the two potential EkMasc<sup>ms</sup> proteins is identical to the main EkMasc protein for the region indicated by dashed lines. This potential protein also contains a complete bipartite nuclear localization signal (green/brown box), the first half of which is identical to the EkMasc protein. The second potential EkMasc<sup>ms</sup> protein is a short version of the EkMasc protein sharing complete amino acid identity with the C-terminus of the EkMasc protein, except for the first residue. The same pattern is observed for EkMascB and potential EkMascB<sup>ms</sup> proteins.

**S5 Fig. Protein alignment of EkMasc and EkMascB of *Ephestia kuehniella*.** Indicated in the alignment are the conserved bipartite nuclear localization signal (bNLS; green) and the masculinizing domain (MD; blue). Also indicated are deletions (red arrows) and insertions (cyan arrows) in either EkMasc or EkMascB based on comparison to the Masc protein sequence of the closely related *Plodia interpunctella*. A single grey arrow (at amino acid 31) indicates an indel between EkMasc and EkMascB that cannot be categorized as a deletion or an insertion in either protein sequence based on the Masc protein sequence of *P. interpunctella*.

**S6 Fig. Genetic sexing of *Ephestia kuehniella*.** Note the two bands in female samples, the lower band at approximately 500 bp corresponding to the degenerated *EkMasc* copies identified on the W chromosome. Indicated are marker (M) in kb, no template control (NTC), female (♀) and male (♂) samples.

**S7 Fig. Illustrative set-up of the RNAi experiments in *Ephestia kuehniella*.** Indicated in the figure are the degenerated zinc finger motifs (pink boxes) upstream of the open reading frame, the bipartite nuclear localization signal (green box), the male determining region (blue box), the open reading frame (yellow) and the two siRNAs targeting *EkMasc* and *EkMascB* (red dashed lines). Also shown is an exon representation of the genes.

**S8 Fig. Splicing pattern of *Ekdsx* in *Ephestia kuehniella*.** (A) Schematic representation of female- (pink line) and male-specific (blue line) splicing patterns of *Ekdsx*. Indicated below the figure are the targets of the primers used throughout the article (F and R). (B) The two dominant female- and the dominant male-specific transcripts and their predicted respective open reading frames (in grey/pink). (C) Sex-specific splicing of *Ekdsx* as detected by the primers indicated in A. A currently uncharacterized third female-specific splice variant is also visible (the top arrow). (D) Sex-specific splicing pattern of *Ekdsx* during early development in WZ (left) and ZZ (right) individuals. Note transitions of *Ekdsx* splicing from female-specific to male-specific in ZZ individuals only, 16–18 hours post oviposition (hpo). Indicated are marker (M) in kb, and no template control (NTC).

**S9 Fig. Cytological examination of the testes of surviving adult males of *Ephestia kuehniella* after knock-down of the *EkMasc* and *EkMascB* genes by injecting siRNAs into embryos.** (A, D, G) Control: 4 males of the WT-C02 strain examined. (B, E, H) Injection with siMasc\_VII: 5 males stored in the freezer examined. (C, F, I) Injection with siMasc\_II: 3 males stored in the freezer examined. (A–C) Images of dissected testes in physiological solution. Bar = 0.5 mm. (D–F) Nuclei of eupyrene sperm stained with lactic acetic orcein observed using a phase contrast microscope. Bar = 10 μm. (G–I) Nuclei of eupyrene sperm stained with DAPI (4',6-diamidino-2-phenylindole) observed using a fluorescence microscope.

Bar = 10  $\mu\text{m}$ . **D** and **F** show part of the sperm bundle head; **E**, **H**, and **I**, show the complete head of the sperm bundle (arrows indicate directions of sperm tails); **G** shows spread nuclei of eupyrene sperm.

(TIF)

**S10 Fig. Alignment of functionally confirmed Masc proteins in six lepidopteran species.**

The alignment shows low levels of amino acid identity even within the functional domains. Indicated below the alignment are the conserved zinc finger domains (ZF1 & ZF2; pink), the bipartite nuclear localization signal (bNLS; green) and the masculinizing domain (MD; blue). Aligned are the Masc protein sequences of *Bombyx mori* (Bm), *Trilocha varians* (Tv), *Agrotis ipsilon* (Ai), *Ostrinia furnacalis* (Of), *Plutella xylostella* (Pxy), and *Ephestia kuehniella* (Ek).

(TIF)

**S11 Fig. Sliding window analysis of proline content for EkMasc and EkMascB proteins in *Ephestia kuehniella*.** Note that the proline distribution in both proteins is very similar.

