BMC Genomics

RESEARCH Open Access

Genome-wide cataloging and orthology analysis of long noncoding RNA expression in three species of *Anopheles* mosquito

Waseem Akram Malla[1*], Kuldeep Singh[2,3], Niha Ayman[4], Chandini Boro[1], Naorem Chaoba Devi[1], Priyanka Mech[1], Nida Siddiqui[1], Anupkumar R. Anvikar[2,3] and Praveen Kumar Bharti[2,3*]

Abstract

Background Long noncoding RNAs (lncRNAs) are versatile regulatory molecules that affect cellular phenotypes through context-specific expression. While their role in controlling cellular pathways is well-established in insects, investigating lncRNA expression in *Anopheles* in a tissue- and species-specific manner could add to our understanding of malaria transmission by this important vector.

Methodology We performed *de novo* transcriptome assembly of *Anopheles minimus*, *Anopheles albimanus*, and *Anopheles arabiensis* utilising publicly available RNA-Seq datasets of male reproductive tissues, male carcasses, female reproductive tissues, and female carcasses. Various bioinformatics tools were subsequently used for lncRNA identification, conservation analysis, and differential expression analysis across sexes and tissues.

Results We identified 9331, 5372, and 5256 lncRNA transcripts in *An. albimanus*, *An. arabiensis*, and *An. minimus*, respectively. Compared with *An. albimanus* lncRNAs, conservation analysis revealed that a total of 1964 and 1400 lncRNAs were conserved in *An. arabiensis* and *An. minimus*; however, only 283 and 253 lncRNAs presented sequence-level conservation. Differential expression (DE) analysis revealed that the carcasses presented the lowest difference in lncRNA expression, whereas in each comparison, the reproductive tissues (whether male or female) presented relatively high levels of differential expression. Additionally, 69 lncRNAs were found to be conserved at the sequence level in all 3 species. These lncRNAs were almost exclusively upregulated in males (reproductive tissue as well as carcasses) and exclusively downregulated in female carcasses.

Conclusions The genes in the vicinity of differentially expressed, conserved lncRNAs were found to be involved in critical pathways such as nuclear structure, chromatin remodelling, protein and RNA metabolism, and cell cycle in each of these species. Future studies on these lncRNAs can provide useful insights into how these functions control sexually-dimorphic physiological phenomena such as host-seeking and biting behaviour of female mosquitoes, blood meal metabolism, reproductive behaviour, and disease-carrying capacity.

Keywords LncRNA, *Anopheles*, *Anopheles minimus*, *Anopheles albimanus*, *Anopheles arabiensis*, Noncoding RNA

*Correspondence:
Waseem Akram Malla
malla.wa@icmr.gov.in; drwaseem2kx@gmail.com
Praveen Kumar Bharti
saprapbs@yahoo.co.in

Full list of author information is available at the end of the article

Background

Long noncoding RNAs (lncRNAs) are a highly heterogeneous class of regulatory RNAs that are usually longer (> 200 nucleotides) than other types of regulatory RNAs. LncRNAs can directly interact with the chromatin, recruit or act as decoys for chromatin modifiers, bind to RNA-binding proteins, and function as competitive endogenous RNAs by sponging miRNAs [1]. These interactions modulate gene expression and thereby influence diverse physiological phenomena, such as cell proliferation, differentiation, cell death, disease progression, and the immune response [1, 2]. While most studies on lncRNAs have been conducted in human disease models (especially various types of malignancies) and model organisms, it has been established that these RNAs nevertheless play regulatory roles in biological processes in other life forms as well [3–6]. Among insects, the long noncoding transcriptome of *Drosophila melanogaster* (fruit fly) is among the most well-characterized. lncRNAs have also been documented in other insect species, where they affect evolutionary adaptation (e.g., *ivory* lncRNA regulates wing pigmentation in the buckeye butterfly *Junonia coenia* [7]), social immunity (e.g., XR_001766094.2 lncRNA was found to be enriched in *Apis cerana* uncapping worker bees, potentially modulating hygienic behaviour [8]), pathogen-carrying capability (e.g., lncRNAs Zinc1, Zinc2, and Zinc22 affect the permissiveness of *Aedes aegypti* to Zika virus infection [9], *Wolbachia*-induced lncRNAs regulate intracellular ROS stress and activate the antidengue Toll pathway in *Ae. aegypti* [10]), reproductive behaviour (e.g., lncRNAs Zinc9 and Zinc22 in *Ae. aegypti* affect fecundity and oviposition [9]), insecticide resistance (e.g., Cry1Ac resistance in the bollworms *Pectinophora gossypiella* and *Helicoverpa zea* is thought to be driven by various lncRNAs [11, 12]), and several other functions [13, 14]. The high-throughput nature of second- and third-generation sequencing technologies enables the reliable quantification of relatively rare transcripts as well as facilitates the discovery of novel transcripts. These technologies have been employed to study the genomes and transcriptomes of numerous mosquito species [15–18]. The mosquito genus *Anopheles* has been extensively studied, as more than 40 species in this genus have been identified as vectors of the *Plasmodium* parasite that causes malaria [19]. However, limited information is available on the annotation or expression of various classes of noncoding RNAs. Among the various vector mosquito species, the expression of lncRNAs and their regulatory effects on reproductive competence and pathogen-carrying ability have been studied in *Ae. aegypti* and *Ae. albopictus* [9, 20, 21]. Similarly, more than 21,000 lncRNA transcripts have been identified in various developmental stages of *Ae. albopictus*, where differential lncRNA expression potentially affects metamorphosis at various developmental time points [22]. Within the *Anopheles* genus, approximately 3000 lncRNAs whose expression varies in a context-specific manner, similar to mRNA expression, across various life stages and sexes of *An. gambiae* have been identified [23].

In this study, we used publicly available RNA-Seq datasets from *An. albimanus*, *An. arabiensis*, and *An. minimus* for *de novo* transcriptome assembly of each species. *An. albimanus* is the main anopheline species in northern South America, Central America, and the Caribbean islands [24], *An. arabiensis* is a sibling species of *An. gambiae* and is widely distributed throughout several African countries [25], and *An. minimus* has been reported to be widely distributed in several Southeast Asian countries, where it has been recognized as an important malaria vector [19, 26]. At present, lncRNA annotation using *de novo* transcriptome assemblies presents various challenges such as classification of spurious transcripts as lncRNAs, misidentification of already annotated lncRNAs as coding genes, or misidentification of unannotated coding genes as lncRNAs. To this effect, after assembling the transcriptome, we identified the potential lncRNA transcripts using various filtering criteria and compared their expression between male and female reproductive tissues and the remaining carcass. Previous studies across diverse species have demonstrated a 'sexually dimorphic' pattern of gene expression, characterised by a higher expression of male- or female-biased genes in the corresponding sex [27–29]. Since only the female *Anopheles* mosquitoes transmit malaria, it is important to study how sexually dimorphic lncRNAs regulate female-specific behavioural patterns and biological functions relevant to disease transmission. We also compared the lncRNAs of *An. arabiensis*, and *An. minimus* with *An. albimanus* lncRNAs to find transcripts that may be conserved across these species. The cataloguing of the lncRNA components in *Anophelines* is a crucial step toward the biological characterization and annotation of this class of regulatory RNAs.

Materials and methods

Data retrieval

We obtained a publicly available dataset with the accession number SRP083856 from the NCBI Sequence Read Archive. It consists of paired-end data from 36 samples, with 12 samples each belonging to *An. albimanus*, *An. arabiensis*, and *An. minimus*. For each species, the data consisted of 3 replicates each from male reproductive tissue (testis and accessory glands), female reproductive tissue (ovaries and common oviduct), the remaining male carcass (head, thorax, and abdomen), and the remaining female carcass (head, thorax, and abdomen). While the data have been previously analysed for differential mRNA

expression, the present study used these datasets to study differential lncRNA expression between different samples within each species and lncRNA conservation/orthology analysis between the three species [29].

Transcriptome assembly and alignment

The *de novo* transcriptome assembly pipeline is shown in Fig. 1. LncRNA expression analysis was carried out according to the LncEvo pipeline [30] separately for each species. However, considering the available annotation of each genome, the computational power available, and the feasibility of the bioinformatics tools, we used a modified protocol as outlined below.

First, the reference genome (in the fa.gz format) of each species and its corresponding annotation file (in the GTF format) were obtained from the Ensembl Metazoa genome browser. The reference genome builds and annotations used included *Anopheles arabiensis* (Dongola; AaraD1; release 59), *Anopheles minimus* (MINIMUS1; AminM1; release 59), and *Anopheles albimanus* (STECLA; AalbS2; release 59). Next, the paired-end reads were trimmed and quality-filtered with Trimmomatic (version 0.39) [31]. The reads were trimmed when the quality score dropped below 25 within a sliding window of 4 nucleotides. Any reads shorter than 50 nucleotides were dropped from further analysis.

