Title: A maleness gene in the malaria mosquito *Anopheles gambiae*

**Authors:** Elzbieta Krzywinska1†, Nathan J. Dennison2,3, †, Gareth J. Lycett2, Jaroslaw Krzywinski1,2, ,\*

**Affiliations:**

1The Pirbright Institute, Ash Road, Pirbright, Surrey, GU24 0NF, United Kingdom.

2Vector Biology Department, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, United Kingdom.

3Department of Life Sciences, Imperial College, South Kensington Campus, London SW7 2AZ, United Kingdom.

\*Correspondence to: Email: jaroslaw.krzywinski@pirbright.ac.uk.

† These authors contributed equally.

**Abstract**: The molecular pathways controlling gender are highly variable and have been identified in only a few non-mammalian model species. In many insects, maleness is conferred by a Y chromosome-linked M factor of unknown nature. We have isolated and characterized a gene, *Yob*, for the M factor in the malaria mosquito *Anopheles gambiae*. *Yob*, activated at the beginning of zygotic transcription and expressed throughout a male’s life, controls male-specific splicing of the *doublesex* gene. Silencing embryonic Yob expression is male-lethal, whereas ectopic embryonic delivery of *Yob* transcripts yields male-only broods. This female-killing property may be an invaluable tool for creation of *conditional* male-only transgenic *Anopheles* strains for malaria control programs.

**One Sentence Summary:** A maleness gene in *Anopheles gambiae*.

**Main Text:** Animals have evolved a variety of mechanisms for sex determination. In insects, a common pattern requires a primary sex determining signal that in early embryos activates a short cascade of sex-specifically spliced genes. The products of these genes modulate differentiation of an individual into either male or female, depending on the sex determination system ([*1*](#_ENREF_1)). In the absence of the primary signal, a default pattern of gene splicing leads to the development of the opposite sex. Only the final element of the cascade, the *doublesex* gene, is well conserved, while upstream genes are highly evolutionarily labile ([*2*](#_ENREF_2)), and are difficult to identify using comparative genomics tools.

*Drosophila melanogaster* has an XX/XY chromosomal sex determination system, in which a double dose of X chromosome-linked signal elements (XSE) represents the primary signal that initiates female-specific pre-mRNA splicing in a cascade of subordinate genes *sex lethal* (*Sxl*), *transformer* (*tra*), and *doublesex* (*dsx*) ([*3*](#_ENREF_3)*,* [*4*](#_ENREF_4)). The resulting transcripts yield functional female-specific proteins, of which SXL regulates splicingof *tra*, and TRA, along with its cofactor TRA2, promotes splicing of *dsx*. In males, a single dose of XSE is insufficient to initiate this female-specific cascade, and the three genes are spliced by a default male-specific mode, with only *dsx* encoding a functional protein. The male and female forms of DSX regulate sex-specific morphologies by modulating the expression of genes involved in downstream sexual differentiation processes.

The dipterans *Ceratitis capitata*, *Lucilia cuprina* and *Musca domestica*, and the beetle *Tribolium castaneum*, also possess an XX/XY chromosomal system. However, in these insects an unidentified Y chromosome-linked M factor constitutes a primary signal that determines maleness ([*5-8*](#_ENREF_5)). In none of these species is *Sxl* involved in sex determination. Instead, maternal *tra* mRNA is sufficient to direct default female development in the absence of the Y-linked factor, whereas in males the M may be preventing establishment of *tra* function by an unknown mechanism ([*9-12*](#_ENREF_9)). The XX/XY system incorporating a dominant Y-linked M factor is common in many insect groups ([*13-15*](#_ENREF_13)). However, with the exception of *Nix* (a homologue of *tra2*) from a mosquito *Aedes aegypti* ([*16*](#_ENREF_16)), genes encoding M factor have remained enigmatic.

In anophelinemosquitoes, the only described element of the (morphological) sex determination cascade is *dsx* ([*17*](#_ENREF_17)). To elucidate the initial steps of sex determinationin the African malaria mosquito, *Anopheles gambiae*, we analyzed transcriptomes of male and female embryos ([*18*](#_ENREF_18)), whose sexual identity was determined by polymerase chain reaction (PCR) (fig. S1). Separate pools of mRNA were sequenced, yielding approximately 500,000 Roche 454 reads from each sex. From the male pool, 21 reads uniquely mapped to the previously characterized scaffold AAAB01008227 derived from the *A. gambiae* Y chromosome ([*19*](#_ENREF_19)).