(TIF)

**S12 Fig. *EkMasc* and *EkMascB* splicing at two time points during embryogenesis in *Ephestia kuehniella*.** RT-PCR using cDNA obtained from pooled unsexed eggs samples of 16–17 hours post oviposition (hpo) and 40–64 hpo. At 16–17 hpo, only the splice variant containing exon 7 is visible, whereas at 40–64 hpo an additional band corresponding to the *EkMasc<sup>ms</sup>* and *EkMascB<sup>ms</sup>* splice variants is visible. Indicated is marker (M) in base pairs (bp).

(TIF)

**S13 Fig. Protein alignment of zinc finger protein 2 (ZNF-2) of *Bombyx mori* (Bm) and the predicted ortholog in *Ephestia kuehniella* (Ek).** Indicated in pink are the two zinc finger domains (ZF1 and ZF2), and in blue two cysteine amino acids separated by two amino acids similar to the masculinizing domain in lepidopteran Masc proteins.

(TIF)

**S1 Methods. Details of methods used.** Identification and isolation of *EkMasc* and *EkMascB* sequences. Assessment of copy number and localization of *Masc* using Southern hybridization. Z-linkage of *EkMasc* and *EkMascB* by quantitative real-time PCR (qPCR). Tissue-specific splicing of *Masc*. Expression analysis of *EkMasc* and *EkMascB* during embryogenesis. Back extraction DNA isolation protocol. Functional analysis of *EkMasc* and *EkMascB*.

(PDF)

## Acknowledgments

We thank Marie Korchová for excellent technical assistance and Joanna Kotwica-Rolinska for providing technical help with microinjections. We also thank Arjen van't Hof and Atsuo Yoshido for their valuable comments on the manuscript. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

## Author Contributions

**Conceptualization:** Sander Visser, Eveline C. Verhulst, František Marec.

**Data curation:** Sander Visser.

**Formal analysis:** Sander Visser.

**Funding acquisition:** František Marec.

**Investigation:** Sander Visser.

**Methodology:** Eveline C. Verhulst.

**Resources:** Anna Voleníková, Petr Nguyen.

**Supervision:** František Marec.

**Writing – original draft:** Sander Visser, Eveline C. Verhulst, František Marec.

**Writing – review & editing:** Sander Visser, Anna Voleníková, Petr Nguyen, Eveline C. Verhulst, František Marec.

## References

1. Nöthiger R, Steinmann-Zwicky M. A single principle for sex determination in insects. *Cold Spring Harbor Symp Quant Biol.* 1985; 50:615–621. <https://doi.org/10.1101/sqb.1985.050.01.074> PMID: 3938366
2. Wilkins AS. Moving up the hierarchy: A hypothesis on the evolution of a genetic sex determination pathway. *BioEssays.* 1995; 17(1):71–77. <https://doi.org/10.1002/bies.950170113> PMID: 7702596
3. Verhulst EC, van de Zande L, Beukeboom LW. Insect sex determination: it all evolves around transformer. *Curr Opin Genet Dev.* 2010; 20(4):376–383. <https://doi.org/10.1016/j.gde.2010.05.001> PMID: 20570131