The cleaned and trimmed reads were used for alignment to the reference genome and transcriptome assembly. The reads were mapped to the indexed reference genome (of the corresponding species) via HISAT2 (version 2.2.1) software with default parameters [32]. The mapped reads were obtained as BAM files and used as input for *de novo* transcriptome assembly via StringTie (version 2.2.1) [33], with the corresponding annotation files used as guides. For each sample, a GTF file representing a custom transcriptome was generated. These sample-specific transcriptomes were subsequently merged to create a common transcriptome for each species. For each species, the merged transcriptome (GTF) was compared to the corresponding reference annotation, probable erroneous transcripts were removed, and the transcriptome sequence (FASTA) was generated from the reference genome. Next, the merged transcriptome of each species was indexed, and transcript quantification was carried out by aligning the trimmed reads to the indexed transcriptome via Salmon (version 1.4.0) [34].

Transcript filtering and classification as lncRNAs

After quantification, the transcriptome was again compared against the reference annotation and ENSEMBL transcript biotype information, followed by coding potential analysis via the CPC2 [35] standalone version and TransDecoder (https://github.com/TransDecoder/TransDecoder) with default parameters. Only transcripts that were longer than 200 nucleotides and identified as noncoding by CPC and TransDecoder were retained. Finally, a set of lncRNAs was obtained for each species in the GTF and FASTA formats.

LncRNA orthology analysis

Conservation and orthology analyses of lncRNAs were carried out in a pairwise manner. *An. arabiensis* and *An.*

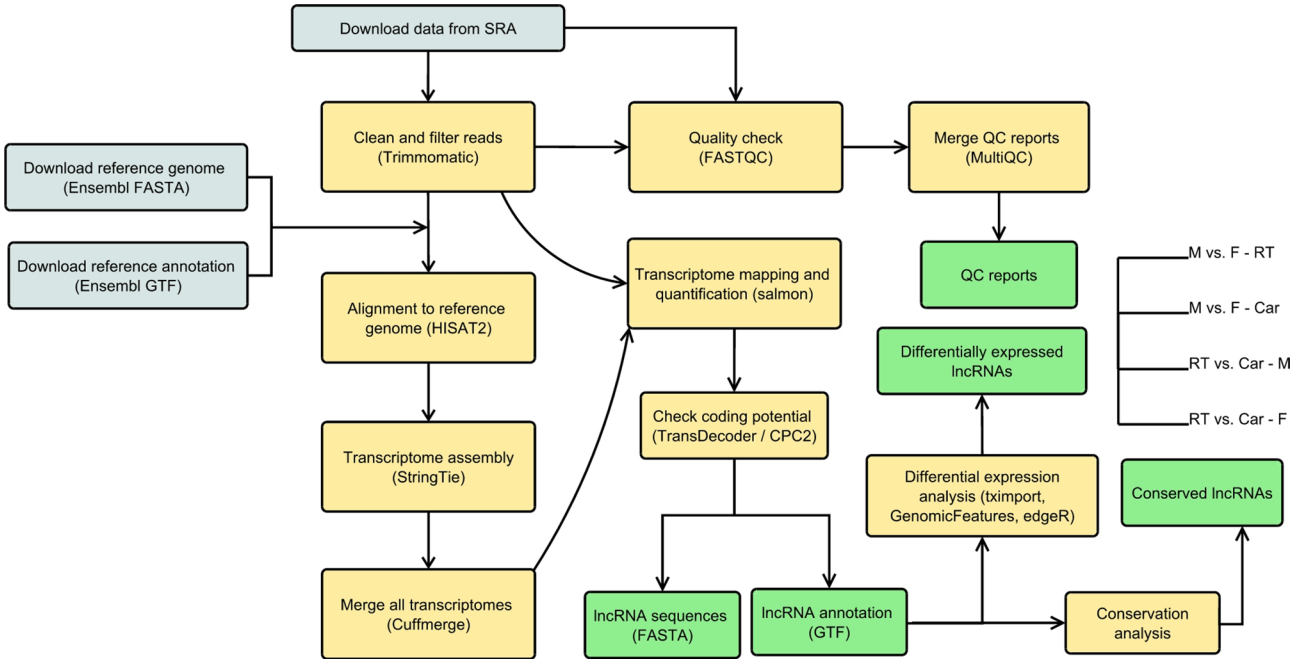


Fig. 1 LncRNA assembly, conservation, and expression analysis pipeline (*Abbreviations*– Car: carcass; RT: reproductive tissue; M: male; F: female)

minimus lncRNAs were used as 'query' sequences, and *An. albimanus* lncRNAs as 'target' sequences (due to its chromosome-level genome assembly) to search for conserved lncRNAs. While the parameters recommended in the LncEvo pipeline for various tools were used, specific adjustments wherever required are mentioned below.

First, the lncRNA annotation (GTF) generated for each species was converted into an Ensembl-compatible format and merged with the downloaded reference annotation. For cross-species alignment, the target reference genome FASTA was first converted into a 2-bit format and then indexed using lastdb with a sliding window of 99 (-W99) and the seeding scheme as NEAR (-uNEAR). Suitable substitution and gap scores for aligning the query and target sequences were determined via last-train, followed by alignment via lastal, with the maximum initial matches per query position set to the default value (-m10) [36]. The MAF (multiple alignment format) files generated were converted into PSL (pairwise sequence local) format and then into alignment chain files using axtChain with a query 2-bit genome, a target 2-bit genome, and a substitution matrix generated with last-train as input. The parameters were set as -minScore = 5000 -linearGap = medium.

In the final step, lncRNA transcripts from Ensemble-compatible lncRNA annotation (obtained above) of target species were converted into a bed file and used as a reference set for comparison between the target and query species. For each comparison, two files were obtained: (a) an orthologs.txt containing information about all orthologue pairs between the query and target species and (b) an orthologs.top.txt file containing only the best alignments on the basis of exonic identity for each lncRNA [37]. Circos plots were used to plot the genomes and visualize the lncRNA conservation in each comparison [38].

Differential expression analysis of lncRNAs

A count matrix of potential lncRNA genes for each species was generated by reading the corresponding 12 Salmon output files and summarizing transcript level counts to gene level counts with the tximport (version 1.30.0) and GenomicFeatures (version 1.54.4) packages in RStudio. Any lncRNA genes that were expressed in less than half of the samples (i.e., zero counts in more than 6 samples) were removed from further analysis. The resulting normalized count matrix containing transcripts per million (TPM) for each lncRNA was subjected to differential expression analysis using edgeR (version 4.0.16) [39] by fitting a negative binomial generalized log-linear model. A total of 4 comparisons were carried out for each species: (a) male reproductive tissue vs. female reproductive tissue, (b) male carcass vs. female carcass, (c) female reproductive tissue vs. female carcass, and (d) male reproductive tissue vs. male carcass. All lncRNAs with an absolute log 2-fold change ($|\log2FC| > 2$ and false discovery rate (FDR) < 0.05 were considered significantly differentially expressed lncRNAs (DE-lncRNAs). Venn diagrams were generated via Venny 2.1 (https://bioinfogp.cnb.csic.es/tools/venny/) to identify unique and overlapping DE-lncRNAs between comparisons.

Gene ontology (GO) analysis

To identify the potential functions regulated by the differentially expressed and conserved lncRNAs, we identified their neighbouring protein-coding genes as per the protocol outlined in Kang and Liu [40]. Briefly, neighbouring genes were identified within a 100 kb window either upstream or downstream of any lncRNA using 'windowBed' function in bedtools. Next, Pearson's correlation co-efficients (r) between the expression levels of the lncRNAs and their neighbouring protein-coding genes were calculated, and only lncRNA-gene pairs with a correlation less than minus 0.90 ($r < -0.9$) were retained for GO analysis. GO analysis was performed using g:GOSt function of g:Profiler [41].

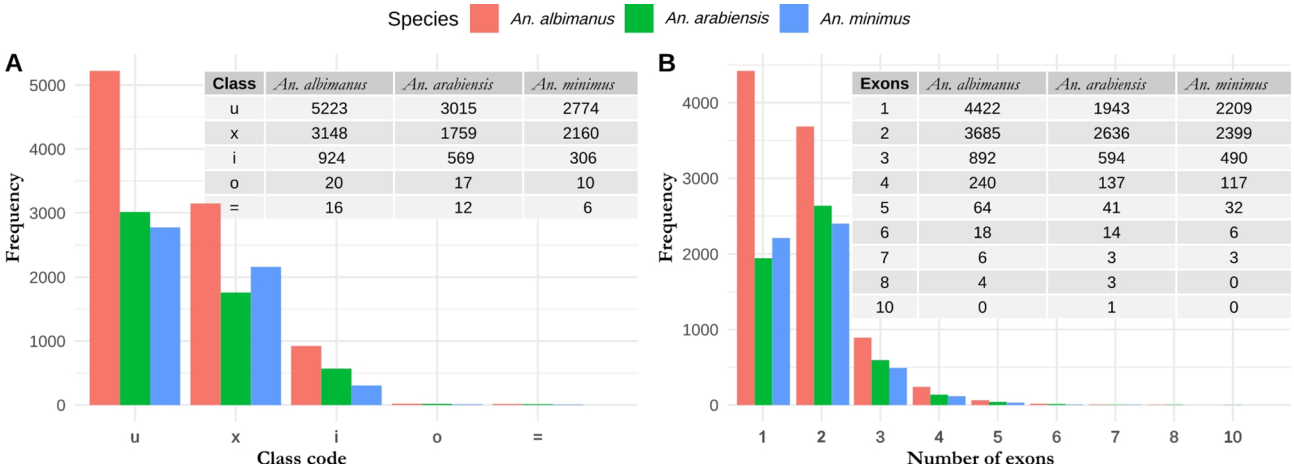
Results

Transcriptome assembly, LncRNA identification, and LncRNA characteristics

A total of 9331, 5372, and 5256 transcripts were classified as lncRNAs in *An. albimanus*, *An. arabiensis*, and *An. minimus*, respectively. As the data on lncRNA annotation in these species are limited, the proportion of transcripts classified as known lncRNAs (class code '=') or those having exons overlapping with a reference transcript (class code 'o') was less than 1% in all the species. The classwise distributions of all the lncRNAs are shown in Fig. 2 (A). Most of the identified lncRNA transcripts belonged to one of the 3 classes, viz.:

