RESEARCH ARTICLE

REVISED Single nucleotide polymorphism (SNP) in the *doublesex* (*dsx*) gene splice sites and relevance for its alternative splicing in the malaria vector *Anopheles gambiae* [version 3; peer review: 2 approved, 1 approved with reservations]

Oswald Djihinto[1], Helga D.M. Saizonou[1], Luc S. Djogbenou[1-3]

[1]Tropical Infectious Diseases Research Centre (TIDRC), University of Abomey-Calavi, Abomey-Calavi, 01BP526 Cotonou, Benin
[2]Institut Régional de Santé Publique, University of Abomey-Calavi, Ouidah, BP 384 Ouidah, Benin
[3]Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

v3 **First published:** 28 Jan 2022, **7**:31
https://doi.org/10.12688/wellcomeopenres.17572.1
Second version: 19 Dec 2022, **7**:31
https://doi.org/10.12688/wellcomeopenres.17572.2
Latest published: 06 Feb 2023, **7**:31
https://doi.org/10.12688/wellcomeopenres.17572.3

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, involved in mosquitos' somatic sexually dimorphic traits determination. However, the pathways that trigger the signal of *dsx* gene exon skipping alternative splicing mechanism in anopheline mosquitoes are not well characterized. This study aims to screen the *An. gambiae dsx* gene splice site sequences for single-nucleotide polymorphisms (SNPs) that could be critical to its alternative splicing.
Methods: 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*).
Results: SNPs were found in both donor and acceptor sites of the *Agdsx*. No splice-relevant SNPs were identified in the female-specific intron 4 acceptor site and the corresponding region in males. Two SNPs (rs48712947, rs48712962) were found in the 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.

Open Peer Review

Approval Status ? ✓ ✓

	1	2	3
version 3 (revision) 06 Feb 2023		✓ view	✓ view
version 2 (revision) 19 Dec 2022		✓ view	
version 1 28 Jan 2022	? view	✗ view	

1. **Tony Nolan**, Liverpool School of Tropical Medicine, Liverpool, UK
2. **Yoosook Lee**, University of Florida, Vero Beach, USA
3. **Moussa W Guelbeogo**, Centre National de Recherche et Formation sur le Paludisme (CNRFP), Ouagadougou, Burkina Faso

Any reports and responses or comments on the article can be found at the end of the article.

Conclusions: 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.

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

Corresponding author: Luc S. Djogbenou (ldjogbenou22002@yahoo.fr)

Author roles: Djihinto O: Conceptualization, Data Curation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Saizonou HDM**: Data Curation, Writing – Review & Editing; **Djogbenou LS**: Funding Acquisition, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This study was supported by Wellcome [109917/Z/15/Z, to LSD] under the Intermediate Fellowship; bioinformatics training was provided to OYD and HDMS as a PhD student; OYD was a fellow of the Beninese government fellowship scheme "Programme Appui aux doctorants" 2018-2019.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Djihinto O, Saizonou HDM and Djogbenou LS. **Single nucleotide polymorphism (SNP) in the *doublesex* (*dsx*) gene splice sites and relevance for its alternative splicing in the malaria vector *Anopheles gambiae* [version 3; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2023, **7**:31 https://doi.org/10.12688/wellcomeopenres.17572.3

First published: 28 Jan 2022, **7**:31 https://doi.org/10.12688/wellcomeopenres.17572.1

REVISED **Amendments from Version 2**

We thank the reviewer for the consideration of the revised manuscript.

In this updated version, we addressed the new reviewer comments as below:

Reviewer: "Moreover, the epigenetic system was also reported to regulate the alternative splicing in mammalian and other insects cells." -> change to 'other insect cells'.

Reply: Done

Reviewer: Some citations (e.g. World Health Organization fact sheet) seems like they should be in a format consistent with other references (hyper link text to numbers).

Reply: Done.

Any further responses from the reviewers can be found at the end of the article

List of abbreviations

Agdsx*: *Anopheles gambiae doublesex gene

dsx: doublesex gene

ESE: Exonic Splicing Enhancers

ESI: Exonic Splicing Silencers

Fle: Femaleness gene

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

Introduction

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 (World Health Organization)[2] and continues to take a heavy toll on communities, especially in African 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 the 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].

The widespread of insecticide resistance in natural vector populations has intensified researches on alternative malaria vectors control strategies. Alternative tools for vector control have included technologies such as cytoplasmic incompatibility with the use of natural *Wolbachia* bacteria infection[10,11]; repressible dominant lethal systems in *Aedes aegypti*[12,13]; Y-chromosome shredding gene drive (gene drive cassette that also incorporates a programmable endonuclease that shreds the Y chromosome, converting XY males into fertile XO females)[14]; and the genetic sterilisation of *Anopheles* sp., known as Sterile Insect Techniques (SIT)[15]. The SIT technique, as firstly developed, 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[16,17]. Indeed, mated females will not produce viable offspring, resulting in reduced population numbers or even local 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[15].

In *An. gambiae*, one of the major malaria vectors, population suppression strategies are already under investigation by targeting the gender determination genes such as the *doublesex* (*dsx*) transcription factor gene[18,19]. Therefore, the *Anopheles gambiae doublesex* gene (*Agdsx*) represents 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 results in female and male-specific transcripts necessary for gender determination in this species[20]. 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 genetic SIT technologies. However, the molecular mechanisms underlying gender determination are highly variable.

The only well-known model of the *dsx* splicing comes from the fly *Drosophila melanogaster* sex determination pathway[21]. The *dsx* gene acts as a transcription factor targeting several genes which have mostly sex- and tissue-specific functions that determine somatic sexual dimorphism traits in later stages of sexual development[22,23]. Transformer (*TRA*) and Transformer 2 (*TRA2*) are the key regulatory factors of the female-specific alternative splicing of *dsx* pre-mRNA (*dsxF* isoform) under the control of the Sex lethal gene (*Sxl*) product while the absence of *TRA* (non-productive form) leads to the male-specific splicing (*dsxM* isoform) in fruit fly[22]. Unfortunately, *An. gambiae dsx* gene (*Agdsx*) has different gene organization and regulatory elements positions suggesting that *Agdsx* gender-specific splicing event is caused by a mechanism different from that of the *D. melanogaster dsx*[20,24]. Recently, it was reported that in *An. gambiae*, *femaleness* gene (*Fle*) is necessary for the splicing of *dsx* into the female-specific mRNA

(*AgdsxF*)[25]. However, *Fle* is not involved in the *dsx* splicing into the male-specific transcript (*AgdsxM*)[25]. Indeed, *Yob1* gene (Y-linked) which is activated at earlier stage of zygotic transcription and expressed all throughout a male's life, regulates male-specific *dsx* splicing[26].

Agdsx is located in the 17C band of the chromosome 2R (2R: 48703664 - 48788460) on the reverse strand. The gene is 84.8 kb long and encodes *AgdsxM* and *AgdsxF* transcripts. *AgdsxM* transcript (6975 bp) is shorter than that of *AgdsxF* (8667 bp). The difference between the two gender-specific transcripts is due to the alternative splicing of exon 5. The latter is a cassette exon, which is retained in female and skipped in male transcript. The whole sequence of the female-specific exon 5 is included in the male intron 4 region and is spliced out. This gene structure causes a shift in intron/exon number in male. Thus, although male and female share the same exon/intron or intron/exon boundaries, they have common and specific splice sites (Figure 1). Though it was demonstrated that *Fle* and *Yob1* genes control respectively *AgdsxF* and *AgdsxM* specific splicing, the pathways triggering the signal of *dsx* gene exon skipping alternative splicing mechanism in *An. gambiae* are not well characterized.

The exon definition by the spliceosome requires interplays between splice sites on either side of the exon. 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[27]. In mammalian cells, the presence of genetic variations such as single nucleotide polymorphisms (SNPs) within the donor and acceptor splice sites is susceptible to influence the splicing and might lead to changes in normal splicing pattern[28–30]. The presence of SNP at the acceptor splice site of several genes is reported in human and lead to the alternative splicing of the corresponding genes[28]. Indeed, in humans, splicing signals are a common point of mutations. Most of the splicing mutations analysed so far directly influence the conventional consensus splicing sequence, and consequently lead to skipping of the adjacent exon[29]. Lamba *et al.*, revealed that a nonsynonymous SNP (15631G>T), which disrupted an exonic splicing enhancer (in exon 4), and a SNP (15582C>T) in an intron-3 branch

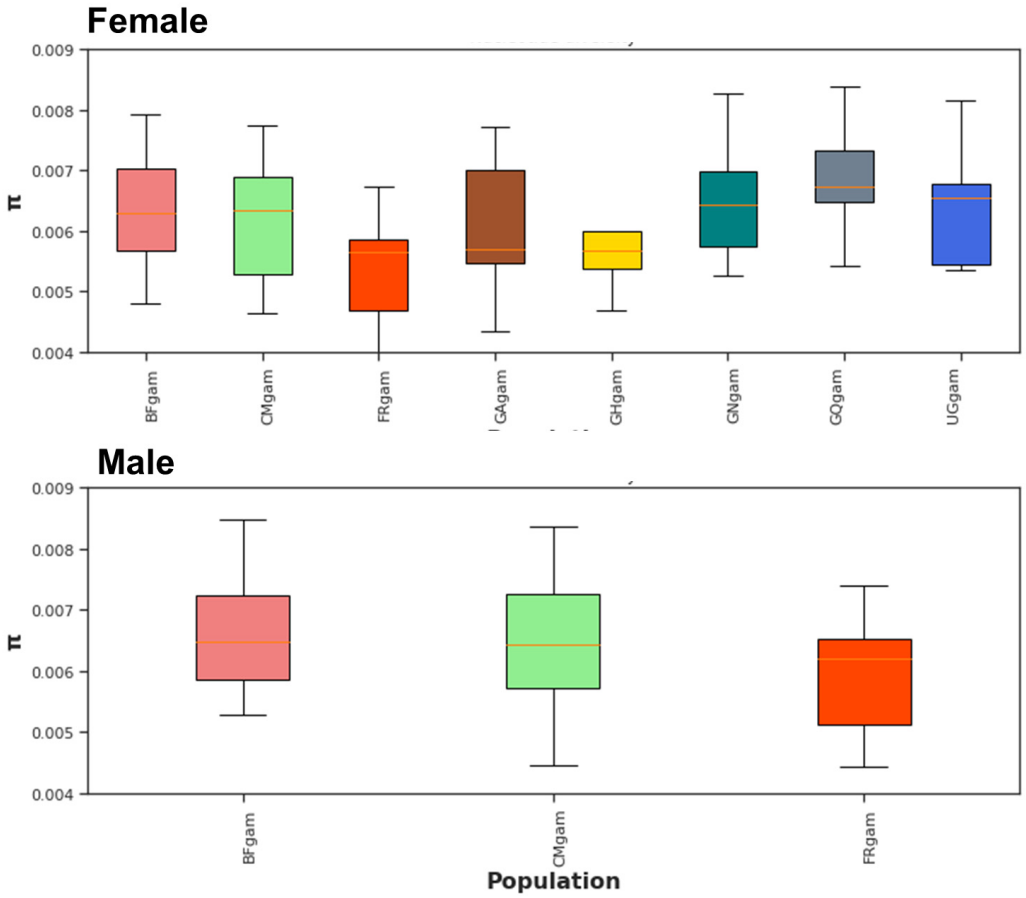

Figure 1. Schematic of common and specific splice sites of *dsx* in *An. gambiae* males and females. Exons are represented irrespective of their size by black boxes. Straight black lines indicate introns. Coloured lines above exon/intron junction indicate donor splice sites. Coloured lines above intron/exon junction indicate acceptor splice sites. Same coloured line above exon/intron or intron/exon boundaries indicate common splice sites in both *AgdsxF* and *AgdsxM* transcripts. The red and green splice sites are specific to *AgdsxF* transcript.

