

Current Biology

***femaleless* Controls Sex Determination and Dosage Compensation Pathways in Females of *Anopheles* Mosquitoes**

Highlights

- *fle* is a new sex determination pathway element conserved in *Anopheles* mosquitoes
- *fle* may have originated in the *Anopheles* lineage and is highly conserved in *Anopheles*
- *fle* suppresses activation of dosage compensation in females
- Depletion of *fle* transcripts is lethal or otherwise deleterious to females

Authors

Elzbieta Krzywinska, Luca Ferretti, Jianwei Li, Jian-Chuan Li, Chun-Hong Chen, Jaroslaw Krzywinski

Correspondence

jaroslaw.krzywinski@pirbright.ac.uk

In Brief

Krzywinska et al. identify a new element of the sex determination pathway in *Anopheles*. *femaleless* (*fle*), in addition to controlling splicing of *dsx* and *fru*, is essential for suppression of dosage compensation and viability of females. Female-deleterious effects upon *fle* knockdown make *fle* a promising target for control of malaria mosquitoes.



Report

femaleless Controls Sex Determination and Dosage Compensation Pathways in Females of *Anopheles* Mosquitoes

Elzbieta Krzywinska,¹ Luca Ferretti,² Jianwei Li,¹ Jian-Chiuan Li,³ Chun-Hong Chen,^{3,4} and Jaroslaw Krzywinski^{1,5,*}

¹The Pirbright Institute, Ash Road, Pirbright, Surrey GU24 0NF, UK

²Big Data Institute, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK

³National Institute of Infectious Diseases and Vaccinology, National Health Research Institutes, 35 Keyan Road, Zhunan, Miaoli 350401, Taiwan

⁴National Mosquito-Borne Disease Control Research Center, National Health Research Institutes, 35 Keyan Road, Zhunan, Miaoli 350401, Taiwan

⁵Lead Contact

*Correspondence: jaroslaw.krzywinski@pirbright.ac.uk

<https://doi.org/10.1016/j.cub.2020.12.014>

SUMMARY

The insect sex determination and the intimately linked dosage compensation pathways represent a challenging evolutionary puzzle that has been solved only in *Drosophila melanogaster*. Analyses of orthologs of the *Drosophila* genes identified in non-drosophilid taxa^{1,2} revealed that evolution of sex determination pathways is consistent with a bottom-up mode,³ where only the terminal genes within the pathway are well conserved. *doublesex* (*dsx*), occupying a bottom-most position and encoding sex-specific proteins orchestrating downstream sexual differentiation processes, is an ancient sex-determining gene present in all studied species.^{2,4,5} With the exception of lepidopterans, its female-specific splicing is known to be regulated by *transformer* (*tra*) and its co-factor *transformer-2* (*tra2*).^{6–20} Here we show that in the African malaria mosquito *Anopheles gambiae*, a gene, which likely arose in the *Anopheles* lineage and which we call *femaleless* (*fle*), controls sex determination in females by regulating splicing of *dsx* and *fruitless* (*fru*; another terminal gene within a branch of the sex determination pathway). Moreover, *fle* represents a novel molecular link between the sex determination and dosage compensation pathways. It is necessary to suppress activation of dosage compensation in females, as demonstrated by the significant upregulation of the female X chromosome genes and a correlated female-specific lethality, but no negative effect on males, in response to *fle* knockdown. This unexpected property, combined with a high level of conservation in sequence and function in anopheline mosquitoes, makes *fle* an excellent target for genetic control of all major vectors of human malaria.

RESULTS AND DISCUSSION

fle Is a Sex Determination Pathway Component in *Anopheles*

In the mosquitoes from the genus *Anopheles*, sex is chromosomally determined, with the XX individuals developing as females and the XY individuals developing into males. Three components from the molecular pathway controlling this process have been identified in the African malaria mosquito *Anopheles gambiae*: a Y-chromosome-linked primary signal gene, *Yob*, conferring maleness^{21,22} and, on the opposite end, *dsx*²³ and *fru*.²⁴ To search for molecules transducing the sex-determining instruction, we queried the *A. gambiae* genome with the *D. melanogaster* Tra2 sequence using a BLAST algorithm. Tra2 is an RNA-binding protein with a characteristic structure consisting of an RNA recognition motif (RRM) flanked by two arginine-rich/serine-rich (RS) domains containing multiple serine-arginine dipeptides.²⁵ In all significant BLAST hits, sequence similarity

was limited to the RRM region. The top two hits, AGAP029421 (e value = 1e–28) and AGAP006798 (e value = 1e–25), encoding proteins with a structure typical of Tra2, are regarded here as the *tra2* homologs, named, respectively, *tra2b* and *tra2a*. To assess whether they are involved in sex determination, we investigated the effect of their transient knockdown on *dsx* splicing in the *A. gambiae* Sua5.1 female-like cell line. Transfection experiments using the *in-vitro*-synthesized double-stranded RNA (dsRNA) of either *tra2a* or *tra2b* did not yield the expected change in the *dsx* splicing pattern from the female to the male mode, indicating that neither has the sex-determining role (Figures S1A and S1B), consistent with the findings that *tra2* homologs have no role in *dsx* splicing in a culicine mosquito, *Aedes albopictus*.²⁶ In contrast, knockdown of the third-best hit, AGAP013051 (e value = 5e–18), caused a clear decrease of the female, and gain of the male, *dsx* transcript isoforms, as compared with the control non-transfected Sua5.1 cells (Figures 1A, 1B, and S1C). This result, further validated in transgenic



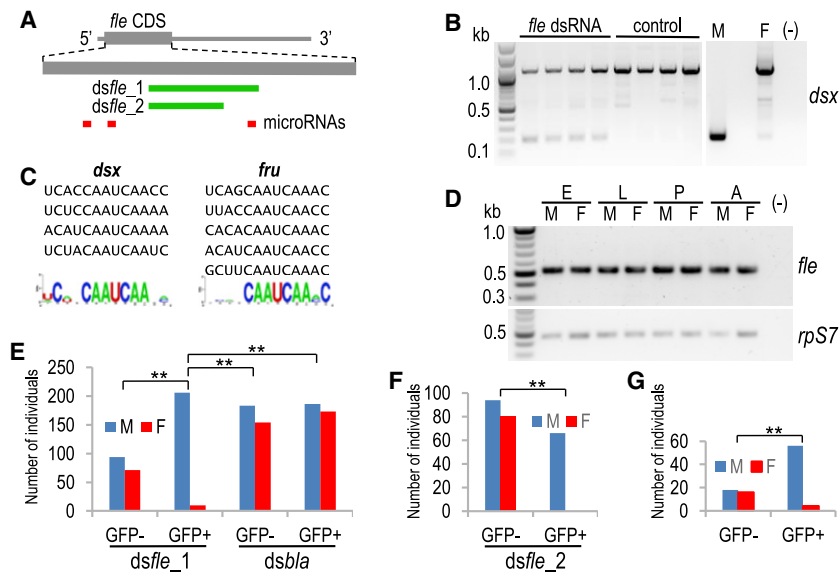


Figure 1. *fle* Is a Sex Determination Gene Necessary for *dsx* Splicing and Survival in *Anopheles* Females

(A) A schematic of the *A. gambiae fle* transcript, with the location of RNAi targets within the coding sequence.

(B) RT-PCR analysis of the *dsx* splicing pattern in *A. gambiae* Sua5.1 cells transfected with *in vitro*-synthesized *dsfle_1* dsRNA, as compared with control non-transfected cells.

(C) Sequences of the putative Fle binding sites in the *A. gambiae dsx* and *fru* female-specific exons. Presented are 13-nt fragments with similarity to the *Drosophila* repeat elements [TC(T/A)(T/A)CAATCAACA].

(D) RT-PCR analysis of *fle* transcription during *A. gambiae* development. E, embryos; L, larvae; P, pupae; A, adults; (-), negative control; M, male; F, female. Ribosomal protein S7 (*rpS7*) transcript levels were used as a gel loading control.

(E and F) Knockdown of *fle* expression causes female-specific lethality in *A. gambiae* embryos. A summary of three independent microinjection experiments using *dsfle_1* (E) and *dsfle_2* (F) dsRNA

fragments, with *bla* dsRNA used as control. GFP-/+ denote cohorts of individuals scored during the first larval instar as GFP-negative or GFP-positive. ***p* < 0.0001; Fisher's exact test.

(G) Knockdown of *fle* ortholog expression causes female-specific lethality in *A. stephensi* embryos. A summary of two independent *fle* dsRNA microinjection experiments. GFP-/+ as in (E). ***p* < 0.0001; Fisher's exact test.

See also Figures S1–S3.

A. gambiae lines with a stable AGAP013051 knockdown (see below), demonstrated that AGAP013051 represents a sex determination pathway element regulating *dsx* splicing in *A. gambiae* females.

Apart from similarity within the RRM, AGAP013051 substantially differs from insect *tra2* orthologs in the length of the coding region (420 amino acids compared to 232–299 amino acids in the described *Tra2* proteins)^{10,11,14,16,27–30} and structure of the encoded protein, including lack of typical RS domains and presence of two additional putative RRM regions in the N-terminal half of the protein (Figures S2A and S2B). Combined with the results of a phylogenetic analysis, in which AGAP013051 does not cluster with *Tra2* proteins but is grouped with a putative splicing factor, *Nix*, functioning as the male determiner in *Aedes aegypti*³¹ (Figures S2C and S2D), the above indicates that AGAP013051 and *tra2* may be only distantly related. We named the gene *femaleless (fle)*, to reflect the associated knockdown phenotypes described below.