a) Intergenic (class code 'u'; lincRNAs): 56% in *An. albimanus* as well as in *An. arabiensis*, 53% in *An. minimus*;
b) Exonic overlaps with the reference on the opposite strand (class code 'x'; exonic lncRNAs): 34% in *An. albimanus*, 33% in *An. arabiensis*, 41% in *An. minimus*; and.
c) Entirely intronic (class code 'i'; intronic lncRNAs): 10% in *An. albimanus*; 11% in *An. arabiensis*, 6% in *An. minimus*.

More than 80% of the lncRNA transcripts had either one or two exons (Fig. 2 (B)). Since lncRNAs typically have fewer exons, they are generally shorter than protein-coding genes (e.g., 4.6 exons per protein-coding gene in *An. darlingi* and 4.4 in *An. gambiae* [42]). The distribution of lncRNA lengths vs. lengths of the neighbouring genes in



Class	*An. albimanus*	*An. arabiensis*	*An. minimus*
u	5223	3015	2774
x	3148	1759	2160
i	924	569	306
o	20	17	10
=	16	12	6

Exons	*An. albimanus*	*An. arabiensis*	*An. minimus*
1	4422	1943	2209
2	3685	2636	2399
3	892	594	490
4	240	137	117
5	64	41	32
6	18	14	6
7	6	3	3
8	4	3	0
10	0	1	0

Fig. 2 Distribution of lncRNAs by class code **(A)** and number of exons **(B).** The inset tables show the number of lncRNAs for the class code in (**A**) and for the number of exons per transcript in (**B**) in each species. *An. albimanus* is represented by salmon bars, *An. arabiensis* by green bars, and *An. minimus* by blue bars. Class codes: 'u' – intergenic lncRNAs (lincRNA), 'x' – exonic lncRNAs on the opposite strand, 'i' – intronic lncRNAs, 'o' – generic exonic overlap with a reference transcript, and '=' – known lncRNAs

Table 1 Summary of the transcript lengths of LncRNAs and their host/neighbouring genes

Species	Feature	Minimum Length (nt)	Median Length (Q1, Q3; nt)	Mean Length (nt)	Maximum Length (nt)
An. albimanus	lncRNA	200	499 (293, 969)	899.6	16,421
	Neighbouring genes	67	3018 (1425, 5349)	3928	45,495
An. arabiensis	lncRNA	200	619.5 (350, 1227)	1145.1	22,579
	Neighbouring genes	48	3175 (1458, 5971)	4270	50,767
An. minimus	lncRNA	200	665 (368.8, 1488)	1338.7	21,396
	Neighbouring genes	72	2106 (1058, 4229)	3071	45,474

each species confirms this correlation, as summarized in Table 1 and shown in Fig. 3. The lncRNA annotation files for each species are provided in Supplementary files S1 (*An. albimanus*), S2 (*An. arabiensis*), and S3 (*An. minimus*).

LncRNA orthology analysis

Conservation analysis revealed that there were 1964 orthologous lncRNA pairs between *An. arabiensis* and *An. albimanus* and 1400 orthologous pairs between *An. minimus* and *An. albimanus*. Among the conserved lncRNAs, most of the lncRNA pairs were of the intergenic-intergenic type in both species, as shown in Fig. 4A. Additionally, the highest number of the conserved *An. arabiensis* lncRNAs mapped to the 2R chromosome of *An. albimanus*, whereas the highest number of *An. minimus* lncRNAs mapped to the 3L chromosome. In each of these species, the least number of conserved lncRNAs mapped to the X chromosome of *An. albimanus*. A summary of the chromosomewise mapping is shown in Fig. 4B.

Next, we compared the locusID and exonID scores generated by slncky and investigated how the lncRNA pairs were conserved at the sequence level. In the *An. arabiensis* vs. *An. albimanus* comparison, 283 lncRNA transcript pairs were found to have nonzero exon locus identity scores (i.e., exonID and locusID > 0), whereas in the *An. minimus* vs. *An. albimanus* comparison, 253 lncRNA pairs had nonzero exon locus identity scores (i.e., exonID and locusID > 0). Among these lncRNAs, 69 lncRNAs presented sequence-level conservation in all three species (i.e., these lncRNA triads presented nonzero exonID and locusID scores in all the species). A list of these conserved triads is provided in Supplementary file S4. The sequences of these lncRNAs in FASTA format are provided in Supplementary files S5 (*An. albimanus*), S6 (*An. arabiensis*), and S7 (*An. minimus*).

Identification of differentially expressed lncRNAs

In the differential expression analysis, the lncRNA transcripts mapped to 8097, 4601, and 4616 lncRNA genes in *An. albimanus*, *An. arabiensis*, and *An. minimus*, respectively. We carried out a specieswise differential expression analysis, and the findings are as follows:

An. albimanus

Among 8097 lncRNA genes, only 2998 were expressed in at least 6 samples and were retained for differential expression analysis. There were 289 significant

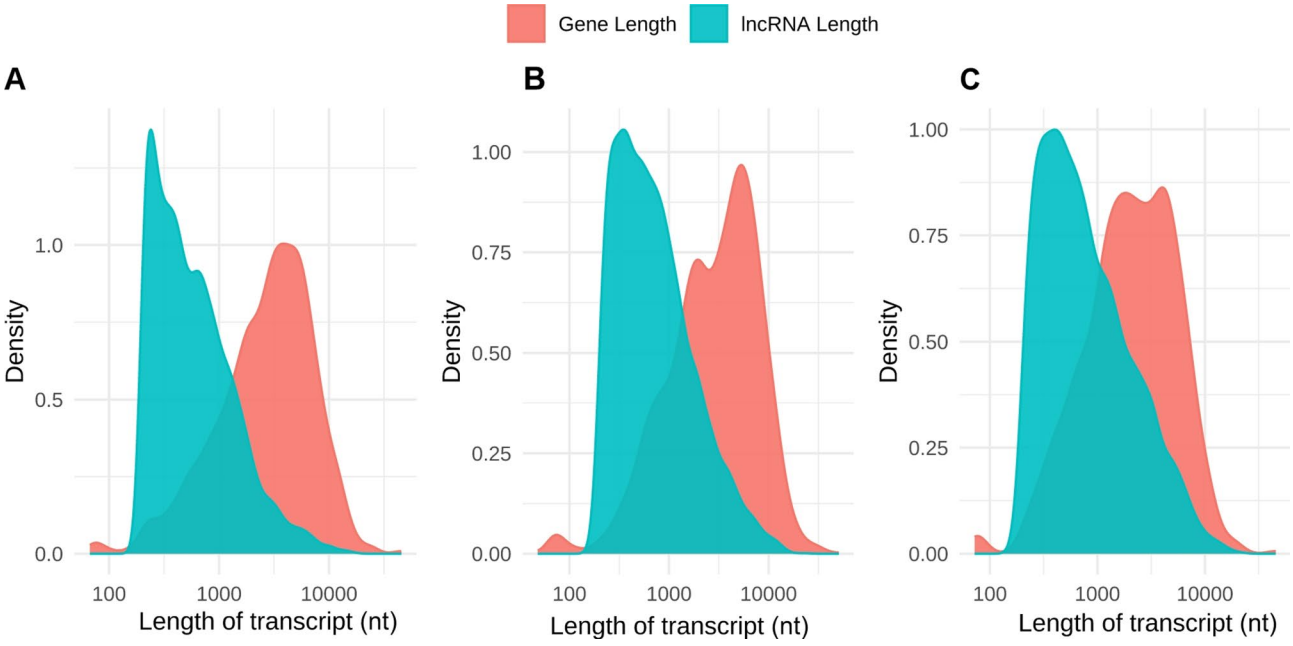


Fig. 3 Density plots showing the distribution of the length of the assembled transcripts. The length of the lncRNA transcripts (in sea-green) and their neighbouring genes (in salmon) in *An. albimanus*, *An. arabiensis*, and *An. minimus* are shown in **A**, **B**, and **C**, respectively. In each plot, the length on the x-axis is plotted on a log scale (log10)

