RNA BIOLOGY
2025, VOL. 22, NO. 1, 1–16
https://doi.org/10.1080/15476286.2025.2463812

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Analysis of somatic piRNAs in the malaria mosquito *Anopheles coluzzii* reveals atypical classes of genic small RNAs

Sergei Funikov[a], Alexander Rezvykh[a], Natalia Akulenko[b#], Jiangtao Liang[c], Igor V. Sharakhov [c,d,e], and Alla Kalmykova[f]

[a]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia; [b]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia; [c]Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; [d]The Center for Emerging, Zoonotic, and Arthropod-Borne Pathogens, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; [e]Department of Genetics and Cell Biology, Tomsk State University, Tomsk, Russia; [f]Koltzov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia

ABSTRACT

Piwi-interacting small RNAs (piRNA) play a key role in controlling the activity of transposable elements (TEs) in the animal germline. In diverse arthropod species, including the pathogen vectors mosquitoes, the piRNA pathway is also active in nongonadal somatic tissues, where its targets and functions are less clear. Here, we studied the features of small RNA production in head and thorax tissues of an uninfected laboratory strain of *Anopheles coluzzii* focusing on the 24–32-nt-long RNAs. Small RNAs derived from repetitive elements constitute a minor fraction while most small RNAs process from long noncoding RNAs (lncRNAs) and protein-coding gene mRNAs. The majority of small RNAs derived from repetitive elements and lncRNAs exhibited typical piRNAs features. By contrast, majority of protein-coding gene-derived 24–32 nt small RNAs lack the hallmarks of piRNAs and have signatures of nontemplated 3' end tailing. Most of the atypical small RNAs exhibit female-biased expression and originate from mitochondrial and nuclear genes involved in energy metabolism. We also identified atypical genic small RNAs in *Anopheles gambiae* somatic tissues, which further validates the noncanonical mechanism of their production. We discuss a novel mechanism of small RNA production in mosquito somatic tissues and the possible functional significance of genic small RNAs.

ARTICLE HISTORY
Received 25 June 2024
Revised 28 January 2025
Accepted 3 February 2025

KEYWORDS
Anopheles coluzzii; atypical small RNAs; somatic piRNA; transposable elements; mitochondrial small RNAs; piRNA processing

Introduction

Small RNAs (18–32 nucleotides in length) are the primary players in RNA interference (RNAi) processes, which are based on the complementary recognition of RNA targets by a complex of small RNAs and Argonaute proteins. Small RNAs represent the basis of the diverse defence pathways directed against viruses and transposable elements (TEs). Small interfering RNAs (siRNAs) 21-nt in length are generated from RNA duplexes of exogenous and endogenous origins. Piwi-interacting RNAs (piRNAs), a different class of small RNAs that are 24–32 nucleotides long, are primarily responsible for repressing TEs, the endogenous counterpart of viruses, in the gonads of multicellular animals [1,2]. piRNAs in a complex with proteins of the PIWI subfamily of the Argonaute family suppress the activity of their targets at different levels, which leads to efficient and inherited epigenetic TE silencing. Unlike siRNAs generated from double-stranded RNA, piRNAs are processed from single-stranded precursors of endogenous origin. Dedicated sites in the genome, called piRNA clusters, specialize in piRNA production [3]. Transcripts generated from these regions are recognized by the piRNA processing system. piRNA clusters are represented by extended pericentromeric regions enriched in degraded TEs [3], full-length copies of active euchromatic TEs [4], telomeric regions [5], and unique regions of the genome, including both genes and intergenic regions [6,7]. Processing of piRNA precursors occurs in the perinuclear structure of 'nuage' and on the outer mitochondrial membrane [8,9]. piRNAs are formed as a result of two nucleolytic mechanisms: ping-pong (slicing) and subsequent processing of the RNA cleavage products (phasing). The coordinated action of two proteins of the Piwi subfamily – Aubergine and Ago3 — involving transcripts of piRNA clusters and TE-derived mRNA leads to the multiplication of sense and antisense piRNAs with 10 nt overlap, in the process called the ping-pong cycle [3,10]. Cleavage products are a substrate for Zucchini endonuclease, which produces piRNAs during sequential cleavages from the 5' to 3' end of RNA (phasing)

CONTACT Alla Kalmykova allakalm@idbras.ru Koltzov Institute of Developmental Biology, Russian Academy of Sciences, Moscow 119334, Russia; Igor V. Sharakhov igor@vt.edu Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
#Present address: National Research Centre 'Kurchatov Institute', Moscow 123182, Russia.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/15476286.2025.2463812.

[11,12]. piRNAs have specific nucleotide bias as a consequence of the enzymatic activity and target preference of the ribonucleases involved in the piRNA biogenesis in *Drosophila*. The 1U bias (preference for the uridine at the 5' end) is related to the activity of endonuclease Zucchini localized on mitochondria, which cleaves long piRNA precursors upstream of U [12]. 1U at the 5' end is characteristic of antisense piRNAs bound to the Piwi proteins (Piwi and Aubergine). 10A, i.e. a preference for the adenine in position 10 of piRNA, is a signature of the ping-pong processing since 5' uridine of piRNA determines the position of the tenth nucleotide of the complementary piRNA. Exonucleolytic trimming of the 3' end of long piRNA intermediates is essential for piRNA maturation; this function is performed by exoribonuclease Nibbler [13].

piRNAs are produced not only from TE transcripts, but also from a number of protein-coding gene mRNAs, which suggests the potential role of the piRNA-mediated pathway in the biogenesis of protein-coding mRNAs. Genic piRNAs are predominantly generated from the 3' untranslated regions (UTRs) of the protein-coding gene mRNAs in different species [6,14]. Generation of genic piRNAs can be a result of the presence of TE fragments in the 3'UTRs which are recognized by endogenous TE-derived piRNAs, leading to mRNA cleavage and processing of genic piRNAs [4,15]. In ovaries of the malaria mosquito *Anopheles gambiae*, endogenous piRNAs derived from ancient genomic repeats may target mRNAs leading to the production of genic piRNAs that in turn may direct piRNA production from other transcripts [16]. The functional significance of this cascade of concerted piRNA-mediated slicing of RNAs remains unclear. In mice and chickens, piRNA processing of mRNA comprising TE-fragments in the 3' UTR is coupled with translation and affects protein production [17] that highlights the functional significance of the piRNA-mediated pathway in the regulation of mRNA expression.

The piRNA pathway functions not only in gonads but also in somatic tissues of many arthropod species [18]. Of particular interest is the fact that the somatic piRNA pathway is active in the species of the mosquito genera *Anopheles*, *Culex* and *Aedes* [19–23] . Mosquitoes are one of the most societally relevant insects due to their role in disease transmission and their large and continually expanding habitats. In the genomes of various mosquito species, 3–8 proteins of the PIWI subfamily are encoded [24], some of which are expressed in somatic tissues [25–27]. piRNAs play a role in both the modulation of TE activity and the antiviral response in somatic tissues [18,28,29]. The latter is mediated by viral genome fragments, endogenous viral elements (EVEs), which were discovered in TE-enriched piRNA clusters [22].

A distinct feature of the piRNA composition in mosquitoes of the *Anopheles* genus is the presence of a larger number of piRNAs derived from protein-coding genes compared with *Drosophila* [30]. In the ovaries of *An. gambiae*, 11% of all piRNAs correspond to gene regions, mainly to the 3'UTRs. One of these genes (AGAP003387) produces 8% of all piRNAs [16,30]. Genic piRNAs are also highly represented in the piRNA libraries prepared from whole mosquitoes [19] and from isolated somatic tissues (abdomen, midgut, carcasses) of *An. gambiae* [20]. Analysis of transposon and viral small RNAs in mosquito cell cultures and somatic tissues of four mosquito species revealed that most of them showed patterns of both ping-pong and phasing piRNA biogenesis mechanisms [21]. Growing evidence suggests that the somatic piRNA pathway in mosquito vectors evolved as a response to a persistent viral infection [22,31]. The endogenous targets of the piRNA machinery and the functional importance of piRNA synthesis from cellular gene transcripts in somatic tissues, however, remain largely unknown.

In this study, we analysed small RNAs produced in the head and thorax of males and females of uninfected laboratory strain of the African malaria mosquito *An. coluzzii* and described strong female bias in the production of protein-coding gene-derived piRNAs. We discovered that siRNAs and piRNAs are derived from a variety of transcripts in somatic tissues, including TEs, lncRNAs, and protein-coding genes. Despite the diversity of transcripts that generate small RNAs, piRNAs account for only a small percentage of all small RNAs found in *An. coluzzii* somatic tissues. We found atypical 24–32-nt-long small RNAs processed from protein-coding gene transcripts that are lacking signatures of piRNAs. Gene ontology (GO) analysis demonstrates that atypical genic small RNAs are mostly produced from transcripts of the mitochondrial genome and mRNAs of nuclear genes involved in energy metabolism. Moreover, these atypical small RNAs are subjected to nontemplated nucleotide addition at their 3' end. We discuss the possible mechanisms that determine the generation of these novel types of small RNAs and their potential functional roles in the mosquito somatic tissues.

Results

Characterization of small RNAs from male and female head and thorax of An. coluzzii

To characterize the genomic origin of small RNAs in mosquito somatic tissues, we prepared small RNA libraries, each in three biological replicates, from the heads and thoraxes of males and females of the MOPTI *An. coluzzii* strain. The performed correlation and principal component analyses demonstrate a high degree of similarity between the biological replicates of the studied samples (Supplementary Figure S1).

Pre-processed small RNA reads filtered from all ribosomal RNAs (rRNAs), small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), transfer RNAs (tRNAs) were mapped to the *An. coluzzii* MOPTI genome (VectorBase rel. 57) [32]. For the preliminary analysis of small RNAs, we counted all reads that mapped on the *An. coluzzii* genome with up to 3 mismatches.

miRNAs were always the major fraction of the small RNA libraries, representing ~ 94–97% of somatic small RNAs (Figure 1A). Approximately 2.5–4% of somatic small RNAs were derived from repeats, exons of protein-coding genes and lncRNAs (Figure 1A). The remaining ~ 1–2% of small RNA reads mapped to unannotated intergenic regions of the genome and introns of genes (Figure 1A). Except for miRNAs, lncRNA-derived small RNAs were most abundant whereas repeat-derived small RNAs represented a minor fraction of

A The percentage of genome mappers

	microRNAs	Repeats	lncRNAs	Protein-coding genes	Others
Female head	94.6	0.4	2.3	1.3	1.3
Female thorax	94.5	0.6	2.1	1.3	2.3
Male head	96.6	0.3	1.8	0.5	0.9
Male thorax	95.2	0.5	1.7	0.9	1.7