point are responsible for the skipping of exons 4, 5, and 6 of cytochrome 2B6 (*CYP2B6*) in females human liver[31]. Furthermore, coding single-nucleotide polymorphisms (cSNPs) are thought to have the same effect on splicing[29]. Moreover in animals, especially in cattle, the *ectodysplasin 1* gene (*ED1*) produces two isoforms that result from alternative splicing. It was reported that this alternative splicing event in *ED1* mRNA is caused by a point mutation found in the 5′ splice donor site of intron 8[32]. Gargani *et al.*, have also showed that another single nucleotide polymorphism (SNP) in the exonic splicing enhancer (ESEs) of the exon 8 in *ED1* mRNA leads to the exon skipping in cattle[33]. In the marine fish *Trachinotus anak*, Fan *et al.*, have identified a single intronic SNP at the first exon/intron boundary of the gender-determining gene *17b-hydroxysteroid dehydrogenase 1* (*hsd17b1*)[34]. Specifically, this SNP was demonstrated to affect *hsd17b1* splicing, leading to female development[34].

Taking together these observations in humans and animal models, we hypothesized that the same events could be possible in insects and that SNPs could occur in acceptor and/or donor splice sites in mosquitoes that might result in the splice variation. The current report seeks then to screen *Anopheles gambiae doublesex* gene (*Agdsx*) splice site sequences for single-nucleotide polymorphisms (SNPs) that could be associated with alternative splicing.

Methods

Sequence data and mosquito samples

Genomic sequences used in this study came from the Anopheles 1000 genomes (Ag1000G) project phase 2 released in 2017[24]. 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 mosquito whole genomes and that pass the variant filtering process described by 24. Only *Anopheles gambiae s.s.* 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 the Ag1000G phase 2 project.

Country	Site	Year	Geographic coordinate		Number of species	Number of female	Number of male
			Latitude	Longitude			
Burkina Faso	Bana	2012	11.2330	-4.4720	20	3	17
	Pala	2012	11.1500	-4.2350	46	38	8
	Souroukoudinga	2012	11.2350	-4.5350	26	26	0
Cameroon	Daiguene	2009	4.7770	13.8440	96	81	15
	Gado Badzere	2009	5.7470	14.4420	73	58	15
	Mayos	2009	4.3410	13.5580	105	91	14
	Zembe Borongo	2009	5.7470	14.4420	23	23	0
Equatorial Guinea	Bioko	2002	3.7000	8.7000	9	9	0
France (Mayotte)	Bouyouni	2011	-12.7378	45.1417	1	1	0
	Combani	2011	-12.7787	45.1429	5	2	3
	Karihani Lake	2011	-12.7965	45.1217	3	3	0
	Mont Benara	2011	-12.8570	45.1552	2	1	1
	Mtsamboro Forest Reserve	2011	-12.7027	45.0811	1	1	0
	Mtsanga Charifou	2011	-12.9907	45.1557	8	3	5
	Sada	2011	-12.8521	45.1039	4	1	3
Gabon	Libreville	2000	0.3840	9.4550	69	69	0
Ghana	Madina	2012	5.6685	-0.2193	12	12	0
Guinea	Koraboh	2012	9.2500	-9.9170	22	22	0
	Koundara	2012	8.5000	-9.4170	18	18	0
Uganda	Tororo (Nagongera)	2012	0.7700	34.0260	112	112	0

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 (https://vcftools.github.io/index.html)[35] was used to extract the SNPs within the genomic region corresponding to the *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 (https://tassel.bitbucket.io/)[36]. 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 six nucleotides of an exon. In donor splice sites, SNPs were identified within in the last six nucleotides of an exon and the first 16 nucleotides in an intron. The average nucleotide diversity at the *dsx* locus between male and female was calculated using scikit-allel version 1.2.1 (https://scikit-allel.readthedocs.io/en/stable/)[37] in order to determine whether SNPs density at the *dsx* locus differed between the two genders.

Results

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

Male and female mosquitoes share exon 1, 2, 3, 4 and 6 donor splice sites while exon 5 donor site is specific to female as it is only recognized by the spliceosome in females (Table 2). Similarly, both male and female share intron 1, 2, 3, and 6 acceptor sites. Male intron 4 and female intron 5 share the same 3' end as the female, and exon 5 is included in the male intron 4 sequence. However, females have the intron 4 specific acceptor site, as the cassette exon 5 definition is not established in males (Table 2).

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

Along the *Agdsx* gene, 17,196 polymorphic sites were identified. Wherever both male and female mosquitoes are present (in Burkina Faso, Cameroon and Mayotte), the nucleotide diversity is similar between both genders (Figure 2). This was expected as male and female in each country make up a single population. In addition, no difference in the nucleotide diversity was observed between male populations from the three countries (Burkina Faso, Cameroon and Mayotte) (Figure 2, top panel). The same trend was observed between female populations as well (Burkina Faso, Cameroon, Mayotte, Gabon, Ghana, Guinea, Equatorial Guinea and Uganda).

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 the

Table 2. Splice donor and acceptor sites within the double sex (*dsx*) gene *of Anopheles gambiae*.

Splice donor sites							
Gender	**Exon**	**Size**	**Exon position**		**Splice site sequence**	**Site position**	
			Start	**End**		**Start**	**End**
Male/Female	1	1415	48788460	48787046	tatttg/gtaagtaaatatgcaa	48787051	48787030
Male/Female	2	1445	48785629	48784185	TGGGAG/gtaagtacgatcatgc	48784190	48784169
Male/Female	3	45	48747737	48747693	TACCTG/gtaagtaaatataatt	48747698	48747677
Male/Female	4	135	48715295	48715161	ACGAAG/gtaagctggcgatgat	48715166	48715145
Female	5	1692	48714648	48712957	cagaag/gtatggtaagacggcc	48712962	48712941
Male/Female	6	1267	48712794	48711528	aaaaag/gtaagtgtgggtagta	48711533	48711512
Male/Female	7	2668	48706331	48703664	None		
Splice acceptor sites							
Gender	**Intron**	**Size**	**Intron position**		**Splice site sequence**	**Site position**	
			Start	**End**		**Start**	**End**
Male/Female	1	1416	48787045	48785630	gtacgtttgattgcag/atctcc	48785645	48785624
Male/Female	2	36447	48784184	48747738	ttgctctccttttcag/CTACTC	48747753	48747732
Male/Female	3	32397	48747692	48715296	ttccgccccgtttcag/ACGACG	48715311	48715290
Female	4	512	48715160	48714649	tttatgtttaacacag/GTCAAG	48714664	48714643
Male	4	2366	48715160	48712795	tgtaacccccaaaaag/gtaaac	48712810	48712789
Female	5	162	48712956	48712795			
Male/Female	6	5196	48711527	48706332	cgcttcctcaaaatag/atcgat	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.

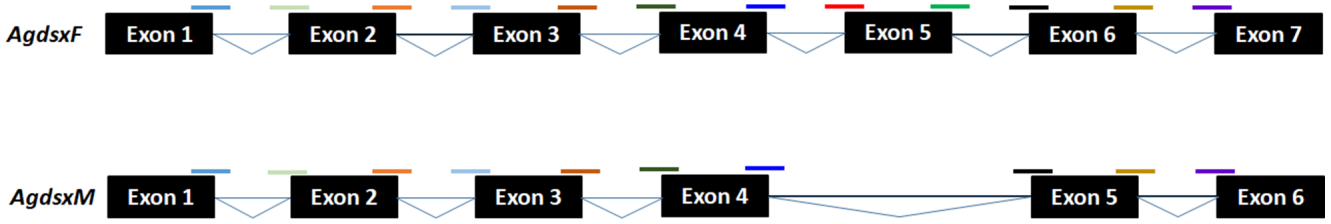

Figure 2. Nucleotide diversity (π) at *Agdsx* locus within mosquito populations. BF: Burkina Faso; CM: Cameroon; FR: Mayotte; GA: Gabon; CH: Ghana; GN: Guinea; GQ: Equatorial Guinea; UG: Uganda. No differences in nucleotide diversity were observed within male or female mosquito populations.

acceptor sequence of female-specific intron 4 nor in the corresponding male region (Figure 3). However in the female-specific exon 5 donor site, two SNPs (rs48712947, rs48712962) were found. Nevertheless, they were not specific to females as the rs48712947 was found in Cameroon female mosquitoes and in both males and females from Burkina Faso (Figure 4). The rs48712962 was absent in the male mosquito population, while it was found only in females in Cameroon. The minor allele frequencies (MAF) of both SNPs identified were very low in each population. The MAF of rs48712947 and rs48712962 amounted to less than 1% in each female population, and only 2% of Burkina Faso male carried the rs48712947.

SNPs in other splice sites of Agdsx

The other splice sites were also examined for identification of gender-specific SNPs. No SNP was found in the shared exon 1 donor, introns 1. No splice-relevant SNP was found in the other donor (Figure 5A, Figure 6, and Figure 7B) and acceptor (Figure 5B, and Figure 7A) splice sites. The highest number of SNPs (7) was found in the common intron 3 acceptor site sequence (rs48715291, rs48715294, rs48715302, rs48715306, rs48715307, rs48715308, rs48715309) (Figure 8). However, each of these SNPs occurred in a 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 genetic SIT strategy[18,19,25]. The translation and the success of using *dsx* in SIT methodology require a clearer understanding of the genetic bases of the gender determination pathway. This study screened the *Agdsx* donor and acceptor splice sites for identification of splice-relevant SNPs.

The alternative splicing of *Agdsx* gene is governed by exon 5 skipping in male mosquitoes[20] suggesting a silencing mechanism of the female-specific splice sites recognition (intron 4 acceptor and exon 5 donor sites) by the splicing machinery in males. 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 males and females. The SNPs rs48712947 and rs48712962 identified in female-specific exon 5 donor site were neither splice-relevant nor gender-specific. They appeared only is 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 causes 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*[21]. In *Anopheles gambiae* the number of pyrimidine (8) in the 12bp 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.*[20]. 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[20]. Overall, these findings add further 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 tans-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 enhancers; ESI: exonic splicing silencers; ISE: intronic splicing enhancers; ISI: intronic splicing silencers)[38–40]. Similar regulatory *cis*-elements were found in *Drosophila melanogaster* female-specific exon and putative homologs were identified in *An. gambiae* female-specific exon 5[20]. Therefore, further molecular analyses are needed toward characterizing these regulatory sequence and their binding trans-factors in order to underpin the somatic sex determination in *An. gambiae*.