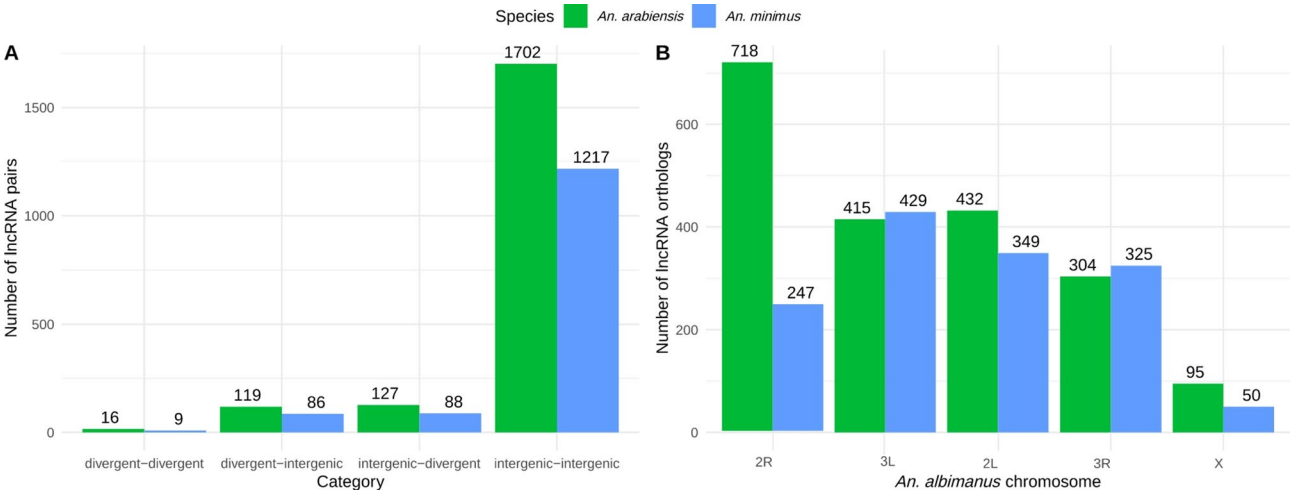


Fig. 4 Distribution of conserved lncRNA pairs. The number of lncRNA pairs in each category is shown in **A**, and the distribution of conserved query lncRNAs across the target genome (*An. albimanus*) is shown in **B**. The green bars represent *An. arabiensis* lncRNAs and the blue bars represent *An. minimus* lncRNAs. The figures on the top of each bar represent the number of lncRNAs of a particular category in (**A**) or lncRNAs mapping to a target chromosome in (**B**) in each species

DE-lncRNAs between male and female carcasses, 1704 between male reproductive tissue and female reproductive tissue, 840 between male reproductive tissue and male carcass, and 1085 between female reproductive tissue and female carcass. The dysregulated genes are represented by volcano plots in Fig. 5, Panel A. Among the dysregulated genes, 28 unique DE-lncRNAs were identified when the male and female carcasses were compared, whereas 304 DE-lncRNAs were found when male and female reproductive tissues were compared. Additionally, 77 and 228 unique DE-lncRNAs were observed in the male reproductive tissue vs. male carcass and female reproductive tissue vs. female carcass comparisons, respectively. The overlap between various comparisons is shown in Fig. 6 (A).

An. arabiensis

Among the 4601 lncRNA genes, only 2546 were expressed in at least 6 samples and were retained for differential expression analysis. There were 81 significant

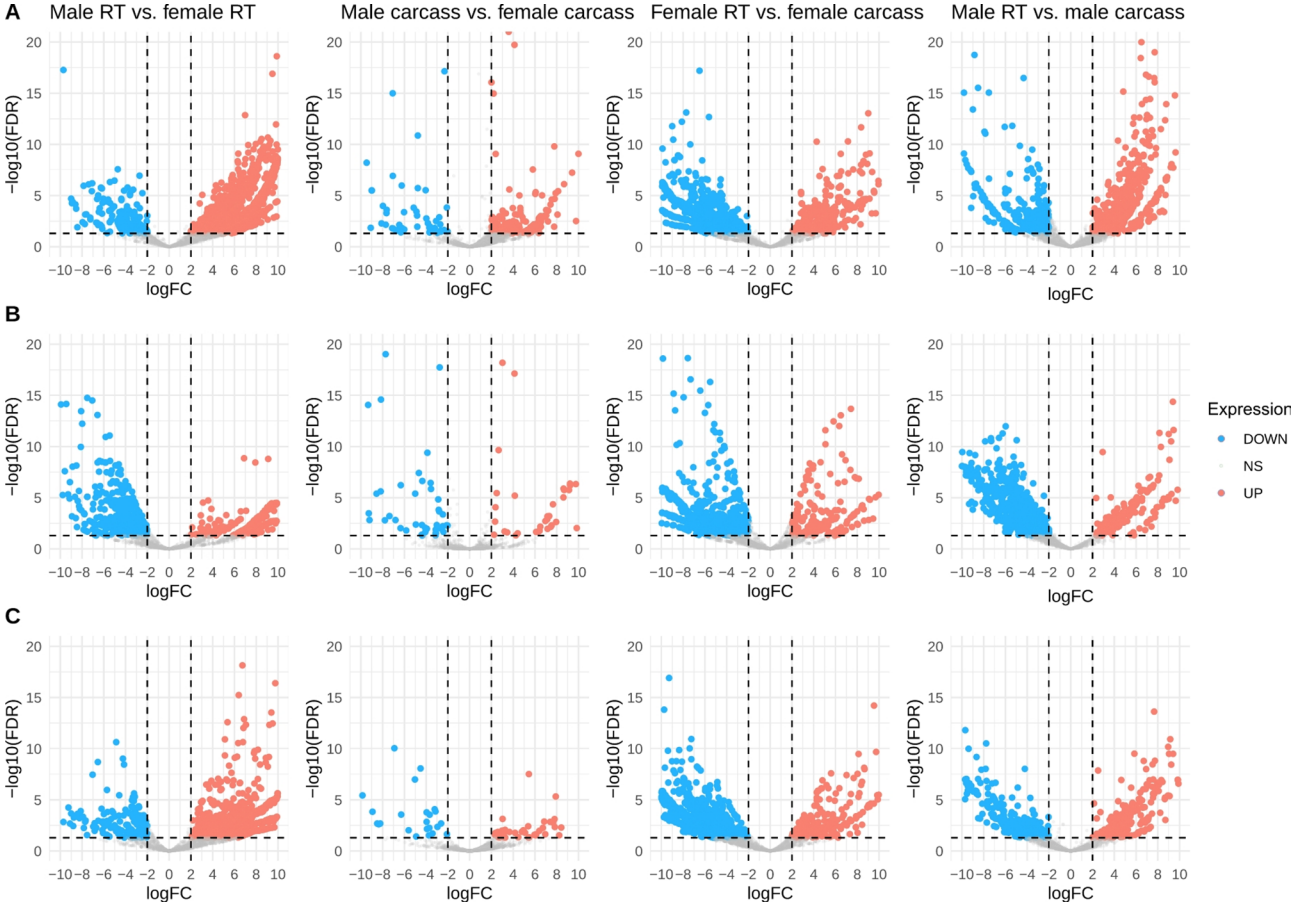


Fig. 5 Volcano plots depicting DE-lncRNAs across three *Anopheles* species. Panel **A** presents DE-lncRNAs in *An. albimanus* for the four comparison groups (labels on top), Panel **B** shows the DE-lncRNAs in *An. arabiensis*, and Panel **C** displays DE-lncRNAs in *An. minimus*. The vertical dotted lines represent the log2-fold-change (FC) cut-off for significance, whereas the horizontal dotted lines represent the false discovery rate (FDR) cut-off for significance. Colours of dots: Salmon, significantly upregulated lncRNAs (log2FC > + 2 and FDR < 0.05); blue, significantly downregulated lncRNAs (log2FC < -2 and FDR < 0.05); grey, lncRNAs that do not meet either cut-off (nonsignificant: +2 < log2FC < -2 and FDR > 0.05). For consistency, the x-axis has been restricted to represent log2FC values between – 10 and 10, and the y-axis has been restricted to represent -log10(FDR) values between 0 and 20. (*Abbreviations* – RT: reproductive tissue)

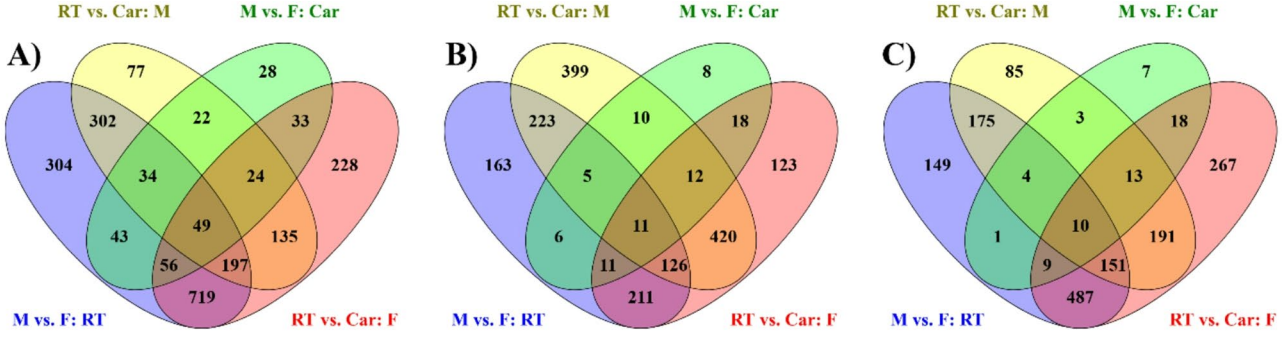


Fig. 6 Venn diagrams representing the overlap of DE-lncRNA expression between various comparisons. **A** represents the overlap in *An. albimanus* tissues, **B** in *An. arabiensis*, and **C** in *An. minimus*. (*Abbreviations* – Car: carcass, RT: reproductive tissue, M: male, F: female)