B Length profiles for small RNAs

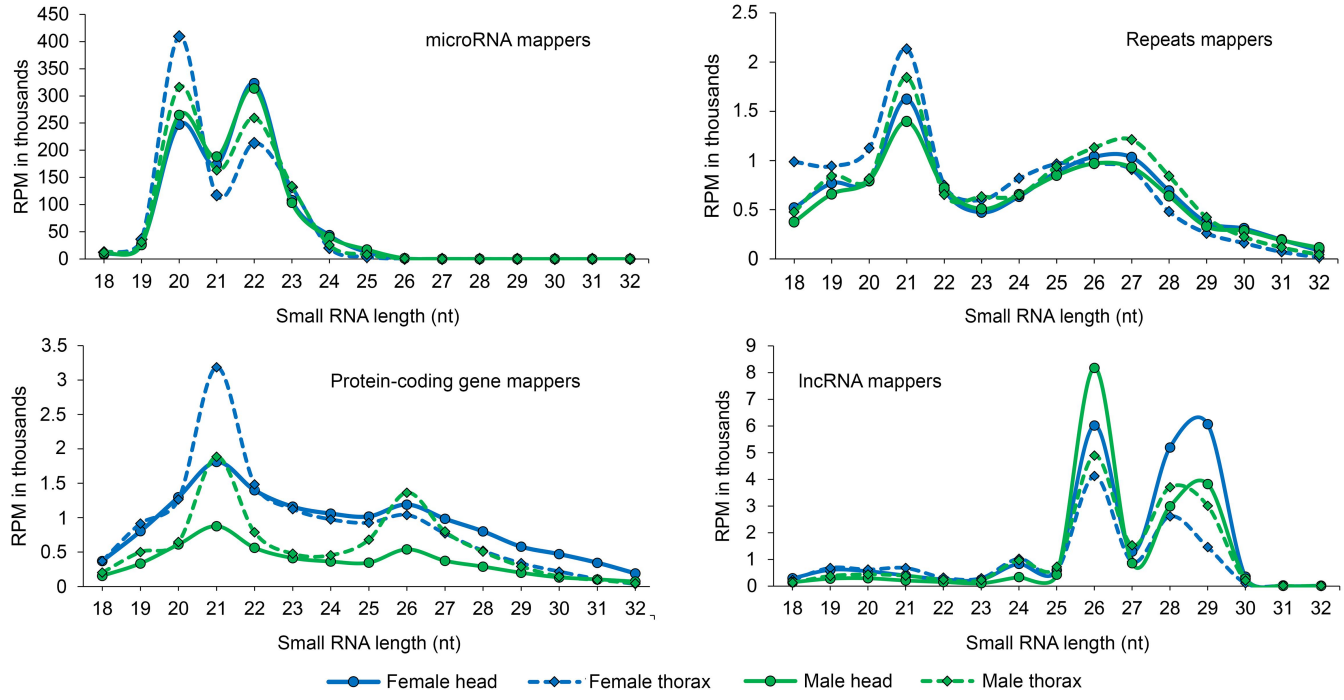


Figure 1. Annotation of small RNAs from somatic tissues of *An. coluzzii*. (A) Percentage of mapped small RNA reads in heads and thoraxes of female and male mosquitos. miRNAs, exons of protein-coding genes, lncRNAs and repeat-derived reads that mapped to the *An. coluzzii* genome with up to 3 mismatches were considered. For lncRNA mapping, we used a combined list of annotated ncRNA sequences of *An. coluzzii* (excluding rRNA, snoRNA, snRNA, tRNA and miRNA) and lncRNAs described for *An. gambiae*. Only exon mappers were considered for protein-coding genes. Other mappers represent reads that mapped to intergenic unannotated regions and introns of genes. (B) Length distribution of small RNAs from mosquito male and female heads and thoraxes mapped to miRNAs, repeats, lncRNAs and protein-coding genes (exons). RPM - reads per million mapped genomic reads. The resulting value for sexes and tissues is calculated based on three biological replicates.

somatic small RNAs in contrast to ovarian tissues [30] (Figure 1A). Thus, small RNA pathways in mosquito somatic tissues did not primarily target TEs, which is likely due to the low expression levels and repression of TEs in these tissues. Analysis of length distribution revealed differences between analysed sequences with a more pronounced peak of presumably piRNAs, in length between 24–32 nts for lncRNAs (Figure 1B). The small RNAs with a length of 21 nt that belong to siRNAs were found to dominate among repeat-derived and protein-coding gene small RNA populations (Figure 1B). Intergenic small RNAs were predominantly 22 nt in length (Supplementary Figure S2). Despite the predominance of sense mappers in a pool of protein-coding gene-derived siRNAs, antisense siRNAs still constituted 19–29% of all protein-coding gene mappers (Supplementary Figure S3). It remains to be determined whether the dsRNA precursors of mosquito siRNAs are produced by antisense transcription or by the generation of intramolecular hairpin RNAs, as has been discovered in other species [33]. We also mapped the obtained libraries to known sequences of mosquito viruses but our effort failed to find any virus mappers in the somatic small RNA libraries.

Protein-coding gene-derived somatic small RNAs of An. coluzzii *lack the typical piRNA nucleotide bias*

To define the putative somatic piRNA population of *An. coluzzii*, we selected small RNAs with lengths ranging from 24 to 32 nts. From this step onwards, we considered reads that mapped uniquely on the *An. coluzzii* genome with up to 3 mismatches for analysis of protein-coding genes and lncRNAs. For the analysis of repeat-derived small RNAs, all reads that mapped to the consensus sequences of TEs and other genomic repeats with 0–3 mismatches were taken into account. As observed for total small RNAs, repeat-derived piRNAs constituted a minor fraction (2–3% in male and female tissues) while most piRNAs were processed from lncRNAs (from 50 to 70%) and protein-coding gene mRNAs

(12–23%) (Figure 2A). The remaining 16–24% of 24–32 nt small RNAs were produced by intergenic regions (Figure 2A). The majority of protein-coding genes and lncRNAs produced predominantly sense piRNAs (~98% on average of all somatic piRNAs in both sexes) (Figure 2A, Supplementary Table S1). In contrast, the majority of repeat-derived piRNAs were produced from the antisense strand (~80% of antisense piRNAs on average) (Figure 2A, Supplementary Table S1). The majority of piRNAs derived from lncRNAs and repeats exhibited a strong preference for uridine at the first position, 1 U bias, which is typical for ovarian piRNAs loaded into Piwi proteins [3]. However, for 24–32 nt small RNAs that originated from protein-coding genes, the number of 1 U-biased reads was significantly lower. Only ~ 30% of genic 24–32 nt small RNAs were found to have uridine at their 5' ends demonstrating that the most of these molecules have atypical for piRNAs nucleotide bias (Figure 2A).

The antisense piRNAs targeting repeats demonstrated a pattern typical for ovarian transposon-derived piRNAs with length distribution from 24 to 32 nts and a peak at 26

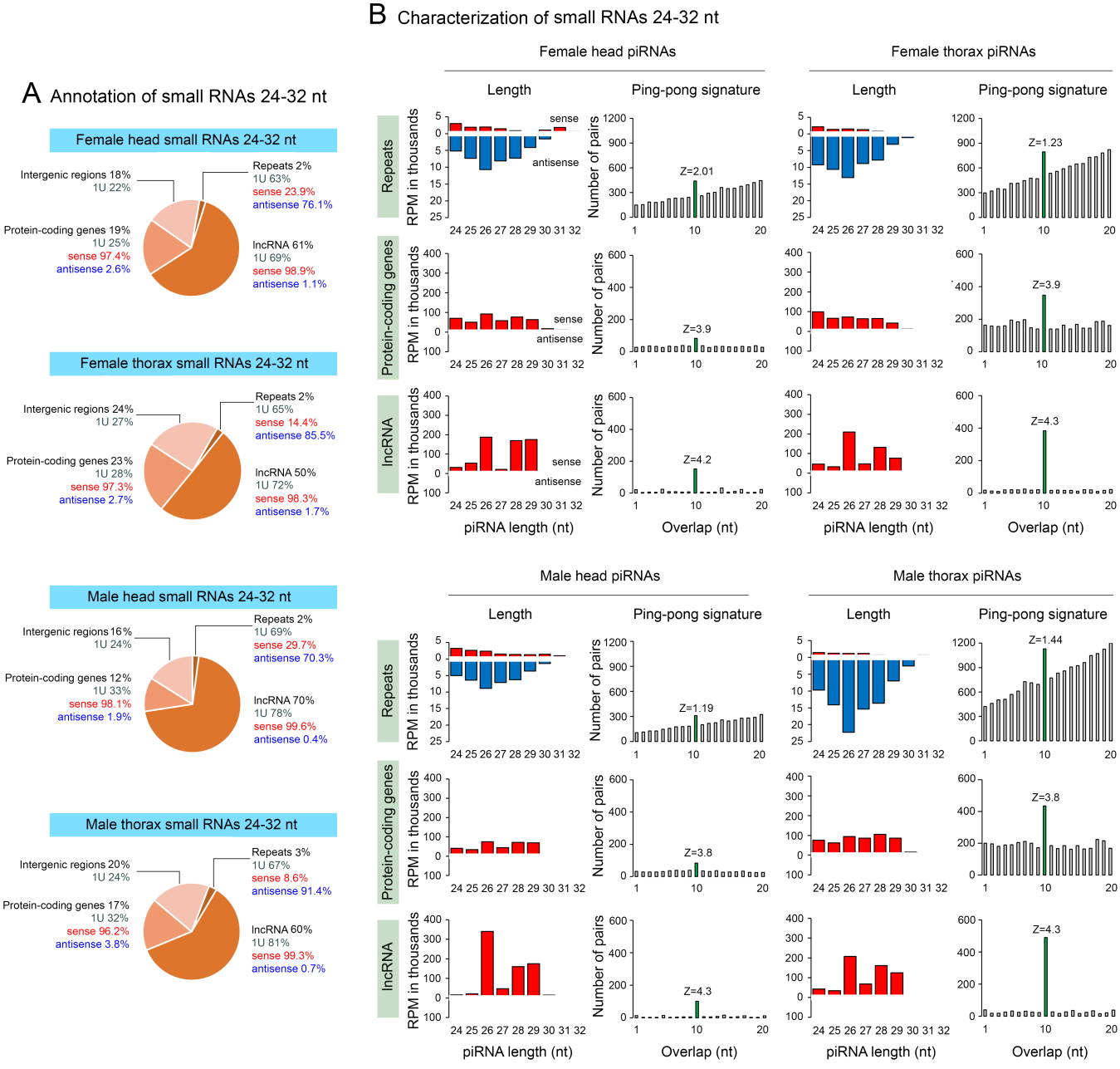


Figure 2. Analysis of small RNA fraction of 24–32 nt in length from somatic tissues of *An. coluzzii*. (A) Annotation of 24–32 nt small RNAs. 1U nucleotide bias, sense and antisense reads (percentage) for 24–32 nt small RNAs are shown for each gene category. (B) Characterization of small RNAs (24–32 nt) mapping to repeats, lncRNAs and protein-coding genes. The left panels are the size distribution of mapped small RNAs. Red and blue bars show small RNAs mapped in sense and antisense orientation, respectively. The right panels are the number of partially complementary piRNA pairs showing enrichment for 10 nt overlap between sense and antisense sRNAs (i.e. ping-pong signature). Ping-pong Z-scores are shown above the plots. For lncRNAs and transcripts of protein-coding genes only uniquely mapped small RNAs with up to 3 mismatches were considered. For repeats, all small RNAs mapped with up to 3 mismatches to known consensus sequences of TEs. Other repeats of *An. coluzzii* were taken into account. RPM - reads per million. Three biological replicates were used to calculate expression values for sexes and tissues.

nt (Figure 2B). The lncRNA sense piRNAs exhibited size distribution with peaks at 26 and 28–29 nt in all samples (Figure 2B). The length of protein-coding gene-derived small RNAs was randomly distributed from 24 to 30 nts (without an obvious peak) in mosquito somatic tissues (Figure 2B). The observed variations in the length distribution across the analysed sequences most likely point to distinct processes for the production of TE- and gene-specific piRNAs.

piRNAs derived from repeats, protein-coding genes and lncRNAs show a ping-pong signature, an enrichment of a 10-nt overlap between the 5' ends of complementary piRNA pairs (Figure 2B). Notably, the number of complementary pairs for repeat-derived piRNAs showed a gradual increase from 1-nt overlap to 20-nt overlap with a peak at 10 nt (Figure 2B). This pattern suggests that many complementary piRNA pairs are generated not only by a ping-pong amplification loop, like in the germline, but also by primary piRNA processing of both sense and antisense transcripts generated by bidirectional transcription of genomic loci occupied by repetitive elements.