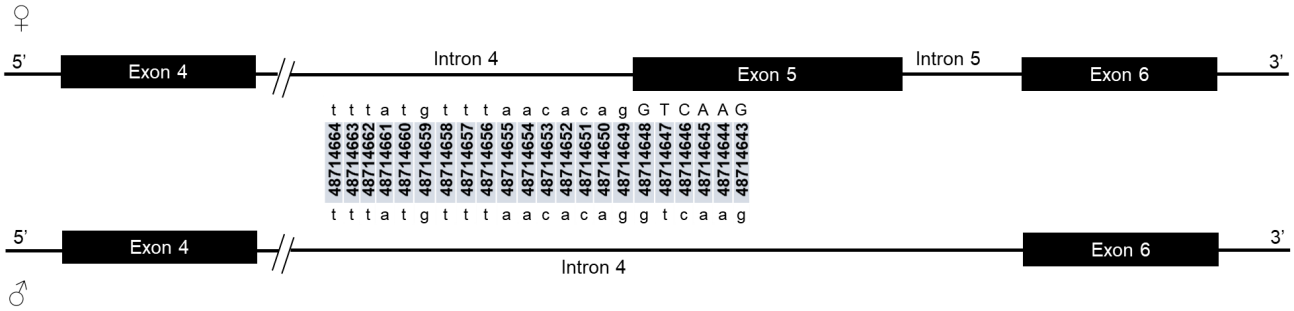

Figure 3. Female-specific intron 4/exon 5 junction and the corresponding region in male. Intron 4 splice acceptor sequence is indicated in female with the genomic positions of each nucleotide. Female-specific exon 5 is included in male intron 4 sequence. The corresponding region of the female intron 4 acceptor site is indicated within male intron 4. Exonic coding sequences are shown in uppercase letters, and non-coding regions are in lowercase letters. No SNP was neither found in the female intron 4 splice acceptor site or in the corresponding region in male.

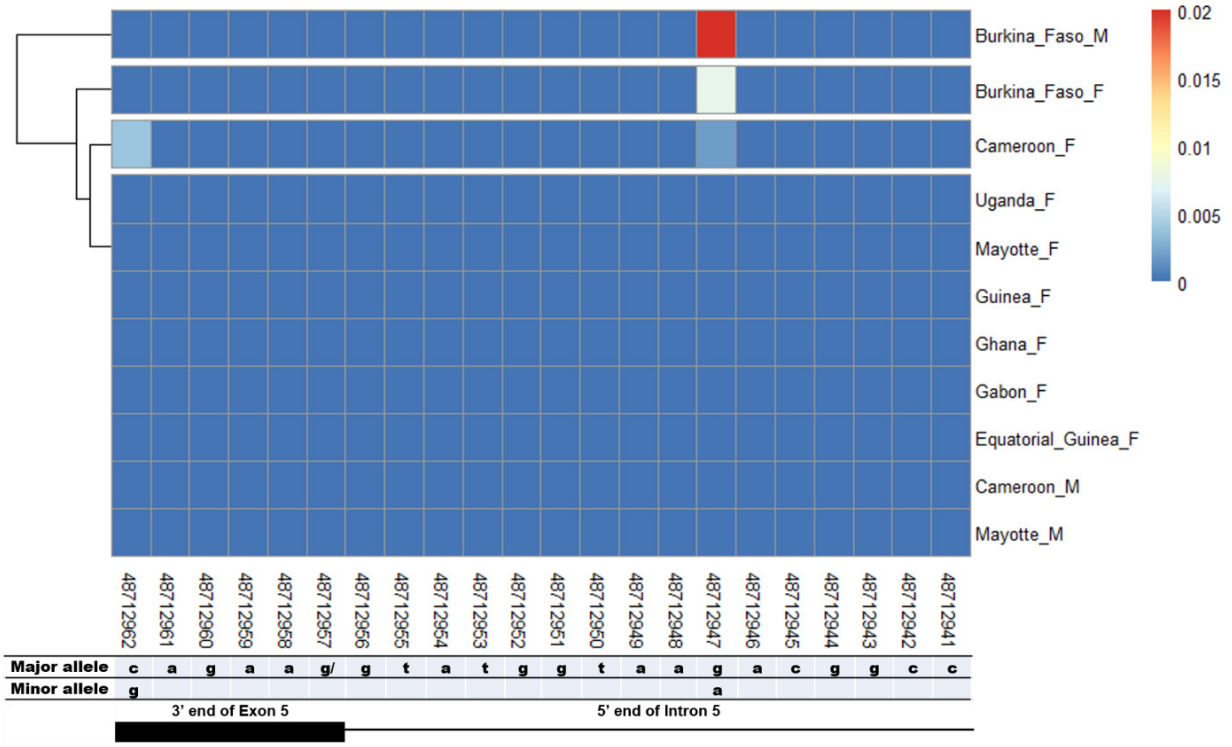

Figure 4. SNP within female specific exon 5 donor splice site and the corresponding region in male. The coloured and dark blue squares denote respectively the presence or absence of SNP. Each row represents male (**M**) or female (**F**) mosquito population. At the bottom, the numbers are the genomic positions of each nucleotide. The major and minor allele (where applicable) are indicated for each nucleotide position. The key colour is scaled to the minor allele frequency. The black box and the black line respectively depict an exonic and intronic regions covered by the splice site in females. The corresponding region of this female donor splice site within male intron 4 was analysed. SNPs were found at very low frequencies at position 48,712,947 in Burkina Faso and Cameroon female and Burkina Faso male populations. SNP at 48,712,962 was found only in Cameroon female population. No gender-specific SNP was identified.

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[41–43]. Also histone post-translational modifications (PTMs) such as lysine acetylation and methylation were associated to the alternative splicing event[44–46]. 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*[47,48]. 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

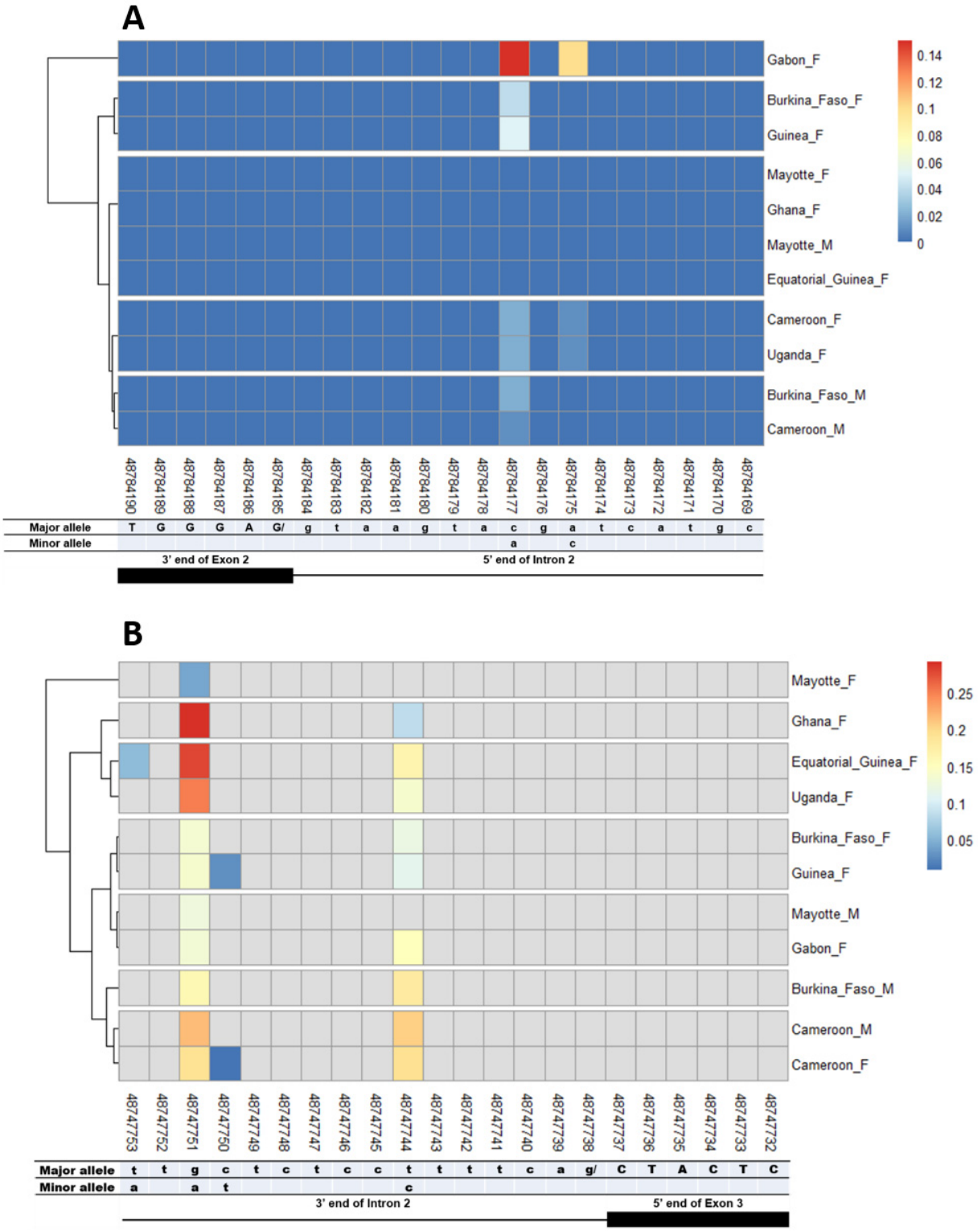

Figure 5. Single nucleotide polymorphism within exon 2 and intron 2 splice sites. A) SNPs within exon 2 donor splice site between male and female *An. gambiae* mosquitoes. Each row represents male (**M**) or female (**F**) mosquito population. At the bottom, the numbers are the genomic positions of each nucleotide. The major and minor allele (where applicable) are indicated for each nucleotide position. The key colour is scaled to the minor allele frequency. The black box and the black line respectively depict an exonic and intronic regions covered by the splice site in females. The uppercase and lowercase letters denote respectively coding and non-coding region. **B**) SNPs within intron 2 acceptor splice site between male and female *An. gambiae* mosquitoes. The uppercase and lowercase letters denote respectively coding and non-coding region.