DE-lncRNAs between the male carcass and female carcass, 756 between the male reproductive tissue and female reproductive tissue, 1206 between the male reproductive tissue and male carcass, and 932 between the female reproductive tissue and female carcass. The dysregulated genes are represented by volcano plots in Fig. 5, Panel B. Among the dysregulated genes, there were 8 unique DE-lncRNAs when male and female carcasses were compared, 163 when male and female reproductive tissues were compared, 399 when male reproductive

Table 2 Summary of differential lncRNA expression in the 3 species. ('+': number of upregulated lncRNA genes; '−': number of downregulated lncRNA genes; 'T': total number of dysregulated genes)

Comparison	*An. albimanus*			*An. arabiensis*			*An. minimus*		
	−	+	T	−	+	T	−	+	T
Male carcass vs. female carcass	52	237	289	45	36	81	26	39	65
Male reproductive system vs. female reproductive system	206	1498	1704	425	331	756	161	825	986
Male reproductive system vs. male carcass	297	543	840	1011	195	1206	365	267	632
Female reproductive system vs. female carcass	1085	356	1441	731	201	932	890	256	1146

Table 3 Conserved LncRNA counts by comparison and upregulation. (*Abbreviations* – Car: Carcass, RT: reproductive tissue, M: male, F: female)

Species	Comparison	Upregulated in	Number of conserved lncRNAs
An. albimanus	M_Car_vs_F_Car	Male_Car	3
	M_RT_vs_F_RT	Female_RT	4
	M_RT_vs_F_RT	Male_RT	26
	F_RT_vs_F_Car	Female_RT	2
	M_RT_vs_M_Car	Male_Car	2
An. arabiensis	M_RT_vs_F_RT	Female_RT	4
	M_RT_vs_F_RT	Male_RT	15
	F_RT_vs_F_Car	Female_RT	2
	M_RT_vs_M_Car	Male_Car	18
	M_RT_vs_M_Car	Male_RT	1
An. minimus	M_RT_vs_F_RT	Male_RT	25
	F_RT_vs_F_Car	Female_RT	2
	M_RT_vs_M_Car	Male_Car	4

tissue and carcass were compared, and 123 when the female reproductive tissue and carcass were compared. The overlap between various comparisons is shown in Fig. 6 (B).

An. minimus

Among the 4616 lncRNA genes, only 2608 were expressed in at least 6 samples and were retained for differential expression analysis. There were 65 significant DE-lncRNAs between male and female carcasses, 986 between male reproductive tissue and female reproductive tissue, 632 between male reproductive tissue and male carcass, and 1146 between female reproductive tissue and female carcass. The dysregulated genes are represented by volcano plots in Fig. 5, Panel C. Among the dysregulated genes, 7, 149, 85, and 267 unique DE-lncRNAs were identified in the male carcass vs. female carcass, male reproductive tissue vs. female reproductive tissue, male reproductive tissue vs. male carcass, and female reproductive tissue vs. female carcass comparisons, respectively. The overlap between various comparisons is shown in Fig. 6 (C).

A summary of the results is provided in Table 2. The results indicate that, in each species, the carcasses differ the least in terms of lncRNA expression, whereas the reproductive tissues show a greater level of differential expression in each comparison. For each species and each comparison, the top 15 upregulated and top 15 downregulated genes plotted as heatmaps are available in Supplementary file S8.

Differentially expressed, sequence-conserved lncRNAs

Next, we analysed how these lncRNAs with sequence conservation were expressed across comparisons in each species. We found that 37, 40, and 31 lncRNAs were differentially expressed in *An. albimanus*, *An. arabiensis*, and *An. minimus*, respectively. Interestingly, the conserved lncRNAs were almost exclusively upregulated in male tissues (reproductive tissue or carcass). Additionally, across comparisons, these conserved lncRNAs were found to be exclusively downregulated in the female carcass. The results are summarized in Table 3. Finally, the overall summarized results of the analysis are presented as Circos plots in Fig. 7 (*An. arabiensis* vs. *An. albimanus*) and Fig. 8 (*An. minimus* vs. *An. albimanus*). The genome assemblies available for all the three species are contig-level, with *An. albimanus* having a better annotation (5 chromosomes and 196 contigs) than *An. arabiensis* (1217 contigs) and *An. minimus* (678 contigs). For the sake of clarity, only contigs longer than 300 kilobasepairs are shown in the Circos plots.

Functional analysis of differentially expressed, sequence-conserved lncRNAs

In *An. albimanus*, 1312 neighbouring genes were located, and GO analysis found these to be involved in functions related to nucleic acid metabolism, chromatin expression and remodelling, and cell cycle. In *An. arabiensis*, 470 neighbouring genes were identified and found to be involved in terms related to RNA biogenesis, protein synthesis, and cell cycle. Similarly, in *An. minimus*, 451 neighbouring genes were identified and found to be involved in protein metabolism, nucleic acid metabolism, and cell cycle. A complete list of the enriched GO terms is given in Supplementary file S9. A list of the significantly enriched, driver GO terms is provided in Table 4.

Discussion

There is a limited amount of data available on the annotation of lncRNAs in anopheline species. LncRNAs are an area of special interest because they can regulate the