The RRM is a highly abundant motif in various eukaryotic proteins. It folds into a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ sandwich topology, with the β_1 and β_3 sheets encompassing, respectively, ribonucleoprotein 2 (RNP2) and RNP1 sequence elements that are required for specific binding to the RNA sequences.^{25,32,33} In *Drosophila*, *Tra2* molecules bind six 13-nt repeat elements [UC(U/A)(U/A)CAUCAACA], clustered on the *dsx* pre-mRNA within the untranslated region (UTR) of the female-specific *dsx* exon 4, to promote the use of an adjacent upstream weak splice acceptor site, leading to splicing of *dsx* into the female form.^{8,34,35} *fle* may directly target a (U/A)C(U/A)(U/C/A)CAAUCA(U/C/A)(C/A) sequence, which forms a four-repeat cluster within the UTR of the *A. gambiae* female-specific *dsx* exon 5 (Figure 1C). These putative *fle* targets map to regions of very low nucleotide diversity in natural *A. gambiae* populations³⁶ and coincide with blocks of

increased conserved sequence across *Anopheles* species (data not shown), which lends support to the notion of their functional significance. The *A. gambiae fru* gene, whose sex-specific splicing is also *fle* dependent (this study; see below), contains three putative *fle* binding sites clustered at the 5' end of the female-specific exon 3 and two binding sites in the sex-specifically processed exon 4. Their consensus sequence shares the same invariant core (CAATCAA) with the *A. gambiae dsx* repeat elements, as well as with the counterparts from the *Drosophila dsx* and *fru* (Figure 1C). Considerable variation within the nucleotide positions flanking the core suggests that only the core may be important for efficient *fle* binding.

We queried the NCBI whole-genome shotgun contigs database with the *fle* amino acid sequence using tBLASTn to evaluate the phylogenetic distribution of the *fle* orthologs. Significant hits were identified only in the genomes of genus *Anopheles* representatives, with highly conserved 1-to-1 orthologs detected in each case (Figure S3A), indicating that *fle* has been subject to strong functional constraints throughout approximately 100 million years of *Anopheles* lineage evolution.^{37,38} A lack of discernible orthologs beyond anopheline genomes suggests that *fle* may have originated in a recent ancestor of the *Anopheles* lineage. To test whether *fle* could have arisen through a gene duplication, we searched the *A. gambiae* annotated proteins for sequences similar to Fle. The top hit was AGAP001643 (the gene restricted to the family Culicidae [mosquitoes]), which at the amino acid level shares a significant similarity (BLASTp, *e* value = 4e–20) to the N-terminal half of Fle and a lower similarity to the C-terminal RRM (Figures S3B and S3C). The first intron in AGAP001643 and the intron in *fle* split a codon located at the putative homologous alignment position, further suggesting a common origin of the two genes. Thus, *fle* may represent an ancient paralog of AGAP001643, which after

a gene duplication rapidly diverged to assume essential developmental functions. In that case, the similarity between Tra2 and Fle within the RRM could represent a case of convergent evolution driven by the adaptation of *fle* to efficiently bind highly conserved target sequences. Alternatively, *fle* could have originated from a fusion between an AGAP001643 copy and a *tra2* copy, followed by subsequent rearrangements. A likely evolutionary scenario cannot be reliably inferred because the sequences in question are highly divergent.

Transient *fle* Knockdown Kills *Anopheles* Female Embryos

The *fle* gene produces a single transcript and is constitutively expressed throughout development in both *A. gambiae* sexes (Figure 1D). To assess whether a transient *fle* knockdown has an effect on *A. gambiae* early development, we injected the *fle* dsRNA into non-sexed preblastoderm embryos using an established protocol.²¹ The injection mix contained a control plasmid with a green fluorescent protein (GFP) expression cassette, which allowed for an easy identification of individuals that received sufficient amounts of nucleic acids. Surviving first-instar larvae were sorted into GFP-negative and GFP-positive cohorts, and at the pupal stage mosquitoes were sexed. A very strong male bias was observed among the GFP-positive mosquitoes (Figures 1E and 1F), whereas in the GFP-negative group and in the control group of *A. gambiae* embryos injected with heterologous *bla* dsRNA the sex ratio was not significantly biased. A random sample of individuals (n = 20) from the strong GFP cohort were sexed using PCR²¹ and confirmed to have the XY karyotype (data not shown). Similar results, with a strongly male-biased survival, were obtained after microinjection of *Anopheles stephensi fle* dsRNA into embryos of *A. stephensi* (Figure 1G), a species that diverged from the *A. gambiae* lineage over 40 million years ago (mya).^{37,39} These results indicated that embryonic knockdown of *fle* is lethal to genetic females but apparently has no discernible effect on the development of genetic males. The female death was occurring during the embryo stage, because the numbers of hatched GFP-positive larvae were comparable to the numbers of pupating individuals.

fle Links the Sex Determination and Dosage Compensation Pathways

Using piggyBac-based transgenesis, we generated six *A. gambiae* lines that stably produce a polycistronic transcript encoding three microRNAs designed to silence expression of *fle* (Figure 2A). In each line, *fle* knockdown affected the development of females, with phenotypic effects defined by the genomic location of the transgene insertions (none known to interrupt a gene; Table S1). The abnormalities in sexually dimorphic characters ranged from mild, such as asymmetric development or loss of cerci (terminal appendages of the female abdomen) in line 4M6, to an extensive masculinization manifested in the development of a nearly normal male copulatory apparatus and male-like head appendages in line 4M4B (Figures 2B and 2C). The internal reproductive system was similarly affected, including underdeveloped or atrophied ovaries, atrophied spermatheca, partially developed male accessory glands, and a rudimentary ejaculatory pump (Figure 2D). The level of masculinization of transgenic females was correlated with a substantially altered splicing of *dsx* (similar to the effects of *fle*

knockdown in Sua5.1 cells) and, additionally, *fru* (Figure 2E), whose sex-specific splicing is conserved between *Anopheles* and *Drosophila*.²⁴ In *Drosophila*, female-specific *fru* transcripts do not code for a functional protein, whereas the male-specific isoforms encode a protein produced in a small subset of neurons in the central nervous system, where it regulates male sexual behavior⁴⁰ and specifies the development of a male-specific muscle of Lawrence (MOL) in the fifth abdominal segment.⁴¹ The *fru* ortholog apparently performs the same function in *A. gambiae*.²⁴ Consistent with that notion, we found a male-like, bilaterally paired muscle reminiscent of MOL in the fifth abdominal segment of the *A. gambiae* transgenic females (Figure 2F). Moreover, in none of the lines were transgenic females attracted to a blood source, likely because they produce the male form of *fru*.

Beyond masculinization, *fle* depletion caused partial or complete female lethality in some lines (Figure 2G). In line 4M2B, only one-third of hatched larvae at each generation were transgenic. They developed exclusively as males, which indicated that transgenic females were dying during embryonic stage, consistent with the female-specific embryonic lethality in transient *fle* knockdown experiments. In line 4M4B, female death was occurring during late developmental stages; a small proportion of individuals that survived to adulthood died shortly (up to 3 days) after eclosion. The majority of females dying as pupae developed a distinctive black tumor at the abdominal terminus (Figure 2H). In contrast to females, males appeared not to be affected in any of the transgenic lines.

The severity of the aberrant female development in transgenic lines was roughly correlated with the level of expression of the GFP transformation marker (Figure 2I) and with the extent of the *fle* knockdown (Figure 2J) in pupae. Line 4M2B, with an invariable female embryonic-lethal phenotype but with a relatively low GFP expression and a low *fle* knockdown level in male pupae, represented a notable exception, suggesting an involvement of an embryonic enhancer driving higher microRNA expression and a resulting increased *fle* knockdown during early development. Overall, the observed levels of knockdown were surprisingly low, even in a line exhibiting strong masculinization (Figure 2J), suggesting that *fle* is a haploinsufficient gene. Presumably, *fle* knockdown below a certain critical threshold leads to female lethality.

Sex-specific lethality caused by loss-of-function mutations or knockdown of sex determination genes results from misregulation of dosage compensation in *Drosophila* and a silk moth, *Bombyx mori*.^{19,42–44} In *Drosophila*, dosage compensation relies on a two-fold upregulation of transcription from the single male X chromosome to the levels of expression from both X chromosomes in females.⁴⁵ Dosage compensation machinery is not assembled in *Drosophila* females, because SXL, a female-specific protein involved in sex determination, prevents translation of MSL-2, a key protein of the dosage compensation complex.⁴⁶ Mutations in *Sxl*, or in genes involved in *Sxl* regulation, lead to overexpression of X linked genes and female death during embryogenesis.^{42,43,47,48} In *A. gambiae*, dosage compensation also operates by upregulation of the X chromosome in males and that process is controlled by the primary sex determiner gene *Yob*.^{21,49} Similarly, in *A. stephensi*, the X chromosome dosage in males is regulated by the Y linked maleness gene *guy-1*.⁵⁰ Female-specific lethality observed in this study

