Single Nucleotide Polymorphism (SNP) in splice sites and prediction for the doublesex (dsx) gene alternative splicing in the malaria vector Anopheles gambiae.

Oswald Djihinto  
Université d'Abomey-Calavi

Helga Saizonou  
Université d'Abomey-Calavi

Luc Djogbenou  
Université d'Abomey-Calavi

Research Article

Keywords: SNP, alternative splicing, dsx gene, Anopheles gambiae, malaria

DOI: https://doi.org/10.21203/rs.3.rs-118845/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background: Malaria burden continues to be significant in tropical regions, and conventional vector control methods are faced with challenges such as insecticide resistance. To overcome these challenges, additional vector control interventions are vital and include modern genetic approaches as well as classical methods like the sterile insect technique (SIT). In the major human malaria vector Anopheles gambiae, a candidate gene favourable for sterility induction is the doublesex (dsx) gene, encoding somatic sexually dimorphic traits in mosquitoes. However, the mechanism that regulates the expression of this gene in anopheline mosquitoes is poorly understood. This study aims to screen the An. gambiae dsx gene for single-nucleotide polymorphisms (SNPs) that could be critical to its alternative splicing.

Result: Variant annotation data from Ag1000G project phase 2 was analysed, in order to identify splice-relevant SNPs within acceptor and donor splice sites of the An. gambiae dsx gene (Agdsx). SNPs were found in both donor and acceptor site of the Agdsx. No splice-relevant SNPs were identified in female-specific intron 4 acceptor site and the corresponding region in males. Two SNPs (rs48712947, rs48712962) were found in female-specific donor site of exon 5. They were not specific to either males or females. as the rs48712947 was found in female mosquitoes from Cameroon, and in both males and females from Burkina Faso. In the other splice sites, the intron 3 acceptor site carried the greatest abundance of SNPs.

Conclusion: There were no gender association between the identified SNPs and the random distribution of these SNPs in mosquito populations. The SNPs in Agdsx splice sites are not critical for the alternative splicing. Other molecular mechanisms should be considered and investigated.

Background

Malaria is a vector-borne infectious disease caused by the protozoan parasite belonging to the Plasmodium genus (1). The transmission occurs among humans through the bite of the female Anopheles mosquito. This disease is among the top ten causes of death in low-income countries (2) and continues to take a heavy toll on communities, especially in Africa regions. The malaria transmission cycle involves four major elements: the host (human), the parasite, the vector and the environment (3). In the absence of effective vaccine or sustainable treatment options, vector control is the cornerstone of malaria management and is based on prevention of human-host contact and reduction in vector population density (1, 4). The traditional vector control strategies rely on long-lasting insecticidal net (LLIN) distribution and indoor residual sprays (IRS) which have contributed to the decreasing malaria cases and mortality (5, 6). However, vector resistance against the existing insecticides is increasing in natural mosquito populations (7–9).

In the last decade, scientific advances in additional tools for vector control include technologies such as cytoplasmic incompatibility with the use of Wolbachia infection (10); repressible dominant lethal systems in Aedes aegypti (11, 12); Y-chromosome shredding gene drive (13); and the genetic sterilisation of Anopheles sp., known as Sterile Insect Techniques (SIT) (14). The latter technique, SIT, is based on the repeated, high-density release of radio-sterilized males, through gamma radiation, into the environment in order to compete with wild males for mating with the native female anopheles mosquitoes hindering the production of offspring (14). Indeed, mated females will not produce viable offspring, resulting in reduced population numbers or even elimination of the target species. However, instead of exposing males to a source of radiation, sterility could be induced by genetic modification of the mosquito genome and may improve the effectiveness of classical SIT-based approaches (14).