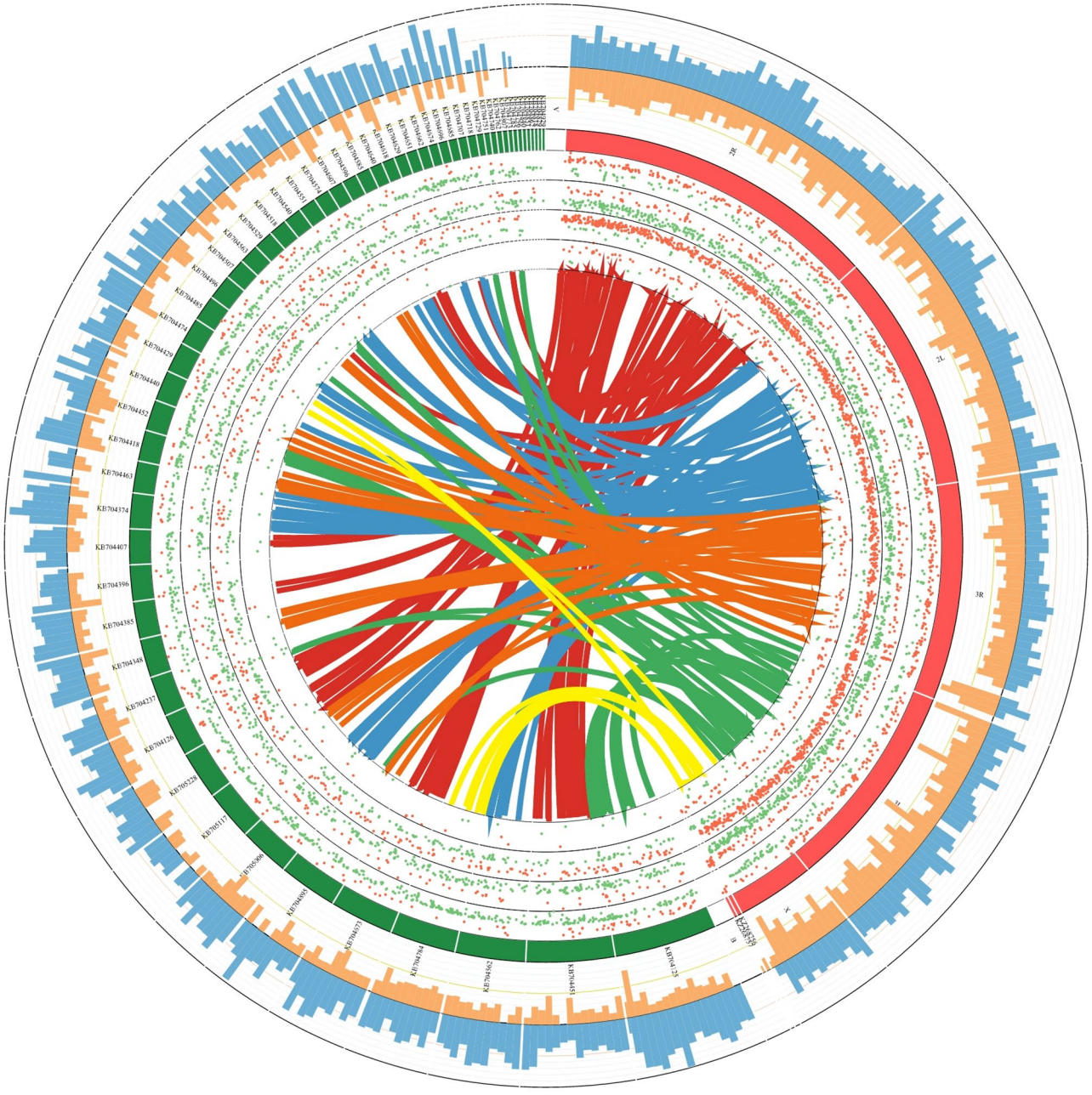


Fig. 7 Circos plot representing lncRNA conservation between *An. arabiensis* and *An. albimanus*. From periphery to centre, the first track (blue) represents the distribution of protein-coding genes (reference annotation) across the genomes of these 2 species; the second track (orange; inverted) represents the distribution of *de novo* assembled lncRNA transcripts across the genomes; the third track represents the ideograms of *An. albimanus* (A to B; clockwise; salmon) and *An. arabiensis* (B to A; clockwise; green); the next four tracks depict the DE-lncRNAs (upregulated in red and downregulated in green) in the 4 comparisons (from innermost to outer: male vs. female carcass, male vs. female reproductive tissue, female reproductive tissue vs. female carcass, and male reproductive tissue vs. male carcass); and the innermost links track depicts the conservation of lncRNAs between the two species. For ideograms, only contigs longer than 300 kbp are represented. For conserved lncRNAs, only the topmost orthologues with sequence-level conservation (i.e., locusID and exonID scores greater than 0) are depicted. In the innermost track, the colored links connect *An. arabiensis* lncRNAs with the corresponding location of their ortholog on *An. albimanus* chromosome (red: 2R, blue: 2L, orange: 3R, green: 3L, yellow: X)

expression of genes across various physiological conditions of the vector [14]. In this study, we used publicly available datasets to discover novel lncRNA transcripts in 3 anopheline species, their expression under different conditions (sex: male vs. female; tissue: reproductive tissue vs. carcass), and their conservation across these species. A comparison of lncRNA expression between male and female mosquitoes can provide useful insights into how these transcripts control the biting behaviour of female mosquitoes and affect their reproductive

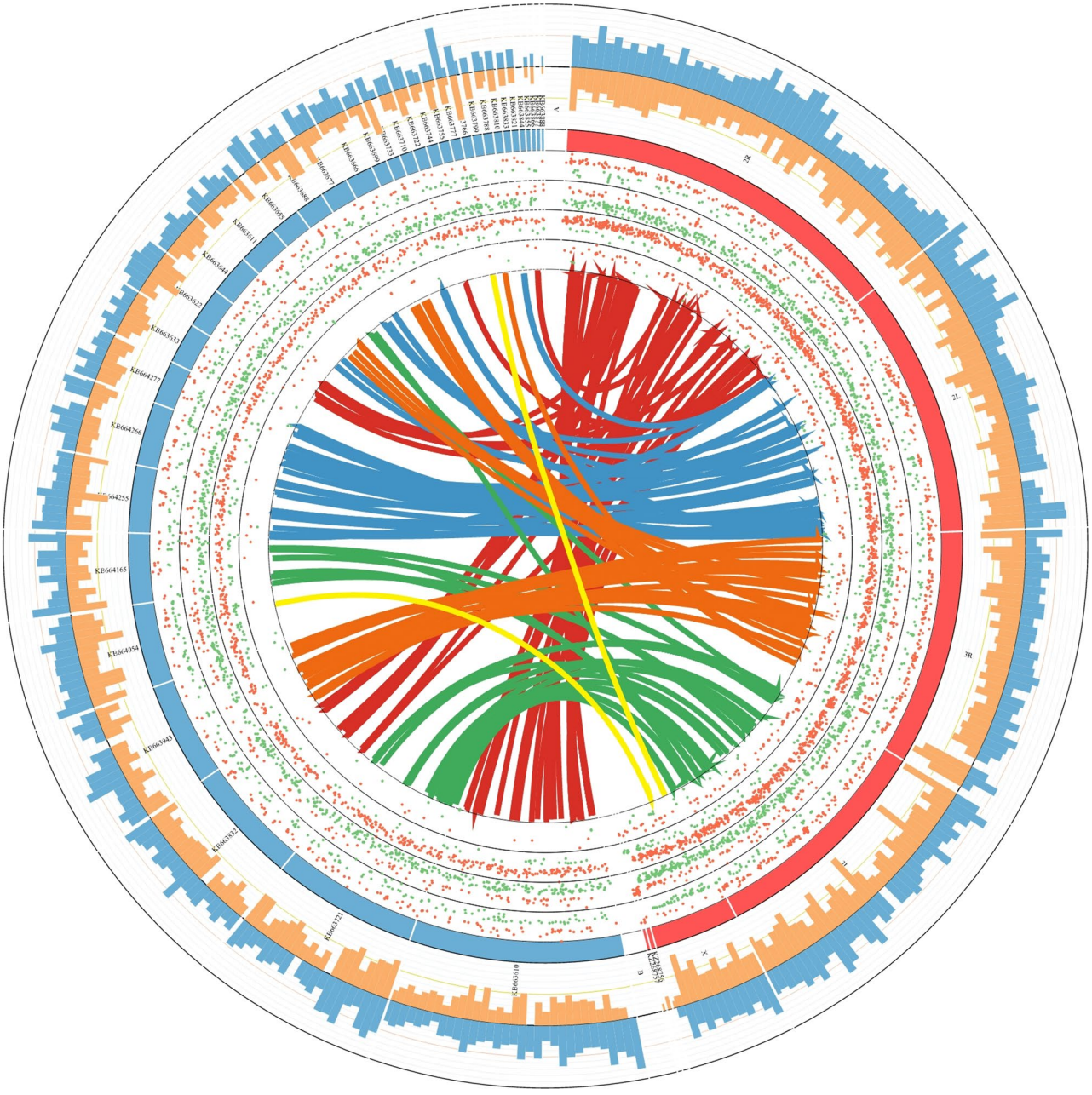


Fig. 8 Circos plot representing lncRNA conservation between *An. minimus* and *An. albimanus*. From periphery to centre, the first track (blue) represents the distribution of protein-coding genes (reference annotation) across the genomes of these 2 species; the second track (orange; inverted) represents the distribution of *de novo* assembled lncRNA transcripts across the genomes; the third track represents the ideograms of *An. albimanus* (A to B; clockwise; salmon) and *An. minimus* (B to A; clockwise; blue); the next four tracks depict the DE-lncRNAs (upregulated in red and downregulated in green) in the 4 comparisons (from innermost to outer: male vs. female carcass, male vs. female reproductive tissue, female reproductive tissue vs. female carcass, and male reproductive tissue vs. male carcass); and the innermost links track depicts the conservation of lncRNAs between the two species. For ideograms, only contigs longer than 300 kbp are represented. For conserved lncRNAs, only the topmost orthologues with sequence-level conservation (i.e., locusID and exonID scores greater than 0) are depicted. In the innermost track, the colored links connect *An. minimus* lncRNAs with the corresponding location of their ortholog on *An. albimanus* chromosome (red: 2R, blue: 2L, orange: 3R, green: 3L, yellow: X)

behaviour and, consequently, their disease-carrying capacity. However, this study was limited to only cataloguing lncRNAs; therefore, how the differential expression (or absence) and conservation of these lncRNAs affect cellular biology is beyond the scope of this manuscript. Moreover, we did not investigate the conserved lncRNAs that were not differentially expressed, and their presence cannot be ruled out.

In previous studies in *An. gambiae*, lncRNAs have been reported to range in length from 200 to 7000 nts in

Table 4 List of significantly enriched driver terms identified by GO analysis of neighbouring protein-coding genes