Overall, 24–32 nt small RNAs in somatic tissues of *An. coluzzii* demonstrated piRNA-specific features in terms of length distribution and 1 U nucleotide bias for repeat and lncRNA-derived small RNAs. In *Drosophila* ovaries, 1 U bias is a signature of Zucchini endonuclease which cleaves the piRNA precursor upstream of a U residue generating both 3' and 5' ends of piRNAs [12]. Protein-coding gene derived 24–32 nt small RNAs, on the other hand, lack a preference for a 5' terminal uridine, implying a different mechanism of processing.

Features of the genomic loci that generate the majority of somatic piRNAs in An. coluzzii

For further analysis of the somatic piRNA of mosquitoes, we combined sequenced libraries from all tissues and both sexes to increase the number of reads mapped to the genome. The genomic sequences with at least 10 piRNA mapped were considered. As was previously described, most of the lncRNAs and repeat-derived piRNAs demonstrated pronounced 1 U nucleotide bias (Figure 2A). Our findings revealed that 1 U-biased piRNAs (with a 1 U percentage of greater than 50%, twice more than a random distribution of nucleotides) covered 73 repeats and 37 lncRNAs (Figure 3A). These piRNAs constituted 98.4% and 84.6% of all piRNAs mapped to lncRNAs and repeats, respectively (Figure 3A). In contrast, only 35.2% of all 24–32 nt small RNAs derived from protein-coding genes showed 1 U bias, while 64.8% of small RNAs derived from 3135 genes do not demonstrate 1 U bias (Figure 3A). These findings point to a new class of small RNAs, possibly specific to mRNAs, that are 24–32 nt long and lack a 1 U signature.

Noteworthy, some genomic loci produced more piRNAs than any others, such as lncRNAs (*ACMO_005053*, *Merged.4615.3*), TEs (*Acol_Tc1_Ele1*, *Acol_gypsy_Ele60*) and protein-coding genes (*ACMO_004083*, *ACMO_003949*) (Figure 3A). Next, to characterize top piRNA producers in mosquito somatic tissues we calculated the percentage of 1 U-biased and 10A-biased piRNAs for sense and antisense mappers, estimated the overall small RNA expression levels as well as the percentage of sequence covered by piRNAs. Interestingly, only two lncRNAs (Supplementary Table S1), *ACMO_005053* and *Merged.4615.3*, were found to produce more than 40% of all somatic piRNAs (Figure 3B). The orthologue sequence of *ACMO_005053* in *An. gambiae*, protein-coding gene *AGAP003387*, has been previously identified as one of the most potent piRNA producers in mosquito ovaries [30].

The coverage profiles of 24–32 nt small RNAs demonstrated some distinct patterns. The most notable pattern was associated with repetitive sequences (e.g. *Acol_Tc1_Ele1*, *Acol_gypsy_Ele36*) and is characterized by the production of piRNAs with the 1 U signature predominantly from the antisense strand (Figure 3C). The second distinct small RNA pattern is related to genic piRNAs, which mapped almost exclusively to the sense strand and lack the robust piRNA nucleotide preference. These small RNAs cover entire transcript regions (Figure 3D) and are obviously produced by an unknown mechanism from genic mRNAs (Figure 3C). The distinct feature of lncRNA small RNAs is that they were clustered at restricted gene regions (Figure 3C).

The mechanism of processing of *ACMO_005053*/*AGAP003387* small RNAs, which account for more than 23% of all somatic piRNAs, was then addressed. It was previously reported that the generation of genic piRNAs in *An. gambiae* ovaries is caused by phased piRNA biogenesis started by a trigger piRNA [16]. In *Drosophila*, an initial Ago3-bound piRNA, known as a 'trigger'-piRNA, targets a complementary transcript, resulting in the production of 'responder'-piRNAs, which are predominantly Aub-bound and have a 10 nt 5'-overlap with the trigger-piRNAs. Following that, piRNA biogenesis spreads downstream, producing 'trail'-piRNAs that bind to the Piwi protein [12]. To find trigger-responder piRNA pairs for *An. coluzzii* transcripts demonstrating phased piRNA biogenesis, we aligned piRNAs to the annotated transcripts and repeats allowing up to 6 mismatches. Next, we select only trigger piRNAs with a high likelihood of their 5' ends aligning precisely 10 nucleotides away from the responder-piRNA 5' ends. Using this approach, we could identify 47 and 118 trigger-responder pairs for protein-coding genes and lncRNAs, respectively (Supplementary Figure S4A). Analysis of 3' end to 5' end distance for piRNAs derived from these genes revealed a signature of phasing pattern (Supplementary Figure S4B). Interestingly, we observed plenty of trigger-responder pairs for several protein-coding and lncRNA transcripts indicating that phased piRNA biogenesis can be initiated at the multiple sites of mRNA (Supplementary Figure S4C).

Notably, sequences unrelated to the one where phased piRNA production will begin are typically the source of trigger-piRNAs. As a result, genic piRNAs are involved in the network of piRNA production in *An. gambiae* ovaries [16]. One of these networks involves the production of trigger-piRNA from protein-coding gene *AGAP011923* that targets *AGAP003387* transcripts and initiates trail-piRNA production. We observed the same piRNA network operates in the somatic tissue of *An. coluzzii* (Supplementary Figure S4D, S4E). Trigger-piRNA derived from the *ACMO_008460* locus (ortholog of protein-coding gene *AGAP011923*) targets lncRNA *ACMO_005053* (ortholog of

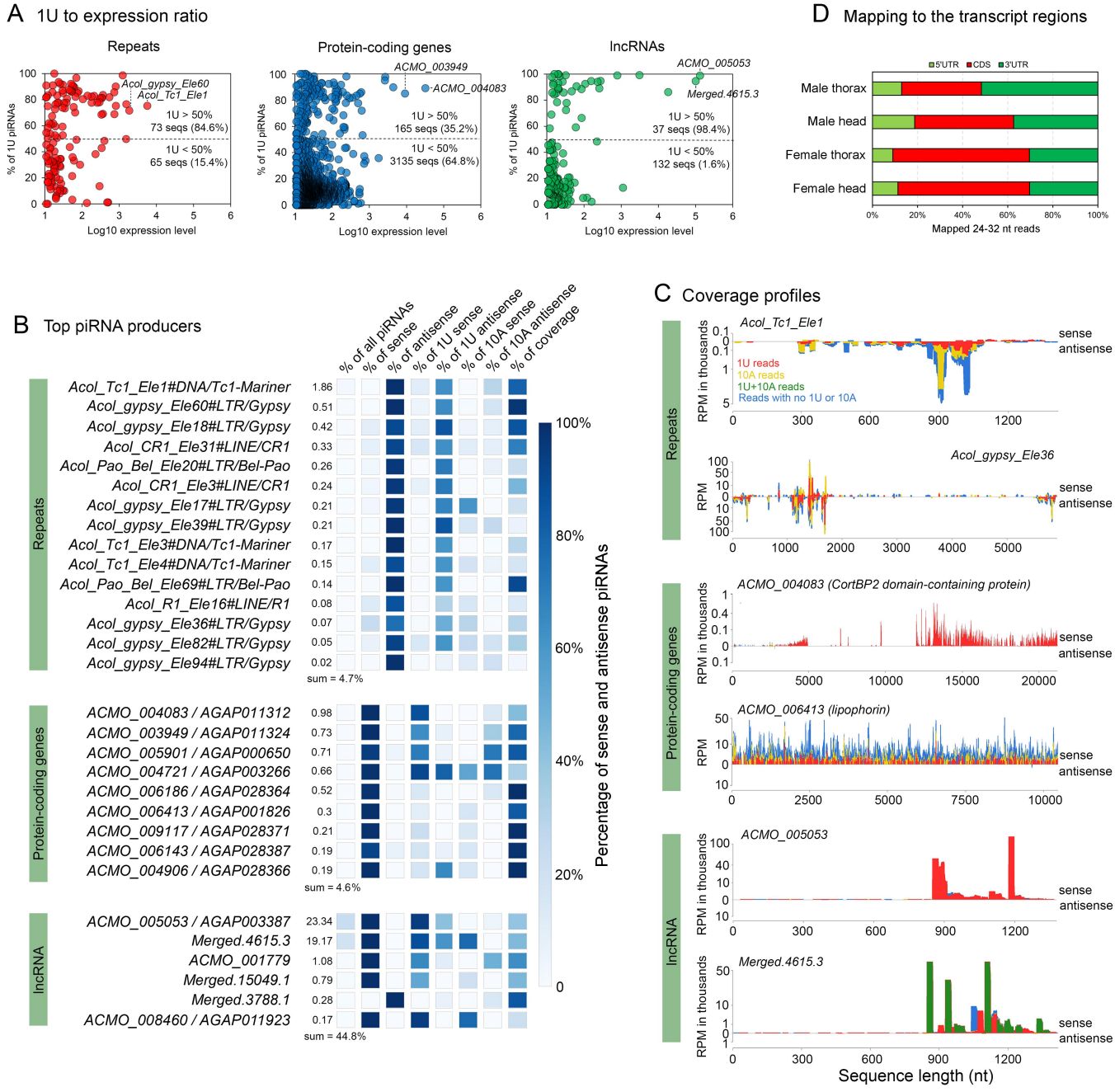


Figure 3. Top sequences processed into piRNAs in somatic tissues of *An. coluzzii*. (A) Ratio of 1U piRNAs to expression levels of all 24–32 nt small RNAs for repeats, protein-coding genes and lncRNAs. Analysis was performed using combined libraries of all tissues and sexes sequenced. (B) Top sequences that were processed into piRNAs. The numbers show the percentage of piRNAs for each gene or repeat calculated by all genomic piRNA mappers, the heatmaps show the percentage of sense and antisense piRNAs, the percentage of 1U and 10A for sense and antisense piRNAs as well as the percentage of sequence covered by piRNAs per specific gene. The color key for heatmaps is indicated to the right. *An. gambiae* orthologs are indicated after the slash. (C) The 24–32 nt small RNA coverage profiles for the indicated loci are shown. Small RNA reads with different characteristics are colored differently (the legend is in the upper plot). RPM - reads per million. (D) Percentage of uniquely mapped 24–32 nt small RNAs to mRNA regions in heads and thoraxes of female and male mosquitoes (the average of three biological replicates). UTR - untranslated region, CDS - coding region.