In An. gambiae, one of the major malaria vectors, population supression strategies are already under investigation by targeting the sex determination genes such as the doublesex (dsx) transcription factor gene (15, 16). Therefore, the Anopheles gambiae doublesex gene (Agdsx) represent a useful candidate gene for genetic manipulation and
improvement of the alternative mosquito control technologies. Interest in this gene comes from the fact that it undergoes alternative splicing and result in female and male-specific transcripts necessary for sex determination in this species (17). The use of transgenic tools in anopheline mosquitoes through targeting the dsx gene could improve the sterility induction and genetic sexing which are major requirements for SIT technologies. However, the molecular mechanisms underlying gender determination are highly variable. Though it was demonstrated that Yob1 gene (Y-linked) is one of the determining factors of the male sex (18), the molecular pathways controlling the signal of somatic sexual commitment (dsx splicing and regulation) in An. gambiae are not well understood. The only well-known model of the dsx splicing comes from the fly Drosophila melanogaster sex determination pathway (19). The dsx gene acts as a transcription factor targeting several genes which have mostly sex- and tissue-specific functions in fly (20, 21). Transformer (TRA) and Transformer 2 (TRA2) are the key regulatory factors of the female-specific alternative splicing of dsx pre-mRNA while the absence of TRA lead to the male-specific splicing (20). Both TRA and TRA2 are downstream targets of the Sex lethal gene (Sxl) product (19). Unfortunately, An. gambiae dsx gene (Agdsx) has a different structure suggesting that Agdsx sex-specific splicing event is caused by a mechanism different from that of the D. melanogaster dsx (17, 22).

In mammalian cells, the presence of genetic variations such as single nucleotide polymorphisms (SNPs) within the donor and acceptor splice sites could influence splicing and might lead to changes in normal splicing pattern (23–25). Donor sites (5’-splice site) are defined by GT dinucleotide at the 5’ end of exon-intron border, while AG dinucleotide defined acceptor sites (3’-splice site) at the 3’ end of intron-exon border (26). Thus, we hypothesize that SNPs could occur in the Agdsx acceptor and donor splice sites that might result in the splice variation With this in mind, the current report seeks to screen Agdsx for single-nucleotide polymorphisms (SNPs) that could be associated with alternative splicing.

Results

Identification of An. gambiae dsx gene (Agdsx) donor and acceptor splice sites sequence.

Agdsx is located in the band 17C of the chromosome 2R (2R: 48703664–48788460) on the reverse strand. The gene is 84.8 kb long and encodes the male and female-specific transcript. The male transcript (6975 bp) is shorter than that of female (8667 bp). The difference between the two sex-specific transcripts is due to the alternative splicing of exon 5. This latter is a cassette exon, which is retained in female and skipped in male transcript. The whole sequence of female specific exon 5 is included in male intron 4 region and is spliced out. This gene structure causes a shift in intron/exon number in male. Thus, male and female have common and specific splice sites.

Male and female mosquitoes share exon 1, 2, 3, 4 and 6 donor splice sites while exon 5 donor site is specific to female (Table 2). Similarly, both sexes share intron 1, 2, 3, and 6 acceptor sites. Male intron 4 and female intron 5 share the same 3’ end as the female exon 5 is included in the male intron 4 sequence. However females have the intron 4 specific acceptor site (Table 2).
Table 2

Splice donor and acceptor sites within the double sex (dsx) gene of *Anopheles gambiae*.

| Splice donor sites |
|--------------------|
| Sex | Exon | Size | Exon position | Splice site sequence | Site position |
| --- | --- | --- | --- | --- | --- |
| Male/Female 1 | 1 | 1415 | 48788460 | 48787046 | tatttg/gtaagtaaatatgcaa | 48787051 48787030 |
| Male/Female 2 | 2 | 1445 | 48785629 | 48784185 | TGGGAG/gtaagctggcgatgat | 48784190 48784169 |
| Male/Female 3 | 3 | 45 | 48747737 | 48747693 | TACCTG/gtaagtaaatataatt | 48747698 48747677 |
| Male/Female 4 | 4 | 135 | 48715295 | 48715161 | ACGAAG/gtaagctggcgatgat | 48715166 48715145 |
| Female 5 | 5 | 1692 | 48714648 | 48712957 | cagaag/gtatggtaagacggcc | 48712962 48712941 |
| Male/Female 6 | 6 | 1267 | 48712794 | 48711528 | aaaaag/gtatggtggtgtagta | 48711533 48711512 |
| Male/Female 7 | 7 | 2668 | 48706331 | 48703664 | None | |