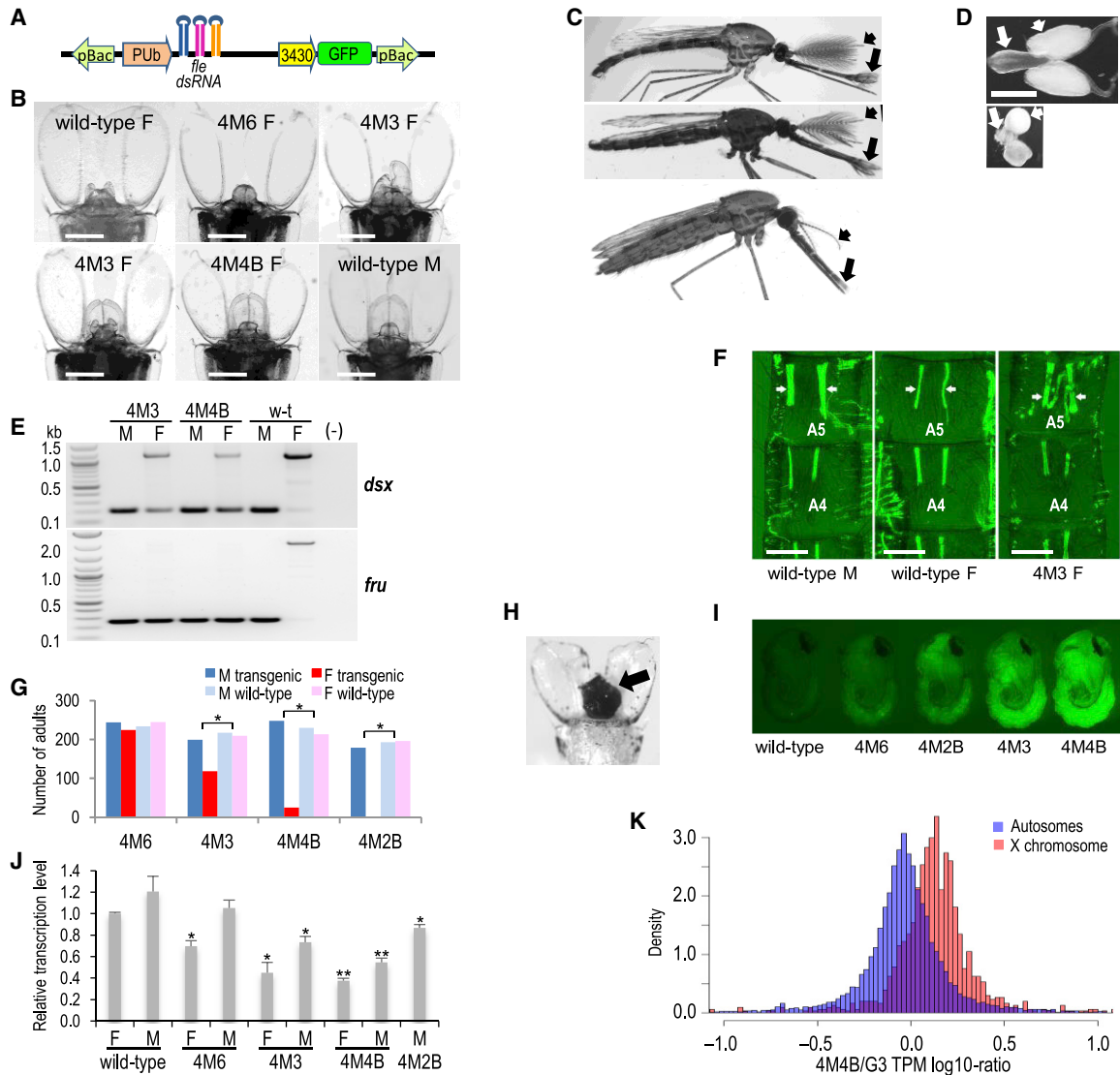


Figure 2. Stable *fle* Knockdown in *A. gambiae* Causes Female Masculinization and Death through Abnormal Upregulation of Transcription from the X Chromosome

(A) Schematic representation of the transgenic *fle* knockdown construct.

(B) Pupal abdominal termini of the wild-type and *fle* knockdown transgenic mosquito lines. Scale bars, 200 μ m.

(C) Adults. From top: wild-type male, 4M4B line female, wild-type female. Note male-like antennae (short arrows) and palps (long arrows), and a slender abdomen with a male-like uninverted copulatory apparatus.

(D) Reproductive system: ejaculatory pump (long white arrows) and accessory glands (short white arrows) in a wild-type male and 4M4B female. Scale bar, 100 μ m.

(E) RT-PCR analysis of *dsx* and *fru* splicing patterns in females from selected transgenic lines.

(F) A sexually dimorphic pair of dorsal muscles (indicated by white arrows) in the fifth segment of the abdomen (A5), likely homologous to the *Drosophila* muscle of Lawrence, is of similar thickness in wild-type males and transgenic females, and considerably thicker than in wild-type females. Scale bars, 200 μ m.

(G) Proportions of individuals reaching adulthood in selected transgenic lines. The lines have been maintained by backcrossing of transgenic males with the wild-type line females at every generation, which should yield 50% of wild-type progeny and equal proportions of the sexes in the transgenic and wild-type groups. However, apart from line 4M6, a highly significant deficiency of transgenic females was observed ($p < 0.0001$; chi-square test). In line 4M2B, no transgenic females developed beyond the embryonic stage.

(H) Pupal abdominal terminus with a characteristic black tumor (arrow) in a 4M4B female.

(I) Fluorescent marker expression in selected transgenic lines.

(J) Levels of *fle* transcripts in the pupae of the wild-type line G3 and in selected transgenic lines quantified by qRT-PCR and normalized relative to the *fle/rpS7* transcription level in wild-type females. The error bars represent standard deviations. $*p < 0.01$, $**p < 0.001$; Student's t test.

(K) Comparison of transcription levels from the autosomes and the X chromosomes in transgenic line 4M4B and wild-type line G3 females. Shown are transcripts per kilobase million (TPM) ratios. The median expression ratio between the 4M4B and G3 lines for the autosomes (0.92) and for the X chromosomes (1.34) indicates that the X chromosome genes in the transgenic line females are significantly overexpressed; $p < 10^{-15}$; Mann-Whitney U test. A similar trend for the median shrunken fold change for the autosomal (0.985) and for the X linked (1.22) transcripts confirms that the effect is statistically robust.

See also Table S1.

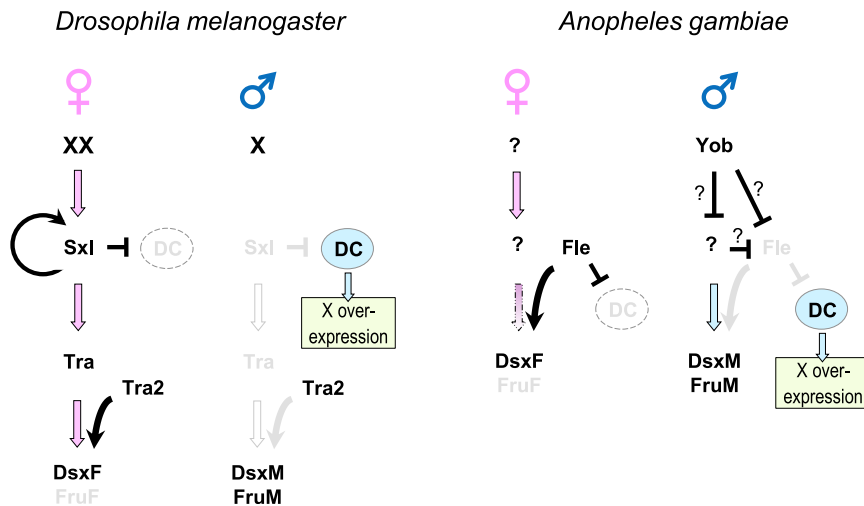


Figure 3. Proteins Determining Sex in Somatic Tissues of *D. melanogaster* and *A. gambiae*

In fruit fly females, a double dose of X linked elements activates productive transcription of *Sxl*, leading to establishment of a positive feedback loop of *Sxl* production. *Sxl* controls splicing of *tra* into a productive form, which, together with *Tra2*, is necessary for splicing of *dsx* into a productive form and *fru* into a non-productive form. *Sxl* also prevents assembly of the dosage compensation (DC) complex by blocking translation of *msl2*, a critical component of the complex. In fruit fly males, a single dose of X linked proteins is insufficient to initiate a productive transcription of *Sxl*. As a result, dosage compensation is activated, *tra* is spliced into a non-productive form, and the default splicing of *dsx* and *fru* productive male forms occurs. In *Anopheles* females, *Fle* is necessary for the splicing of *dsx* and *fru* into productive and non-productive forms, respectively,

as well as for repression of dosage compensation. The upstream sex-determining molecules remain to be identified; it is unclear whether *Fle* requires a cofactor to promote splicing of *dsx* and *fru*. In males, the primary sex determiner *Yob* triggers the sex determination pathway and inactivates *Fle* by a yet unknown mechanism, allowing activation of dosage compensation. *Fle* does not take part in the splicing of *dsx* or *fru* into male forms.

suggests that dosage compensation is activated in *Anopheles* females in response to depletion of *fle* transcripts. To evaluate whether misregulation of the X chromosome transcription is indeed involved, we compared transcriptomes of female pupae from a wild-type line and from the 4M4B transgenic line, in which female lethal effects occur during late stages of development. Relative to the autosomes, transcription from the X chromosomes in transgenic females was significantly upregulated (Mann-Whitney U test, $p < 10^{-15}$), with an overall transcription increase by more than 40% (Figure 2K), which is apparently toxic and leads to preimaginal death. The X-chromosome-wide overtranscription caused by misregulation of dosage compensation mechanisms mimics X chromosome aneuploidy, and is known to lead to tumorigenesis in *Drosophila* and cancers in mice,^{51,52} and apparently is the cause of abdominal tumors developing in the 4M4B female pupae (Figure 2G). However, tumors in the genital area were also observed in a moth, *Plutella xylostella*, female mutants with disrupted *dsxF* transcripts.⁵³

Control of the *dsx* and *fru* female-specific splicing by a non-sex-specifically expressed *fle* provokes a question about the mechanism of that process. In most studied holometabolous insects, the female specificity of *dsx* splicing relies on *Tra*, which is produced only in females, and whose interaction with *Tra2* via RS domains is necessary to stabilize the spliceosome assembly on the weak acceptor splice site.^{8,9,11,14,16,25} Considering an apparent absence of *tra* in the *A. gambiae* genome,^{2,54} as well as a larger size and a structural dissimilarity of *Fle* as compared to *Tra2*, it is conceivable that *Fle* does not have an obligatory *Tra*-like partner and that binding of *Fle* alone to its putative targets (Figure 1C) is sufficient to promote the spliceosome assembly in females, whereas in males that process is interrupted by *Yob* (Figure 3).