Species	Source	Term_name	Term_ID	Adjusted *p*-value
Anopheles albimanus	GO: MF	nucleic acid binding	GO:0003676	2.23E-27
	GO: MF	catalytic activity, acting on a nucleic acid	GO:0140640	2.23E-10
	GO: MF	structural constituent of nuclear pore	GO:0017056	6.55E-04
	GO: MF	isomerase activity	GO:0016853	2.51E-02
	GO: MF	ATP-dependent chromatin remodeler activity	GO:0140658	4.19E-02
	GO: MF	pseudouridine synthase activity	GO:0009982	4.63E-02
	GO: MF	methyltransferase activity	GO:0008168	4.78E-02
	GO: BP	nucleic acid metabolic process	GO:0090304	5.37E-30
	GO: BP	cell cycle	GO:0007049	2.04E-09
	GO: BP	RNA localization	GO:0006403	9.86E-07
	GO: BP	nucleocytoplasmic transport	GO:0006913	3.21E-06
	GO: BP	cell division	GO:0051301	3.59E-05
	GO: BP	chromatin remodeling	GO:0006338	3.67E-05
	GO: BP	deoxyribonucleotide metabolic process	GO:0009262	1.29E-03
	GO: BP	protein-RNA complex organization	GO:0071826	2.39E-02
	GO: CC	nucleus	GO:0005634	1.36E-48
Anopheles arabiensis	GO: MF	organic cyclic compound binding	GO:0097159	2.95E-09
	GO: MF	catalytic activity, acting on a nucleic acid	GO:0140640	1.78E-04
	GO: MF	aminoacyl-tRNA ligase activity	GO:0004812	3.98E-02
	GO: BP	ribonucleoprotein complex biogenesis	GO:0022613	9.01E-07
	GO: BP	cellular response to stress	GO:0033554	2.83E-03
	GO: BP	cell division	GO:0051301	2.20E-02
	GO: CC	intracellular anatomical structure	GO:0005622	1.07E-14
Anopheles minimus	GO: MF	catalytic activity, acting on a nucleic acid	GO:0140640	1.20E-11
	GO: MF	nucleic acid binding	GO:0003676	8.48E-10
	GO: MF	unfolded protein binding	GO:0051082	8.54E-04
	GO: MF	structural constituent of nuclear pore	GO:0017056	1.71E-02
	GO: BP	nucleic acid metabolic process	GO:0090304	1.00E-14
	GO: CC	intracellular anatomical structure	GO:0005622	1.48E-18
	GO: CC	preribosome	GO:0030684	5.00E-02

mosquito haemocytes [15], with a greater proportion of 'i' type lncRNAs than 'x' type lncRNAs [43]. Similar studies in *Ae. aegypti* have reported long intergenic noncoding RNA (lincRNA) lengths ranging from 200 to 9000 nts [44]. The lncRNAs identified in our study ranged in length from 200 nts to ~23,000 nts, with most of them having one or two exons. This finding aligns with other studies reporting that lncRNAs tend to be shorter in length and have fewer exons than mRNAs do [4, 15, 43–45]. We also found that most of the novel lncRNAs belong to one of 3 classes: intergenic ('u') (53–56%), exonic ('x') (33–41%), or entirely intronic ('i') (6–11%). Only a small proportion (less than 1%) of the lncRNAs identified in our analysis were identical to or had exonic overlaps with known reference genes for the reasons mentioned earlier.

LncRNAs vary in their nucleotide sequence more than protein-coding genes do. However, functional orthology between lncRNAs termed as syntenic homologs can exist at the structural level in a sequence-independent manner [3, 30, 37, 45–47]. Comparative analyses of insect lincRNAs have revealed similarly low levels of sequence conservation across species [48]. In each of the two comparisons in this study, more than 80% of the conserved transcripts showed syntenic homology. The conservation of lncRNAs at the structural level has previously been reported within the Gambiae complex [43]. These lncRNAs, usually enriched in *cis*-regulatory repressor functions and located in the proximity of protein-coding genes [49], could play a role in regulating similar or identical genes and/or pathways across the three species.

We found that the male and female carcasses of the three species presented similar lncRNA expression patterns, with lower levels of differential expression in the carcasses than in the reproductive tissues. Previous studies comparing lncRNA expression in multiple sex-specific developmental stages of *D. melanogaster* and *D. pseudoobscura* have yielded similar results [50]. In the original paper (for the datasets used in this study), Papa et al. reported a greater correlation for gene (mRNA) expression between the carcasses than between the reproductive tissues [29].

Sexually-dimorphic expression of lncRNAs has been previously reported in various species of plants as well as in animals [51]. In animals, lncRNAs have been shown to regulate sex determination, sex chromosome dosage compensation, gonadogenesis, sex hormone responses, gametogenesis, etc [50–53]. In plants, lncRNAs regulate floral transition, pollen and anther development, photoperiodism, seed development, and male fertility [51, 54, 55]. In our study, sixty-nine lncRNAs presented some degree of sequence-level conservation and were differentially expressed in at least one comparison group. GO analysis of the genes in the vicinity of these lncRNAs

revealed that the neighbouring protein-coding genes are involved in pathways that regulate nucleic acid and protein metabolism, chromatin remodelling, nuclear structure, and cell cycle. Since most of these lncRNAs were upregulated in male reproductive tissue as well as in carcasses, they potentially negatively regulate the expression of genes involved in female-specific behavioural patterns and biological functions, thus governing sexual dimorphism.

Conclusion

This study provides the first comprehensive catalogue of lncRNA expression and conservation in anophelines, revealing sex- and species-specific expression patterns. We found that, compared with carcass tissues, male and female reproductive tissues presented higher variation in lncRNA expression. While the number of conserved lncRNAs was high across species, only a small number of lncRNAs presented sequence-level conservation most of which were upregulated only in male tissues. The present study is limited by the fact that the authors did not study how the differences in lncRNA expression could impact biological functions. Future validation of differential lncRNA expression and functional assays, co-expression analysis, and structural modelling of conserved lncRNAs can help answer important questions in mosquito biology. Further areas of interest in this direction include studying lncRNAs in infected and noninfected *Anopheles* mosquitoes to decipher how parasites affect lncRNA expression, as has been documented in other mosquito species [9, 44]. Such studies could unravel the dynamics of disease transmission as well as provide targets for interrupting parasite-mosquito interactions. There is also potential for species- and sex-specific lncRNAs to be used as tools for vector control through transgenesis [56]. Exploring the role of lncRNAs in insecticide resistance, such as identifying specific pathways or resistance-associated genes regulated by lncRNAs could be another area of interest from the vector control perspective. Nevertheless, this genome-wide discovery and orthology analysis of long non-coding RNA expression across three *Anopheles* species provides a foundational framework for understanding the evolution, conservation, and functional potential of lncRNAs, paving the way for future studies on their roles in mosquito biology, vector control, and disease transmission.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11687-7.

Supplementary Material 1: S1: Annotation file (in.gtf format) of lncRNAs assembled in ***An. albimanus***.

Supplementary Material 2: S2: Annotation file (in.gtf format) of lncRNAs assembled in ***An. arabiensis***.

Supplementary Material 3: S3: Annotation file (in.gtf format) of lncRNAs assembled in ***An. minimus***. Please note that MSTRG IDs in any one file have no relation with the MSTRG IDs in the other two files.

Supplementary Material 4: S4: List of sequence-based conserved orthologous lncRNA triads.

Supplementary Material 5: S5: FASTA file containing the nucleotide sequence of lncRNAs showing sequence-based conservation in ***An. albimanus***.

Supplementary Material 6: S6: FASTA file containing the nucleotide sequence of lncRNAs showing sequence-based conservation in ***An. arabiensis***.

Supplementary Material 7: S7: FASTA file containing the nucleotide sequence of lncRNAs showing sequence-based conservation in ***An. minimus***.

Supplementary Material 8: S8: Specieswise heatmaps showing the top 15 upregulated and top 15 downregulated lncRNA genes in each comparison. The colors denote the z-score (row-normalised expression value) for each gene in various samples.

Supplementary Material 9: S9: List of enriched Gene Ontology Molecular Function (GO:MF), Biological Processes (GO:BP), and Cellular Component (GO:CC) terms identified for neighbouring genes in each species.

Acknowledgements
We thank the Director NIMR for providing the facilities required to carry out the work.

Author contributions
W.A.M., A.R.A., and P.K.B. conceptualized the study. P.K.B. and W.A.M. developed the methodology. W.A.M. and K.S. conducted the formal analysis. N.A. and W.A.M. created the visualizations. W.A.M. and P.K.B. prepared the original draft of the manuscript. K.S., N.A., C.B., N.C.D., P.M., and N.S. reviewed and edited the manuscript. A.R.A. and P.K.B. supervised the research. All the authors have read and approved the final manuscript.

Funding
None.

Data availability
Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]ICMR-National Institute of Malaria Research, Field Unit Guwahati, Guwahati, Assam 781036, India
[2]ICMR-National Institute of Malaria Research, New Delhi 110077, India
[3]Faculty of Biological Sciences, Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh 201002, India
[4]Faculty of Veterinary Sciences and Animal Husbandry, SKUAST-K, Srinagar, Shuhama, Jammu and Kashmir 190006, India