| Splice acceptor sites |
|-----------------------|
| Sex | Intron | Size | Intron position | Splice site sequence | Site position |
| --- | --- | --- | --- | --- | --- |
| Male/Female 1 | 1 | 1416 | 48787045 | 48785630 | gtacgtttgatt | 48785645 48785624 |
| Male/Female 2 | 2 | 36447 | 48784184 | 48747738 | ttgctctccttt | 48747753 48747732 |
| Male/Female 3 | 3 | 32397 | 48747692 | 48715296 | ttccgccccgtt | 48715311 48715290 |
| Female 4 | 4 | 512 | 48715160 | 48714649 | tttatgttaacacag | 48714664 48714643 |
| Male 4 | 4 | 2366 | 48715160 | 48712795 | tgaaccccccaaaaag | 48712810 48712789 |
| Female 5 | 5 | 162 | 48712956 | 48712795 | |
| Male/Female 6 | 6 | 5196 | 48711527 | 48706332 | cgctcctcaaaaaatag | 48706347 48706326 |

Splice site sequences are given in 5'→3' direction on the reverse strand. Exonic coding sequences are shown in uppercase letters, and non-coding regions are in lowercase letters. The 12 bp preceding the 3' splice-acceptor site (NYag) is indicated, where Y = T or C and N = any nucleotide.

**SNPs in female-specific intron 4 acceptor and exon 5 donor splice sites.**

Along the *Agdsx* gene, 17196 polymorphic sites were identified. The nucleotide diversity is similar between male and female mosquitoes (Fig. 1). The potential splice-relevant SNPs that could trigger the female-specific exon 5 skipping should be in the intron 4 acceptor and exon 5 donor sites. However, there was no SNP in acceptor sequence of female-specific intron 4 nor in the male corresponding region (Fig. 2). However in female specific exon 5 donor site, two SNPs (rs48712947, rs48712962) were found. Nevertheless, they were not specific to female as the rs48712947 was found in Cameroon female mosquitoes and in both males and females from Burkina Faso (Fig. 3). The rs48712962 was absent in male mosquito population, while it was found only in Cameroon females. The minor allele frequencies (MAF) of both SNPs identified were very low in each population. The MAF of rs48712947 and rs48712962 were less than 1% in each female population and only 2% of Burkina Faso male carried the rs48712947. Moreover, none of these SNPs were associated with the sex phenotype (rs48712947: \( p = 0.32 \); rs48712962: \( p = 0.68 \)).

**SNPs in other splice sites of Agdsx.**
The other splice sites were also examined for identification of sex-specific SNPs (see additional Figs. 1–3). No SNP was found in the shared exon 1 donor, introns 1 and 6 acceptor sites. The highest number of SNPs (07) was found in the common intron 3 acceptor site sequence (rs48715291, rs48715294, rs48715302, rs48715306, rs48715307, rs48715308, rs48715309) (Fig. 4). However, each of these SNPs occurred in non-specific manner in both male and female populations with variable minor allele frequencies.

Discussion

The *An. gambiae doublesex* (*Agdsx*) gene is a candidate gene of interest for sterile insect technique (SIT), as a candidate for genetic modifications (15, 16). The translation and the success of using *dsx* in SIT methodology require a clearer understanding of the genetic bases of the sex determination pathway. This study screened the *Agdsx* donor and acceptor splice sites for identification of splice-relevant SNPs.

According to the *D. melanogaster* model (19), the alternative splicing of *Agdsx* gene is governed by exon 5 skipping in male mosquitoes (17) suggesting a silencing mechanism of the female-specific splice sites recognition (intron 4 acceptor and exon 5 donor sites) by the splicing machinery in male. Such silencing mechanism could be due to changes in splice site sequence. However, female-specific intron 4 acceptor site sequence is present within male intron 4 and no SNP was found in this sequence in both sex. The SNPs rs48712947 and rs48712962 identified in female-specific exon 5 donor site were neither splice-relevant nor sex-specific. They appeared only in two mosquito populations (Burkina Faso and Cameroon) over the eight populations considered. In each population where these SNPs have been identified, they appeared in very few individuals, less than 1% in females and no more than 2% in males. These observations suggest that the *Agdsx* cassette exon 5 was not associated with changes in splice site patterns due to the presence of SNPs. The presence of SNPs in the other splice sites had also different distribution and were non-specific to the gender of the mosquitoes.