4. Gempe T, Beye M. Function and evolution of sex determination mechanisms, genes and pathways in insects. *BioEssays.* 2011; 33(1):52–60. <https://doi.org/10.1002/bies.201000043> PMID: 21110346
5. Bopp D, Saccone G, Beye M. Sex determination in insects: variations on a common theme. *Sex Dev.* 2014; 8(1–3):20–28. <https://doi.org/10.1159/000356458> PMID: 24335049
6. Geuverink E, Beukeboom LW. Phylogenetic distribution and evolutionary dynamics of the sex determination genes doublesex and transformer in insects. *Sex Dev.* 2014; 8(1–3):38–49. <https://doi.org/10.1159/000357056> PMID: 24401160
7. Nguantad S, Chumnpanpuen P, Thancharoen A, Vongsangnak W, Sriboonlert A. Identification of potential candidate genes involved in the sex determination cascade in an aquatic firefly, *Sclerotia aquatilis* (Coleoptera, Lampyridae). *Genomics.* 2020; 112(3):2590–2602. <https://doi.org/10.1016/j.ygeno.2020.01.025> PMID: 32061895
8. Zou Y, Geuverink E, Beukeboom LW, Verhulst EC, van de Zande L. A chimeric gene paternally instructs female sex determination in the haplodiploid wasp *Nasonia*. *Science.* 2020; 370(6520):1115–1118. <https://doi.org/10.1126/science.abb8949> PMID: 33243892
9. Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW. The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell.* 2003; 114(4):419–429. [https://doi.org/10.1016/s0092-8674\(03\)00606-8](https://doi.org/10.1016/s0092-8674(03)00606-8) PMID: 12941271
10. Hasselmann M, Gempe T, Schiøtt M, Nunes-Silva CG, Otte M, Beye M. Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature.* 2008; 454(7203):519–522. <https://doi.org/10.1038/nature07052> PMID: 18594516
11. Jia LY, Xiao JH, Xiong TL, Niu LM, Huang DW. The transformer genes in the fig wasp *Ceratosolen solmsi* provide new evidence for duplications independent of complementary sex determination. *Insect Mol Biol.* 2016; 25(3):191–201. <https://doi.org/10.1111/imb.12210> PMID: 26748889
12. Geuverink E, Kraaijeveld K, van Leussen M, Chen F, Pijpe J, Linskens MHK, et al. Evidence for involvement of a transformer paralogue in sex determination of the wasp *Leptopilina clavipes*. *Insect Mol Biol.* 2018; 27(6):780–795. <https://doi.org/10.1111/imb.12522> PMID: 30039559
13. Sharma A, Heinze SD, Wu Y, Kohlbrenner T, Morilla I, Brunner C, et al. Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*. *Science.* 2017; 356(6338):642–645. <https://doi.org/10.1126/science.aam5498> PMID: 28495751
14. Furlong MJ, Wright DJ, Dosdall LM. Diamondback moth ecology and management: problems, progress, and prospects. *Annu Rev Entomol.* 2013; 58:517–541. <https://doi.org/10.1146/annurev-ento-120811-153605> PMID: 23020617
15. Suzuki MG. Sex determination: insights from the silkworm. *J Genet.* 2010; 89(3):357–363. <https://doi.org/10.1007/s12041-010-0047-5> PMID: 20877002
16. Katsuma S, Kawamoto M, Kiuchi T. Guardian small RNAs and sex determination. *RNA Biol.* 2014; 11(10):1238–1242. <https://doi.org/10.1080/15476286.2014.996060> PMID: 25588029

17. Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, et al. A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature*. 2014; 509(7502):633–636. <https://doi.org/10.1038/nature13315> PMID: 24828047
18. Sugano Y, Kokusho R, Ueda M, Fujimoto M, Tsutsumi N, Shimada T, et al. Identification of a bipartite nuclear localization signal in the silkworm Masc protein. *FEBS Lett*. 2016; 590(14):2256–2261. <https://doi.org/10.1002/1873-3468.12246> PMID: 27277067
19. Katsuma S, Sugano Y, Kiuchi T, Shimada T. Two conserved cysteine residues are required for the masculinizing activity of the silkworm Masc protein. *J Biol Chem*. 2015; 290(43):26114–26124. <https://doi.org/10.1074/jbc.M115.685362> PMID: 26342076
20. Kiuchi T, Sugano Y, Shimada T, Katsuma S. Two CCCH-type zinc finger domains in the Masc protein are dispensable for masculinization and dosage compensation in *Bombyx mori*. *Insect Biochem Mol Biol*. 2019; 104:30–38. <https://doi.org/10.1016/j.ibmb.2018.12.003> PMID: 30529581
21. Lee J, Kiuchi T, Kawamoto M, Shimada T, Katsuma S. Identification and functional analysis of a Masculinizer orthologue in *Trilocha varians* (Lepidoptera: Bombycidae). *Insect Mol Biol*. 2015; 24(5):561–569. <https://doi.org/10.1111/imb.12181> PMID: 26154510
22. Fukui T, Kawamoto M, Shoji K, Kiuchi T, Sugano S, Shimada T, et al. The endosymbiotic bacterium *Wolbachia* selectively kills male hosts by targeting the masculinizing gene. *PLoS Pathog*. 2015; 11(7): e1005048. <https://doi.org/10.1371/journal.ppat.1005048> PMID: 26172536
23. Fukui T, Kiuchi T, Shoji K, Kawamoto M, Shimada T, Katsuma S. In vivo masculinizing function of the *Ostrinia furnacalis* Masculinizer gene. *Biochem Biophys Res Commun*. 2018; 503(3):1768–1772. <https://doi.org/10.1016/j.bbrc.2018.07.111> PMID: 30064911
24. Wang YH, Chen XE, Yang Y, Xu J, Fang GQ, Niu CY, et al. The Masc gene product controls masculinization in the black cutworm, *Agrotis ipsilon*. *Insect Sci*. 2019; 26(6):1037–1044. <https://doi.org/10.1111/1744-7917.12635> PMID: 30088858
25. Harvey-Samuel T, Norman VC, Carter R, Lovett E, Alphey L. Identification and characterization of a Masculinizer homologue in the diamondback moth, *Plutella xylostella*. *Insect Mol Biol*. 2020; 29(2):231–240. <https://doi.org/10.1111/imb.12628> PMID: 31793118
26. Katsuma S, Shoji K, Sugano Y, Suzuki Y, Kiuchi T. Masc-induced dosage compensation in silkworm cultured cells. *FEBS Open Bio*. 2019; 9(9):1573–1579. <https://doi.org/10.1002/2211-5463.12698> PMID: 31294930
27. Robinson R. *Lepidoptera genetics*. 1st ed. Oxford: Pergamon Press; 1971.
28. Traut W, Sahara K, Marec F. Sex chromosomes and sex determination in Lepidoptera. *Sex Dev*. 2007; 1(6):332–346. <https://doi.org/10.1159/000111765> PMID: 18391545
29. Champ BR, Dyte CE. FAO global survey of pesticide susceptibility of stored grain pests. *FAO Plant Protection Bulletin*. 1977; 25(2):49–67.
30. Vangansbeke D, Nguyen DT, Audenaert J, Verhoeven R, Gobin B, Tirry L, et al. Performance of the predatory mite *Amblydromalus limonicus* on factitious foods. *BioControl*. 2014; 59(1):67–77. <https://doi.org/10.1007/s10526-013-9548-5>
31. St-Onge M, Cormier D, Todorova S, Lucas É. Conservation of *Ephestia kuehniella* eggs as hosts for *Trichogramma ostrinae*. *J Appl Entomol*. 2016; 140(3):218–222. <https://doi.org/10.1111/jen.12227>
32. Bueno VHP, Montes FC, Sampaio MV, Calixto AM, van Lenteren JC. Performance of immatures of three Neotropical Miridae at five different temperatures, reared on *Ephestia kuehniella* eggs on tobacco plants. *Bull Insectology*. 2018; 71(1):77–87.
33. Cox PD, Mfon M, Parkin S, Seaman JE. Diapause in a Glasgow strain of the flour moth, *Ephestia kuehniella*. *Physiol Entomol*. 1981; 6(4):349–356. <https://doi.org/10.1111/j.1365-3032.1981.tb00650.x>
34. Cerutti F, Bigler F, Eden G, Bosshart S. Optimal larval density and quality control aspects in mass rearing of the Mediterranean flour moth, *Ephestia kuehniella* Zell. (Lep., Phycitidae). *J Appl Entomol*. 1992; 114(1–5):353–361. <https://doi.org/10.1111/j.1439-0418.1992.tb01139.x>
35. Xu J, Wang Q. Male moths undertake both pre- and in-copulation mate choice based on female age and weight. *Behav Ecol Sociobiol*. 2009; 63(6):801–808. <https://doi.org/10.1007/s00265-009-0713-x>
36. Traut W, Rathjens B. Das W-Chromosom von *Ephestia kuehniella* (Lepidoptera) und die Ableitung des Geschlechtschromatins. *Chromosoma*. 1973; 41(4):437–446. <https://doi.org/10.1007/BF00396501>
37. Nguyen P, Sýkorová M, Šichová J, Kůta V, Dalíková M, Čapková Frydrychová R, et al. Neo-sex chromosomes and adaptive potential in tortricid pests. *Proc Natl Acad Sci U S A*. 2013; 110(17):6931–6936. <https://doi.org/10.1073/pnas.1220372110> PMID: 23569222
38. Van't Hof AE, Nguyen P, Dalíková M, Edmonds N, Marec F, Saccheri IJ. Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism. *Heredity*. 2013; 110(3):283–295. <https://doi.org/10.1038/hdy.2012.84> PMID: 23211790