The corresponding gene (hereafter called *Yob*)consists of three exons and produces transcripts differing in size because of intron retention (fig. S2). *Yob* likely represents the same gene as *gYG2* ([*20*](#_ENREF_20)), which partially overlaps *Yob*,but its annotation lacks over half of the *Yob* gene sequence. Among the *A. gambiae* Y-linked genes, only the *Yob*/*gYG2* locus is Y-linked across the examined species from the *A. gambiae* complex ([*21*](#_ENREF_21)), indicative of an important sexual identify function. Consistent with the chromosomal location, transcription of *Yob* is limited to males. Transcription begins in embryos between 2 h and 2.5 h after oviposition, prior to blastoderm formation ([*22*](#_ENREF_22)) and expression of other Y chromosome genes ([*20*](#_ENREF_20)*,* [*21*](#_ENREF_21)), and continues throughout life of the mosquito (Fig. 1A and fig. S3A). Onset of the *Yob* transcription is nearly contemporaneous with that of orthologs of *even skipped* (AGAP010279) and *hunchback* (AGAP010777) (fig. S3B), whose transcripts mark the beginning of zygotic expression in *Drosophila (*[*23*](#_ENREF_23)*,* [*24*](#_ENREF_24)*)*. We determined that *Yob* activation precedes establishment of sex-specific splicing of *dsx* by up to 6 hours*.* The female isoform of *dsx* is maternally deposited, but largely degraded in male embryos within 4 hours of oviposition, and only after complete degradation of the female isoform in males is a persistent pattern of sex-specific *dsx* splicing established (Fig. 1*B*). These findings are consistent with *Yob* acting upstream of *dsx* in the sex-determining hierarchy. If true, introduction of *Yob* transcripts into female cells should cause an alteration of *dsx* splicing pattern.

We investigated the effect of *in vitro* synthesized mRNA corresponding to the shortest, presumably mature *A. gambiae Yob* transcript isoform on *dsx* splicing in the *An. gambiae* Sua5.1 cell line, which lacks the Y chromosomeand *Yob* expression (fig. S4). We observed a significant loss of the female and gain of the male *dsx* transcript isoforms incells transfected with *Yob* mRNA, as compared with control non-transfected cells (Fig. 2A, B, and D), or cells transfected with non-productive forms of *Yob* (see below; Fig. 2D). Transfection experiments in larvae suggest that *Yob* exerts the same effect on *dsx* *in vivo* (fig. S5). This confirms that *Yob* is involved in the sex determination pathway as a direct or an indirect upstream regulator of *dsx* splicing.

In the vast majority of protein coding genes, different evolutionary constraints on individual codon positions result in synonymous substitutions observed with much higher frequency than the non-synonymous ones ([*25*](#_ENREF_25)). Based on this principle,we evaluated the protein coding potential of *Yob* by comparing its sequence to PCR-isolated orthologous sequences from *Anopheles arabiensis* and *Anopheles quadriannulatus*,two members of the *Anopheles gambiae* complex. The analysed region encompasses two open reading frames (ORF) longer than 50 codons, of which only the shorter open reading frame (ORF) bears a substitution pattern indicative of purifying selection (fig. S6). Guided by these results we experimentally tested for the protein-coding potential of the shorter ORF. We transfected the Sua5.1 cells with two modified *Yob* transcripts containing putative non-synonymous point mutations r.1A>C and [r.5U>A; 6U>G] and investigated the effects of the deduced amino acid changes Met1Leu and Phe2Stop, respectively, on *dsx* splicing. No shift in *dsx* splicing pattern was observed in cells transfected with transcripts either lacking an initiation codon or containing a premature stop codon, unlike the positive control cells transfected in parallel with the wild-type *Yob* transcripts (Fig. 2D and fig. S7). These results indicate that *Yob* encodes a short, 56 amino acid protein that may contain a helix-loop-helix motif (fig. S8). A protein with the same length and similar secondary structure, but non-conserved primary sequence, was proposed to be encoded by an early zygotically expressed Y-linked gene *GUY1*, potentially involved in sex determination in *Anopheles stephensi (*[*26*](#_ENREF_26)*)*.