Another factor for exon skipping is the pyrimidine content of the polypyrimidine tract in acceptor splice sequence. Indeed a poor polypyrimidine tract cause a shift of the splicing machinery to the next acceptor site, leading to the skipping as the case of exon 4 skipping in male *Drosophila* (19). In *Anopheles gambiae* the number of pyrimidine (8) in the 12 bp preceding the acceptor site pattern (acag) (Table 2) in the female-specific intron 4 is the same in the male corresponding region. The same number of pyrimidines in this acceptor sequence was reported by Scali et al. (17). Furthermore, the authors found that this number did not differ from the consensus number of pyrimidines (8.69) in *An. gambiae* splice acceptor sites, and concluded that the intron 4 site may not be a weak acceptor site (17). The Drosophila *dsx* splicing regulation involves the products of *transformer* (*Tra*) and *transformer 2* (*Tra2*) genes (19). Although ortholog of *Drosophila Tra2* is present in *An. gambiae* (accession number: AGAP006798), no ortholog of *Tra* was found in *An. gambiae* genome (www.flybase.org). This suggests that *Tra* may be missing in *An. gambiae* or its splicing regulatory function could be ensured by another gene. Overall, these findings add another evidence that other mechanisms underlie the alternative splicing in *An. gambiae* and open perspective for further investigation on the molecular mechanisms of *Agdsx* splicing.

It was known that the regulation of alternative splicing evolved trans-acting splicing factors, such as serine-arginine-rich (SR) family proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) that bind to the auxiliary silencer and enhancer *cis*-element (ESE: exonic splicing enhancing; ESI: exonic splicing silencing; ISE: intronic splicing enhancing; ISI: intronic splicing silencing) (27–29). Similar regulatory *cis*-elements were found in *Drosophila melanogaster* female-specific exon and putative homologs were identified in *An. gambiae* female-specific exon 5 (17). Therefore, further molecular analysis are needed toward characterizing these regulatory sequence and their binding trans-factors in order to underpin the somatic sex determination in *An. gambiae*. 
Moreover, the epigenetic system was also reported to regulate the alternative splicing in mammalian and other insect cells. Indeed, it was demonstrated that changes in DNA cytosine methylation on the gene body in honey bees may lead to alternative splicing (30–32). Also histone post-translational modifications (PTMs) such as lysine acetylation and methylation were associated to the alternative splicing event (33–35). Consequently, similar mechanisms could happen in the malaria vector *An. gambiae* to regulate gene alternative splicing. However, no significant DNA methylation was reported in Diptera including *An. gambiae* (36, 37). Then, the only epigenetic modifications that could be linked to the alternative splicing in this species are histone PTMs. Indeed, the methylation and acetylation of lysines 4, 9 and 29 of histone H3 were reported in *An. gambiae* (38, 39). Then, it will be interesting to evaluate whether such histone modifications enrichment in *Agdsx* between male and female mosquitoes could be critical for *dsx* alternative splicing.

**Conclusion**

Sustainable vector control strategies will rely on the integrated use of chemicals and biological vector control. Given the potential of the *Anopheles gambiae dsx (Agdsx)* gene for sterile insect technique (SIT), the understanding of mechanisms of it regulation could help to improve the tools engineering targeting this locus. SNPs were identified within the *Agdsx* and their putative association with the *dsx* alternative splicing was analysed. No splice-relevant SNP was found in the specific male and female splice site. The SNPs were distributed in few proportion of individuals in the populations where they were identified. With the advances in genetic biotechnology, other mechanisms remain to be explored for providing solid background on somatic sexual fate determination in *Anopheles gambiae*. This will pave the way to find new biochemical and genetics target for malaria vector control.

**Method**

**Sequence data and mosquito samples**

Genomic sequences used in this study came from the Anopheles 1000 genomes (Ag1000G) project phase 2 released in 2017 (22). The SNP annotation was downloaded (ag1000g.phase2.ar1.variants.pass.2R.vcf.gz, November 11, 2019) from the Malaria Genomic Epidemiology Network (MalariaGEN) website. This file contain all SNPs identified in mosquitoes whole genome and that pass the variant filtering process described by (22). Only *Anopheles gambiae* samples were considered in our study. These mosquito samples were collected from natural populations from 2002 to 2012 in eight African countries (Table 1). The reference sequence of *Agdsx* (AGAP004050) was also downloaded from Vectorbase website.
**Table 1**
Sampling locations of *An. gambiae* mosquitoes from Ag1000G phase 2 project.