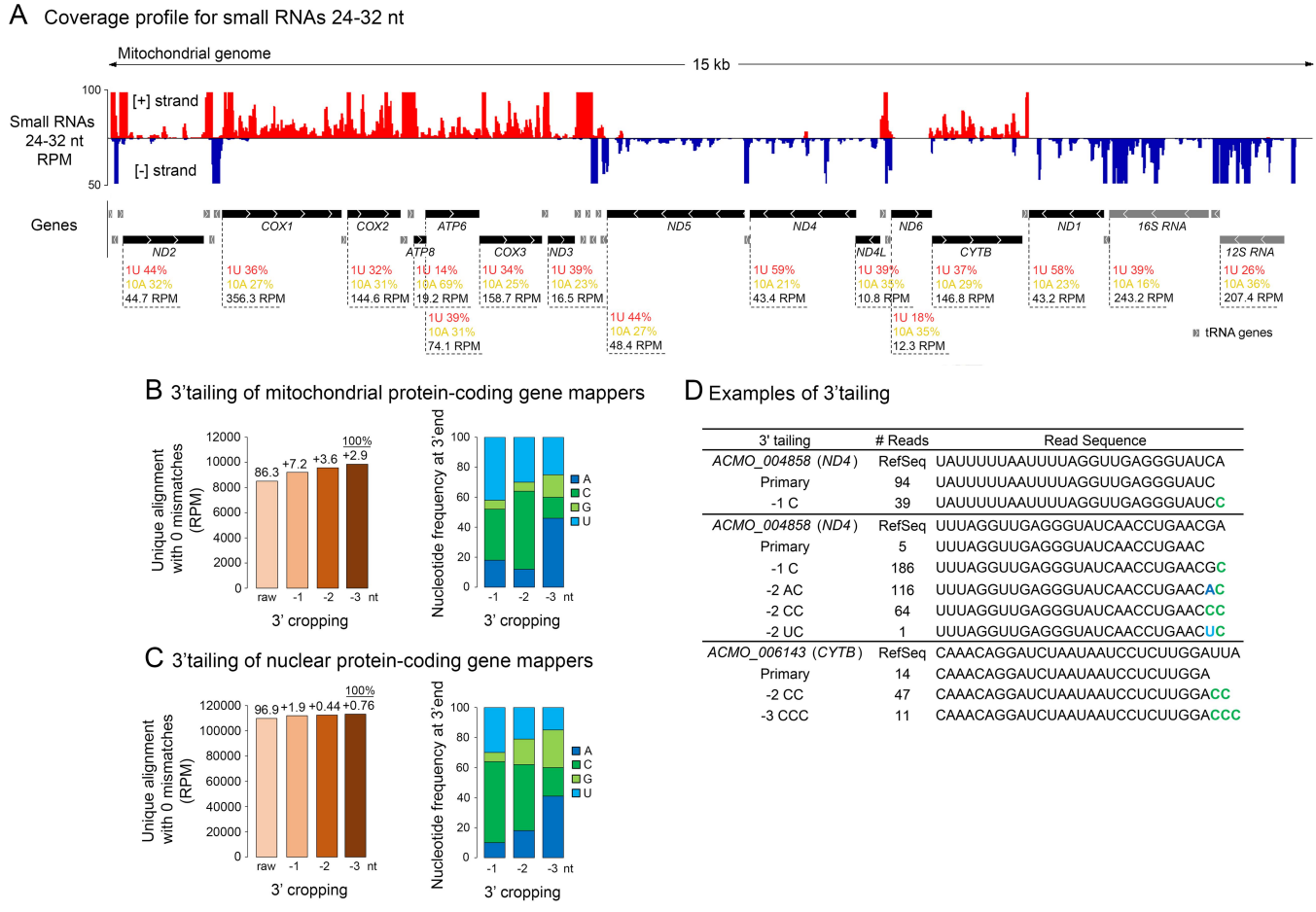
AGAP003387) at 3 different sites and initiates the production of 1 U-biased responder-piRNAs (Supplementary Figure S4D, S4E). Northern blot analysis of small RNAs from *An. coluzzii* carcasses confirmed the presence of abundant trigger piRNA derived from *ACMO_008460* and responder piRNA from *ACMO_005053* (Supplementary Figure S4F). Therefore, a network of inter-regulated transcripts involved in concerted piRNA slicing is a conserved mechanism with unknown biological function.

Small RNAs derived from mitochondrial and nuclear protein-coding genes have signatures of nontemplated 3' tailing

We discovered that mRNAs from all protein-coding genes in the mitochondrial genome, as well as rRNA and tRNA, produce 24–32 nt small RNAs. It is highly probable that these RNAs originate from the mitochondrial genome, as no mitochondrial sequences were discovered in the current assembly

of the *A. coluzzii* genome. This is consistent with the observation that nuclear mitochondrial DNA segments (NUMTs) are small and rare in the genomes of other *Anopheles* species [34]. The majority of mitochondrial small RNAs lack 1U signature, except for small RNAs derived from the NADH dehydrogenase subunit 1 (*ND1*) and NADH dehydrogenase subunit 4 (*ND4*) genes, which demonstrated 1U bias (Figure 4A). Mitochondrial small RNAs were generated from the entire mRNAs according to the mapping profile. This is the most intriguing point because it immediately raises numerous questions about the mysterious formation of these mitochondrial small RNAs, which will be discussed later.

Little is known about 3' end formation of mosquito small RNAs. Formation of 3' ends of small RNAs involves such processes as methylation, trimming and tailing. Nibbler is involved in the 3' trimming of long pre-piRNAs, however *Anopheles* genome lacks the gene encoding Nibbler [13]. 3' tailing is a characteristic of mammalian oocyte-specific piRNAs lacking 3' methyl group [35,36]. To test the possible 3' tailing of mosquito small RNAs we first selected 24–32 nt reads which are uniquely mapped to the genome with 0–3 mismatches and then aligned them to mitochondrial protein-coding genes with 0 mismatches. After that, all unaligned reads were subjected to the 3' cropping of 1, 2 or 3 nucleotides and then aligned again to the mitochondrial genes uniquely with 0 mismatches. We noticed that each time we cropped a nucleotide from the 3' end, the number of aligned small RNA reads increased (Figure 4B). The percentage of single-mapped reads prior to 3' cropping was 86.3%, calculated by taking all uniquely mapped 24–32 nt reads with 0–3 mismatches for 100% (8501 RPM) (Figure 4B). After cropping of 1 nt, 2nt and 3nt the percentage of mapped mitochondrial reads was increased by 7.2% (705 RPM), 3.6% (357.5 RPM) and 2.9% (281.5 RPM) respectively (Figure 4B). Next, we tested if the same pattern applies to small RNAs derived from nuclear protein-coding genes and again we observed better alignment of 3'-trimmed small RNAs (Figure 4C). Trimming the 5' end nucleotides did not result in a similar increase in mapped reads (Supplementary Figure S5), demonstrating the specificity of 3' end tailing. The most frequent



3' tailing	# Reads	Read Sequence
ACMO_004858 (ND4)	RefSeq	UAUUUUUAAUUUUAGGUUGAGGGUAUCA
Primary	94	UAUUUUUAAUUUUAGGUUGAGGGUAUC
-1 C	39	UAUUUUUAAUUUUAGGUUGAGGGUAUCC
ACMO_004858 (ND4)	RefSeq	UUUAGGUUGAGGGUAUCAACCUGAACGA
Primary	5	UUUAGGUUGAGGGUAUCAACCUGAAC
-1 C	186	UUUAGGUUGAGGGUAUCAACCUGAACGC
-2 AC	116	UUUAGGUUGAGGGUAUCAACCUGAACAC
-2 CC	64	UUUAGGUUGAGGGUAUCAACCUGAACCC
-2 UC	1	UUUAGGUUGAGGGUAUCAACCUGAACUC
ACMO_006143 (CYTB)	RefSeq	CAAACAGGAUCUAAUAAUCCUCUUGGAUUA
Primary	14	CAAACAGGAUCUAAUAAUCCUCUUGGA
-2 CC	47	CAAACAGGAUCUAAUAAUCCUCUUGGACC
-3 CCC	11	CAAACAGGAUCUAAUAAUCCUCUUGGACCC

Figure 4. Small RNAs derived from mitochondrial gene transcripts are produced in somatic tissues of *An. coluzzii*. (A) Coverage profile of uniquely mapped small RNAs 24–32 nt with up to 3 mismatches at the mitochondrial genome of *An. coluzzii*. Red and blue mappers are shown according to "+" and "-" DNA strands, respectively. The direction of gene transcription is shown by arrows. The percentage of 1U, 10A and expression levels of small RNAs are shown for each mitochondrial gene. Black color in genome annotation depicts protein-coding genes and grey color depicts tRNAs and rRNAs. Analysis was performed using combined libraries of all tissues and sexes sequenced. RPM - reads per million. (B, C) Alignment of 24–32 nt small RNAs to mitochondrial (B) and nuclear protein-coding genes (C), respectively, after cropping 1–3 nucleotides from the 3' end of small RNAs. Plots to the left show the number of reads uniquely aligned with 0 mismatches after sequential cropping 1, 2 and 3 nucleotides from the 3' end of small RNAs. "Raw" represents mapping before cropping. Plots to the right demonstrate the frequency of the nontemplated nucleotides. RPM - reads per million. (D) Examples of 3' tailed small RNAs. Reference mRNA sequences (RefSeq) are shown above the small RNAs.

nontemplated nucleotides at the 3' end of 24–32 nt small RNAs were U, C and A (Figure 4B–D). The presence of 3' tailing in genic small RNAs suggests that they lack 2'-O-methylation, which protects against tailing and trimming and stabilizes small RNAs [37].

The majority of An. coluzzii *somatic 24–32 nt small RNAs are derived from genes involved in energy metabolism*

To look into differences in 24–32 nt small RNA expression across tissues and sexes, we selected genomic loci that produced more than 50 RPM of small RNAs in at least one sample. We chose 248 protein-coding genes, 13 lncRNAs, and 53 repeats with the specified expression levels (Supplementary Table S2). Analysis of differential expression using one-way ANOVA with the linear model showed large differences in protein-coding gene-derived 24–32 nt small RNAs expression between females and males (FDR ≤0.05, Figure 5A, Supplementary Table S2). In particular, we observed 152 protein-coding genes (~61% of all protein-coding genes producing 24–32 nt small RNAs with more than 50 RPM) that exhibit female-biased expression (Log2 Fold change > 2 in