In summary, in addition to a vital role in the splicing of *dsx* and *fru* into a female form, *fle* represses dosage compensation in *A. gambiae* females through a yet unknown mechanism. As such, *fle* is only the second known, after *Sxl* in *Drosophila*,

molecular link between the sex determination and dosage compensation pathways in insects (Figure 3). Female lethality in response to *fle* knockdown indicates that dosage compensation is regulated by *fle* also in *A. stephensi*.

fle as a Target Molecule for *Anopheles* Control

Mosquitoes from the genus *Anopheles* are the exclusive vectors of *Plasmodium* parasites that cause human malaria, a disease infecting annually nearly 230 million people and causing over 400,000 deaths worldwide.⁵⁵ Over 90% of malaria cases occur in sub-Saharan Africa, where *A. gambiae* is the primary vector. Control of the disease relies heavily on the use of insecticides, which are increasingly ineffective because of the emergence and spread of insecticide resistance in natural *Anopheles* populations.^{55,56} Various genetic control approaches, proposed to complement the existing insecticidal tools,^{57,58} rely on mass releases of irradiated or otherwise genetically modified males, which through mating with wild-type females spread desirable traits, such as sterility, female lethality, or inability to transmit pathogens, to cause mosquito population suppression or modification. The requirement for male-only releases is dictated by ethical and safety considerations, because only females feed on blood, and released modified females could contribute to biting and parasite transmission. To date, maleness genes *Yob* and *guy-1* have been found to have female-killing properties,^{21,59} and thus could be used to conditionally eliminate females of the *A. gambiae* species complex and *A. stephensi*, respectively, in genetic control operations.⁵⁸ This study, in addition to advancing our understanding of sex determination and dosage compensation regulation in insects, identifies *fle* as a universal molecule, conserved in sequence and function in anopheline mosquitoes, that could be targeted in genetic control to eliminate females of all major malaria vector species. The apparent haploinsufficiency makes *fle* unsuitable as a target of homing gene drives; however, female-specific deleterious effects could be exploited in conditional *fle* knockdown transgenic lines to suppress populations of various *Anopheles* vector species.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Mosquito strains
 - Mosquito cells
- METHOD DETAILS
 - Sequence analyses
 - dsRNA synthesis
 - Transfection experiments
 - Analysis of sex-specific splicing
 - Transient gene silencing
 - Transgenic construct plasmid and transgenesis
 - Identification of transgene insertion sites
 - Abdominal musculature
 - Real-time PCR
 - RNA-seq analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2020.12.014>.

ACKNOWLEDGMENTS

We thank Neil Hobbs and Aqib Ali for their technical support in maintenance of mosquito colonies. This study was supported by funds from UK Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/P019269/1 (to J.K.), and through strategic BBSRC grants BBS/E/1/00007033, BBS/E/1/00007038, and BBS/E/1/00007039 to the Pirbright Institute.

AUTHOR CONTRIBUTIONS

E.K. and J.K. conceived the study; E.K. performed molecular biology, cell line, and mosquito experiments, with contributions from J.K.; J.L. designed and conducted qRT-PCR experiments; E.K., J.-C.L., and C.-H.C. designed and generated plasmid constructs; L.F. performed RNA-sequencing (RNA-seq) data analysis; E.K. and J.K. wrote the paper with input from the other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 5, 2020

Revised: November 9, 2020

Accepted: December 11, 2020

Published: January 7, 2021

REFERENCES

1. Saccone, G., Pane, A., and Polito, L.C. (2002). Sex determination in flies, fruitflies and butterflies. *Genetica* *116*, 15–23.
2. Geuverink, E., and Beukeboom, L.W. (2014). Phylogenetic distribution and evolutionary dynamics of the sex determination genes *doublesex* and *transformer* in insects. *Sex Dev.* *8*, 38–49.
3. Wilkins, A.S. (1995). Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *BioEssays* *17*, 71–77.
4. Wexler, J., Delaney, E.K., Belles, X., Schal, C., Wada-Katsumata, A., Amicucci, M.J., and Kopp, A. (2019). Hemimetabolous insects elucidate the origin of sexual development via alternative splicing. *eLife* *8*, e47490.
5. Verhulst, E.C., and van de Zande, L. (2015). Double nexus—*Doublesex* is the connecting element in sex determination. *Brief. Funct. Genomics* *14*, 396–406.
6. Amrein, H., Maniatis, T., and Nöthiger, R. (1990). Alternatively spliced transcripts of the sex-determining gene *tra-2* of *Drosophila* encode functional proteins of different size. *EMBO J.* *9*, 3619–3629.
7. Hedley, M.L., and Maniatis, T. (1991). Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to *tra-2* protein *in vitro*. *Cell* *65*, 579–586.
8. Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H., and Shimura, Y. (1992). Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of *doublesex* primary transcript for sex-specific RNA processing. *Proc. Natl. Acad. Sci. USA* *89*, 8092–8096.
9. Pane, A., Salvemini, M., Delli Bovi, P., Polito, C., and Saccone, G. (2002). The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. *Development* *129*, 3715–3725.
10. Burghardt, G., Hediger, M., Siegenthaler, C., Moser, M., Dübendorfer, A., and Bopp, D. (2005). The *transformer2* gene in *Musca domestica* is required for selecting and maintaining the female pathway of development. *Dev. Genes Evol.* *215*, 165–176.
11. Salvemini, M., Robertson, M., Aronson, B., Atkinson, P., Polito, L.C., and Saccone, G. (2009). *Ceratitis capitata transformer-2* gene is required to establish and maintain the autoregulation of *Cctra*, the master gene for female sex determination. *Int. J. Dev. Biol.* *53*, 109–120.
12. Hediger, M., Henggeler, C., Meier, N., Perez, R., Saccone, G., and Bopp, D. (2010). Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. *Genetics* *184*, 155–170.
13. Verhulst, E.C., Beukeboom, L.W., and van de Zande, L. (2010). Maternal control of haplodiploid sex determination in the wasp *Nasonia*. *Science* *328*, 620–623.
14. Geuverink, E., Rensink, A.H., Rondeel, I., Beukeboom, L.W., van de Zande, L., and Verhulst, E.C. (2017). Maternal provision of *transformer-2* is required for female development and embryo viability in the wasp *Nasonia vitripennis*. *Insect Biochem. Mol. Biol.* *90*, 23–33.
15. Shukla, J.N., and Palli, S.R. (2012). Sex determination in beetles: production of all male progeny by parental RNAi knockdown of *transformer*. *Sci. Rep.* *2*, 602.
16. Shukla, J.N., and Palli, S.R. (2013). *Tribolium castaneum* Transformer-2 regulates sex determination and development in both males and females. *Insect Biochem. Mol. Biol.* *43*, 1125–1132.
17. Xu, J., Chen, S., Zeng, B., James, A.A., Tan, A., and Huang, Y. (2017). *Bombyx mori* P-element somatic inhibitor (*BmPSI*) is a key auxiliary factor for silkworm male sex determination. *PLoS Genet.* *13*, e1006576.
18. Suzuki, M.G., Imanishi, S., Dohmae, N., Nishimura, T., Shimada, T., and Matsumoto, S. (2008). Establishment of a novel *in vivo* sex-specific splicing assay system to identify a *trans*-acting factor that negatively regulates splicing of *Bombyx mori dsx* female exons. *Mol. Cell. Biol.* *28*, 333–343.
19. Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., et al. (2014). A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* *509*, 633–636.
20. Zheng, Z.Z., Sun, X., Zhang, B., Pu, J., Jiang, Z.Y., Li, M., Fan, Y.J., and Xu, Y.Z. (2019). Alternative splicing regulation of *doublesex* gene by RNA-binding proteins in the silkworm *Bombyx mori*. *RNA Biol.* *16*, 809–820.
21. Krzywinska, E., Dennison, N.J., Lycett, G.J., and Krzywinski, J. (2016). A maleness gene in the malaria mosquito *Anopheles gambiae*. *Science* *353*, 67–69.