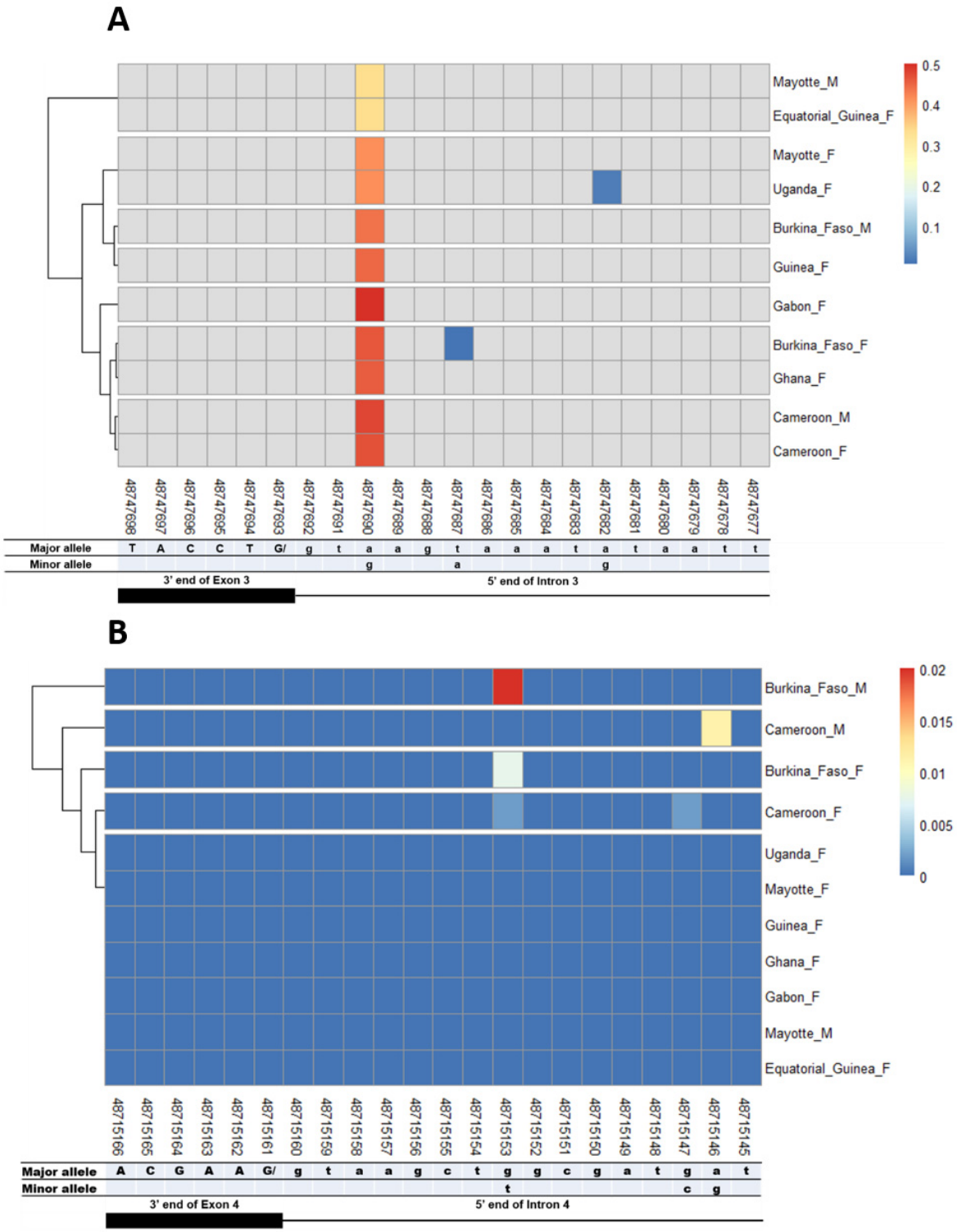

Figure 6. Single nucleotide polymorphism within exon 3 and exon 4 splice sites. A) SNPs within exon 3 donor splice site between male and female *An. gambiae* mosquitoes. Each row represents male (**M**) or female (**F**) mosquito population. At the bottom, the numbers are the genomic positions of each nucleotide. The major and minor allele (where applicable) are indicated for each nucleotide position. The key colour is scaled to the minor allele frequency. The black box and the black line respectively depict an exonic and intronic regions covered by the splice site in females. The uppercase and lowercase letters denote respectively coding and non-coding region. **B**) SNPs within exon 4 donor splice site between male and female *An. gambiae* mosquitoes. The uppercase and lowercase letters denote respectively coding and non-coding region.

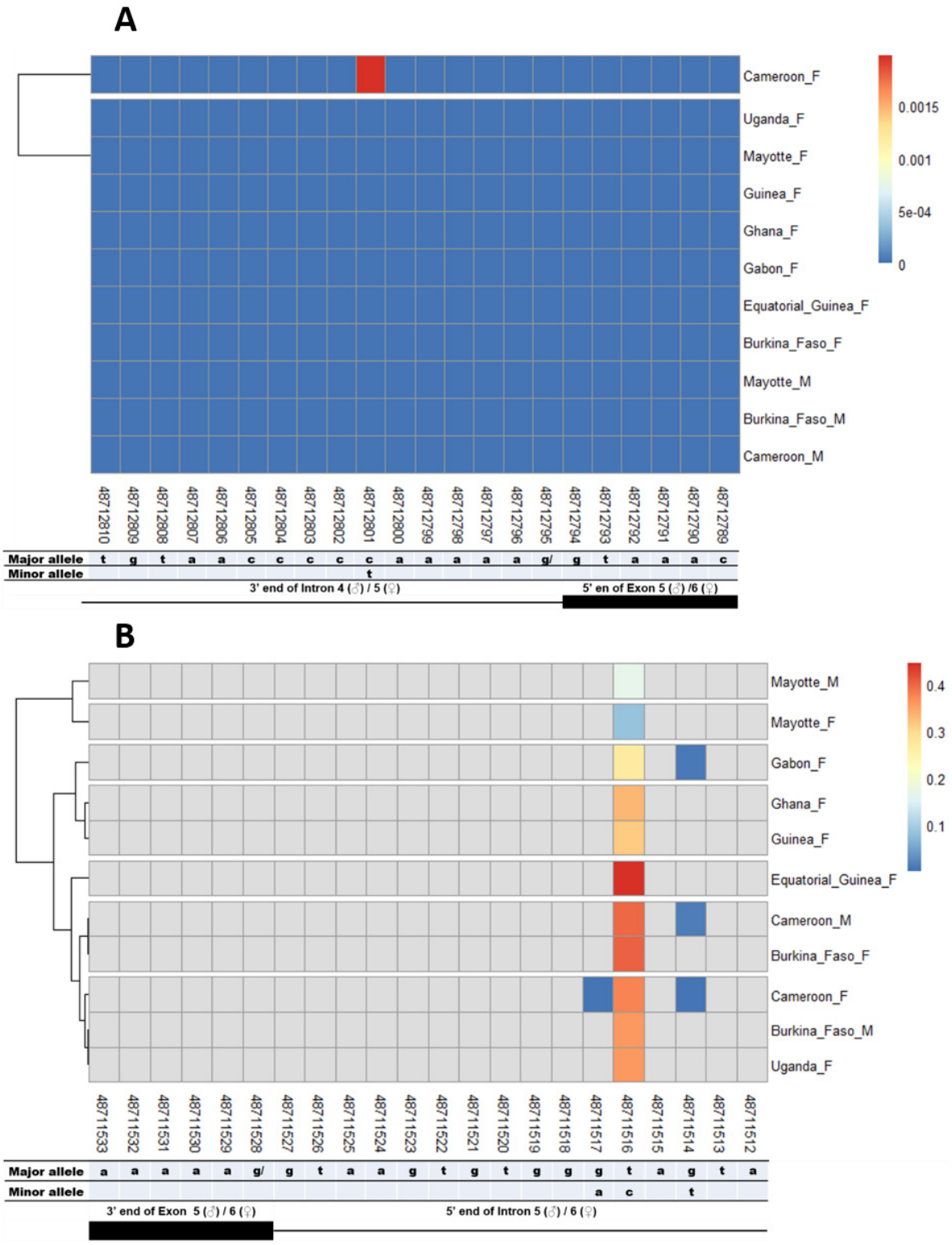

Figure 7. Single nucleotide polymorphism in the last *Agdsx* acceptor and donor splice sites. A) SNPs within the common acceptor site (Intron 4/5) between male and female *An. gambiae* mosquitoes. The female (♀) specific exon 5 is included in the male (♂) intron 4 sequence making a shift in exon number is male. Thus, the male intron 4 and female intron 5 share the same 3′ end. Similarly, the male exon 5 and female exon 6 share the same 5′ end region. **B**) SNPs within the common donor site (Exon 5/6) between male and female *An. gambiae* mosquitoes. The female (♀) specific exon 5 is included in the male (♂) intron 4 sequence making a shift in exon number is male. Thus, the male exon 5 and female exon 6 share the same 3′ end. Similarly, the male intron 5 and female intron 6 share the same 5′ end region.

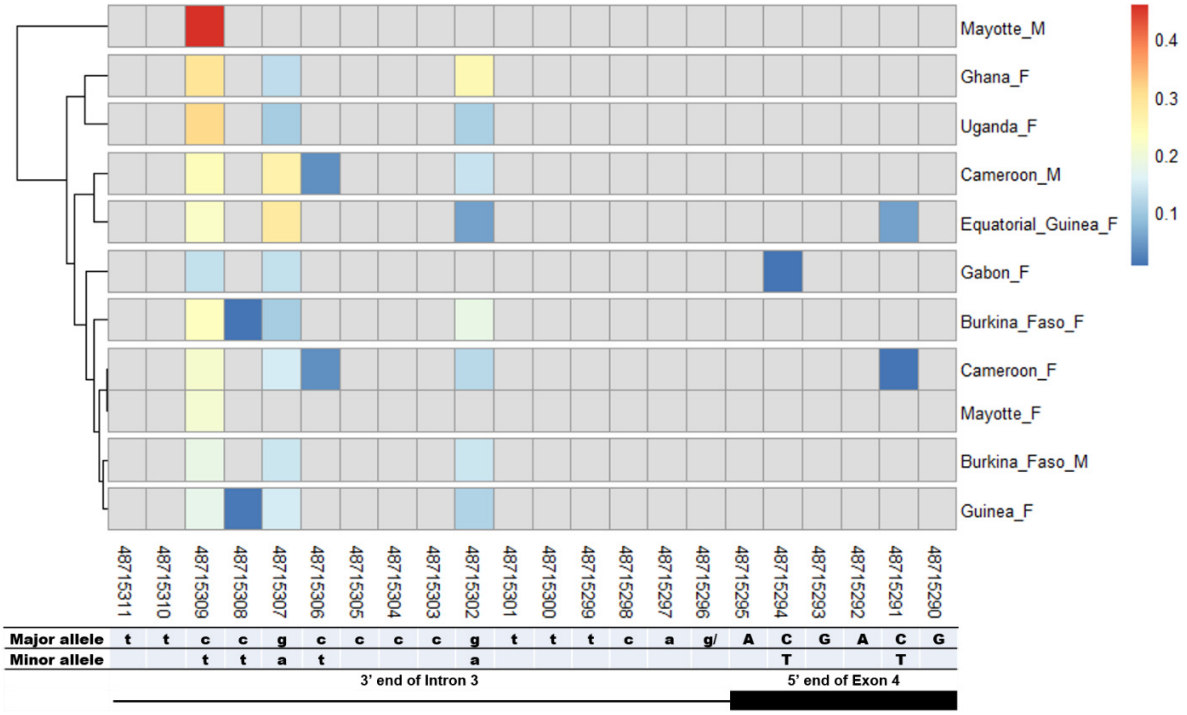

Figure 8. SNP within intron 3 acceptor splice site between *An. gambiae* male and female mosquitoes. The uppercase and lowercase letters denote respectively coding and non-coding region. The coloured and grey squares denote respectively the presence or absence of SNP. Each row represents male (**M**) or female (**F**) mosquito population. At the bottom, the numbers are the genomic positions of each nucleotide. The major and minor allele (where applicable) are indicated for each nucleotide position. The key colour is scaled to the minor allele frequency. The black box and the black line depict respectively an exonic and intronic regions covered by the splice site in each mosquito gender. No gender-specific SNP was identified.