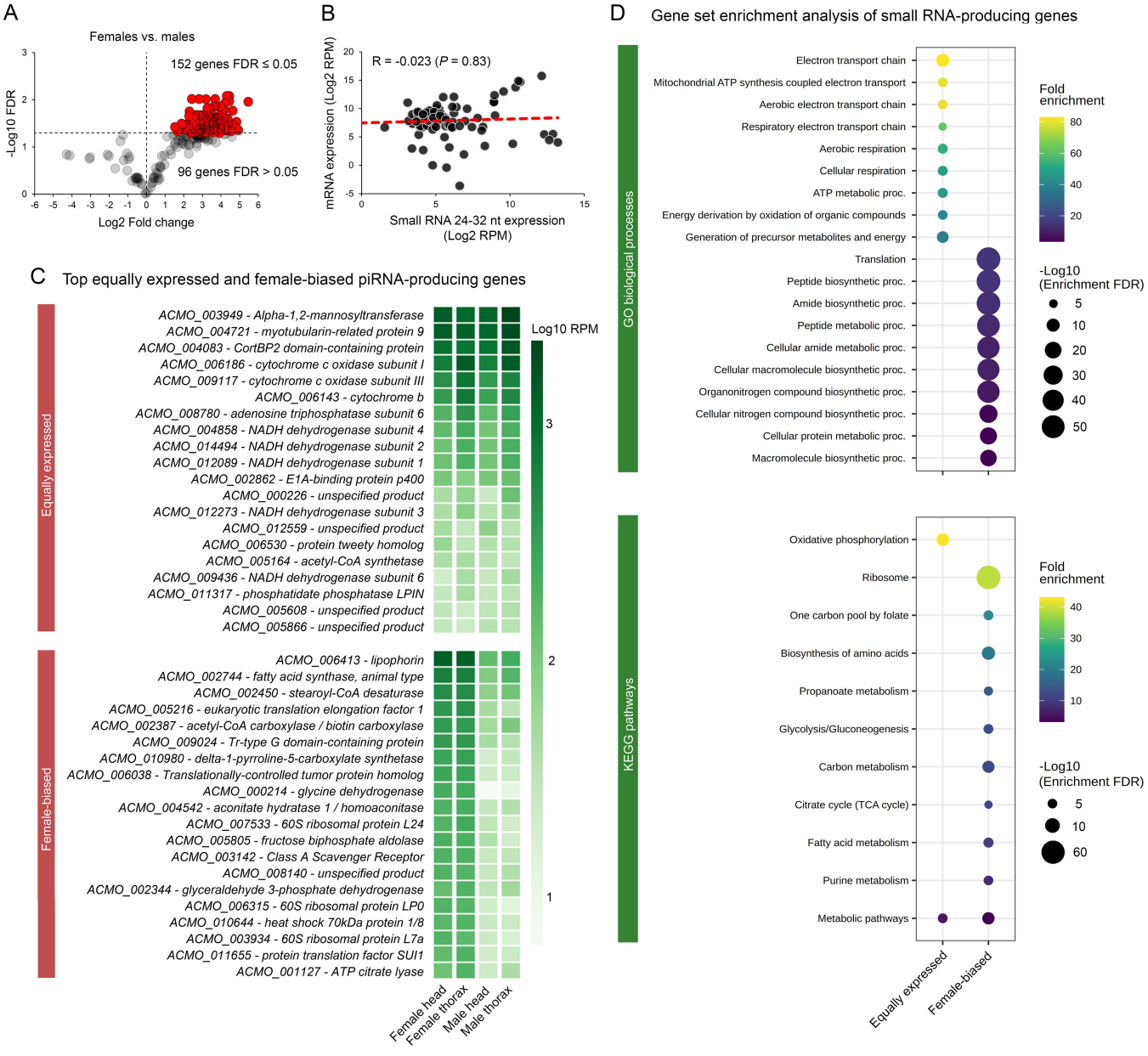


Figure 5. Differences in small RNA expression between male and female somatic tissues of *An. coluzzii*. (A) The volcano plot shows sex-biased expression of 24–32 nt small RNAs mapped to transcripts of protein-coding genes. Horizontal dotted lines show a statistical significance threshold FDR ≤ 0.05 (equal to Log10 transformed value 1.3). Only uniquely mapped small RNAs with 0–3 mismatches and expression level ≥ 50 RPM were considered. Three biological replicates were used to calculate expression changes. (B) The scatter plot depicts the relationship between the expression of 24–32 nt small RNAs and mRNA levels of protein-coding genes with equal small RNA production in males and females (96 genes). Spearman's rank test was used to calculate the correlation coefficient (R). (C) Heatmaps include the female-biased and equally expressed top protein-coding genes producing 24–32 nt small RNAs (Log2 Fold Change < 0.5, FDR > 0.05, $n = 3$). Expression levels are average values for three biological replicates. (D) The gene set enrichment analysis (GSEA) of protein-coding genes with equally expressed and female-biased small RNA production was performed using the biological processes terms of Gene Ontology (top) and KEGG pathway database (bottom). Only terms with FDR < 0.05 were considered statistically significant.

females compared to males) (Figure 5A). The expression of lncRNA and repeat-derived piRNAs showed only minor sex-specific variations (two lncRNA and only one repeat, Supplementary Table S2). Unlike sex-biased expression, expression analysis between heads and thoraxes revealed no differences (Supplementary Table S2). Only 1% of the unique 24–32 nt reads mapped to Y chromosome contigs in the current genome assembly. These reads come from repeats and not from annotated genes, implying that the Y chromosome does not contribute significantly to the total pool of genic piRNAs. Furthermore, no correlation was observed between the expression of 24–32 small RNAs and the mRNA of protein-coding genes, with equal amounts of small RNA produced in males and females (Spearman's correlation coefficient (R) = −0.023, $p = 0.83$), suggesting that non-specific mRNA degradation is unlikely to be the cause of the observed small RNA production (Figure 5B). In summary, our findings show that somatic 24–32 nt small RNA expression derived from protein-coding genes is higher in females than in males, possibly due to the sex-specific expression pattern of specific mRNAs serving as small RNA precursors in mosquito female somatic tissues compared to males.Which genes are involved in small RNA production? To annotate the genes according to the engagement of their products in cellular biological processes we performed gene set enrichment analysis for genes with female-biased and ubiquitous small RNA expression (log2 fold change < 0.5, FDR > 0.05, Figure 5C, Supplementary Table S2).

The performed analysis showed that female-biased small RNA-producing genes are involved in a variety of metabolic processes including amide and peptide biosynthesis and translation according to Gene Ontology biological processes (Figure 5D). Enrichment of female-biased small RNA-producing genes using KEGG database demonstrates their involvement in ribosome biogenesis, glycolysis, biosynthesis of amino acids, purine metabolism and citrate cycle (Figure 5D). Female mosquitoes are larger than males and require metabolic reserves for reproduction. In female somatic tissues, piRNA biogenesis is most likely fuelled by abundant transcripts essential for growth and protein biosynthesis. We then focused on genes that produce similar levels of small RNAs in different sexes and tissues to identify common targets of the piRNA machinery in somatic tissues. Interestingly, these genes demonstrated enrichment in electron transport chain and oxidation phosphorylation according to GO and KEGG databases, respectively (Figure 5D). These data indicate that somatic small RNAs are generated mostly from transcripts of mitochondrial genes and nuclear genome genes involved in ATP synthesis, i.e. in energy metabolism. Therefore, somatic 24–32 nt small RNAs are produced from the mRNAs of genes related to mitochondrial processes.

Next, we examined previously published *An. gambiae* somatic small RNA data to address the existence of atypical genic small RNAs in other mosquito species [20]. The mapping profile and characteristic features of 24–32 nt small RNAs from *An. gambiae* female head/thorax samples are largely similar to those observed in *An. coluzzii*. Repeat-derived piRNAs in *An. gambiae* are predominantly antisense and constitute a minor fraction in somatic tissue (~6%) while most small RNAs were mapped in sense orientation to lncRNAs (49%) and protein-coding gene mRNAs (~18%) (Supplementary Figure S6A). Importantly, the majority of piRNAs derived from lncRNAs and repeats exhibited a pronounced 1U bias (87%). In contrast, 24–32 nt small RNAs that are mapped to protein-coding genes demonstrate 1U preference for only 23% of mappers (Supplementary Figure S6A). A strong preference for a 5' terminal uridine (>50% of small RNAs) was found only for a subset of genes, and such small RNAs constitute 7.6% of the protein-coding gene-derived 24–32 nt small RNAs (Supplementary Figure S5B). We discovered that the *An. gambiae* mitochondrial genome produces abundant small RNAs of 24–32 nt that lack 1U bias (Supplementary Figure S6C). After 3' cropping of 1 nt, 2 nt and 3 nt the percentage of mapped mitochondrial reads was increased by 1.7%, 1.5% and 11.3%, respectively, indicating that these atypical 24–32 nt small RNAs in *An. gambiae* also undergone nontemplated 3' tailing (Supplementary Figure S6D). According to GO and KEGG databases, atypical somatic small RNAs in *An. gambiae* were generated mostly from transcripts of nuclear genes involved in energy production and biosynthesis of major cellular metabolites (Supplementary Figure S6E). These findings suggest that the mechanism underlying the generation of atypical small RNAs from gene transcripts is manifested in the somatic tissues of various mosquitos of *Anopheles gambiae* species complex.

Discussion

Accumulating evidence suggests that the piRNA pathway, which is involved in anti-transposon defence in animal gonads, is also active in nongonadal somatic tissues [18], raising the question of the functional significance of the somatic branch of this small RNA pathway. In mosquitoes, which are carriers of many pathogenic viruses, somatic piRNAs are involved in anti-virus activity [38–41]. At the same time, endogenous sources and targets of the somatic piRNAs have remained largely unexplored. In this study, we analysed small RNAs from male and female somatic tissues of an uninfected laboratory strain of *An. coluzzii*.

Here we show that TE-derived piRNAs represent a minor fraction (2–3%) of somatic piRNAs (Figure 2A). They are generated from all classes of TEs including LINE and LTR retrotransposons and DNA transposons. TE-derived piRNAs demonstrate typical piRNA signatures and are most likely produced by a ping-pong mechanism mediated by Piwi proteins. The analysis of TE-specific piRNA expression revealed no tissue- or sex-specificity. The majority of somatic 24–32 nt small RNAs are produced from lncRNAs (50–70%) and protein-coding gene mRNAs (12–23%). Most of the genic putative piRNAs are generated from the entire mRNA. Surprisingly, they lack typical characteristics of piRNAs such as a 1U bias. Somatic genic small RNAs have a length that ranges randomly from 24 to 32 nt, with no particular 25–27 nt peak that is normally characteristic of TE-derived piRNAs (Figure 2B). Because of their unique characteristics, genic

mosquito somatic 24–32 nt small RNAs identified in both *An. coluzzii* and *An. gambiae* species can be classified as a distinct type of piRNA-like small RNAs.

Since piRNA pathway is highly adaptable to new targets [24,42] it is not surprising that various piRNA biogenesis mechanisms have been reported in different species. An unusual 19–20-nt long piRNA family, whose biogenesis differs from that of typical piRNAs, was discovered in mammalian oocytes [35,43]. *Drosophila eugracilis* was shown to be lacking Ago 3, which is typically involved in ping-pong piRNA amplification; piRNAs however are efficiently produced by the mitochondrial endonuclease Zucchini that performs 5'-3' fragmentation of piRNA precursors [44]. *Anopheles* mosquitoes lack exonuclease Nibbler, suggesting alternative ways of the piRNA 3' end maturation [13]. It is noteworthy that in the germ cells of silkworms, flies and mice, numerous protein-coding gene mRNAs produce low amounts of small RNAs [45]. These small RNAs are processed independently of the known small RNA pathways, most likely through random fragmentation.

Genic small RNA expression in *An. coluzzii* exhibits a significant bias towards females, which is probably due to the specificity of metabolic pathway genes that are expressed in response to female mosquito blood feeding. Gene ontology analysis identified unique biological activities of genes that produce such atypical small RNAs in *An. coluzzii* and *An. gambiae*, allowing us to speculate about the mechanism of their generation. Somatic small RNAs are derived from transcripts encoding proteins involved in energy metabolism, mitochondrial functions and major biosynthetic processes. According to Liu et al., piRNA-generating genes in *Aedes albopictus* are involved in 270 pathways, mainly related to protein processing in the endoplasmic reticulum, glycolysis/gluconeogenesis, endocytosis, and fatty acid metabolism [46]. This suggests that the genic small RNA repertoires are similar in diverse mosquito species. Moreover, we found that transcripts of all protein-coding genes in the mitochondrial genome generate 24–32-nt-long small RNAs. Small RNAs derived from mitochondrial genome have been found in mouse germ cells, zygote, gonads and human cells and tissues [47–52]. How did the mitochondrial mRNAs get into the cytoplasm, where small RNA processing occurs? According to a recent study, mitochondria release RNA, and the accumulation of mitochondrial RNAs in the cytoplasm signals the activation of innate immunity [53]. In addition, mitochondrial transcripts could appear in the cytoplasm during the elimination of damaged organelles known as mitophagy [54]. Most likely, mitochondrial RNAs are released from mitochondria by an unknown mechanism or appear in the cytoplasm as a result of mitophagy, and are processed into small RNAs at the outer mitochondria membrane. Growing evidence suggests that nuclear transcripts encoding mitochondrial proteins are also targeted to the outer membrane of the mitochondria, where they are translated to perform the locally required functions [55–59]. The alternative nuclear export machinery assists in the localization of mRNAs to the mitochondria. UAP56, a component of the RNA export TREX complex, mediates mitochondria-directed export of mRNAs encoding mitochondrial proteins [60]. This suggests that those mosquito genic transcripts that are cleaved into small RNAs are compartmentalized at mitochondria. Interestingly UAP56 is also required for the export of piRNA precursors to a perinuclear piRNA-processing compartment in *Drosophila* [8,61]. In *Arthropods*, piRNA processing factors are located on the outer mitochondrial membrane [9,62,63].