22. Krzywinska, E., and Krzywinski, J. (2018). Effects of stable ectopic expression of the primary sex determination gene *Yob* in the mosquito *Anopheles gambiae*. *Parasit. Vectors* *11* (Suppl 2), 648.
23. Scali, C., Catteruccia, F., Li, Q., and Crisanti, A. (2005). Identification of sex-specific transcripts of the *Anopheles gambiae doublesex* gene. *J. Exp. Biol.* *208*, 3701–3709.
24. Gailey, D.A., Billeter, J.C., Liu, J.H., Bauzon, F., Allendorfer, J.B., and Goodwin, S.F. (2006). Functional conservation of the *fruitless* male sex-determination gene across 250 Myr of insect evolution. *Mol. Biol. Evol.* *23*, 633–643.
25. Amrein, H., Hedley, M.L., and Maniatis, T. (1994). The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by *Transformer 2*. *Cell* *76*, 735–746.
26. Li, X., Jin, B., Dong, Y., Chen, X., Tu, Z., and Gu, J. (2019). Two of the three *Transformer-2* genes are required for ovarian development in *Aedes albopictus*. *Insect Biochem. Mol. Biol.* *109*, 92–105.
27. Amrein, H., Gorman, M., and Nöthiger, R. (1988). The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA binding protein. *Cell* *55*, 1025–1035.
28. Martín, I., Ruiz, M.F., and Sánchez, L. (2011). The gene *transformer-2* of *Sciara* (Diptera, Nematocera) and its effect on *Drosophila* sexual development. *BMC Dev. Biol.* *11*, 19.
29. Nissen, I., Müller, M., and Beye, M. (2012). The *Am-tra2* gene is an essential regulator of female splice regulation at two levels of the sex determination hierarchy of the honeybee. *Genetics* *192*, 1015–1026.
30. Concha, C., and Scott, M.J. (2009). Sexual development in *Lucilia cuprina* (Diptera, Calliphoridae) is controlled by the *transformer* gene. *Genetics* *182*, 785–798.
31. Hall, A.B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V.A., Biedler, J.K., Sharakhova, M.V., Elahi, R., Anderson, M.A., Chen, X.G., et al. (2015). A male-determining factor in the mosquito *Aedes aegypti*. *Science* *348*, 1268–1270.
32. Dreyfuss, G., Swanson, M.S., and Piñol-Roma, S. (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.* *13*, 86–91.
33. Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* *272*, 2118–2131.
34. Burtis, K.C., and Baker, B.S. (1989). *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* *56*, 997–1010.
35. Lynch, K.W., and Maniatis, T. (1995). Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.* *9*, 284–293.
36. *Anopheles gambiae* 1000 Genomes Consortium (2020). Genome variation and population structure among 1142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*. *Genome Res.* *30*, 1533–1546.
37. Neafsey, D.E., Waterhouse, R.M., Abai, M.R., Aganezov, S.S., Alekseyev, M.A., Allen, J.E., Amon, J., Arcà, B., Arensburger, P., Artemov, G., et al. (2015). Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* *347*, 1258522.
38. Krzywinski, J., Grushko, O.G., and Besansky, N.J. (2006). Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol. Phylogenet. Evol.* *39*, 417–423.
39. Moreno, M., Marinotti, O., Krzywinski, J., Tadei, W.P., James, A.A., Achee, N.L., and Conn, J.E. (2010). Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. *Malar. J.* *9*, 127.
40. Ryner, L.C., Goodwin, S.F., Castrillon, D.H., Anand, A., Villella, A., Baker, B.S., Hall, J.C., Taylor, B.J., and Wasserman, S.A. (1996). Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* *87*, 1079–1089.
41. Gailey, D.A., Taylor, B.J., and Hall, J.C. (1991). Elements of the *fruitless* locus regulate development of the muscle of Lawrence, a male-specific structure in the abdomen of *Drosophila melanogaster* adults. *Development* *113*, 879–890.
42. Lucchesi, J.C., and Skripsky, T. (1981). The link between dosage compensation and sex differentiation in *Drosophila melanogaster*. *Chromosoma* *82*, 217–227.
43. Gergen, J.P. (1987). Dosage compensation in *Drosophila*: evidence that *daughterless* and *Sex-lethal* control X chromosome activity at the blastoderm stage of embryogenesis. *Genetics* *117*, 477–485.
44. Cline, T.W., and Meyer, B.J. (1996). Vive la différence: males vs females in flies vs worms. *Annu. Rev. Genet.* *30*, 637–702.
45. Lucchesi, J.C., Kelly, W.G., and Panning, B. (2005). Chromatin remodeling in dosage compensation. *Annu. Rev. Genet.* *39*, 615–651.
46. Bashaw, G.J., and Baker, B.S. (1997). The regulation of the *Drosophila msl-2* gene reveals a function for *Sex-lethal* in translational control. *Cell* *89*, 789–798.
47. Cline, T.W. (1978). Two closely linked mutations in *Drosophila melanogaster* that are lethal to opposite sexes and interact with *daughterless*. *Genetics* *90*, 683–698.
48. Hilfiker, A., Amrein, H., Dübendorfer, A., Schneiter, R., and Nöthiger, R. (1995). The gene *virilizer* is required for female-specific splicing controlled by *Sxl*, the master gene for sexual development in *Drosophila*. *Development* *121*, 4017–4026.
49. Rose, G., Krzywinska, E., Kim, J., Revuelta, L., Ferretti, L., and Krzywinski, J. (2016). Dosage compensation in the African malaria mosquito *Anopheles gambiae*. *Genome Biol. Evol.* *8*, 411–425.
50. Qi, Y., Wu, Y., Saunders, R., Chen, X.G., Mao, C., Biedler, J.K., and Tu, Z.J. (2019). *Guy1*, a Y-linked embryonic signal, regulates dosage compensation in *Anopheles stephensi* by increasing X gene expression. *eLife* *8*, e43570.
51. Clemente-Ruiz, M., Murillo-Maldonado, J.M., Benhra, N., Barrio, L., Pérez, L., Quiroga, G., Nebreda, A.R., and Milán, M. (2016). Gene dosage imbalance contributes to chromosomal instability-induced tumorigenesis. *Dev. Cell* *36*, 290–302.
52. Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). *Xist* RNA is a potent suppressor of hematologic cancer in mice. *Cell* *152*, 727–742.
53. Wang, Y., Chen, X., Liu, Z., Xu, J., Li, X., Bi, H., Andongma, A.A., Niu, C., and Huang, Y. (2019). Mutation of *doublesex* induces sex-specific sterility of the diamondback moth *Plutella xylostella*. *Insect Biochem. Mol. Biol.* *112*, 103180.
54. Salvemini, M., D'Amato, R., Petrella, V., Aceto, S., Nimmo, D., Neira, M., Alphey, L., Polito, L.C., and Saccone, G. (2013). The orthologue of the fruit fly sex behaviour gene *fruitless* in the mosquito *Aedes aegypti*: evolution of genomic organisation and alternative splicing. *PLoS ONE* *8*, e48554.
55. World Health Organization (2019). World malaria report 2019. <https://apps.who.int/iris/handle/10665/330011>.
56. Ranson, H., and Lissenden, N. (2016). Insecticide resistance in African *Anopheles* mosquitoes: a worsening situation that needs urgent action to maintain malaria control. *Trends Parasitol.* *32*, 187–196.
57. Knippling, E.F. (1959). Sterile-male method of population control. *Science* *130*, 902–904.
58. Alphey, L. (2014). Genetic control of mosquitoes. *Annu. Rev. Entomol.* *59*, 205–224.
59. Criscione, F., Qi, Y., and Tu, Z. (2016). GUY1 confers complete female lethality and is a strong candidate for a male-determining factor in *Anopheles stephensi*. *eLife* *5*, e19281.
60. Müller, H.M., Dimopoulos, G., Blass, C., and Kafatos, F.C. (1999). A hemocyte-like cell line established from the malaria vector *Anopheles gambiae* expresses six prophenoloxidase genes. *J. Biol. Chem.* *274*, 11727–11735.
61. Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., et al. (2016). A

- CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* **34**, 78–83.
62. Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. USA* **104**, 3312–3317.
 63. Volohonsky, G., Terenzi, O., Soichot, J., Naujoks, D.A., Nolan, T., Windbichler, N., Kapps, D., Smidler, A.L., Vittu, A., Costa, G., et al. (2015). Tools for *Anopheles gambiae* transgenesis. *G3 (Bethesda)* **5**, 1151–1163.
 64. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
 65. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394.
 66. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.
 67. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874.
 68. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.
 69. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675.
 70. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527.
 71. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
 72. Zhu, A., Ibrahim, J.G., and Love, M.I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092.
 73. Benedict, M.Q. (1997). Care and maintenance of *Anopheline* mosquito colonies. In *The Molecular Biology of Insect Disease Vectors*, J.M. Crampton, C.B. Beard, and C. Louis, eds. (Chapman & Hall), pp. 3–12.
 74. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* **48** (D1), D265–D268.
 75. Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.
 76. Adolfi, A., Pondeville, E., Lynd, A., Bourgoïn, C., and Lycett, G.J. (2018). Multi-tissue GAL4-mediated gene expression in all *Anopheles gambiae* life stages using an endogenous polyubiquitin promoter. *Insect Biochem. Mol. Biol.* **96**, 1–9.
 77. Chen, C.H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L.V., Guo, M., and Hay, B.A. (2007). A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* **316**, 597–600.
 78. Lombardo, F., Lycett, G.J., Lanfrancotti, A., Coluzzi, M., and Arcà, B. (2009). Analysis of a pyrase 5′ upstream region validates improved *Anopheles gambiae* transformation technique. *BMC Res. Notes* **2**, 24.
 79. Mattox, W., McGuffin, M.E., and Baker, B.S. (1996). A negative feedback mechanism revealed by functional analysis of the alternative isoforms of the *Drosophila* splicing regulator transformer-2. *Genetics* **143**, 303–314.
 80. Potter, C.J., and Luo, L. (2010). Splinkerette PCR for mapping transposable elements in *Drosophila*. *PLoS ONE* **5**, e10168.
 81. Lycett, G.J., Amey, D., and Lynd, A. (2012). The *Anopheles gambiae* alpha-tubulin-1b promoter directs neuronal, testes and developing imaginal tissue specific expression and is a sensitive enhancer detector. *Insect Mol. Biol.* **21**, 79–88.
 82. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{−(Delta Delta C(T))} method. *Methods* **25**, 402–408.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
<i>E. coli</i> DH10B ElectroMax	Thermo Fisher Scientific	Cat#18290015
Chemicals, Peptides, and Recombinant Proteins		
Lipofectamine 2000 Transfection Reagent	Invitrogen	Cat#11668019
Schneider's Modified Medium	Lonza	Cat#04-351Q
Critical Commercial Assays		
PureLink RNA Micro Kit	Invitrogen	Cat#12183016
ActinGreen 488 ReadyProbes Reagent	ThermoFisher Scientific	Cat#R37110
LunaScript RT SuperMix Kit	New England Biolabs	Cat#E3010S
Luna Universal qPCR Master Mix	New England Biolabs	Cat#M3003S
DNeasy Blood & Tissue Kit	QIAGEN	Cat#69504
MEGAscript RNAi T7 Kit	Life Technologies	Cat#AM1626
One-Step RT-PCR System	Invitrogen	Cat#12574026
Deposited Data		
RNA-Seq data	This study	ENA: PRJEB38605
Experimental Models: Cell Lines		
<i>A. gambiae</i> Sua5.1	⁶⁰	N/A
Experimental Models: Organisms/Strains		
<i>A. gambiae</i> G3 strain	BEI Resources	MRA-112
<i>A. stephensi</i> AF-SDA500 strain	Infravec2	V.2.3.1.L.FR.1.0
Oligonucleotides		
See Table S2 for detailed information on used oligonucleotides	This study	N/A
Recombinant DNA		
p165	⁶¹	GenBank: KU189142
pUAST-attB	⁶²	GenBank: EF362409
pBac_pattPs_fle_3miR	This study	GenBank: MW147152
pENTR R4-vas2-Transposase-R3	⁶³	N/A
pGEM-T Easy Vector System	Promega	Cat#A1360
Software and Algorithms		
BLAST	⁶⁴	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Jpred-4	⁶⁵	http://www.compbio.dundee.ac.uk/jpred/
ClustalX2	⁶⁶	https://clustalx.software.informer.com/2.1/
MEGA v. 7.0.26	⁶⁷	https://www.megasoftware.net/
WebLogo	⁶⁸	https://weblogo.berkeley.edu/logo.cgi
ImageJ	⁶⁹	https://imagej.nih.gov/ij/
Kallisto v0.46	⁷⁰	https://pachterlab.github.io/kallisto/
DESeq2	⁷¹	http://bioconductor.org/packages/release/bioc/html/DESeq2.html
apeGLM	⁷²	https://bioconductor.org/packages/release/bioc/html/apeglm.html
Other		
QuantStudio 3 Real-Time PCR System	Applied Biosystems	Cat#A28136

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jaroslaw Krzywinski (jaroslaw.krzywinski@pirbright.ac.uk).