We injected *Yob* mRNA into non-sexed preblastoderm embryos of *A. gambiae* and its sibling species *A. arabiensis* to assess whether ectopic *Yob* transcripts affect mosquito sex ratios. To control for successful injection, we co-injected a plasmid with a green fluorescent protein (GFP) expression cassette (embryos that receive sufficient nucleic acids develop into larvae transiently expressing GFP; fig. S9). Surviving individuals were sorted at the larval stage into a GFP-positive and a GFP-negative group (Fig. 3A), and at the pupal stage mosquitoes were sexed. In both species, all the GFP-positive individuals developed as phenotypic males, while in the GFP-negative group and in the control group of *A. gambiae* embryos injected with GFP plasmid only, the sex ratio was unbiased (Fig. 3B). The GFP-positive males had the XY karyotype, as indicated by PCR; moreover, they had normally developed reproductive organs, produced motile sperm and were fertile (table S1). Thus, ectopic delivery of *Yob* mRNA is lethal to genetic female embryos, but has no discernible effect on the sexual development of genetic males.

In *D. melanogaster*,female-specific embryonic lethality is caused by loss-of-function mutations in genes located at the top of the sex determination cascade, and invariably results from a misregulation of dosage compensation ([*3*](#_ENREF_3)*,* [*27-29*](#_ENREF_27)). The dosage compensation mechanism in that species equalizes X-linked gene products in XX and XY individuals by upregulation of transcription from the male X chromosome ([*30*](#_ENREF_30)). The process is tightly linked to the sex determination through female-specific repression by SXL ([*31*](#_ENREF_31)*,* [*32*](#_ENREF_32)). Mutations in *Sxl,* or in genes involved in *Sxl* regulation, cause X-linked genes overexpression and female death during embryogenesis ([*3*](#_ENREF_3)*,* [*27-29*](#_ENREF_27)). In femalesof *T. castaneum*, in which *Sxl* has no sex determining role,TRA has been proposed to inhibit dosage compensation ([*12*](#_ENREF_12)). In a lepidopteran, *Bombyx mori*, in which females are heterogametic (ZW) and males homogametic (ZZ), male embryo lethality, observed after a knockdown of a sex determination gene *Masculinizer* (responsible for male-specific splicing of *dsx* and repression of transcription from Z chromosomes), has been similarly attributed to misregulation of dosage compensation ([*33*](#_ENREF_33)).

Dosage compensation in *A.* *gambiae* also relies on hyperactivation of the X chromosome genes in males ([*34*](#_ENREF_34)). Therefore, in the presence of YOB, dosage compensation may be directly or indirectly induced in the XX embryos, leading to their death as a result of abnormal over-transcription from both X chromosomes (fig. S10). Conversely, in the absence of YOB, lack of dosage compensation and concomitant insufficient transcription from the X chromosome should be male-lethal, rather than leading to feminization of the XY individuals. Indeed, we observed highly significant male deficiency in mosquitoes surviving transient knockdown of *Yob* in non-sexed embryos (fig. 3C). All tested female survivors had the XX karyotype.

Involvement of the Y chromosome factor in sex determination in *Anopheles* was first supported by the finding of a single, triploid *A. culicifacies* male with the XXY sex chromosomes ([*15*](#_ENREF_15)). Such a karyotype, apparently extremely rare and not reported elsewhere in numerous mutational and cytogenetic studies in *Anopheles*, seems to counter over-transcription of the X chromosome genes as a cause of female embryo lethality observed in our study. However, the genetic background of this male must have been severely compromised through the mutagenic effects of irradiation on his parents ([*15*](#_ENREF_15)). It thus seems likely that the XXY male may have carried multiple mutations, including those causing loss of function of dosage compensation machinery that allowed his survival to adulthood despite possessing two X chromosomes.