Piwi proteins are not involved in the generation of mosquito genic small RNAs, as no nucleotide signature of Piwi-mediated processing was observed. Genic small RNAs look more like degradation products of mRNAs. However, no correlation was found between the expression of 24–32 small RNAs and the mRNA of protein-coding genes (Figure 5B), implying that non-specific mRNA degradation is unlikely to be the source of the observed small RNA production. The lengths of the reads mapped to protein-coding genes showed peaks at 21 and 26 nt (Figure 1), but were not randomly distributed as would be expected if degradation were occurring. Furthermore, using independently obtained small RNA sequencing data [20], we discovered the generation of atypical small RNAs from gene transcripts in the somatic tissues of the closely related mosquito species *Anopheles gambiae*. Taken together, these findings suggest that genic small RNAs are most likely a unique class of somatic small RNAs, rather than the result of random degradation. Apart from Zucchini, there are no known mitochondrial endonucleases capable of cleaving single-stranded RNA into 20–30 nt fragments. We propose that being localized at the mitochondria, protein-encoding transcripts are cleaved by endonuclease Zucchini, related to a superfamily of mitochondrial phospholipase D (PLD) protein, which participates in piRNA biogenesis. Zucchini creates phased piRNAs by cleaving the RNA upstream of uridine and employing as a substrate RNAs bearing 5' phosphate from prior Piwi-mediated cleavage events [12]. In flies, the 3' ends of Zucchini-generated pre-piRNAs are trimmed by exonuclease Nibbler and methylated by the methyl transferase Hen1 [64] to produce mature piRNAs. However, Nibbler was not found in the *Anopheles* genome, suggesting an alternative piRNA maturation mechanism. Intriguingly, in contrast to the canonical export pathway requiring splicing and functional cap, the alternative export pathway targeting nuclear mRNAs to mitochondria does not require splicing or a 5' cap [56,65]. Such mRNAs will probably be recognized as Zucchini substrates once localized on the mitochondrial membrane and will be processed into piRNAs without the assistance of Piwi proteins which generate 5' phosphate. Another characteristic of somatic genic small RNAs is that they do not exhibit the typical 1 U bias, indicating that the mosquito Zucchini does not have a preference for certain sequences. 1 U bias is also absent in small RNAs derived from the mitochondrial genome of diverse species studied so far [47–52]. In the absence of a trimming enzyme, Nibbler, this specificity was likely lost by Zucchini in *Anopheles* species. In flies, mature siRNAs and piRNAs have 2'-O-methylation at the 3' ends that protect them from degradation [37]. The process of small RNA tailing or nontemplated nucleotide addition is used for 3' end processing and is thought to be linked to small RNA instability [36,66–68]. In *Caenorhabditis elegans*, a nucleotidyltransferase CDE-1 catalyses 3' uridylation of siRNAs [69]. The addition of

uridine and adenosine to the 3' ends are typical for the piRNA populations during mammalian oogenesis [36]. In human oocyte, 20-nt long piRNAs associated with HIWI3 were found to be tailed by mono- or oligonucleotides and lack 2'-O-methylation at their 3' end [35]. We found that 24–32 nt small RNAs derived from mitochondrial and nuclear genes have signatures of nontemplated 3' end tailing resulting in the lengthening of small RNAs by 1–3 nucleotides. Template-independent RNA polymerases that add ribonucleotides to the 3' ends of RNA are a diverse family of enzymes involved in the mRNA and small RNA quality control and stability [70]. Future research is needed to determine the role of Zucchini or other yet unknown factors in the novel mechanism underlying the generation of small RNAs from mitochondrial and nuclear transcripts in mosquito somatic tissues (Figure 6).

Is small RNA processing from mRNA of genes involved in mitochondrial functions a byproduct of their location on the outer mitochondrial membrane or does it have functional significance? Zucchini, also known as MitoPLD, is a PLD superfamily member involved in the control of mitochondria biogenesis. The canonical activity of PLD is the hydrolysis of the phosphodiester bond in phospholipids but MitoPLD/Zucchini can also hydrolyse the phosphodiester bond of RNA demonstrating nuclease activity essential in the piRNA biogenesis [71]. MitoPLD demonstrates phospholipase activity to generate phosphatidic acid, which is involved in the dynamics of mitochondrial division and fusion [72]. How these functionally distinct activities are balanced remains unclear. Regulation of mitochondria number and shape is aimed to provide energy and stimulate metabolic processes in stress conditions. It was reported that infection induces a strong upregulation of metabolic gene expression which contributes to organismal response and survival [73]. Glycolysis/gluconeogenesis and fatty acid biosynthesis are among the pathways that were significantly differentially regulated in CHIKV and DENV-infected *Aedes aegypti* female mosquitoes [74]. When MitoPLD/Zuc catalytic activity is switched to phospholipid hydrolysis, its ribonuclease activity should be reduced, resulting in the accumulation of mitochondrial transcripts. If this is true, the levels of proteins involved in mitochondrial biogenesis will rise, boosting energy metabolism. It is tempting to speculate that MitoPLD/Zuc proteins play an ancient role in regulating mitochondrial transcript levels by cleaving them into small RNAs in order to fine-tune the cellular bioenergetic and metabolic status.

Mitochondria in mosquito midguts play a crucial role in regulating mosquito resistance to infection, metabolism, lifespan, and reproduction – the phenotypes that are relevant in vector control strategies [75]. It has been shown that energy metabolism affects the susceptibility of *An. gambiae* mosquitoes to *Plasmodium* infection [76]. Furthermore, a study has demonstrated that disease-refractory *An. gambiae* females exhibit an impaired mitochondrial respiratory state and a heightened rate of mitochondrial electron leak associated with a reduced lifespan [77]. The regulation of mitochondrial transcript levels by the small RNAs machinery may have implications for transmission of malaria and other pathogens. Other research

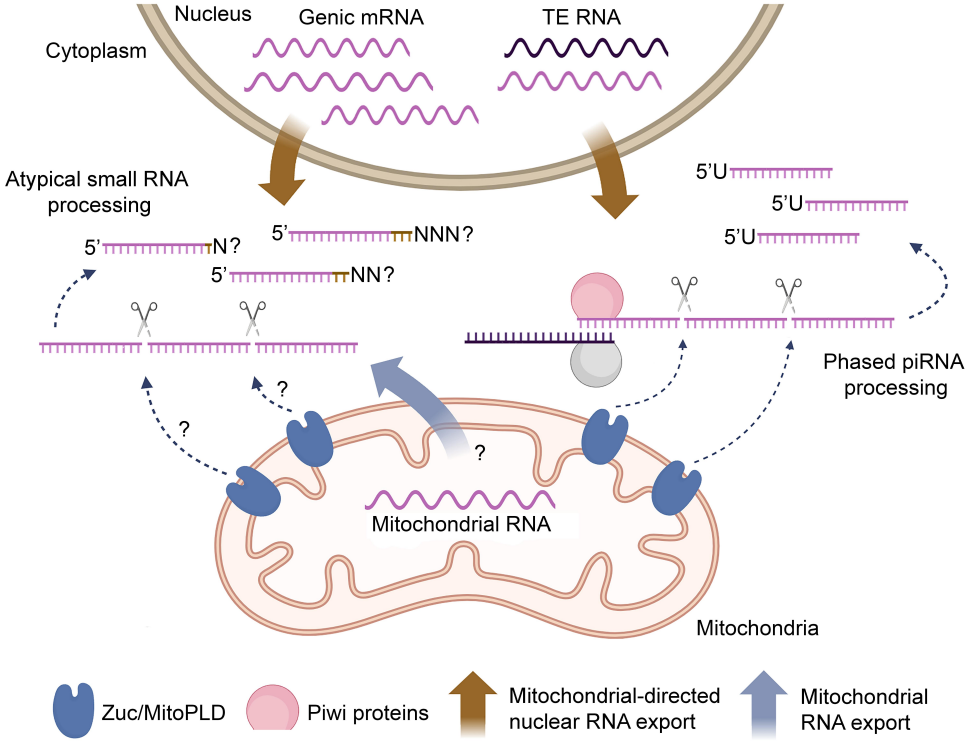


Figure 6. A scheme illustrating somatic piRNA processing in *An. coluzzii*. The production of TE-specific piRNAs via ping-pong processing and Zucchini-mediated cleavage is shown on the right. The hypothetical mechanism of genic small RNA processing is shown on the left. The question marks represent an unknown mitochondrial RNA export mechanism and the enzymes involved in small RNA processing and tailing.

should be done to assess the difference in small RNAs aligned to mitochondria-related genes using infected and non-infected mosquitos.

Methods

Preparation of small RNA libraries

Total RNA was isolated from the heads and thoraxes of 2–5-day-old virgin adult males and females of the *An. coluzzii* MOPTI (MRA-763) laboratory colony using TRIzol Reagent (Invitrogen) extraction in three biological replicates. Females used for RNA isolation were nonblood fed. Thirty individual mosquitoes were used per each sample. Small RNA (18–32 nt) libraries were generated as previously described [78]. Sequencing was performed at the Duke Center for Genomics and Computational Biology on the Illumina HiSeq4000 to obtain 50-bp single reads. About 18.1 Gb of total data or ~ 1.5 Gb of data on average for each replicate were obtained. Raw small RNA sequence data were deposited in NCBI GEO under the number GSE251974. Small RNAs from female somatic tissues of *An. gambiae* (head/thorax samples) were fetched from NCBI BioProject number PRJNA630738.

Small RNA-seq data processing and analysis

Pre-processing of sequenced small RNAs was performed by Trim_Galore script (https://github.com/FelixKrueger/TrimGalore) and included 3'-adapter trimming, filtration of reads by length (>18 nt) and quality (80% of nt have ≥ 20 Phred score). Pre-processed reads were further subjected to subtraction of reads matching to rRNA, snoRNA, snRNA, tRNA (combined list of annotated *An. coluzzii* sequences and the SILVA ribosomal RNA gene database) [79] and miRNA sequences (*An. gambiae* and *Aedes aegypti* microRNAs fetched from miRbase) [80]. rRNA and tRNA mappers were taken into consideration only for the analysis of small RNAs from the mitochondrial genome which encodes both of these RNA types.

The filtered and pre-processed small RNA reads were mapped to *An. coluzzii* MOPTI genome (VectorBase rel. 57), the canonical sequences of *An. coluzzii* TEs and other repeats [81], exons of annotated protein-coding genes or their transcripts, lncRNAs (combined list of annotated genomic sequences and lncRNAs described for *An. gambiae*) [82] and known *Anopheles* viral sequences [38,83] by bowtie allowing up to 3 mismatches using the following parameters 'bowtie – all – tryhard -v 3 –best – strata – quiet' including option '–m 1' for unique mapping [84]. For analysis of small RNAs obtained from *An. gambiae* somatic tissues pre-processed reads were aligned to *An. gambiae* PEST genome (VectorBase rel. 54). The obtained SAM files were subsequently converted to BAM files, sorted by coordinate and indexed using SAMtools software [85]. To reduce the redundancy of repeats and lncRNA sequences we used the CD-HIT program with the following parameters 'cdhit-est -c 0.8 -n 5 -d 0' [86]. Also, to reduce the number of reads mapped to multiple locations we selected only the longest transcripts (splice variants) for every protein-coding gene annotated for the MOPTI genome (VectorBase rel. 57). For analysis of lncRNAs and protein-coding genes we counted uniquely mapped reads. For repeats multimapped reads were used. FeatureCounts utility was used to obtain the number of reads for annotated sequences [87]. To identify the genomic loci that generate the mosquito somatic small RNAs, we combined sequenced libraries from both sexes and all tissues to increase the amount of reads mapped to the genome. For analysis of combined libraries of all tissues, sequences that have at least 10 uniquely mapped reads were considered. We considered the fraction of small RNAs with a length of 24–32 nt as putative piRNAs, 21 nt as putative siRNAs. Mapped small RNA reads were normalized to the sequencing depth.