Materials Availability

All relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon a reasonable request.

Data and Code Availability

The RNA-seq data generated during this study are available at the European Nucleotide Archive with the accession number PRJEB38605.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mosquito strains

A. gambiae G3 strain, transgenic lines generated on the *A. gambiae* G3 strain background, and *A. stephensi* AF-SDA500 strain were reared at 28°C and 80% humidity on 12 h: 12 h light:dark cycle, following the standard protocol.⁷³ Larvae were reared in plastic trays filled with 1 L of deionised water and provided with ground TetraMin tropical fish food flakes (Tetra). The adults were kept in BugDorm-1 (30 × 30 × 30 cm) or BugDorm-4 (24.5 × 24.5 × 24.5 cm) cages (BugDorm) and provided with 10% sucrose solution *ad libitum*. A 1:1 mixture of time-expired human red blood cells and plasma sourced from a blood bank was used for feeding females through the Hemotek membrane feeder.

Mosquito cells

The *A. gambiae* female-like Sua5.1 cells⁶⁰ were cultured at 28°C in Schneider's Modified medium (Lonza) supplemented with 10% fetal bovine serum (PAA), and 100 U/ml Penicillin and 100 µg/ml Streptomycin (Life Technologies).

METHOD DETAILS

Sequence analyses

All similarity searches were conducted using BLAST,⁶⁴ with the word size = 2. The search for the *A. gambiae dsx* splicing factors within the *A. gambiae* PEST strain genome was conducted using a BLASTp and tBLASTn algorithm, with the expect value 1e-7, and the *D. melanogaster* Tra2 (FlyBase IDs CG10128-PA) or *Aedes aegypti* Nix (a distant homolog of *D. melanogaster tra2*; GenBank accession AHW46195.1) protein sequences used as query. The phylogenetic distribution of the *fle* orthologs was evaluated through a tBLASTn search against the NCBI whole-genome shotgun contigs (wgs) database using as query the amino acid translation of *fle*. The search for a putative *fle* paralog was conducted using a local BLASTp search against a database of the *A. gambiae* AgamP4.12 peptides downloaded from Vectorbase, and for other *Anopheles* species by implementing BLASTp search against a protein database for a respective species in Vectorbase. Structure of the proteins was derived by searches of the Conserved Domain Database⁷⁴ and through structure prediction using Jpred-4.⁶⁵ Sequence alignments were conducted using ClustalX2.⁶⁶ Maximum likelihood phylogenetic analysis was conducted in MEGA7⁶⁷ using JTT matrix-based model⁷⁵ with Gamma distributed evolutionary rate differences among sites. The sequence logos were generated using WebLogo.⁶⁸

dsRNA synthesis

Open reading frame fragments of the *A. gambiae* or *A. stephensi* genes were amplified through PCR from a genomic DNA template or in a one-step RT-PCR (Invitrogen) reaction from the pupal total RNA templates using gene-specific primers flanked at their 5' ends with the T7 promoter sequence (for details on primers see Table S2). The resulting products, cloned into pGEM T-easy vector (Promega), were used directly, or after reamplification with the same primer pairs, as a template to synthesize double-stranded RNA (dsRNA) using the MEGAscript RNAi T7 kit (Life Technologies) according to manufacturer's recommendation. Similarly, a fragment of β -lactamase (*bla*) gene was amplified by PCR from the pGEM-T Easy vector using *bla*-specific primers, each containing the T7 promoter sequence at the 5' end, and dsRNA was synthesized as described above.

Transfection experiments

Prior to transfection, the *A. gambiae* Sua5.1 cells were split into new culture flasks and transfection experiments were performed when cells' confluency reached 60%–80%. Approximately 1×10^6 cells per well were seeded onto 24 well plates and transfected in suspension, using 3 µl of Lipofectamine 2000 transfection reagent (Life Technologies) and 1.5 µg of dsRNA of a tested gene per well. In parallel, cells in a separate set of wells were transfected with a plasmid (0.3 µg per well) containing the eGFP open reading frame under the control of *A. gambiae* polyubiquitin promoter as a control of transfection efficiency. In addition, non-transfected control cells were cultured in a set of wells in each experiment. After approximately 24 hours, the transfection efficiency was evaluated

using fluorescence microscopy. If at least 30% of the plasmid control cells per well were GFP-positive on a given plate, experimental and non-transfected control cells from that plate were harvested 48 hours post transfection to isolate total RNA with PureLink Micro kit (Life Technologies). Transfection experiments were repeated 3 times.

Analysis of sex-specific splicing

The effect of knockdown of the analyzed genes on the pattern of the *dsx* splicing was evaluated through RT-PCR using total RNA templates from the transfected Sua5.1 cells and primers *dsxF2* and *dsxR5m*. Similarly, RT-PCR was used to analyze the effect of a stable *fle* knockdown on the splicing pattern of *dsx* and *fru* (the latter using primers *Aga_fruF* and *Aga_fruR*) in pupae of selected transgenic *A. gambiae* lines.

Transient gene silencing

Transient gene silencing experiments were conducted as described earlier.²¹ Briefly, early preblastoderm *A. gambiae* or *A. stephensi* embryos of unknown sex were microinjected with a solution containing dsRNA (1–1.5 $\mu\text{g}/\mu\text{l}$) of either *fle* or *bla* (as a control) gene and a plasmid (0.2 $\mu\text{g}/\mu\text{l}$) with the GFP gene downstream of the *Drosophila melanogaster actin 5C* promoter, or with plasmid alone as a control. Surviving first instar larvae were screened for the presence of GFP marker in the midgut cells using an M165 FC microscope equipped with a GFP filter. The larvae were sorted into GFP-negative and GFP-positive groups (cf. Figure S9 in Krzywinska et al.²¹). After pupation, sex of individuals was determined based on morphological characters. The experiments in *A. gambiae* were repeated three times with each of the two dsRNAs (cf. Figure 2A), and in *A. stephensi* were repeated two times.

Transgenic construct plasmid and transgenesis

To create the transgenic construct we used the p165 plasmid⁶¹ backbone by digesting the plasmid with *MluI* and *NotI* and ligating a linker containing the *SfiI* site and compatible *MluI/NotI* ends. The resulting plasmid was digested with *SfiI* and *NotI*, and two *in vitro* synthesized inserts were incorporated in a single ligation reaction. One insert, with *SfiI* and *NheI* ends, encoded the puromycin resistance gene *pac* under the control of the AGAP004395 promoter; the other insert, with *NheI/NotI* ends, encoded SV40 terminator, followed by GFP under the control of the AGAP003430 promoter (both promoters drive expression in the Sua5.1 cells, and the AGAP003430 promoter *in vivo* throughout development in gastric caeca, anterior and posterior stomach, Malpighian tubules and rectum, and at lower level in the brain, thoracic muscles and anal papillae). The resulting plasmid was digested with *MluI* and *SfiI* to clone a PCR-generated *A. gambiae* polyubiquitin promoter⁷⁶ with *MluI/BsaI-FseI* ends and a PCR-generated SV40 terminator with *FseI/SfiI* ends. Finally, that plasmid was digested with *BsaI* and *FseI* to clone a *fle_3miR* fragment encoding a polycistronic transcript designed to silence expression of *fle*. The *fle_3miR* fragment contained three microRNAs targeting *fle* and was constructed by annealing overlapping oligonucleotides (Table S2) and PCR amplification, as previously described.⁷⁷ After cloning into pUAST-attB vector,⁶² the *fle_3miR* fragment was released using *BsaI* and *FseI*. Thus engineered construct was excised using *MluI* and *AsiSI* and cloned into a *MluI* and *AsiSI*-cut p165-based plasmid backbone flanked by piggyBac arms and a $\phi\text{C31 attB}$ site added to each end in opposite orientation, to create a transformation plasmid pBac_attBs_fle-3miR (GenBank: MW147152).

Early preblastoderm *A. gambiae* embryos were microinjected with a solution of pBac_attBs_fle_3miR (0.4 $\mu\text{g}/\mu\text{l}$) and a helper plasmid pENTR R4-vas2-Transposase-R3 (0.2 $\mu\text{g}/\mu\text{l}$), containing piggyBac transposase reading frame under the control of the *vas2* regulatory sequences,⁶³ following previously described methods.⁷⁸ Injection of approximately 1200 embryos yielded 252 G0 larvae, of which 108 individuals, that exhibited a transient fluorescent marker expression, reached the pupal stage. The emerging adults (59 males and 49 females) were placed in 9 same-sex pools for crosses with the wild-type G3 strain mosquitoes. Over 120 transgenic G1 mosquitoes were recovered from four pools of male founders, whereas no transgenic G1 individuals were produced by the female founders. Selected G1 males originating from different founder cages or exhibiting different intensity of fluorescent marker expression (if originating from the same cage; Figure 2I) were crossed to wild-type females to establish 7 independent lines. Progeny from these crosses were screened for inheritance of fluorescent marker to evaluate transgene copy number. In three lines approximately 50% of G2 individuals were transgenic, indicative of single insertions. Four other lines exhibited 60%–89% transgene inheritance, with transgenic G2 individuals representing up to three discernible classes of fluorescence intensity or pattern per line, indicative of multiple insertions. In an effort to isolate single insertion sub-lines, males derived from multiple-insertion lines and representing different fluorescence classes were backcrossed with wild-type females at consecutive generations and the number of fluorescence phenotypes was monitored at each generation. Sub-lines that, after 6 generations, produced more than one fluorescence phenotype, or in which more than 50% of the individuals inherited the transgene, were eliminated. Finally, molecular characterization was used to confirm that each line from the final set possessed a single, unique genomic transgene integration site.