39. Dalíková M, Zrzavá M, Hladová I, Nguyen P, Šonský I, Flegrová M, et al. New insights into the evolution of the W chromosome in Lepidoptera. *J Hered.* 2017; 108(7):709–719. <https://doi.org/10.1093/jhered/esx063> PMID: 28992287
40. Suzuki MG, Ohbayashi F, Mita K, Shimada T. The mechanism of sex-specific splicing at the doublesex gene is different between *Drosophila melanogaster* and *Bombyx mori*. *Insect Biochem Mol Biol.* 2001; 31(12):1201–1211. [https://doi.org/10.1016/s0965-1748\(01\)00067-4](https://doi.org/10.1016/s0965-1748(01)00067-4) PMID: 11583933
41. Shukla JN, Jadhav S, Nagaraju J. Novel female-specific splice form of *dsx* in the silkworm, *Bombyx mori*. *Genetica.* 2011; 139(1):23–31. <https://doi.org/10.1007/s10709-010-9479-3> PMID: 20714790
42. Chang AYF, Liao BY. Reduced translational efficiency of eukaryotic genes after duplication events. *Mol Biol Evol.* 2020; 37(5):1452–1461. <https://doi.org/10.1093/molbev/msz309> PMID: 31904835
43. Zhao Q, Li J, Wen MY, Wang H, Wang Y, Wang KX, et al. A novel splice variant of the masculinizing gene *Masc* with piRNA-cleavage-site defect functions in female external genital development in the silkworm, *Bombyx mori*. *Biomolecules.* 2019; 9(8):318. <https://doi.org/10.3390/biom9080318> PMID: 31366115
44. Gopinath G, Arunkumar KP, Mita K, Nagaraju J. Role of *Bmzmf-2*, a *Bombyx mori* CCCH zinc finger gene, in masculinisation and differential splicing of *Bmtra-2*. *Insect Biochem Mol Biol.* 2016; 75:32–44. <https://doi.org/10.1016/j.ibmb.2016.05.008> PMID: 27260399
45. Gerber HP, Seipel K, Georgiev O, Höfferer M, Hug M, Rusconi S, et al. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science.* 1994; 263(5148):808–811. <https://doi.org/10.1126/science.8303297> PMID: 8303297
46. Williamson MP. The structure and function of proline-rich regions in proteins. *Biochem J.* 1994; 297(2):249–260. <https://doi.org/10.1042/bj2970249> PMID: 8297327
47. Marec F. Genetic control of pest Lepidoptera: induction of sex-linked recessive lethal mutations in *Ephestia kuehniella* (Pyralidae). *Acta Entomol Bohemoslov.* 1990; 87(6):445–458.
48. Fuková I, Nguyen P, Marec F. Codling moth cytogenetics: karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. *Genome.* 2005; 48(6):1083–1092. <https://doi.org/10.1139/g05-063> PMID: 16391677
49. Challis RJ, Kumar S, Dasmahapatra KK, Jiggins CD, Blaxter M. Lepbase: the Lepidopteran genome database. *bioRxiv.* 2016. <https://doi.org/10.1101/056994>
50. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012; 28(12):1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
51. Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A.* 1988; 85(23):8998–9002. <https://doi.org/10.1073/pnas.85.23.8998> PMID: 2461560
52. Ferguson KB, Visser S, Dalíková M, Provozničková I, Urbaneja A, Pérez-Hedo M, et al. Jekyll or Hyde? The genome (and more) of *Nesidiocoris tenuis*, a zoophytophagous predatory bug that is both a biological control agent and a pest. *Insect Mol Biol.* 2021; 30(2):188–209. <https://doi.org/10.1111/imb.12688> PMID: 33305885
53. Buntrock L, Marec F, Krueger S, Traut W. Organ growth without cell division: somatic polyploidy in a moth, *Ephestia kuehniella*. *Genome.* 2012; 55(11):755–763. <https://doi.org/10.1139/g2012-060> PMID: 23057509
54. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017; 27(5):722–736. <https://doi.org/10.1101/gr.215087.116> PMID: 28298431
55. Frydrychová R, Marec F. Repeated losses of TTAGG telomere repeats in evolution of beetles (Coleoptera). *Genetica.* 2002; 115(2):179–187. <https://doi.org/10.1023/a:1020175912128> PMID: 12403172
56. Dalíková M, Zrzavá M, Kubičková S, Marec F. W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). *Chromosome Res.* 2017; 25(3–4):241–252. <https://doi.org/10.1007/s10577-017-9558-8> PMID: 28500471
57. Hejničková M, Koutecký P, Potocký P, Provozničková I, Voleníková A, Dalíková M, et al. Absence of W chromosome in Psychidae moths and implications for the theory of sex chromosome evolution in Lepidoptera. *Genes.* 2019; 10(12):1016. <https://doi.org/10.3390/genes10121016> PMID: 31817557
58. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
59. Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschli T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* 2001; 20(23):6877–6888. <https://doi.org/10.1093/emboj/20.23.6877> PMID: 11726523

60. Pei Y, Tuschl T. On the art of identifying effective and specific siRNAs. *Nat Methods*. 2006; 3(9):670–676. <https://doi.org/10.1038/nmeth911> PMID: [16929310](https://pubmed.ncbi.nlm.nih.gov/16929310/)
61. Wickham H. *ggplot2: Elegant graphics for data analysis*. 2nd ed. New York: Springer-Verlag; 2016.