| Country          | Site                  | Year | Geographic coordinate | Number of species |
|------------------|-----------------------|------|-----------------------|-------------------|
|                  |                       |      | Latitude | Longitude |                  |
| Burkina Faso     | Bana                  | 2012 | 11.2330 | -4.4720  | 20                |
|                  | Pala                  | 2012 | 11.1500 | -4.2350  | 46                |
|                  | Souroukoudinga        | 2012 | 11.2350 | -4.5350  | 26                |
| Cameroon         | Daiguene              | 2009 | 4.7770  | 13.8440  | 96                |
|                  | Gado Badzere          | 2009 | 5.7470  | 14.4420  | 73                |
|                  | Mayos                 | 2009 | 4.3410  | 13.5580  | 105               |
|                  | Zembe Borong          | 2009 | 5.7470  | 14.4420  | 23                |
| Equatorial Guinea| Bioko                 | 2002 | 3.7000  | 8.7000   | 9                 |
| France (Mayotte) | Bouyouni              | 2011 | -12.7378| 45.1417  | 1                 |
|                  | Combani               | 2011 | -12.7787| 45.1429  | 5                 |
|                  | Karihani Lake         | 2011 | -12.7965| 45.1217  | 3                 |
|                  | Mont Benara           | 2011 | -12.8570| 45.1552  | 2                 |
|                  | Mtsamboro Forest Reserve| 2011| -12.7027| 45.0811  | 1                 |
|                  | Mtsanga Charifou      | 2011 | -12.9907| 45.1557  | 8                 |
|                  | Sada                  | 2011 | -12.8521| 45.1039  | 4                 |
| Gabon            | Libreville            | 2000 | 0.3840  | 9.4550   | 69                |
| Ghana            | Madina                | 2012 | 5.6685  | -0.2193  | 12                |
| Guinea           | Koraboh               | 2012 | 9.2500  | -9.9170  | 22                |
|                  | Koundara              | 2012 | 8.5000  | -9.4170  | 18                |
| Uganda           | Tororo (Nagongera)    | 2012 | 0.7700  | 34.0260  | 112               |

**Sequence analysis and SNP identification**

From the *Agdsx* reference sequence, the list of genomic positions of donor and acceptor sites was extracted. VCFtools version 0.1.15 (40) was used to extract the SNPs within the genomic region corresponding to *Agdsx* sequence from the SNPs annotation file. The polymorphic nucleotides were then identified within the splice sequences, in comparison to the reference sequence. SNPs were then visualized using TASSEL version 5.2.63 software (41). The genomic position of the acceptor sites was used to select SNPs in the last 12 nucleotides of an intron preceding the 3’ splice pattern NYAG and in the first 6 nucleotides of an exon. In donor splice sites SNPs were identified within in the last 6 nucleotides of an exon and the first 16 nucleotide in an intron. The SNPs association to the sex phenotype (male or female) was evaluated by running the association analysis using the general linear model (GLM) function in TASSEL. The averages nucleotide diversity at the *dsx* locus between male and female was calculated using scikit-allel version 1.2.1 (42) in order to determine whether SNPs density at the *dsx* locus is different between the two genders.
Abbreviations

Agdsx
*Anopheles gambiae doublesex gene*

dsx
doublesex gene

ESE
Exonic Splicing Enhancers

ESI
Exonic Splicing Silencers

hnRNPs
heterogeneous nuclear ribonucleoproteins

ISE
Intronic Splicing Enhancers

ISI
Intronic Splicing Silencers

PTMs
post-translational modifications

SIT
Sterile Insect Techniques

SNP
Single Nucleotide Polymorphism

Sxl
Sex lethal gene

TRA
Transformer transcription factor

TRA2
Transformer 2 transcription factor

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are available from Ag1000G phase 2 via the MalariaGEN website (ftp://ngs.sanger.ac.uk/production/ag1000g/phase2/AR1/variation/main/vcf/pass_variants/)

Competing interests

The authors declare that they have no competing interests

Funding
**Authors’ contributions**

HS retrieved data and extracted data corresponding to the genomic region of interest. OD analysed data and wrote the manuscript. LD contributed to the critical review of the draft manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

The authors are grateful to Dr. Luisa Nardini, Dr. Lizette Koekemoer and Dr David Weetman for their helpful comments on this manuscript. The authors also acknowledge Wellcome Trust for financial support to LSD (grant 109917/Z/15/Z).