No efforts have been made to generate *fle* loss-of-function mutants in this study out of feasibility concerns. *Tra2* in *Drosophila*, in addition to female sex determination, is necessary for male germline development – loss of *Tra2* function leads to male sterility.⁷⁹ *Fle* may perform a similar role during the *Anopheles* spermatogenesis, which, combined with an apparent *fle* haploinsufficiency in females, could make recovering *fle* knockout mutants biologically impossible.

Identification of transgene insertion sites

The integration sites of the piggyBac element within the genome has been identified using the splinkerette PCR protocol⁸⁰ or by inverse PCR. DNA isolated from individual pupae was used for both approaches. For inverse PCR, the DNA was digested with *CviQI*, *HaeIII*, *MspI*, *Sau3AI*, or *TaqI* (NEB), circularized by ligation, and amplified by PCR using primers ITRL1F and ITRL1R for piggyBac left

arm, or ITRR1F⁸¹ and InpBacR2R for piggyBac right arm (Table S2) to isolate flanking genomic regions. The products containing genomic sequences flanking the piggyBac elements were sequenced directly, or after cloning, and genomic location of the integration sites was identified by BLAST search.

Abdominal musculature

Adult mosquito abdomens were dissected in phosphate-buffered saline (PBS) to release tergites with the associated musculature. The tissues were fixed in PBS containing 4% paraformaldehyde for 15 min, washed three times for 5 min in PBS, and incubated in ActinGreen 488 ReadyProbes Reagent containing AlexaFluor 488-conjugated phalloidin. After three short washes the tissues were mounted on slides and photographed with a Leica DFC365 FX camera mounted on a Leica M165 FC microscope equipped with a GFP filter. Images were processed with ImageJ.⁶⁹

Real-time PCR

Total RNA was extracted from individual *A. gambiae* pupae using PureLink RNA Micro Kit (Invitrogen) according to manufacturer's recommendations. For each sample, 500 ng of total RNA was used to synthesize cDNA with LunaScript RT SuperMix Kit (NEB). Quantitative PCR was conducted using primer pairs JK1051/JK1052 and JK1053/JK1054 to amplify, respectively, a fragment of *fle* and of the housekeeping gene encoding ribosomal protein S7 (*rpS7*, AGAP010592) used to normalize the expression. QuantStudio 3 Real-Time PCR System (Applied Biosystems) was employed to run the reactions using Luna Universal qPCR Master Mix (NEB) at annealing temperature of 59°C. Expression levels were calculated using $2^{-\Delta\Delta C_t}$ method,⁸² with triple technical and three biological replicates for each sample, and all data normalized to the relative *fle/rpS7* expression in the samples of the wild-type female pupae.

RNA-seq analysis

Total RNA was extracted using the Trizol method and quality-checked using TapeStation (Agilent). Triplicate samples of female pupae from wild-type G3 line and from transgenic 4M4B line were used for transcriptome sequencing. The TruSeq library preparation protocol (Illumina) was followed by 150 bp paired-end sequencing using NovaSeq 6000 sequencing system (Illumina). The reads were pseudo-aligned to the *A. gambiae* transcriptome genebuild AgamP4.12 using Kallisto v0.46.⁷⁰ Transcripts per kilobase million (TPM) value was quantified for each transcript and averaged across multiple replicates of the same sample. As a further check of statistical robustness, differential expression analysis was performed using DESeq2,⁷¹ with filtering out transcripts covered by less than 10 reads among all samples, and then shrinking log₂-fold changes using apeGLM.⁷²

QUANTIFICATION AND STATISTICAL ANALYSIS

For the transient *fle* experiments, the probability of the observed microinjection results under the null hypothesis that there is no sex bias difference between the GFP-positive and GFP-negative (or control) groups was calculated using Fisher's exact test. For the real-time PCR experiments, Student's t test was used to evaluate statistical differences between the relative *fle* expression levels in the wild-type and transgenic strains after performing a goodness of fit test.

Current Biology, Volume 31

Supplemental Information

***femaleless* Controls Sex Determination
and Dosage Compensation Pathways
in Females of *Anopheles* Mosquitoes**

Elzbieta Krzywinska, Luca Ferretti, Jianwei Li, Jian-Chiuan Li, Chun-Hong Chen, and Jaroslaw Krzywinski

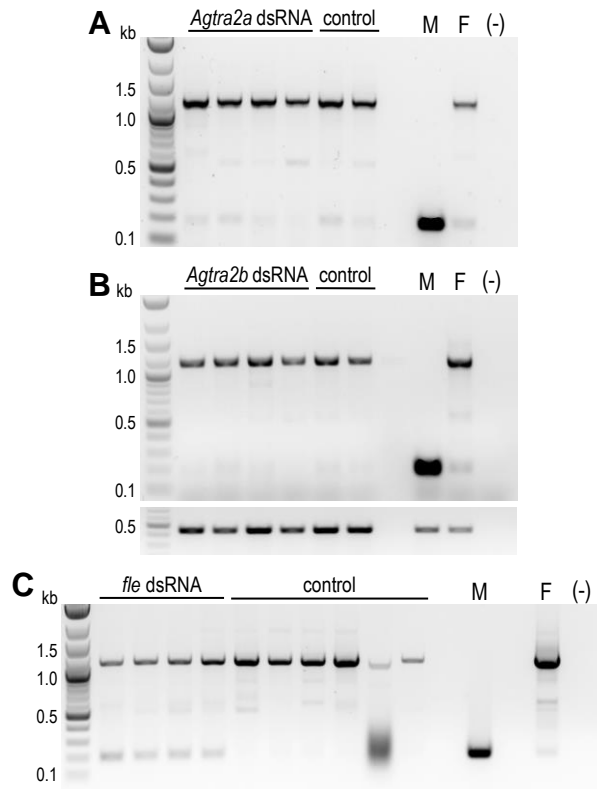


Figure S1. Homologs of *tra2* do not control sex determination in *A. gambiae*. Related to Figure 1. RT-PCR analysis of *dsx* splicing in the *A. gambiae* Sua5.1 cells transfected with in vitro-synthesized dsRNA of *A. gambiae tra2a* (AGAP029421) (A), *tra2b* (AGAP006798) (B), and *fle* (AGAP013051) (C), as compared to the control cells. No difference in *dsx* splicing pattern between the dsRNA-transfected and the control cell samples in A or B indicates that neither of the *tra2* paralogs is involved in the *A. gambiae* sex determination pathway. In contrast, *fle* dsRNA-transfected cells produce fewer female-specific *dsx* transcripts, which coincides with a clear production of male *dsx* transcript forms (C). Cropped elements of the gel image presented in C, with two poor quality RNA control sample lanes omitted, are shown in the main body of the paper as Fig. 1A. M and F, male and female pupae; (-), negative control.