An. gambiae[49,50]. 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 chemical and biological vector control. Given the potential of the *Agdsx* gene for 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.

Data availability

Figshare: Data underlying Single nucleotide polymorphism (SNP) in the doublesex (dsx) gene splice sites and relevance for its alternative splicing in the malaria vector *Anopheles gambiae*. https://doi.org/10.6084/m9.figshare.18589781.v1

This project contains the following underlying data:

Dsx_f.h5. (Data used to plot nucleotide diversity in female populations)

Dsx_f.h5. (Data used to plot nucleotide diversity in male populations)

dsx_variant_seq_norm_F. (VCF format dataset containing SNPs in *doubesex* gene sequence of female mosquitoes)

dsx_variant_seq_norm_M. (VCF format dataset containing SNPs in *doubesex* gene sequence of male mosquitoes)

Female_sample_ID. (Dataset of the accession numbers of females sequences in the dsx_variant_seq_norm_F file)

Female_sample_ID. (Dataset of the accession numbers of males sequences in the dsx_variant_seq_norm_M file).

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.

References

1. Greenwood BM, Bojang K, Whitty CJ, *et al.*: **Malaria.** *Lancet.* 2005; **365**(9469): 1487–1498.
PubMed Abstract | Publisher Full Text

2. World Health Organization: **The top 10 causes of death.** 2018; [cited 6 Jul 2020].
Reference Source

3. Castro MC, Singer BH: **Malaria in the Brazilian Amazon.** *Water and Sanitation-Related Diseases and the Environment: Challenges, Interventions, and Preventive Measures.* John Wiley & Sons; 2011; 401–419.
Publisher Full Text

4. Abdul-Ghani R, Al-Mekhlafi AM, Alabsi MS: **Microbial control of malaria: Biological warfare against the parasite and its vector.** *Acta Trop.* 2012; **121**(2): 71–84.
PubMed Abstract | Publisher Full Text

5. Bhatt S, Weiss D, Cameron E, *et al.*: **The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015.** *Nature.* 2015; **526**(7572): 207–211.
PubMed Abstract | Publisher Full Text | Free Full Text

6. World Health Organization: **World malaria report 2019.** 2019.
Reference Source

7. Kamgang B, Tchapga W, Ngoagouni C, *et al.*: **Exploring insecticide resistance mechanisms in three major malaria vectors from Bangui in Central African Republic.** *Pathog Glob Health.* 2018; **112**(7): 349–359.
PubMed Abstract | Publisher Full Text | Free Full Text

8. Tchouakui M, Fossog BT, Ngannang BV, *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* [version 2; peer review: 2 approved, 1 approved with reservations].** *Wellcome Open Res.* 2019; **4**: 13.
PubMed Abstract | Publisher Full Text | Free Full Text

9. Collins E, Vaselli NM, Sylla M, *et al.*: **The relationship between insecticide resistance, mosquito age and malaria prevalence in *Anopheles gambiae* s.l. from Guinea.** *Sci Rep.* 2019; **9**(1): 8846.
PubMed Abstract | Publisher Full Text | Free Full Text

10. LePage D, Bordenstein SR: ***Wolbachia*: Can we save lives with a great pandemic?** *Trends Parasitol.* 2013; **29**(8): 385–393.
PubMed Abstract | Publisher Full Text | Free Full Text

11. Hughes GL, Koga R, Xue P, *et al.*: ***Wolbachia* Infections Are Virulent and Inhibit the Human Malaria Parasite *Plasmodium Falciparum* in *Anopheles Gambiae*.** *PLoS Pathog.* 2011; **7**(5): e1002043.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Phuc HK, Andreasen MH, Burton RS, *et al.*: **Late-acting dominant lethal genetic systems and mosquito control.** *BMC Biol.* 2007; **5**: 11.
PubMed Abstract | Publisher Full Text | Free Full Text

13. Fu G, Lees RS, Nimmo D, *et al.*: **Female-specific flightless phenotype for mosquito control.** *Proc Natl Acad Sci U S A.* 2010; **107**(10): 4550–4554.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Prowse TA, Adikusuma F, Cassey P, *et al.*: **A Y-chromosome shredding gene drive for controlling pest vertebrate populations.** Messer PW, Tautz D, editors. *eLife.* 2019; **8**: e41873.
PubMed Abstract | Publisher Full Text | Free Full Text

15. Catteruccia F, Crisanti A, Wimmer EA: **Transgenic technologies to induce sterility.** *Malar J.* 2009; **8 Suppl 2**(Suppl 2): S7.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Ricardo Machi A, Rodrigues Mayne R, Adriani Gava M, *et al.*: **Gamma Radiation Sterilization Dose of Adult Males in Asian Tiger Mosquito Pupae.** *Insects.* 2019; **10**(4): 101.
PubMed Abstract | Publisher Full Text | Free Full Text

17. Dyck VA, Hendrichs J, Robinson AS: **Sterile insect technique: principles and practice in area-wide integrated pest management**. Dordrecht, Netherlands: Springer; 2005.
Reference Source

18. Kyrou K, Hammond AM, Galizi R, *et al.*: **A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes.** *Nat Biotechnol.* 2018; **36**(11): 1062–1066.
PubMed Abstract | Publisher Full Text | Free Full Text

19. Taracena ML, Hunt CM, Benedict MQ, *et al.*: **Downregulation of female *doublesex* expression by oral-mediated RNA interference reduces number and fitness of *Anopheles gambiae* adult females.** *Parasit Vectors.* 2019; **12**(1): 170.
PubMed Abstract | Publisher Full Text | Free Full Text

20. Scali C, Catteruccia F, Li Q, *et al.*: **Identification of sex-specific transcripts of the *Anopheles gambiae doublesex* gene.** *J Exp Biol.* 2005; **208**(Pt 19): 3701–3709.
PubMed Abstract | Publisher Full Text | Free Full Text

21. Black DL: **Mechanisms of Alternative Pre-Messenger RNA Splicing.** *Annu Rev Biochem.* 2003; **72**: 291–336.
PubMed Abstract | Publisher Full Text

22. Clough E, Jimenez E, Kim YA, *et al.*: **Sex- and Tissue-Specific Functions of *Drosophila* Doublesex Transcription Factor Target Genes.** *Dev Cell.* 2014; **31**(6): 761–773.
PubMed Abstract | Publisher Full Text | Free Full Text

23. Chatterjee SS, Uppendahl LD, Chowdhury MA, *et al.*: **The female-specific Doublesex isoform regulates pleiotropic transcription factors to pattern genital development in *Drosophila*.** *Development.* 2011; **138**(6): 1099–1109.
PubMed Abstract | Publisher Full Text

24. The Anopheles gambiae 1000 Genomes Consortium, Clarkson CS, Miles A, *et al.*: **Genome variation and population structure among 1,142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*.** *bioRxiv.* 2019; 864314.
Publisher Full Text

25. Krzywinska E, Ferretti L, Li J, *et al.*: ***femaleless* Controls Sex Determination and Dosage Compensation Pathways in Females of *Anopheles* Mosquitoes.** *Curr Biol.* 2021; **31**(5): 1084–1091.e4.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Krzywinska E, Dennison NJ, Lycett GJ, *et al.*: **A maleness gene in the malaria mosquito *Anopheles gambiae*.** *Science.* 2016; **353**(6294): 67–69.
PubMed Abstract | Publisher Full Text

27. Wachutka L, Caizzi L, Gagneur J, *et al.*: **Global donor and acceptor splicing site kinetics in human cells.** Black DL, Manley JL, editors. *eLife.* 2019; **8**: e45056.
PubMed Abstract | Publisher Full Text | Free Full Text

28. Hiller M, Huse K, Szafranski K, *et al.*: **Single-Nucleotide Polymorphisms in NAGNAG Acceptors Are Highly Predictive for Variations of Alternative Splicing.** *Am J Hum Genet.* 2006; **78**(2): 291–302.
PubMed Abstract | Publisher Full Text | Free Full Text

29. Cartegni L, Chew SL, Krainer AR: **Listening to silence and understanding nonsense: exonic mutations that affect splicing.** *Nat Rev Genet.* 2002; **3**(4): 285–298.
PubMed Abstract | Publisher Full Text

30. Garcia-Blanco MA, Baraniak AP, Lasda EL: **Alternative splicing in disease and therapy.** *Nat Biotechnol.* 2004; **22**(5): 535–546.
PubMed Abstract | Publisher Full Text

31. Lamba V, Lamba J, Yasuda K, *et al.*: **Hepatic CYP2B6 Expression: Gender and Ethnic Differences and Relationship to *CYP2B6* Genotype and CAR (Constitutive Androstane Receptor) Expression.** *J Pharmacol Exp Ther.* 2003; **307**(3): 906–922.
PubMed Abstract | Publisher Full Text

32. Drögemüller C, Peters M, Pohlenz J, *et al.*: **A single point mutation within the *ED1* gene disrupts correct splicing at two different splice sites and leads to anhidrotic ectodermal dysplasia in cattle.** *J Mol Med (Berl).* 2002; **80**(5): 319–323.
PubMed Abstract | Publisher Full Text

33. Gargani M, Valentini A, Pariset L: **A novel point mutation within the EDA gene causes an exon dropping in mature RNA in Holstein Friesian cattle breed affected by X-linked anhidrotic ectodermal dysplasia.** *BMC Vet Res.* 2011; **7**: 35.
PubMed Abstract | Publisher Full Text | Free Full Text

34. Fan B, Xie D, Li Y, *et al.*: **A single intronic single nucleotide polymorphism in splicing site of steroidogenic enzyme *hsd17b1* is associated with phenotypic sex in oyster pompano, *Trachinotus anak*.** *Proc Biol Sci.* 2021; **288**(1963): 20212245.
PubMed Abstract | Publisher Full Text | Free Full Text