Received: 20 January 2025 / Accepted: 9 May 2025
Published online: 20 May 2025

References

1. Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22(2):96–118.
2. Jiang N, Zhang X, Gu X, Li X, Shang L. Progress in Understanding the role of LncRNA in programmed cell death. Cell Death Discovery. 2021;7(1):30.
3. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen L-L, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol. 2023;24(6):430–47.
4. Pandey A, Malla WA, Sahu AR, Wani SA, Khan RIN, Saxena S, et al. Differential expression of long non-coding RNAs under peste des petits ruminants virus (PPRV) infection in goats. Virulence. 2022;13(1):310–22.
5. Tanuj GN, Khan O, Malla WA, Rajak KK, Chandrashekar S, Kumar A, et al. Integrated analysis of long-noncoding RNA and circular RNA expression in Peste-des-Petits-Ruminants virus (PPRV) infected marmoset B lymphocyte (B95a) cells. Microb Pathog. 2022;170:105702.
6. Huang J, Zhou W, Zhang X, Li Y. Roles of long non-coding RNAs in plant immunity. PLoS Pathog. 2023;19(5):e1011340.
7. Fandino RA, Brady NK, Chatterjee M, McDonald JMC, Livraghi L, van der Burg KRL, et al. The Ivory LncRNA regulates seasonal color patterns in Buckeye butterflies. Proc Natl Acad Sci U S A. 2024;121(41):e2403426121.
8. Li X, Yang X, You F, Miao C, Li M, Wang K, et al. Differences between uncapping and removal behaviors in Apis cerana from the perspective of long non-coding RNAs. BMC Genomics. 2024;25(1):912.
9. Belavilas-Trovas A, Tastsoglou S, Dong S, Kefi M, Tavadia M, Mathiopoulos KD, et al. Long non-coding RNAs regulate Aedes aegypti vector competence for Zika virus and reproduction. PLoS Pathog. 2023;19(6):e1011440.
10. Mao W, Zeng Q, She L, Yuan H, Luo Y, Wang R et al. Wolbachia utilizes LncRNAs to activate the Anti-Dengue toll pathway and balance reactive oxygen species stress in Aedes aegypti through a competitive endogenous RNA network. Front Cell Infect Microbiol. 2022;11.
11. Lawrie RD, Mitchell RD 3rd, Deguenon JM, Ponnusamy L, Reisig D, Pozo-Valdivia AD et al. Characterization of long Non-Coding RNAs in the bollworm, Helicoverpa Zea, and their possible role in Cry1Ac-Resistance. Insects. 2021;13(1).
12. Li S, Hussain F, Unnithan GC, Dong S, UlAbdin Z, Gu S, et al. A long non-coding RNA regulates Cadherin transcription and susceptibility to Bt toxin Cry1Ac in Pink bollworm, pectinophora Gossypiella. Pestic Biochem Physiol. 2019;158:54–60.
13. Zafar J, Huang J, Xu X, Jin F. Recent advances and future potential of long Non-Coding RNAs in insects. Int J Mol Sci. 2023;24(3).
14. Legeai F, Derrien T. Identification of long non-coding RNAs in insects genomes. Curr Opin Insect Sci. 2015;7:37–44.
15. Saha B, McNinch CM, Lu S, Ho MCW, De Carvalho SS, Barillas-Mury C. In-depth transcriptomic analysis of Anopheles gambiae hemocytes uncovers novel genes and the oenocytoid developmental lineage. BMC Genomics. 2024;25(1):80.
16. Lukyanchikova V, Nuriddinov M, Belokopytova P, Taskina A, Liang J, Reijnders MJMF, et al. Anopheles mosquitoes reveal new principles of 3D genome organization in insects. Nat Commun. 2022;13(1):1960.
17. Nguyen VT, Collier TC, Seok S, Wang X, Mburu MM, Simubali L, et al. The first genome sequence of Anopheles squamous from Macha. Zambia F1000Res. 2023;12:330.
18. Ayala D, Akone-Ella O, Kengne P, Johnson H, Heaton H, Collins J, et al. The genome sequence of the malaria mosquito, Anopheles funestus, Giles, 1900. Wellcome Open Res. 2022;7:287.
19. Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al. A global map of dominant malaria vectors. Parasit Vectors. 2012;5:69.
20. Belavilas-Trovas A, Gregoriou M-E, Tastsoglou S, Soukia O, Giakountis A, Mathiopoulos K. A species-specific LncRNA modulates the reproductive ability of the Asian tiger mosquito. Front Bioeng Biotechnol. 2022;10.
21. Azlan A, Obeidat SM, Yunus MA, Azzam G. Systematic identification and characterization of Aedes aegypti long noncoding RNAs (lncRNAs). Sci Rep. 2019;9(1):12147.
22. Liu W, Cheng P, Zhang K, Gong M, Zhang Z, Zhang R. Systematic identification and characterization of long noncoding RNAs (lncRNAs) during Aedes albopictus development. PLoS Negl Trop Dis. 2022;16(4):e0010245.
23. Xu J, Hu K, Riehle MM, Khadka V. Transcriptome-based identification of long noncoding RNAs (lncRNAs) across the genome of Anopheles gambiae. bioRxiv. 2024:2024.10.02.616138.
24. Terradas G, Novelo M, Metz H, Brustolin M, Rasgon JL. Anopheles albimanus is a potential alphavirus vector in the Americas. Am J Trop Med Hyg. 2023;108(2):412–23.
25. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, et al. The dominant Anopheles vectors of human malaria in Africa, Europe and the middle East: occurrence data, distribution maps and bionomic précis. Parasit Vectors. 2010;3:117.
26. Dev V, Manguin S. Biology, distribution and control of Anopheles (Cellia) minimus in the context of malaria transmission in Northeastern India. Parasites Vectors. 2016;9(1):585.
27. Martynova T, Kamanda P, Sim C. Transcriptome profiling reveals sex-specific gene expressions in pupal and adult stages of the mosquito Culex pipiens. Insect Mol Biol. 2022;31(1):24–32.
28. Yang X, Schadt EE, Wang S, Wang H, Arnold AP, Ingram-Drake L, et al. Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res. 2006;16(8):995–1004.
29. Papa F, Windbichler N, Waterhouse RM, Cagnetti A, D'Amato R, Persampieri T, et al. Rapid evolution of female-biased genes among four species of Anopheles malaria mosquitoes. Genome Res. 2017;27(9):1536–48.
30. Bryzghalov O, Makałowska I, Szcześniak MW. LncEvo: automated identification and conservation study of long noncoding RNAs. BMC Bioinformatics. 2021;22(1):59.
31. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
32. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.
33. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, stringtie and ballgown. Nat Protoc. 2016;11(9):1650–67.
34. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417–9.
35. Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. 2017;45(W1):W12–6.
36. Frith MC, Wan R, Horton P. Incorporating sequence quality data into alignment improves DNA read mapping. Nucleic Acids Res. 2010;38(7):e100–e.
37. Chen J, Shishkin AA, Zhu X, Kadri S, Maza I, Guttman M, et al. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. Genome Biol. 2016;17:19.
38. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009.
39. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
40. Kang C, Liu Z. An Easy-to-Follow pipeline for long noncoding RNA identification: A case study in diploid strawberry Fragaria vesca. Methods Mol Biol. 2019;1933:223–43.
41. Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. 2023;51(W1):W207–12.
42. Marinotti O, Cerqueira GC, de Almeida LGP, Ferro MIT, Loreto ELS, Zaha A, et al. The genome of Anopheles Darlingi, the main Neotropical malaria vector. Nucleic Acids Res. 2013;41(15):7387–400.
43. Jenkins AM, Waterhouse RM, Muskavitch MA. Long non-coding RNA discovery across the genus anopheles reveals conserved secondary structures within and beyond the Gambiae complex. BMC Genomics. 2015;16(1):337.
44. Etebari K, Asad S, Zhang G, Asgari S. Identification of Aedes aegypti long intergenic Non-coding RNAs and their association with Wolbachia and dengue virus infection. PLoS Negl Trop Dis. 2016;10(10):e0005069.
45. Caixia G, Xiuwen Z, Hubo L, Mlekwa UA, Yu G, Jie X. Roles of LncRNAs in rice: advances and challenges. Rice Sci. 2020;27(5):384–95.
46. Ramírez-Colmenero A, Oktaba K, Fernandez-Valverde SL. Evolution of Genome-Organizing long Non-coding RNAs in metazoans. Front Genet. 2020;11.
47. Tsagakis I, Douka K, Birds I, Aspden JL. Long non-coding RNAs in development and disease: conservation to mechanisms. J Pathol. 2020;250(5):480–95.
48. Lopez-Ezquerra A, Harrison MC, Bornberg-Bauer E. Comparative analysis of LincRNA in insect species. BMC Evol Biol. 2017;17(1):155.
49. Ranjan G, Scaria V, Sivasubbu S. Syntenic LncRNA locus exhibits DNA regulatory functions with sequence evolution. Gene. 2025;933:148988.
50. Nyberg KG, Machado CA. Comparative expression dynamics of intergenic long noncoding RNAs in the genus Drosophila. Genome Biol Evol. 2016;8(6):1839–58.

51. Golicz AA, Bhalla PL, Singh MB. LncRNAs in plant and animal sexual reproduction. Trends Plant Sci. 2018;23(3):195–205.
52. Taylor DH, Chu ET, Spektor R, Soloway PD. Long non-coding RNA regulation of reproduction and development. Mol Reprod Dev. 2015;82(12):932–56.
53. Wang K-X, Chen C-B, Wan Q-X, Zha X-F. Long Non-Coding RNA Bmdsx-AS1 effects on male external genital development in silkworm. Insects. 2022;13(2):188.
54. Babaei S, Singh MB, Bhalla PL. Role of long non-coding RNAs in rice reproductive development. Front Plant Sci. 2022;13.
55. Johnson C, Conrad LJ, Patel R, Anderson S, Li C, Pereira A, et al. Reproductive long intergenic noncoding RNAs exhibit male gamete specificity and polycomb repressive complex 2-Mediated repression. Plant Physiol. 2018;177(3):1198–217.
56. Dong S, Dong Y, Simões ML, Dimopoulos G. Mosquito transgenesis for malaria control. Trends Parasitol. 2022;38(1):54–66.

Publisher's note