**Authors’ information (optional)**

Not applicable

**References**

1. Greenwood BM, Bojang K, Whitty CJ, Targett GA. Malaria. The Lancet. 2005 Apr;365(9469):1487–98.
2. World Health Organization. The top 10 causes of death [Internet]. 2018 [cited 2020 Jul 6]. Available from: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death
3. Castro MC, Singer BH. Malaria in the Brazilian Amazon. In: Water and Sanitation-Related Diseases and the Environment: Challenges, Interventions, and Preventive Measures. John Wiley & Sons; 2011. p. 401–19.
4. Abdul-Ghani R, Al-Mekhlafi AM, Alabsi MS. Microbial control of malaria: Biological warfare against the parasite and its vector. Acta Tropica. 2012 Feb;121(2):71–84.
5. Bhatt S, Weiss D, Cameron E, Bisanzio D, Mappin B, Dalrymple U. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526(7572):207–11.
6. World Health Organization. World malaria report 2019 [Internet]. 2019. Available from: https://www.who.int/publications/i/item/world-malaria-report-2019
7. Kamgang B, Tchapga W, Ngoagouni C, Sangbakembi-Ngounou C, Wondji M, Riveron JM, et al. Exploring insecticide resistance mechanisms in three major malaria vectors from Bangui in Central African Republic. Pathogens and Global Health. 2018;112(7):349–59.
8. Tchouakui M, Tene Fossog B, Ngannang BV, Djonabay D, Tchapga W, Njiokou F, et al. Investigation of the influence of a glutathione S-transferase metabolic resistance to pyrethroids/DDT on mating competitiveness in males of the African malaria vector, Anopheles funestus. Wellcome Open Res. 2019 Mar 21;4:13.
9. Collins E, Vaselli NM, Sylla M, Beavogui AH, Orsborne J, Lawrence G, et al. The relationship between insecticide resistance, mosquito age and malaria prevalence in Anopheles gambiae s.l. from Guinea. Sci Rep. 2019 Jun 20;9(1):1–12.
10. LePage D, Bordenstein SR. Wolbachia: Can we save lives with a great pandemic? Trends Parasitol. 2013 Aug;29(8):385–93.
11. Phuc HK, Andreasen MH, Burton RS, Vass C, Epton MJ, Pape G, et al. Late-acting dominant lethal genetic systems and mosquito control. BMC Biol. 2007 Mar 20;5:11.
12. Fu G, Lees RS, Nimmo D, Aw D, Jin L, Gray P, et al. Female-specific flightless phenotype for mosquito control. Proc Natl Acad Sci USA. 2010 Mar 9;107(10):4550–4.
13. Prowse TA, Adikusuma F, Cassey P, Thomas P, Ross JV. A Y-chromosome shredding gene drive for controlling pest vertebrate populations. Messer PW, Tautz D, editors. eLife. 2019 Feb 15;8:e41873.

14. Catteruccia F, Crisanti A, Wimmer EA. Transgenic technologies to induce sterility. Malaria Journal. 2009 Nov 16;8(2):S7.

15. Kyrou K, Hammond AM, Galizi R, Kranjc N, Burt A, Beaghton AK, et al. A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. Nature Biotechnology. 2018 Nov;36(1):1062–6.

16. Taracena ML, Hunt CM, Benedict MQ, Pennington PM, Dotson EM. Downregulation of female doublesex expression by oral-mediated RNA interference reduces number and fitness of Anopheles gambiae adult females. Parasites & Vectors. 2019 Apr 15;12(1):170.

17. Scali C, Catteruccia F, Li Q, Crisanti A. Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene. J Exp Biol. 2005 Oct;208(Pt 19):3701–9.

18. Krzywinska E, Dennison NJ, Lycett GJ, Krzywinski J. A maleness gene in the malaria mosquito Anopheles gambiae. Science. 2016 Jul 1;353(6294):67–9.

19. Black DL. Mechanisms of Alternative Pre-Messenger RNA Splicing. Annu Rev Biochem. 2003 Jun 1;72(1):291–336.

20. Clough E, Jimenez E, Kim Y-A, Whitworth C, Neville MC, Hempel LU, et al. Sex- and Tissue-Specific Functions of Drosophila Doublesex Transcription Factor Target Genes. Developmental Cell. 2014 Dec 22;31(6):761–73.