35. Danecek P, Auton A, Abecasis G, *et al.*: **The variant call format and VCFtools.** *Bioinformatics.* 2011; **27**(15): 2156–2158.
PubMed Abstract | Publisher Full Text | Free Full Text

36. Bradbury P, Zhang Z, Kroon DE, *et al.*: **TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples.** *Bioinformatics.* 2007; **23**(19): 2633–5.
PubMed Abstract | Publisher Full Text

37. Miles A, Harding N: **Cggh/Scikit-Allel: V1.1.0.** *Zenodo.* 2017.
Publisher Full Text

38. Smith CWJ, Valcárcel J: **Alternative pre-mRNA splicing: the logic of combinatorial control.** *Trends Biochem Sci.* 2000; **25**(8): 381–388.
PubMed Abstract | Publisher Full Text

39. Singh R, Valcárcel J: **Building specificity with nonspecific RNA-binding proteins.** *Nat Struct Mol Biol.* 2005; **12**(8): 645–653.
PubMed Abstract | Publisher Full Text

40. Izquierdo JM, Valcárcel J: **A simple principle to explain the evolution of pre-mRNA splicing.** *Genes Dev.* 2006; **20**(13): 1679–1684.
PubMed Abstract | Publisher Full Text

41. Li-Byarlay H: **The Function of DNA Methylation Marks in Social Insects.** *Front Ecol Evol.* 2016; **4**: 57.
Publisher Full Text

42. Lyko F, Foret S, Kucharski R, *et al.*: **The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers.** *PLoS Biol.* 2010; **8**(11): e1000506.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Cingolani P, Cao X, Khetani RS, *et al.*: **Intronic Non-CG DNA hydroxymethylation and alternative mRNA splicing in honey bees.** *BMC Genomics.* 2013; **14**: 666.
PubMed Abstract | Publisher Full Text | Free Full Text

44. Rahhal R, Seto E: **Emerging roles of histone modifications and HDACs in RNA splicing.** *Nucleic Acids Res.* 2019; **47**(10): 4911–4926.
PubMed Abstract | Publisher Full Text | Free Full Text

45. Gunderson FQ, Merkhofer EC, Johnson TL: **Dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements.** *Proc Natl Acad Sci U S A.* 2011; **108**(5): 2004–2009.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Luco RF, Pan Q, Tominaga K, *et al.*: **Regulation of Alternative Splicing by Histone Modifications.** *Science.* 2010; **327**(5968): 996–1000.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Bewick AJ, Vogel KJ, Moore AJ, *et al.*: **Evolution of DNA Methylation across Insects.** *Mol Biol Evol.* 2017; **34**(3): 654–665.
PubMed Abstract | Publisher Full Text | Free Full Text

48. Falckenhayn C, Carneiro VC, de Mendonça Amarante A, *et al.*: **Comprehensive DNA methylation analysis of the *Aedes aegypti* genome.** *Sci Rep.* 2016; **6**: 36444.
PubMed Abstract | Publisher Full Text | Free Full Text

49. Gómez-Díaz E, Rivero A, Chandre F, *et al.*: **Insights into the epigenomic landscape of the human malaria vector *Anopheles gambiae*.** *Front Genet.* 2014; **5**: 277.
PubMed Abstract | Publisher Full Text | Free Full Text

50. Ruiz JL, Yerbanga RS, Lefèvre T, *et al.*: **Chromatin changes in *Anopheles gambiae* induced by *Plasmodium falciparum* infection.** *Epigenetics Chromatin.* 2019; **12**(1): 5.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ? ✓ ✓

Version 3

Reviewer Report 02 August 2023

https://doi.org/10.21956/wellcomeopenres.21053.r58870

✓ **Moussa W Guelbeogo**
Centre National de Recherche et Formation sur le Paludisme (CNRFP), Ouagadougou, Burkina Faso

Review comments

General comments

The authors of this manuscript have addressed a topic issue relating to the need for new vector control tools in regard of the declining of effectiveness of current insecticide-based tools. They are looking how they can increase scientific finding to improve SIT tools to tackle the major vector An gambiae. They aimed to explore the pathways that trigger the signal of dsx gene exon skipping alternative splicing mechanism in anopheline mosquitoes. They screen the An. gambiae dsx gene splice site sequences for single nucleotide polymorphisms (SNPs) that could be critical to its alternative splicing.

The present version of the manuscript is well improved, and the results are well presented. The content fits very well to the title. However, it is remaining minor corrections to fix.

Introduction

1. “This disease is among the top ten causes of death in low-income countries (World Health Organization)2 and continues to take a heavy toll on communities, especially in African regions”. Remove the “World Health Organization” keep only the same type of citation

Methodology

Very suitable to address the research question.

Line” This file contain all SNPs identified in mosquito whole genomes and that pass the variant filtering process described by 24” I suggest to add Clarkson CS and colleague and then put the citation number 24

Table1: Head of column 6 : the change the sentence " Number of species" by 'Number of specimens" the species is An gambiae

Results

Figure 1 : Please, can you explain or give the meaning of BFgam, CMgam, FRgam, GAgam, GHgam, GNgam, GQgam and UGgam? it can help to do the connection with the study countries.

Figure2 : Check the sentence in the title: "BF: Burkina Faso; CM: Cameroon; FR: Mayotte; GA: Gabon; CH: Ghana; GN: Guinea; GQ: Equatorial Guinea; UG: Uganda". I do not understand why this description here, because I don't see any connection between this description and the figure. The description must be deleted.

Discussion

It looks fine.

Conclusion

Why the *An coluzzi* was not included in the study. The result of your study has opened some interesting questions to address these question, but you must include the *An coluzzi* for the next studies because it is one the major vector and widely distributed in western Africa

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Vector bilogist, genetic population of vector

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 06 February 2023

https://doi.org/10.21956/wellcomeopenres.21053.r54729

Yoosook Lee
Florida Medical Entomology Laboratory, Department of Entomology and Nematology, Institute of Food and Agricultural Sciences, University of Florida, Vero Beach, FL, USA

No further edits needed

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 27 January 2023

https://doi.org/10.21956/wellcomeopenres.20704.r53754

Yoosook Lee
Florida Medical Entomology Laboratory, Department of Entomology and Nematology, Institute of Food and Agricultural Sciences, University of Florida, Vero Beach, FL, USA

The paper could be set to 'approved' status for me.

There are some minor edits needed though:

- "*Moreover, the epigenetic system was also reported to regulate the alternative splicing in mammalian and other insects cells.*" -> change to 'other insect cells'.
- Some citations (e.g. World Health Organization fact sheet) seems like they should be in a format consistent with other references (hyper link text to numbers).

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 08 July 2022

https://doi.org/10.21956/wellcomeopenres.19431.r51402

Yoosook Lee
Florida Medical Entomology Laboratory, Department of Entomology and Nematology, Institute of Food and Agricultural Sciences, University of Florida, Vero Beach, FL, USA

Overall comment:

The authors should update the Introduction section to reflect the latest development of sex determination research on *Anopheles gambiae* and related species/genes and set reasonable expectations and hypotheses.

The detailed suggested edits are noted below:

- "*The malaria transmission cycle involves, four major elements: the host (human), the parasite, the vector and the environment*."

  Use oxford style commas. "... the vector, and the environment."

- "*In the last decade, scientific advances in additional tools for vector control have included technologies such as* ..." -> the papers cited for novel vector control tools are over 10 years old. There are newer publications in this realm and should be changed to newer citations or change wording to "In the last two decades, ..."
- "The latter technique, SIT, is based on ..." -> construct of listing more than two things and referencing 'the latter' for the last element is a bit awkward.

- The 2nd paragraph of the introduction needs to be changed. As far as I am concerned there has not been radio-sterilized male SIT for Anopheles species although some people have attempted to test it in lab conditions and prepare for field trials. The paper 13 cited for SIT is introducing a genetic tool for inducing SIT not radiation. So explaining general radiation-based SIT method and citing paper 13 is not appropriate. It is especially concerning as this paper pertains to the *Anopheles gambiae* in Africa. Combined with more-than-a-decade-old citations for the alternative vector control strategies, it is somewhat questionable if the authors have due diligence in background research.

- "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." -> add 'genetic' in front of 'SIT technologies' as radiation-based SIT does not require genetic sexing.

- The manuscript has mixed use of gender and sex. I suggest using the same and consistent terminology throughout the manuscript.

- "*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*" -> word 'structure' in this case is not clear to readers. Structure is a very generic term and

can mean very different things (e.g. RNA structure? Protein structure?). Clarification would be helpful for readers.

- *"Indeed, in humans, splicing signals are a common point of mutations…. Taking together these observations in humans and animal models, we hypothesized that the same events could be possible in insects and that SNPs could occur in acceptor and/or donor splice sites in mosquitoes that might result in the splice variation."*-> There is some logical gap about expecting mutations in splicing regions determining sex locus. It seems like this is such a fundamental biological function that has huge implications in the downstream process so this region would be likely conserved. The authors should find examples of not just having mutations in splicing regions but splicing mutations in sex-determining genes to be relevant and set reasonable hypotheses/expectations.

- Table 1 – Since the study examined the allele frequency between males and females, Table 1 should provide the number of male and female sequences used for the analysis. Some sites of the Ag1000 data only contain females so generalizing the sex-dependent patterns with skewed data can be misleading.

- "*Only Anopheles gambiae samples were considered in our study*." -> Anopheles gambiae s.s. (S form) only? Since the species *gambiae* and *coluzzii* were separated only in 2014, I think it would be best to clarify what you included for the readers.

- The first paragraph of the results section should be in the introduction. Also, readers who are not familiar with the sex determination process would wonder why sex determining gene is located on an autosome, not a sex chromosome. To educate readers, the authors should briefly describe how doublesex on autosome is involved in sex determination of *Anopheles gambiae* based on currently published literature in the introduction.

- Figure 1. It would be best if female and male from the same location is side by side for a direct comparison. Males and females separate with different y-axis scales makes it difficult to compare. FR and GA colors are too similar.

- "*The An. gambiae doublesex (Agdsx) gene is a candidate gene of interest for SIT, as a candidate for genetic modifications*"-> candidate is repeated twice. Perhaps change it to "… is a candidate gene for genetic SIT strategy" or something to that effect?

- Important paper on this topic like https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7955153/ [1] is not cited in this paper.

- "*According to the D. melanogaster model18, the alternative splicing of Agdsx gene is governed by exon 5 skipping in male mosquitoes16 suggesting a silencing mechanism of the female-specific splice sites recognition (intron 4 acceptor and exon 5 donor sites) by the splicing machinery in males*." -> I think there is a new paper on this topic from *Anopheles gambiae* which sequenced female and male transcriptome so the author doesn't have to deduce from the Drosophila sex termination model. Please check the latest papers by Krzywinski for updated literature on *Anopheles gambiae* sex determination.

- "*They appeared only is two mosquito populations (Burkina Faso and Cameroon) over the eight*

populations considered." -> Burkina Faso and Cameroon are far apart so I wouldn't minimize how widespread this mutation is based on the number of countries examined (2 out of 8 is still not a low number).