*A. gambiae* and *A. arabiensis* are the most important African vectors of human malaria. Control of the disease depends heavily on the use of insecticides, but emergence of resistance in mosquito populations severely threatens the effectiveness of these approaches ([*35*](#_ENREF_35)). The sterile insect technique and other genetic control methods have been proposed to complement current efforts to suppress mosquito populations ([*36*](#_ENREF_36)*,* [*37*](#_ENREF_37)). Such programs must incorporate male-only releases, because released females would contribute to pathogen transmission. However, no effective methods to sex the large number of *Anopheles* needed for releases currently exist. The fitness and mating competitiveness of adults is highly dependent on larval density; therefore, removing females from the release generation during the embryonic stage would drastically decrease the costs of rearing of high quality males ([*38*](#_ENREF_38)). *Yob* represents an excellent tool to be utilized in transgenic technology to *conditionally* eliminate female embryos and efficiently produce male-only generations of both malaria-transmitting *Anopheles* species ([*37*](#_ENREF_37)).

References and Notes:

1. T. Gempe, M. Beye, *Bioessays* **33**, 52-60 (2011).

2. A. S. Wilkins, *Bioessays* **17**, 71-77 (1995).

3. T. W. Cline, B. J. Meyer, *Annu. Rev. Genet.* **30**, 637-702 (1996).

4. J. W. Erickson, J. J. Quintero, *PLoS Biol.* **5**, e332 (2007).

5. T. Hiroyoshi, *Genetics* **50**, 373-385 (1964).

6. D. Bedo, G. Foster, *Chromosoma* **92**, 344-350 (1985).

7. U. Willhoeft, G. Franz, *Genetics* **144**, 737-745 (1996).

8. J. J. Stuart, G. Mocelin, *Genome* **38**, 673-680 (1995).

9. A. Pane *et al.*, *Development* **129**, 3715-3725 (2002).

10. C. Concha, M. J. Scott, *Genetics* **183**, 785-798 (2009).

11. M. Hediger *et al.*, *Genetics* **184**, 155-170 (2010).

12. J. N. Shukla, S. R. Palli, *Sci. Rep.* **2**, 602 (2012).

13. I. Marin, B. S. Baker, *Science* **281**, 1990-1994 (1998).

14. M. J. D. White, *Animal Cytology and Evolution*. (Cambridge University Press, Cambridge, ed. 3rd., 1973), pp. 961 pp.

15. R. H. Baker, R. K. Sakai, *J. Hered.* **70**, 345-346 (1979).

16. A. B. Hall *et al.*, *Science* **348**, 1268-1270 (2015).

17. C. Scali, F. Catteruccia, Q. Li, A. Crisanti, *J. Exp. Biol.* **208**, 3701-3709 (2005).

18. Materials and methods are available on *Science* Online.

19. J. Krzywinski, D. R. Nusskern, M. K. Kern, N. J. Besansky, *Genetics* **166**, 1291-1302 (2004).

20. A. B. Hall *et al.*, *BMC Genomics* **14**, 273 (2013).

21. A. B. Hall *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* (2016). 10.1073/pnas.1525164113.

22. J. Juhn, A. A. James, *Insect Mol. Biol.* **15**, 363-372 (2006).

23. D. Tautz *et al.*, *Nature* **327**, 383-389 (1987).

24. D. K. Pritchard, G. Schubiger, *Genes Dev.* **10**, 1131-1142 (1996).

25. W. H. Li, in *Molecular Evolution,* W. H. Li, Ed. (Sinauer Associates, Sunderland, MA., 1997), pp. 177–214.

26. F. Criscione *et al.*, *Insect Mol. Biol.* **22**, 433-441 (2013).

27. J. C. Lucchesi, T. Skripsky, *Chromosoma* **82**, 217-227 (1981).

28. J. P. Gergen, *Genetics* **117**, 477-485 (1987).

29. A. Hilfiker *et al.*, *Development* **121**, 4017-4026 (1995).

30. J. C. Lucchesi, W. G. Kelly, B. Panning, *Annu. Rev. Genet.* **39**, 615-651 (2005).

31. M. E. Gelbart, M. I. Kuroda, *Development* **136**, 1399-1410 (2009).

32. G. J. Bashaw, B. S. Baker, *Cell* **89**, 789-798 (1997).

33. T. Kiuchi *et al.*, *Nature* **509**, 633-636 (2014).

34. G. Rose *et al.*, *Genome Biol. Evol.* **8**, 411-425 (2016).

35. J. Hemingway *et al.*, *Lancet* (2016). 10.1016/S0140-6736(15)00417-1.

36. E. F. Knipling, *Science* **130**, 902-904 (1959).

37. L. Alphey, *Annu. Rev. Entomol.* **59**, 205-224 (2014).

38. P. A. Papathanos *et al.*, *Malaria J.* **8 Suppl 2**, S5 (2009).

39. M. Q. Benedict, in *The Molecular Biology of Insect Disease Vectors,* J. M. Crampton, C. B. Beard, C. Louis, Eds. (Chapman & Hall, London, 1997), pp. 3-12.