21. Chatterjee SS, Uppendahl LD, Chowdhury MA, Ip P-L, Siegal ML. The female-specific Doublesex isoform regulates pleiotropic transcription factors to pattern genital development in Drosophila. Development. 2011 Mar 15;138(6):1099–109.

22. Consortium TA gambiae 1000 G, Clarkson CS, Miles A, Harding NJ, Lucas ER, Battey CJ, et al. Genome variation and population structure among 1,142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii. bioRxiv. 2019 Dec 9;864314.

23. Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, et al. Single-Nucleotide Polymorphisms in NAGNAG Acceptors Are Highly Predictive for Variations of Alternative Splicing. Am J Hum Genet. 2006 Feb;78(2):291–302.

24. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet. 2002 Apr;3(4):285–98.

25. Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. Nat Biotechnol. 2004 May;22(5):535–46.

26. Wachutka L, Caizzi L, Gagneur J, Cramer P. Global donor and acceptor splicing site kinetics in human cells. Black DL, Manley JL, editors. eLife. 2019 Apr 26;8:e45056.

27. Smith CWJ, Valcárcel J. Alternative pre-mRNA splicing: the logic of combinatorial control. Trends in Biochemical Sciences. 2000 Aug 1;25(8):381–8.

28. Singh R, Valcárcel J. Building specificity with nonspecific RNA-binding proteins. Nat Struct Mol Biol. 2005 Aug;12(8):645–53.

29. Izquierdo J-M, Valcárcel J. A simple principle to explain the evolution of pre-mRNA splicing. Genes Dev. 2006 Jan 7;20(13):1679–84.

30. Li-Byarlay H. The Function of DNA Methylation Marks in Social Insects. Front Ecol Evol [Internet]. 2016 [cited 2020 Jun 2];4. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2016.00057/full

31. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. Keller L, editor. PLoS Biol. 2010 Nov 2;8(11):e1000506.

32. Cingolani P, Cao X, Khetani RS, Chen C-C, Coon M, Sammak A, et al. Intronic Non-CG DNA hydroxymethylation and alternative mRNA splicing in honey bees. BMC Genomics. 2013 Sep 30;14:666.
33. Rahhal R, Seto E. Emerging roles of histone modifications and HDACs in RNA splicing. Nucleic Acids Research. 2019 Jun 4;47(10):4911–26.

34. Gunderson FQ, Merkhofer EC, Johnson TL. Dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements. Proc Natl Acad Sci USA. 2011 Feb 1;108(5):2004–9.

35. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of Alternative Splicing by Histone Modifications. Science. 2010 Feb 19;327(5968):996–1000.

36. Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA Methylation across Insects. Mol Biol Evol. 2017 Mar;34(3):654–65.

37. Falckenhayn C, Carneiro VC, de Mendonça Amarante A, Schmid K, Hanna K, Kang S, et al. Comprehensive DNA methylation analysis of the *Aedes aegypti* genome. Scientific Reports. 2016;6:36444.

38. Gomez-Diaz E, Rivero A, Chandre F, Corces VG. Insights into the epigenomic landscape of the human malaria vector *Anopheles gambiae*. Frontiers in Genetics [Internet]. 2014 Aug 15 [cited 2019 Mar 13];5. Available from: http://journal.frontiersin.org/article/10.3389/fgene.2014.00277/abstract

39. Ruiz JL, Yerbanga RS, Lefèvre T, Ouedraogo JB, Corces VG, Gómez-Díaz E. Chromatin changes in *Anopheles gambiae* induced by *Plasmodium falciparum* infection. Epigenetics & Chromatin [Internet]. 2019 Dec [cited 2019 Mar 13];12(1). Available from: https://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-018-0250-9

40. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.

41. Bradbury P, Zhang Z, Kroon D, Casstevens T, Ramdoss Y, Buckler E. TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. Bioinformatics (Oxford, England). 2007 Nov 1;23:2633–5.

42. Miles A, Harding N. Cggh/Scikit-Allel: V1.1.0 [Internet]. Zenodo; 2017 [cited 2020 Jul 9]. Available from: https://zenodo.org/record/815477