References

1. Krzywinska E, Ferretti L, Li J, Li J, et al.: femaleless Controls Sex Determination and Dosage Compensation Pathways in Females of Anopheles Mosquitoes. *Current Biology*. 2021; **31** (5): 1084-1091.e4 Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
No

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Mosquito population genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 03 Dec 2022

Oswald Djihinto

#Reviewer 2. The authors should update the Introduction section to reflect the latest development of sex determination research on Anopheles gambiae and related species/genes and set reasonable expectations and hypotheses.

Reply*: An attempt was made to update the Introduction.*

#Reviewer 2. "The malaria transmission cycle involves, four major elements: the host

(human), the parasite, the vector and the environment." Use oxford style commas. "... the vector, and the environment."

***Reply**: Done*

#Reviewer 2. "In the last decade, scientific advances in additional tools for vector control have included technologies such as ..." -> the papers cited for novel vector control tools are over 10 years old. There are newer publications in this realm and should be changed to newer citations or change wording to "In the last two decades, ..."

***Reply**: The sentence has been changed. Please, see lines 66-67.*

#Reviewer 2. "The latter technique, SIT, is based on ..." -> construct of listing more than two things and referencing 'the latter' for the last element is a bit awkward.

***Reply**: The sentence was changed. Please see lines 72 in the revised manuscript.*

#Reviewer 2. The 2nd paragraph of the introduction needs to be changed. As far as I am concerned there has not been radio-sterilized male SIT for Anopheles species although some people have attempted to test it in lab conditions and prepare for field trials.

***Reply**: The second paragraph was changed. We did not mean that there has been radio-sterilized male SIT for Anopheles species. We did mention "the genetic sterilisation of Anopheles species" (**line 72**). However, we were explaining the principle of SIT approach as firstly developed based on radiation. Please, see lines 72-75.*

#Reviewer 2. The paper 13 cited for SIT is introducing a genetic tool for inducing SIT not radiation. So explaining general radiation-based SIT method and citing paper 13 is not appropriate. It is especially concerning as this paper pertains to the Anopheles gambiae in Africa. Combined with more-than-a-decade-old citations for the alternative vector control strategies, it is somewhat questionable if the authors have due diligence in background research.

***Reply**: The paper 13 was changed were the general radiation-based SIT method is explained. The alternative vector control strategies are based on tools that have been developed more than decade ago. This justifies the date of the citations used.*

#Reviewer 2. "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." -> add 'genetic' in front of 'SIT technologies' as radiation-based SIT does not require genetic sexing.

***Reply**: Done*

#Reviewer 2. *The manuscript has mixed use of gender and sex. I suggest using the same and consistent terminology throughout the manuscript.*

***Reply**: Done*

#Reviewer 2. "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" -> word 'structure' in this case is not clear to readers. Structure is a very generic term and can mean very different things (e.g. RNA structure? Protein structure?). Clarification would be helpful for readers.

***Reply**: Clarification was made as follows: "Unfortunately, An. gambiae dsx gene (Agdsx) has different gene organization and regulatory elements positions…"*

Indeed, Scali et al., (2005) have reported that D. melanogaster female dsx gene has 4 exons while An. gambiae (Agdsx) has 7 exons. The exon 4 is splice out in female Dmdsx transcript while the exon 5 is the cassette exon in Agdsx. In Dmdsx, the inclusion or the exclusion of the female-specific exon depends on the presence of a weak 3' splice site in the preceding intron. Genomic sequence and transcript analysis indicate that in An. gambiae, splicing into sex-specific transcripts does not depend on the choice of an alternative 3' splice site (Scali et al., 2005). The 3' acceptor site preceding exon 5 does not appear to be weak. Although regulatory elements (splice enhancers and potential purine-rich enhancers) are found in the 3' UTR of exon 5 in Agdsx, these elements are found much further downstream from the 3′ splice acceptor site than in Drosophila, immediately upstream of the 5′ donor site of intron 5.
Scali C, Catteruccia F, Li Q, Crisanti A. Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene. J Exp Biol. 2005;208: 3701–3709. doi:10.1242/jeb.01819

#Reviewer 2. "Indeed, in humans, splicing signals are a common point of mutations…. Taking together these observations in humans and animal models, we hypothesized that the same events could be possible in insects and that SNPs could occur in acceptor and/or donor splice sites in mosquitoes that might result in the splice variation."-> There is some logical gap about expecting mutations in splicing regions determining sex locus. It seems like this is such a fundamental biological function that has huge implications in the downstream process so this region would be likely conserved. The authors should find examples of not just having mutations in splicing regions but splicing mutations in sex-determining genes to be relevant and set reasonable hypotheses/expectations.

***Reply**: Example of splicing mutations in sex determining genes were added in the introduction. Please, see lines 135-138.*

#Reviewer 2. *Table 1 – Since the study examined the allele frequency between males and females, Table 1 should provide the number of male and female sequences used for the analysis.*

***Reply**: The number of male and female sequences used for the analysis were provided in Table 1.*

#Reviewer 2. *Some sites of the Ag1000 data only contain females so generalizing the sex-dependent patterns with skewed data can be misleading.*

***Reply**: We thank the authors for his point of view. Our objective was to look for any sex-dependent patterns. Since we did not find any, we didn't generalize. We looked at the whole*

Ag1000 data for Anopheles gambiae (all countries). According to our hypothesis, if there was any sex-dependent pattern present, it should be observed at least in the sites (countries) with both female and male data. Furthermore, we were not expecting the sex-dependent pattern to be country-specific.

#Reviewer 2. "Only Anopheles gambiae samples were considered in our study." -> Anopheles gambiae s.s. (S form) only? Since the species gambiae and coluzzii were separated only in 2014, I think it would be best to clarify what you included for the readers.

***Reply**: In this study we used only Anopheles gambiae s.s. Details were added in the method section. Please, see line 151.*

#Reviewer 2. The first paragraph of the results section should be in the introduction. Also, readers who are not familiar with the sex determination process would wonder why sex determining gene is located on an autosome, not a sex chromosome. To educate readers, the authors should briefly describe how doublesex on autosome is involved in sex determination of Anopheles gambiae based on currently published literature in the introduction.

***Reply**: The first paragraph of the results section has been moved to the introduction. Please, see lines 105-113 in the revised manuscript.*
Details on how doublesex is involved in sex determination of Anopheles gambiae was addressed in the revised manuscript. Please, see lines 91-97.

#Reviewer 2. Figure 1. It would be best if female and male from the same location is side by side for a direct comparison. Males and females separate with different y-axis scales makes it difficult to compare. FR and GA colors are too similar.

***Reply**: The figure 1 was changed and became figure 2 with males and females separately with the same y-axis scales.*

#Reviewer 2. "The An. gambiae doublesex (Agdsx) gene is a candidate gene of interest for SIT, as a candidate for genetic modifications"-> candidate is repeated twice. Perhaps change it to "… is a candidate gene for genetic SIT strategy" or something to that effect?
Important paper on this topic like
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7955153/1 is not cited in this paper.

***Reply**: The sentence was corrected and the paper suggested was cited. Please see line 203.*

#Reviewer 2. "According to the D. melanogaster model18, the alternative splicing of Agdsx gene is governed by exon 5 skipping in male mosquitoes16 suggesting a silencing mechanism of the female-specific splice sites recognition (intron 4 acceptor and exon 5 donor sites) by the splicing machinery in males." -> I think there is a new paper on this topic from Anopheles gambiae which sequenced female and male transcriptome so the author doesn't have to deduce from the Drosophila sex termination model. Please check the latest papers by Krzywinski for updated literature on Anopheles gambiae sex determination.

***Reply**: Done*

#Reviewer 2 "They appeared only is two mosquito populations (Burkina Faso and Cameroon) over the eight populations considered." -> Burkina Faso and Cameroon are far apart so I wouldn't minimize how widespread this mutation is based on the number of countries examined (2 out of 8 is still not a low number).

***Reply**: We thank the reviewer for this notice. However, they appeared in very few individuals, less than 1% in females and no more than 2% in males.*

Competing Interests: None

Reviewer Report 11 March 2022

https://doi.org/10.21956/wellcomeopenres.19431.r48830

?

Tony Nolan
Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK

The report looks at variations around the splicing sites in the gene doublesex, a transcription factor essential in determining male and female sexual dimorphism. It mines data from the AG1000g initiative, which is sampling wild caught *Anopheles* malaria vectors and sequencing their genomes. The approach taken here, to mine the same data and look for insight on genetic variation that would be informative in genetic control approaches, is welcome. The methods and approach taken are generally well described and background information provided on doublesex gives good context. Unfortunately, though, I think there are some issues with the rationale applied and these currently, in my view, would preclude the indexing of this study. These issues are detailed below but I do think the methodology applied if given a different focus – for example in focusing the analysis on target sites around the female-specific splice junction that are used in a specific form of genetic control called gene drive. These are detailed below:

Major Points:

- I'm struggling with the logic that SNPs in a splice site could definitively determine, and be sufficient for, sex-specific splicing. By extension, the fact that the SNPs are found in both sexes is taken to mean that these particular SNPs cannot be critical for alternative, sex-specific splicing. But all genes, and alleles thereof, must be found in males and females at some point unless linked to a sex determining chromosome or locus. Am I missing something? In short, I just don't see how gender-specific SNPs will be helpful or informative (or at all likely) in this instance. The good thing at least is that your results confirm this.

- Gene drives that target the doublesex gene, and specifically, the female-specific exon,

recognise and cleave target sites that are uniquely within the female-specific exon. In the first published example by Kyrou *et al.* in 2018 the target site overlaps the intron4-exon5 boundary (the region of study here). Therefore, a refocusing of the analysis approach used in the current report, but instead to the probability of variants occurring (and being tolerated) that might show positive selection (or not) in the face of a gene drive suppressing the population by targeting this region, would be a way to resolve the current limitations. This obviously would require a re-focusing of the manuscript and a tweaking of the analysis by the authors.

- Another potential approach (unrelated) is to look at the conservation across the introns in general, and how (as I guess it is) this conservation is higher towards the donor and acceptor splice sites (since these are more functionally constrained). This analysis might reveal the extent and nature of the splicing sites/

Minor points:

- the *doublesex* (*dsx*) gene, encoding somatic sexually dimorphic traits – I would tweak this language slightly: dsx does not encode these traits directly

- It is not explicitly stated (yet probably should be) that alternative, sex-specific splicing of dsx is critical for the initiation of downstream transcriptional pathways that underpin sexual dimorphism.

- "Y-chromosome shredding gene drive" could be read ambiguously – the gene drive *is* on the Y chromosome but shreds the X chromosome.

- Though the paragraph leads with 'vector control' in general a reader might mistakenly think that Wolbachia or RIDL strains have been developed for malaria vectors (rather than Aedes vectors of Dengue etc.).