Length distribution, counting of mapped small RNA reads, calculation of nucleotide frequencies (bias), and coverage profiles were performed using custom scripts written in Python and R. Number of overlapping pairs and overlap probabilities (ping-pong signature) were calculated from alignment files using the 'signature.py' Python script [88]. All calculations have been carried out in R environment (https://www.R-project.org/). Heatmaps were created using Plotly (https://plot.ly).

Comparative small RNA expression analysis

To determine differences in small RNA expression between somatic tissues and sexes, all sequences with at least 50 RPM (reads per million mapped reads on the genome) of mapped small RNAs (24–32 nt) per repeats, lncRNA or transcripts of protein-coding genes were tested on differential expression, using one-way ANOVA with linear model using 'tissue' and 'sex' as factorial predictors (formula: RPM ~ sex + tissue). Fifty RPM has been proven to be the optimal threshold for reducing noise and improving sequencing data reliability. As a result, we obtained *P*-values representing the significance of variance in terms of sex- and tissue-biased differences in expression values calculated by dividing averaged RPM between groups and applying the logarithm function. Gene set enrichment analysis was performed using ShinyGO using the 'Biological Process' set of terms by the Gene Ontology consortium and KEGG database [89,90].

Analysis of 3' tailing of small RNAs

To determine the influence of 3' tailing of small RNA 24–32 nt on their mapping we first selected all 24–32 nt small RNAs that mapped uniquely with 0–3 mismatches on the genome of *An. coluzzii* using Bowtie aligner [84]. Next, we consistently trimmed 1, 2 or 3 nucleotides from the 3' end of selected small RNAs and re-aligned the obtained libraries again to the transcripts of protein-coding genes uniquely allowing no mismatches. The ratio of perfectly mapped reads with and without 3' trimming was calculated considering the number of uniquely aligned reads after cropping of 3 nucleotides as 100%. Nucleotide frequency calculation and trimming were performed using homerTools trim and freq functions [91].

Analysis of trigger-responder pairs and phasing signature

To determine the putative trigger piRNAs, libraries filtered from all rRNAs and microRNA sequences were re-aligned to the MOPTI genome with a BWA aligner allowing up to 6 mismatches 'bwa aln -n 6 -k 6 -t 64 -N' [12,92]. Obtained alignment files in SAI format were converted into SAM with bwa samse command. After the conversion of aligned files to bam format using SAMtools, bam files were concatenated with 'samtools merge' to form a single bam file for protein-coding gene transcripts, lncRNAs and TEs. We applied PingPongPro software to detect genomic coordinates of complementary pairs of small RNA 24–32 nt with 10-nt nucleotide overlap [93]. The fraction of complementary reads containing 1 U and 10A were calculated using Rsamtools (https://bioconductor.org/packages/release/bioc/html/Rsamtools.html), GenomicRanges [94] and Biostrings (https://bioconductor.org/packages/release/bioc/html/Biostrings.html) libraries for the R language. Complementary pair-containing regions were taken into account if they met two criteria: 1) more than 10 reads mapped into locus, and 2) greater than 80% of all reads aligned containing 1U in sense reads, and 10A – in antisense. Transcript regions with putative phasing patterns of piRNA processing were manually curated in the IGV browser [85,95]. Phasing signature (3'end to 5'end distance) was calculated using Phaser script (https://www.smallrnagroup.uni-mainz.de/).

RNA-seq analysis

Previously published RNA-seq data for male carcasses of *An. coluzzii* were fetched from the NCBI SRA archive SRP047496 [96]. Raw reads from four biological replicates were trimmed and aligned to *An. coluzzii* MOPTI genome (VectorBase rel. 57) using splice-aware aligner HISAT2 [97]. Mapped reads were assigned to genomic features using the FeatureCounts utility from the Subread package and normalized to the sequencing depth [98].

Northern analysis of small RNAs

Northern analysis of small RNAs was performed as previously described [99]. The lanes were loaded with 15 μg of total RNA extracted from male or female carcasses after gonad dissection. Hybridization with P^{32} 5'-end-labelled oligonucleotides complementary to *ACMO_005053* responder piRNA (5' AGGTTTTATACTAGATGTTACATCCGAAA 3') or *ACMO_008460* trigger piRNA (5' AAACAACCACTAA AATTTCTGATA 3') (Figure S4E) was performed. Hybridization with oligonucleotides complementary to miRNA 13b1 (5' ACTCGTCAAAATGGCTGTGATA 3') or U4 small nuclear RNA (5' CAGGGAGGCTTCATTGGTTACGGTA 3') was used as a loading control. The blots were visualized with the phosphorimager Typhoon FLA 9500 (Amersham).

Acknowledgments

We thank Liudmila Protsenko for assistance with the library processing and critical comments on the manuscript. The following reagent was obtained through the Biodefense and Emerging Infections Research Resources Repository, NIAID, NIH: *An. coluzzii*, Strain MOPTI, Eggs, MRA-763, contributed by Gregory C. Lanzaro.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the IDB RAS Government basic research programme 0088–2024–0010 (AK), the US National Institute of Health (NIH) grant R21AI159382 (IVS). Bioinformatic analysis was supported by Russian Science Foundation grant 22-74-10050 (SF). The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

AK and IVS conceived and designed the project. NA and JL performed the experiments. SF and AR conducted the bioinformatic analyses. AK, SF and IVS wrote the paper. All authors read and approved the final manuscript.

Availability of data and materials

The small RNA-seq data for this paper are available online at the NCBI (GEO accession number: GSE251974). All RPM data and differential expression data are submitted with the manuscript.

ORCID

Igor V. Sharakhov https://orcid.org/0000-0003-0752-3747

References

[1] Czech B, Munafò M, Ciabrelli F, et al. piRNA-Guided genome defense: from biogenesis to silencing. Annu Rev Genet. 2018;52(1):131–157. doi: 10.1146/annurev-genet-120417-031441
[2] Ozata DM, Gainetdinov I, Zoch A, et al. Piwi-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 2019;20(2):89–108. doi: 10.1038/s41576-018-0073-3
[3] Brennecke J, Aravin AA, Stark A, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089–1103. doi: 10.1016/j.cell.2007.01.043
[4] Shpiz S, Ryazansky S, Olovnikov I, et al. Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. PloS Genet. 2014;10(2):e1004138. doi: 10.1371/journal.pgen.1004138
[5] Radion E, Morgunova V, Ryazansky S, et al. Key role of piRNAs in telomeric chromatin maintenance and telomere nuclear positioning in Drosophila germline. Epigenetics Chromatin. 2018;11(1):40. doi: 10.1186/s13072-018-0210-4
[6] Robine N, Lau NC, Balla S, et al. A broadly conserved pathway generates 3'UTR-Directed primary piRNAs. Curr Biol. 2009;19(24):2066–2076. doi: 10.1016/j.cub.2009.11.064
[7] Aravin A, Gaidatzis D, Pfeffer S, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature. 2006;442(7099):203–207. doi: 10.1038/nature04916
[8] Zhang F, Wang J, Xu J, et al. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell. 2012;151(4):871–884. doi: 10.1016/j.cell.2012.09.040
[9] Munafò M, Manelli V, Falconio FA, et al. Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to

the biogenesis machinery. Genes Dev. 2019;33(13–14):844–856. doi: 10.1101/gad.325662.119
[10] Gunawardane LS, Saito K, Nishida KM, et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science. 2007;315(5818):1587–1590. doi: 10.1126/science.1140494
[11] Han BW, Wang W, Li C, et al. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348(6236):817–821. doi: 10.1126/science.aaa1264
[12] Mohn F, Handler D, Brennecke J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348(6236):812–817. doi: 10.1126/science.aaa1039
[13] Hayashi R, Schnabl J, Handler D, et al. Genetic and mechanistic diversity of piRNA 3'-end formation. Nature. 2016;539(7630):588–592. doi: 10.1038/nature20162
[14] Chirn GW, Rahman R, Sytnikova YA, et al. Conserved piRNA expression from a distinct set of piRNA cluster loci in Eutherian mammals. PloS Genet. 2015;11(11):e1005652. doi: 10.1371/journal.pgen.1005652
[15] Yuan J, Zhang P, Cui Y, et al. Computational identification of piRNA targets on mouse mRNAs. Bioinformatics. 2016;32(8):1170–1177. doi: 10.1093/bioinformatics/btv729
[16] Jensen S, Brasset E, Parey E, et al. Conserved small nucleotidic elements at the origin of concerted piRNA biogenesis from genes and lncRNAs. Cells. 2020;9(6):1491. doi: 10.3390/cells9061491
[17] Sun YH, Wang RH, Du K, et al. Coupled protein synthesis and ribosome-guided piRNA processing on mRNAs. Nat Commun. 2021;12(1):5970. doi: 10.1038/s41467-021-26233-8
[18] Lewis SH, Quarles KA, Yang Y, et al. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat Ecol Evol. 2018;2(1):174–181. doi: 10.1038/s41559-017-0403-4
[19] Biryukova I, Ye T. Endogenous siRNAs and piRNAs derived from transposable elements and genes in the malaria vector mosquito anopheles gambiae. BMC Genomics. 2015;16(1):278. doi: 10.1186/s12864-015-1436-1
[20] Bryant WB, Ray S, Mills MK. Global analysis of small non-coding RNA populations across tissues in the malaria vector, anopheles gambiae. Insects. 2020;11(7):406. doi: 10.3390/insects11070406
[21] Ma Q, Srivastav SP, Gamez S, et al. A mosquito small RNA genomics resource reveals dynamic evolution and host responses to viruses and transposons. Genome Res. 2021;31(3):512–528. doi: 10.1101/gr.265157.120
[22] Palatini U, Contreras CA, Gasmi L, et al. Endogenous viral elements in mosquito genomes: current knowledge and outstanding questions. Curr Opin Insect Sci. 2022;49:22–30. doi: 10.1016/j.cois.2021.10.007
[23] Abel SM, Hong Z, Williams D, et al. Small RNA sequencing of field Culex mosquitoes identifies patterns of viral infection and the mosquito immune response. Sci Rep. 2023;13(1):10598. doi: 10.1038/s41598-023-37571-6
[24] Lewis SH, Salmela H, Obbard DJ. Duplication and diversification of dipteran argonaute genes, and the evolutionary divergence of Piwi and Aubergine. Genome Biol Evol. 2016;8(3):507–518. doi: 10.1093/gbe/evw018
[25] Macias V, Coleman J, Bonizzoni M, et al. piRNA pathway gene expression in the malaria vector mosquito Anopheles stephensi. Insect Mol Biol. 2014;23(5):579–586. doi: 10.1111/imb.12106
[26] Papa F, Windbichler N, Waterhouse RM, et al. Rapid evolution of female-biased genes among four species of Anopheles malaria mosquitoes. Genome Res. 2017;27(9):1536–1548. doi: 10.1101/gr.217216.116
[27] Pitts RJ, Rinker DC, Jones PL, et al. Transcriptome profiling of chemosensory appendages in the malaria vector Anopheles gambiae reveals tissue- and sex-specific signatures of odor coding. BMC Genomics. 2011;12(1):271. doi: 10.1186/1471-2164-12-271
[28] Kolliopoulou A, Santos D, Taning CNT, et al. PIWI pathway against viruses in insects. Wiley Interdiscip Rev RNA. 2019;10(6):e1555. doi: 10.1002/wrna.1555
[29] Ter Horst AM, Nigg JC, Dekker FM, et al. Endogenous viral elements are widespread in arthropod genomes and commonly give rise to PIWI-Interacting RNAs. J Virol. 2019;93(6). doi: 10.1128/jvi.02124-18
[30] George P, Jensen S, Pogorelcnik R, et al. Increased production of piRNAs from euchromatic clusters and genes in Anopheles gambiae compared with Drosophila melanogaster. Epigenetics Chromatin. 2015;8(1):50. doi: 10.1186/s13072-015-0041-5
[31] Blair CD. A brief history of the discovery of RNA-Mediated antiviral immune defenses in vector Mosquitos. Microbiol Mol Biol Rev. 2023;87(1):e0019121. doi: 10.1128/mmbr.00191-21
[32] Zamyatin A, Avdeyev P, Liang J, et al. Chromosome-level genome assemblies of the malaria vectors Anopheles coluzzii and Anopheles arabiensis. Gigascience. 2021;10(3). doi: 10.1093/gigascience/giab017
[33] Okamura K, Lai EC. Endogenous small interfering RNAs in animals. Nat Rev Mol Cell Biol. 2008;9(9):673–678. doi: 10.1038/nrm2479
[34] Ding YR, Li B, Zhang YJ, et al. Complete mitogenome of Anopheles sinensis and mitochondrial insertion segments in the nuclear genomes of 19 mosquito species. PLoS One. 2018;13(9):e0204667. doi: 10.1371/journal.pone.0204667
[35] Yang Q, Li R, Lyu Q, et al. Single-cell cas-seq reveals a class of short piwi-interacting RNAs in human oocytes. Nat Commun. 2019;10(1):3389. doi: 10.1038/s41467-019-11312-8
[36] Roovers EF, Rosenkranz D, Mahdipour M, et al. Piwi proteins and piRNAs in mammalian oocytes and early embryos. Cell Rep. 2015;10(12):2069–2082. doi: 10.1016/j.celrep.2015.02.062
[37] Ameres SL, Horwich MD, Hung J-H, et al. Target RNA–Directed trimming and tailing of small silencing RNAs. Science. 2010;328(5985):1534–1539. doi: 10.1126/science.1187058
[38] Belda E, Nanfack-Minkeu F, Eiglmeier K, et al. De Novo profiling of RNA viruses in Anopheles malaria vector mosquitoes from forest ecological zones in Senegal and Cambodia. BMC Genomics. 2019;20(1):664. doi: 10.1186/s12864-019-6034-1
[39] Suzuki Y, Baidaliuk A, Miesen P, et al. Non-retroviral endogenous viral element limits cognate virus replication in Aedes aegypti Ovaries. Curr Biol. 2020;30(18):3495–3506.e6. doi: 10.1016/j.cub.2020.06.057
[40] Whitfield ZJ, Dolan PT, Kunitomi M, et al. The diversity, structure, and function of heritable adaptive immunity sequences in the Aedes aegypti genome. Curr Biol. 2017;27(22):3511–3519.e7. doi: 10.1016/j.cub.2017.09.067
[41] Morazzani EM, Wiley MR, Murreddu MG, et al. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. PloS Pathog. 2012;8(1):e1002470. doi: 10.1371/journal.ppat.1002470
[42] Parhad SS, Tu S, Weng Z, et al. Adaptive evolution leads to cross-species incompatibility in the piRNA transposon silencing machinery. Dev Cell. 2017;43(1):60–70.e5. doi: 10.1016/j.devcel.2017.08.012
[43] Hasuwa H, Iwasaki YW, Au Yeung WK, et al. Production of functional oocytes requires maternally expressed PIWI genes and piRNAs in golden hamsters. Nat Cell Biol. 2021;23(9):1002–1012. doi: 10.1038/s41556-021-00745-3
[44] Chary S, Hayashi R, Ketting RF. The absence of core piRNA biogenesis factors does not impact efficient transposon silencing in Drosophila. PLoS Biol. 2023;21(6):e3002099. doi: 10.1371/journal.pbio.3002099
[45] Shoji K, Tomari Y. A potential role of inefficient and non-specific piRNA production from the whole transcriptome. bioRxiv, 2024. doi: 10.1101/2024.01.24.577019. 2024.2001.2024.577019
[46] Liu P, Dong Y, Gu J, et al. Developmental piRNA profiles of the invasive vector mosquito Aedes albopictus. Parasites Vectors. 2016;9(1). doi: 10.1186/s13071-016-1815-8