40. J. Krzywinski, D. Sangare, N. J. Besansky, *Genetics* **169**, 185-196 (2005).

41. G. J. Lycett, D. Amenya, A. Lynd, *Insect Mol. Biol.* **21**, 79-88 (2012).

42. L. A. Bertucci, M. A. Noor, *Dros. Inf. Serv.* **84**, 166-168 (2001).

43. T. D. Wu, C. K. Watanabe, *Bioinformatics* **21**, 1859-1875 (2005).

44. S. F. Altschul *et al.*, *J. Mol. Biol.* **215**, 403-410 (1990).

45. C. Cole, J. D. Barber, G. J. Barton, *Nucleic Acids Res.* **36**, W197-201 (2008).

46. G. Cheng *et al.*, *PloS ONE* **6**, e22786 (2011).

47. I. Vandenbroucke, J. Vandesompele, A. D. Paepe, L. Messiaen, *Nucleic Acids Res.* **29**, E68-68 (2001).

48. G. J. Lycett, F. C. Kafatos, T. G. Loukeris, *Genetics* **167**, 1781-1790 (2004).

**Acknowledgments:** We thank H. Ranson, A. Wilson, and anonymous reviewers for comments, and L. Revuelta for some technical assistance. This study was supported by funds from the Liverpool School of Tropical Medicine, an MRC-LSTM Research Studentship, and Wellcome Trust grant **089045/Z/09/Z (J.K. and G.J.L.), and funds from** the Pirbright Institute (J.K.). **Transgenic mosquitoes used for sexing were developed with BBSRC grant** BB/F021933/**1 (G.J.L.). J.K. was partly supported by a subcontract CRSC120816JKUND of the NIH contract HHSN272200900039C.** The RNA-seq data have been deposited in the SRA repository under the accession numbers SRR953486.1 and SRR953487.1.

**Fig. 1**. Transcription of *Yob* and *dsx* in *A. gambiae* early embryos. (**A**) Onset of zygotic *Yob* transcription. (**B**) Establishment of the sex-specific expression pattern of *dsx*. Time of RNA sample collection in hours after oviposition as shown in above gel images; (-), negative control; ribosomal protein *S7* (*rpS7*) transcript levels were used as a gel loading control.

**Fig. 2.** *Yob* is a protein-coding sex determination gene controlling *dsx* splicing. (**A**) Reverse transcription (RT)-PCR analysis of *dsx* splicing pattern in the *A. gambiae* Sua5.1cells transfected with *in vitro* synthesized *Yob* mRNA, as compared with control non-transfected cells. M and F, *A. gambiae* male and female pupae; (-), negative control. Similar results were obtained in three independent experiments. (**B**) Digital representation of the relative amounts of *dsx* splice forms shown in panel A. (**C**) RT-PCR analysis of ribosomal protein *S7* transcript levels used as a sample loading control. (**D**) Splicing pattern of *dsx* in Sua5.1 cells transfected with native *Yob* transcripts, non-productive forms of *Yob*, either lacking the methionine start codon, or containing a premature stop codon, and in non-transfected control cells. The data shown are mean + s. d. \**P* < 0.001, one-sided Mann–Whitney test.

**Fig. 3.** Misregulation of *Yob* expression causes sex-specific lethality in embryos.(**A**) Larvae hatched from embryos injected with insufficient amount (left) or successfully delivered (right) nucleic acids, as indicated by lack or presence of GFP expression in the midgut cells. Scale bars = 1 mm. (**B**) A summary of two and four independent *Yob* mRNA microinjection experiments in *A. arabiensis* and *A. gambiae*, respectively. (**C**) A summary of four independent *Yob* double-stranded RNA microinjection experiments in *A. gambiae.* \**P* < 0.0001; Fisher’s exact test.