- I would add the qualifier 'local' to elimination

- "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" I am not quite sure what is meant by this statement – from what I know the overall structure in terms of sequence conservation and splicing structure, as well as last coding exon being sex-specific etc is pretty well conserved between Anopheles and Drosophila.*

- "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." – see my earlier comment. I think this approach needs rethinking.

- "The difference between the two sex-specific transcripts is due to the alternative splicing of exon 5. The latter is a cassette exon, which is retained in female and skipped in male transcript. The whole sequence of the female-specific exon 5 is included in the male intron 4 region and is spliced out. This gene structure causes a shift in intron/exon number in male. Thus, although male and female share the same exon/intron or intron/exon boundaries, they have common and specific splice sites." – this would surely benefit from a figure (even if it's been done before, it would help here in visualisation)

- " *Agdsx* gene, 17,196 polymorphic sites were identified." What is the definition for polymorphic here? A site where at least one individual in the sampled population that has a variant base compared to the published consensus genome?

- "Wherever both male and female mosquitoes are present (in Burkina Faso, Cameroon and Mayotte), the nucleotide diversity is similar between both sexes (Figure 1). This was expected as male and female in each country make up a single population. In addition, no difference in the nucleotide diversity was observed between male populations from the three countries (Burkina Faso, Cameroon and Mayotte) (Figure 1, top panel). The same trend was observed between female populations as well (Burkina Faso, Cameroon, Mayotte, Gabon, Ghana, Guinea, Equatorial Guinea and Uganda)" again, see earlier comment about rationale for looking at sex-specific differences.

- Figure 2 is nice.

- Figure 3 shows a heatmap of SNPs around exon 5 donor site, but the corresponding heatmap for the region around intrn4/exon 5 (for which a nice cartoon was made in Fig 2) is not shown.

- Re: the 'absence' of transformer in *An. gambiae* – there are several competing hypotheses for this, and not all are mentioned here (e.g difficulty with finding it by homology-based approaches due to extremely rapid sequence divergence)

- Throughout the article, it's never really made explicitly clear *how* this information on Dsx, and the splicing details specifically, would be relevant for genetic control, and SIT (which is mentioned frequently) in particular.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: vector control, gene drive, genetic control, molecular biology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 03 Dec 2022

Oswald Djihinto

#Reviewer 1. The report looks at variations around the splicing sites in the gene doublesex, a transcription factor essential in determining male and female sexual dimorphism. It mines data from the AG1000g initiative, which is sampling wild caught Anopheles malaria vectors and sequencing their genomes. The approach taken here, to mine the same data and look for insight on genetic variation that would be informative in genetic control approaches, is welcome. The methods and approach taken are generally well described and background information provided on doublesex gives good context. Unfortunately, though, I think there are some issues with the rationale applied and these currently, in my view, would preclude the indexing of this study. These issues are detailed below but I do think the methodology applied if given a different focus – for example in focusing the analysis on target sites around the female-specific splice junction that are used in a specific form of genetic control called gene drive.

Reply*: We thank the reviewer for suggesting a new orientation for the work. However, analysis of target sites around the female-specific splice junction used for the genetic control called gene drive has already been done by Kyrou et al. (2018). Despite using cage experiments over 4 mosquito generations, the authors revealed low-frequency indels at the target site.*

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;36: 1062–1066. doi:10.1038/nbt.4245

#Reviewer 1. I'm struggling with the logic that SNPs in a splice site could definitively determine, and be sufficient for, sex-specific splicing. By extension, the fact that the SNPs are found in both sexes is taken to mean that these particular SNPs cannot be critical for alternative, sex-specific splicing. But all genes, and alleles thereof, must be found in males and females at some point unless linked to a sex determining chromosome or locus. Am I missing something? In short, I just don't see how gender-specific SNPs will be helpful or informative (or at all likely) in this instance. The good thing at least is that your results confirm this.

Reply*: The reviewer is correct. The presence of SNPs in both sexes means that these particular SNPs cannot be critical for alternative sex-specific splicing. However, this study was aiming to investigate any gender-specific SNP in the target splice site sequences that could be critical for exon 5 skipping splicing mechanism. This objective was driven by reports of SNPs in the splice site that lead to an alternative exon splicing.*

#Reviewer 1. Gene drives that target the doublesex gene, and specifically, the female-

specific exon, recognise and cleave target sites that are uniquely within the female-specific exon. In the first published example by Kyrou et al. in 2018 the target site overlaps the intron4-exon5 boundary (the region of study here). Therefore, a refocusing of the analysis approach used in the current report, but instead to the probability of variants occurring (and being tolerated) that might show positive selection (or not) in the face of a gene drive suppressing the population by targeting this region, would be a way to resolve the current limitations. This obviously would require a re-focusing of the manuscript and a tweaking of the analysis by the authors.

***Reply**: We are grateful to the reviewer for this suggestion. However, a re-focusing of the manuscript to tackle the suggested aspect of the reviewer will lead to rewriting the whole manuscript content, which will not suit the title of the current manuscript. We will address the minor comments in this revision of the manuscript and later on, submit another manuscript on the investigation of a likely positive selection in the female-specific exon.*

#Reviewer 1. Another potential approach (unrelated) is to look at the conservation across the introns in general, and how (as I guess it is) this conservation is higher towards the donor and acceptor splice sites (since these are more functionally constrained). This analysis might reveal the extent and nature of the splicing sites/

***Reply**: We thank the reviewer for this suggestion. We will consider it when investigating the likely positive selection in the female-specific exon.*

#Reviewer 1. The doublesex (dsx) gene, encoding somatic sexually dimorphic traits – I would tweak this language slightly: dsx does not encode these traits directly.

***Reply**: The sentence was rephrased as follow: "... the doublesex (dsx) gene, involved in mosquitos' somatic sexually dimorphic traits determination." Please, see lines 18 and 19.*

#Reviewer 1. "Y-chromosome shredding gene drive" could be read ambiguously – the gene drive is on the Y chromosome but shreds the X chromosome.

***Reply**: Explanation was made as follow: "Y-chromosome shredding gene drive (the gene drive is on the Y chromosome but shreds the X chromosome)". Please, see line 70 and 71.*

#Reviewer 1. Though the paragraph leads with 'vector control' in general a reader might mistakenly think that Wolbachia or RIDL strains have been developed for malaria vectors (rather than Aedes vectors of Dengue etc.).

***Reply**: The paragraph was corrected. But yes Wolbachia has been developed for malaria vectors. Please see lines 66-69*

#Reviewer 1. I would add the qualifier 'local' to elimination

***Reply**: Done*

#Reviewer 1. "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" I am not quite sure what is meant by this statement – from what I know the overall structure in terms of sequence conservation and splicing structure, as well as last coding exon being sex-specific etc is pretty well conserved between *Anopheles* and *Drosophila*.

***Reply**: The structure that we are talking about here is the gene organization and regulatory elements in D. melanogaster dsx (Dmdsx) and Agdsx. Indeed, Scali et al., (2005) have reported that D. melanogaster female dsx gene has 4 exons while An. gambiae (Agdsx) has 7 exons. The exon 4 is splice out in female Dmdsx transcript while the exon 5 is the cassette exon in Agdsx. In Dmdsx, the inclusion or the exclusion of the female-specific exon depends on the presence of a weak 3' splice site in the preceding intron. Genomic sequence and transcript analysis indicate that in An. gambiae, splicing into sex-specific transcripts does not depend on the choice of an alternative 3' splice site (Scali et al., 2005). The 3' acceptor site preceding exon 5 does not appear to be weak. Although regulatory elements (splice enhancers and potential purine-rich enhancers) are found in the 3' UTR of exon 5 in Agdsx, these elements are found much further downstream from the 3′ splice acceptor site than in Drosophila, immediately upstream of the 5′ donor site of intron 5.*
Scali C, Catteruccia F, Li Q, Crisanti A. Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene. J Exp Biol. 2005;208: 3701–3709. doi:10.1242/jeb.01819

#Reviewer 1. "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." – see my earlier comment. I think this approach needs rethinking.

***Reply**: This section was removed from the revised manuscript.*

#Reviewer 1. "The difference between the two sex-specific transcripts is due to the alternative splicing of exon 5. The latter is a cassette exon, which is retained in female and skipped in male transcript. The whole sequence of the female-specific exon 5 is included in the male intron 4 region and is spliced out. This gene structure causes a shift in intron/exon number in male. Thus, although male and female share the same exon/intron or intron/exon boundaries, they have common and specific splice sites." – this would surely benefit from a figure (even if it's been done before, it would help here in visualisation).

***Reply**: A figure was proposed. Please see Figure 1 (line 113). This came in the introduction section according to the reviewer 2.*

#Reviewer 1 " Agdsx gene, 17,196 polymorphic sites were identified." What is the definition for polymorphic here? A site where at least one individual in the sampled population that has a variant base compared to the published consensus genome?

***Reply**: Yes, the polymorphic sites are nucleotide position where at least one individual in the sampled population has a variant base compared to the published AGAMP4 consensus genome.*

#Reviewer 1. "Wherever both male and female mosquitoes are present (in Burkina Faso, Cameroon and Mayotte), the nucleotide diversity is similar between both sexes (Figure 1).

This was expected as male and female in each country make up a single population. In addition, no difference in the nucleotide diversity was observed between male populations from the three countries (Burkina Faso, Cameroon and Mayotte) (Figure 1, top panel). The same trend was observed between female populations as well (Burkina Faso, Cameroon, Mayotte, Gabon, Ghana, Guinea, Equatorial Guinea and Uganda)" again, see earlier comment about rationale for looking at sex-specific differences.

***Reply**: Here we were not looking at sex-specific differences. Instead, we reported the genetic diversity of the dsx gene between mosquito populations. We split the data in female and male because all countries did not have both sex.*

#Reviewer 1. Figure 3 shows a heatmap of SNPs around exon 5 donor site, but the corresponding heatmap for the region around intrn4/exon 5 (for which a nice cartoon was made in Fig 2) is not shown.

***Reply**: The heatmap for the region around intron4/exon 5 was not shown because no SNP was neither found in the female intron 4 splice acceptor site or in the corresponding region in male.*

#Reviewer 1. Re: the 'absence' of transformer in An. gambiae – there are several competing hypotheses for this, and not all are mentioned here (e.g difficulty with finding it by homology-based approaches due to extremely rapid sequence divergence)

***Reply**: We thank the reviewer for this input. This statement was removed from the revised manuscript.*

#Reviewer 1. Throughout the article, it's never really made explicitly clear how this information on Dsx, and the splicing details specifically, would be relevant for genetic control, and SIT (which is mentioned frequently) in particular.

***Reply**: In this manuscript, we highlight that dsx is a promising target for genetic control and SIT. For this reason, we assumed that it is important to understand the mechanisms underlying it alternative splicing in the targeted vectors since the only well-known model comes from D. melanogaster. In An. gambiae, there are still some gap in all components involved in the dsx splicing. The identification of the remaining factors could lead to the characterization of novel genetic and biochemical targets for inducing genetic sterility*

Competing Interests: None