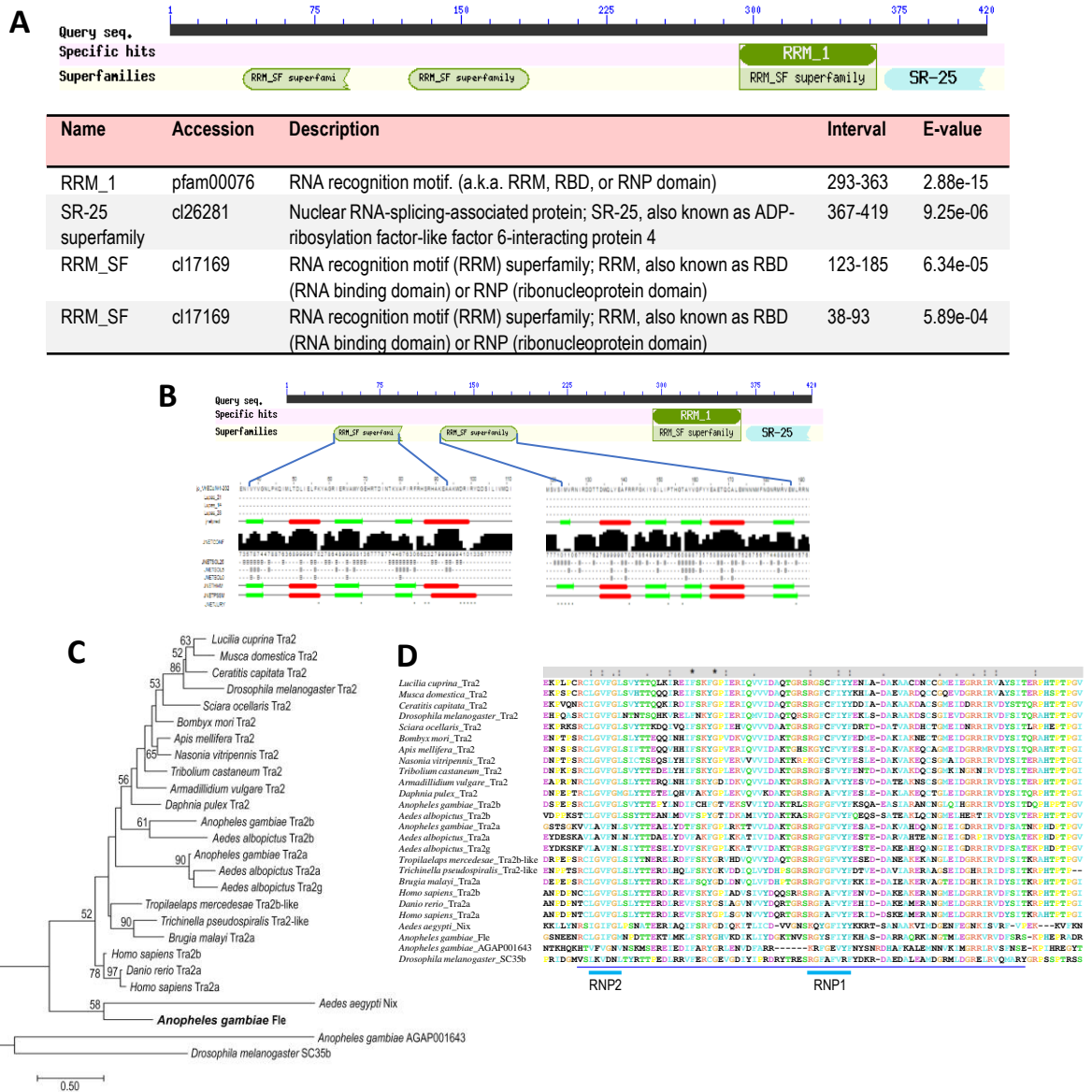


Figure S2. *A. gambiae* Fle is an RRM protein with putative binding sites in *doublesex* and *fruitless*, and is not closely related to TRA2. Related to Figure 1. (A) Conserved domains in the *A. gambiae* Fle protein identified through a search against the Pfam v.31.0 database. (B) Jpred4-based secondary structure prediction of the putative RRM regions in Fle. α -helices are marked in red and β -sheets in green. The truncated first putative RRM lacks the β_4 sheet. (C) The maximum likelihood phylogenetic tree of Fle, representative Tra2, and other RRM proteins inferred with MEGA v. 7.0.26 using the JTT matrix-based model and a discrete Gamma distribution to model evolutionary rate differences among sites. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysed data set consisted of amino acid sequences aligned using ClustalX2. All gapped alignment positions were eliminated prior to analysis; there were 82 positions in the final data set, covering RRM (C-terminus proximal in case of Fle) and its immediate flanks (see D). The tree is rooted with the *D. melanogaster* SR family splicing factor SC35. Numbers above or below branches representing bootstrap support >50% are based on 500 ML bootstrap replications. (D) An alignment of the amino acid region used for the phylogenetic reconstruction. Identical/similar positions are marked above, and the location of the RRM (thin line) and RNP sequence elements below, the alignment.



Figure S3. A highly conserved *fle* may have originated by gene duplication in the recent common ancestor of *Anopheles* species. Related to Figure 1. (A) Amino acid sequence alignment of the *fle* orthologs from *Anopheles* species. The RRM region showing similarity to Tra2 RRM is shaded black, with the RPN-2 and RPN-1 elements in yellow font. Two other putative RRM regions are shaded grey. (B) A comparison of the Fle and AGAP001643 protein structures. Boundaries of a similarity region (BLASTp e -value = $4e-20$) are marked by red lines. (C) Amino acid alignment of the N-terminal and C-terminal part of Fle and AGAP001643, respectively, with the similarity region marked in red, the RRMs shaded green, and the codon interrupted by an intron shaded bright green. Note a higher sequence similarity within the N-terminal region. AGAP001643 orthologs from *A. maculatus* and *A. albimanus* that split from *A. gambiae* approximately 36-70 Mya and 100 Mya, respectively, are also shown. The phylogenetic tree shown in Figure S2C has been inferred using a slightly extended C-terminal region (cf. Figure S2D) than that presented in (C).

Line	pBAC arm	Flanking genomic sequence	Fragment length	% Identity to genome	Chromosomal integration position
3M	L	TTAATAGGCACGTGCCTTCG	171	99	3R: 15803701 3R: 15803704
	R	CCTGCCTGCTCGAGCTTTAA			
4M2B	L only	TTAAAGGCTGTGGTACACTT	294	100 ^a	UNK NW_161049: 12
4M3	L	TTAAGGGTCTGCGCCAGGGG	224	89 ^b	Undetermined
	R	AGGCTTTTTGCTACCATTAA	248	91 ^b	Undetermined
4M4B	L	TTAAATCACCTTGTAACGAT	262	99	2R: 21049616
	R	TTGATTCTTGTTAGCTTTAA	304	98	2R: 21049618
4M4C	L	TTAAACTCAACGATGGCAAT	120	98, 95 ^c	3R: 28923242
	R	TGTCAGGCGGTGGCATTAA	180	98	3R: 28923245
4M6	L	TTAAAAGAGGTGAGCCAGA	459	99	3L: 40448623
	R	ATAATCCATCAACTGTTTAA	169	99	3L: 40448620

Table S1. Transgene integration sites in the *A. gambiae* lines generated in this study. Related to Figure 2.

^a Integration site within the retrotransposon *zanzibar* fragment of unknown chromosomal location; maximum 93% identity within the mapped *An. gambiae* PEST scaffolds.

^b Hit to a highly variable repetitive region differing between the *A. gambiae* strains; 94% identity to the *A. colluzzii* strain M genome.

^c The sequence of a fragment isolated adjacent to the pBac L arm yielded a gapped BLAST result, due to the presence of a Tsessebe transposon insertion in the corresponding region within the *A. gambiae* PEST genome. Ungapped alignments, indicating absence of the Tsessebe insertion, were obtained with four other *A. gambiae* genome assemblies available in the NCBI whole genome shotgun contigs database.

Target/Name	Name/sequence (5' – 3')	
	PCR primers	
<i>A. gambiae fle</i> (AGAP013051)	T7-fle-F2 TAATACGACTCACTATAGGGTGC GCACTGG AAATGAACAA	T7-fle-R2 TAATACGACTCACTATAGGGGTTCTTTTGCCG TCGTAGA
	T7-fle-F2 Sequence as above	T7-fle-R3 TAATACGACTCACTATAGGGACAGCGTTCTCC TCGTTC
<i>A. stephensi fle</i> (ASTE008781)	Aste_fleT7F TAATACGACTCACTATAGGGATATACGGCAT CCTGATACC	Aste_fleT7R TAATACGACTCACTATAGGGGCTTAAAGTAGAT GAACGAG
<i>tra2a</i> (AGAP006798)	T7_Ag6798-F TAATACGACTCACTATAGGGTGCAGTGAAGT GTTGTTATTTTG	T7_Ag6798-R TAATACGACTCACTATAGGGCTTTCGAGTGGGA CCAAACT
<i>tra2b</i> (AGAP029421)	T7_Agtra2b-F TAATACGACTCACTATAGGGTCCAGCAGCCA TTGTGTCTA	T7_Agtra2b-R TAATACGACTCACTATAGGGACTACCGACTTTTC CACCGT
<i>dsx</i> (AGAP004050)	dsxF2 CCAGAACCTGTAAATCTCCTAC	dsxR5m GATGACTTCACCACCGCTTC
<i>fru</i> (AGAP000080)	Aga_fruF GGCAGCATCGGACTTGTTC A	Aga_fruR ACGTCGCACAGTTTCTCATC
<i>rpS7</i> (AGAP010592)	S7F TGCTGCAAACCTTCGGCTAT	S7R CGCTATGGTGTTCGGTTCC
<i>bla</i>	blaF_T7 TAATACGACTCACTATAGGGAACCTTATCCG CCTCCATCC	blaR_T7 TAATACGACTCACTATAGGGGCTATGTGGCGCG GTATTAT
<i>A. gambiae fle</i> RT_q-PCR	JK1051 CGGATACGCTACCAAGACTCAATAC	JK1052 GTCGGTTGTATCGTCGCGT
<i>rpS7</i> RT_q-PCR	JK1053 CAACAACAAGAAGGCGATCATCATC	JK1054 GCGCTCGGCAATGAACAC
Inverse PCR pBac right arm	ITRR1F TACGCATGATTATCTTTAACGTA	InpBacR2R TTGCTACTGACATTATGGCT
	Oligonucleotides for microRNA cistron construction	
Mir6.1_5'EcoRI/ Bsal/BglII	GGCGAATTCATGGTCTCTTATGAGATCTTTTAAAGTCCACAACCTCATCAAGGAAAATGAAAGTCAAA GTTGGCAGCTTACTTAAACTTA	
Mir6.1_3'NotI/Fs el/BamHI	GGCCGCGGCCGCTATGGCCGGCCGGATCCAAAACGGCATGGTTATTCGTGTGCCAAAAAAAAAAAA AAATTAATAATGATGTTAGGCAC	
AGAP013051_ miR-1-1	GGCAGCTTACTTAACTTAATCACAGCCTTTAATGTGCATCGTACAGACATCAACACGTAAGTTAAT ATACCATATC	
AGAP013051_ miR-1-2	AATAATGATGTTAGGCACTTTAGGTACGCATCGTACAGACATCAACACGTAGATATGGTATATTAAC TTACGTG	
AGAP013051_ miR-2-1	GGCAGCTTACTTAACTTAATCACAGCCTTTAATGTGGATACGCTACCAAGACTCAATTAAGTTAATA TACCATATC	
AGAP013051_ miR-2-2	AATAATGATGTTAGGCACTTTAGGTACGGATACGCTACCAAGACTCAATTAGATATGGTATATTAAC TTAATTG	
AGAP013051_ miR-3-1	GGCAGCTTACTTAACTTAATCACAGCCTTTAATGTAAAGACATCAAGCTAATCTACGTAAGTTAATA TACCATATC	
AGAP013051_ miR-3-2	AATAATGATGTTAGGCACTTTAGGTACAAAGACATCAAGCTAATCTACGTAGATATGGTATATTAAC TACGTA	

Table S2. Oligonucleotides used in the study. Related to STAR Methods.