Supplementary Materials:

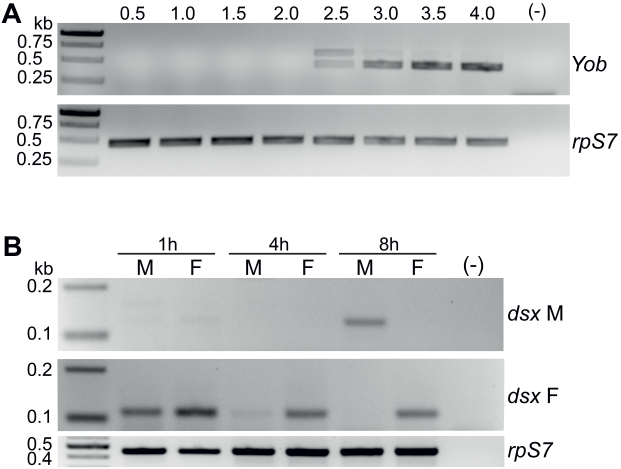
Materials and Methods

Figures S1-S10

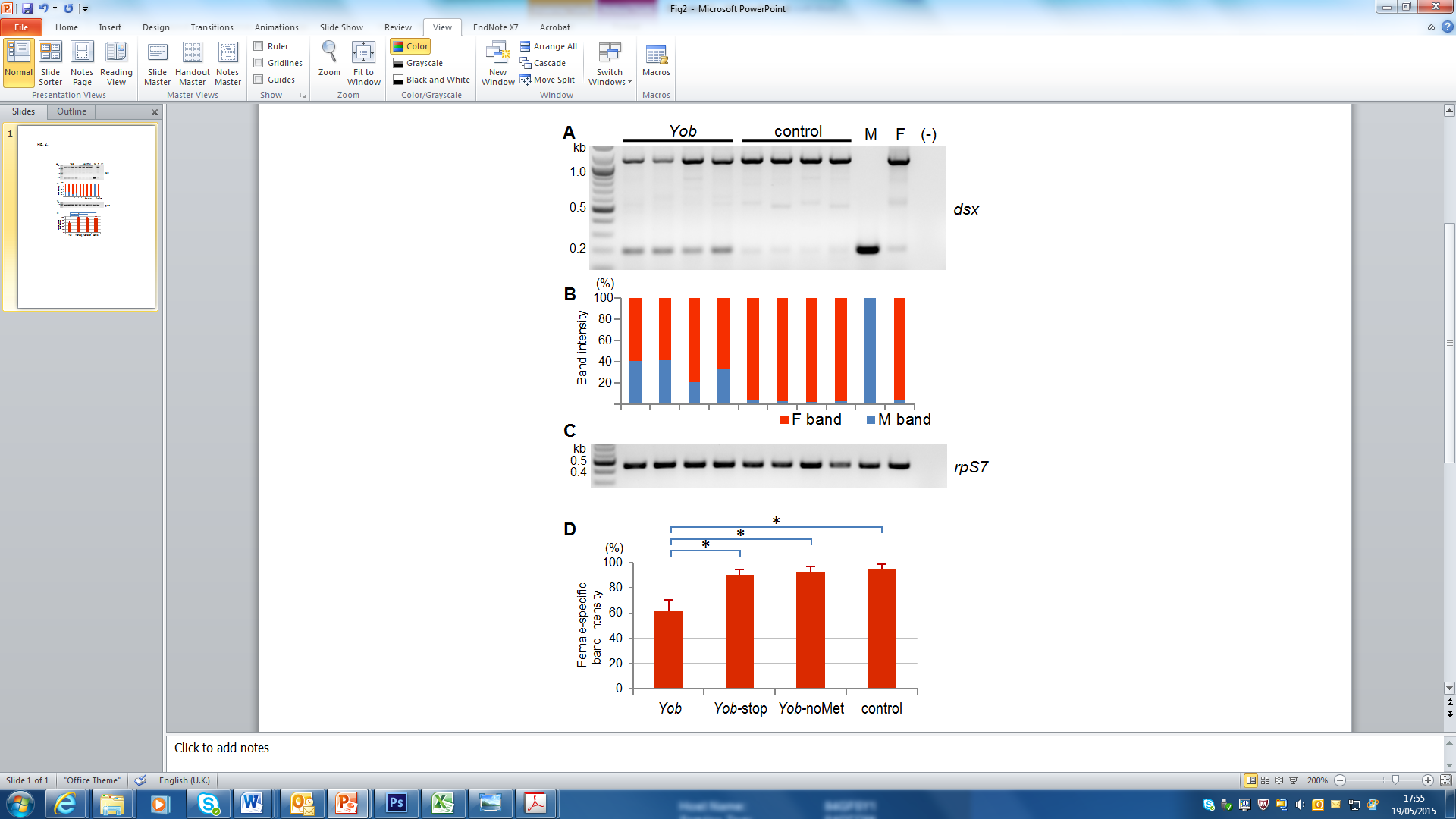
Tables S1-S3

References (*39-48*)

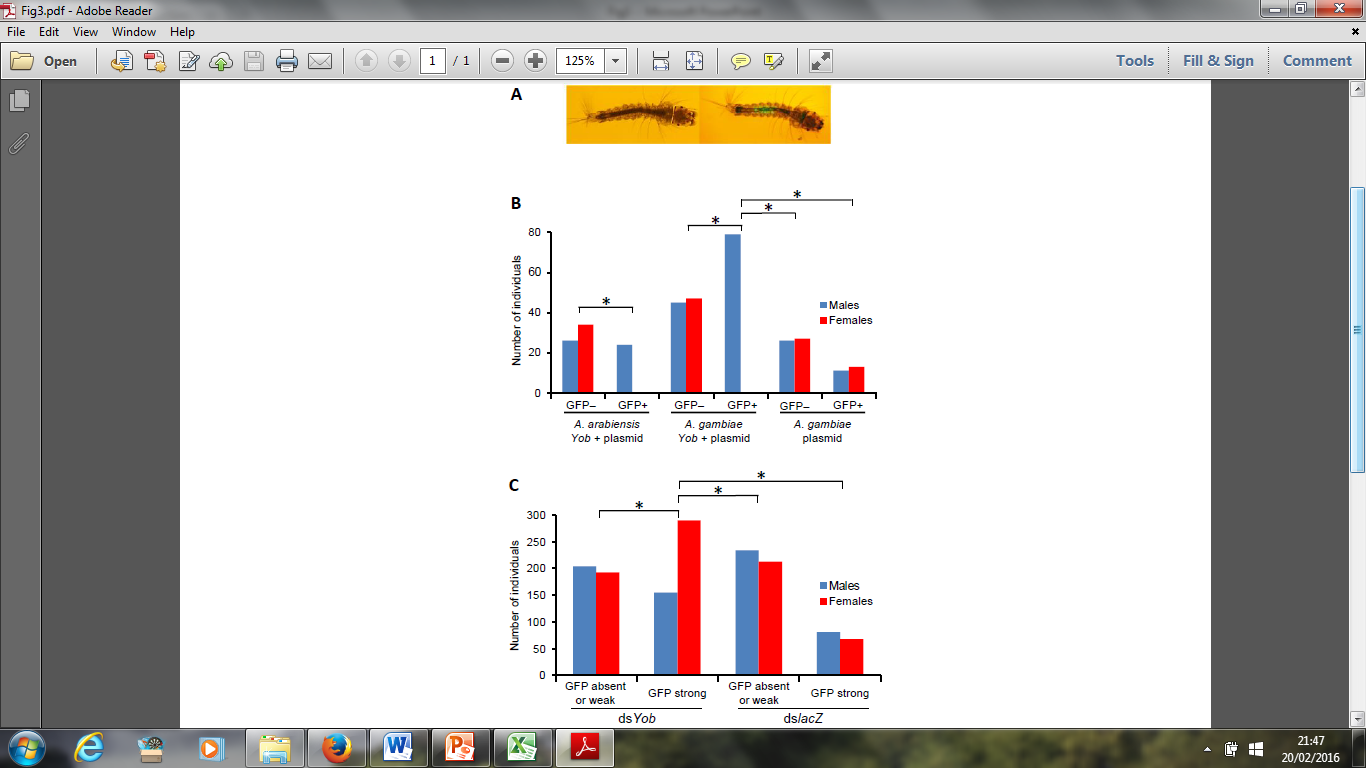
**Fig. 1.**



**Fig. 2.**



**Fig. 3.**

****