[47] Ro S, Ma H-Y, Park C, et al. The mitochondrial genome encodes abundant small noncoding RNAs. Cell Res. 2013;23(6):759–774. doi: 10.1038/cr.2013.37

[48] Larriba E, Rial E, Del Mazo J. The landscape of mitochondrial small non-coding RNAs in the PGCs of male mice, spermatogonia, gametes and in zygotes. BMC Genomics. 2018;19(1):634. doi: 10.1186/s12864-018-5020-3

[49] Barreñada O, Larriba E, Fernández-Pérez D, et al. Unraveling mitochondrial piRNAs in mouse embryonic gonadal cells. Sci Rep. 2022;12(1):10730. doi: 10.1038/s41598-022-14414-4

[50] Kwon C, Tak H, Rho M, et al. Detection of PIWI and piRNAs in the mitochondria of mammalian cancer cells. Biochem Biophys Res Commun. 2014;446(1):218–223. doi: 10.1016/j.bbrc.2014.02.112

[51] Siniscalchi C, Di Palo A, Petito G, et al. A landscape of mouse mitochondrial small non-coding RNAs. PLoS One. 2024;19(1):e0293644. doi: 10.1371/journal.pone.0293644

[52] Sripada L, Tomar D, Prajapati P, et al. Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. PLoS One. 2012;7(9):e44873. doi: 10.1371/journal.pone.0044873

[53] Tigano M, Vargas DC, Tremblay-Belzile S, et al. Nuclear sensing of breaks in mitochondrial DNA enhances immune surveillance. Nature. 2021;591(7850):477–481. doi: 10.1038/s41586-021-03269-w

[54] Ashrafi G, Schwarz TL. The pathways of mitophagy for quality control and clearance of mitochondria. Cell Death Differ. 2013;20(1):31–42. doi: 10.1038/cdd.2012.81

[55] Garcia M, Delaveau T, Goussard S, et al. Mitochondrial presequence and open reading frame mediate asymmetric localization of messenger RNA. EMBO Rep. 2010;11(4):285–291. doi: 10.1038/embor.2010.17

[56] Cenik C, Chua HN, Zhang H, et al. Genome analysis reveals interplay between 5'UTR introns and nuclear mRNA export for secretory and mitochondrial genes. PloS Genet. 2011;7(4):e1001366. doi: 10.1371/journal.pgen.1001366

[57] Sylvestre J, Vialette S, Corral Debrinski M, et al. Long mRNAs coding for yeast mitochondrial proteins of prokaryotic origin preferentially localize to the vicinity of mitochondria. Genome Biol. 2003;4(7):R44. doi: 10.1186/gb-2003-4-7-r44

[58] Fazal FM, Han S, Parker KR, et al. Atlas of subcellular RNA localization revealed by APEX-Seq. Cell. 2019;178(2):473–490.e26. doi: 10.1016/j.cell.2019.05.027

[59] Lashkevich KA, Dmitriev SE. mRNA targeting, transport and local translation in eukaryotic cells: from the classical view to a diversity of new concepts. Mol Biol. 2021;55(4):507–537. doi: 10.1134/s0026893321030080

[60] Akef A, Zhang H, Masuda S, et al. Trafficking of mRNAs containing alrex-promoting elements through nuclear speckles. Nucleus. 2013;4(4):326–340. doi: 10.4161/nucl.26052

[61] Kneuss E, Munafò M, Eastwood EL, et al. Specialization of the Drosophila nuclear export family protein Nxf3 for piRNA precursor export. Genes Dev. 2019;33(17–18):1208–1220. doi: 10.1101/gad.328690.119

[62] Patil AA, Tatsuke T, Mon H, et al. Molecular characterization of mitochondrial Zucchini and its relation to nuage-piRNA pathway components in Bombyx mori ovary-derived BmN4 cells. Biochem Biophys Res Commun. 2017;493(2):971–978. doi: 10.1016/j.bbrc.2017.09.107

[63] Ge DT, Wang W, Tipping C, et al. The RNA-Binding ATPase, Armitage, couples pirna amplification in nuage to phased piRNA production on mitochondria. Mol Cell. 2019;74(5):982–995.e6. doi: 10.1016/j.molcel.2019.04.006

[64] Horwich MD, Li C, Matranga C, et al. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol. 2007;17(14):1265–1272. doi: 10.1016/j.cub.2007.06.030

[65] Palazzo AF, Springer M, Shibata Y, et al. The signal sequence coding region promotes nuclear export of mRNA. PLoS Biol. 2007;5(12):e322. doi: 10.1371/journal.pbio.0050322

[66] Pastore B, Hertz HL, Price IF, et al. Pre-piRNA trimming and 2'-O-methylation protect piRNAs from 3' tailing and degradation in C. elegans. Cell Rep. 2021;36(9):109640. doi: 10.1016/j.celrep.2021.109640

[67] Lee M, Choi Y, Kim K, et al. Adenylation of maternally inherited microRNAs by Wispy. Mol Cell. 2014;56(5):696–707. doi: 10.1016/j.molcel.2014.10.011

[68] Li J, Yang Z, Yu B, et al. Methylation Protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol. 2005;15(16):1501–1507. doi: 10.1016/j.cub.2005.07.029

[69] van Wolfswinkel JC, Claycomb JM, Batista PJ, et al. CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell. 2009;139(1):135–148. doi: 10.1016/j.cell.2009.09.012

[70] Martin G, Keller W. RNA-specific ribonucleotidyl transferases. RNA. 2007;13(11):1834–1849. doi: 10.1261/rna.652807

[71] Gao Q, Frohman MA. Roles for the lipid-signaling enzyme MitoPLD in mitochondrial dynamics, piRNA biogenesis, and spermatogenesis. BMB Rep. 2012;45(1):7–13. doi: 10.5483/bmbrep.2012.45.1.7

[72] Adachi Y, Itoh K, Yamada T, et al. Coincident phosphatidic acid interaction restrains drp1 in mitochondrial division. Mol Cell. 2016;63(6):1034–1043. doi: 10.1016/j.molcel.2016.08.013

[73] Deshpande R, Lee B, Grewal SS, et al. Enteric bacterial infection in Drosophila induces whole-body alterations in metabolic gene expression independently of the immune deficiency signaling pathway. G3 Bethesda. 2022;12(11). doi: 10.1093/g3journal/jkac163

[74] Shrinet J, Bhavesh NS, Sunil S. Understanding oxidative stress in Aedes during chikungunya and dengue virus infections using integromics analysis. Viruses. 2018;10(6):314. doi: 10.3390/v10060314

[75] Luckhart S, Riehle MA. Midgut mitochondrial function as a gatekeeper for malaria parasite infection and development in the mosquito host. Front Cell Infect Microbiol. 2020;10:593159. doi: 10.3389/fcimb.2020.593159

[76] Kumar S, Christophides GK, Cantera R, et al. The role of reactive oxygen species on Plasmodium melanotic encapsulation in Anopheles gambiae. Proc Natl Acad Sci U S A. 2003;100(24):14139–14144. doi: 10.1073/pnas.2036262100

[77] Oliveira JHM, Gonçalves RLS, Oliveira GA, et al. Energy metabolism affects susceptibility of Anopheles gambiae mosquitoes to Plasmodium infection. Insect Biochem Mol Biol. 2011;41(6):349–355. doi: 10.1016/j.ibmb.2011.02.001

[78] Olovnikov I, Ryazansky S, Shpiz S, et al. De Novo piRNA cluster formation in the Drosophila germ line triggered by transgenes containing a transcribed transposon fragment. Nucleic Acids Res. 2013;41(11):5757–5768. doi: 10.1093/nar/gkt310

[79] Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(D1):D590–D596. doi: 10.1093/nar/gks1219

[80] Kozomara A, Birgaoanu M, Griffiths-Jones S. miRbase: from microRNA sequences to function. Nucleic Acids Res. 2019;47(D1):D155–d162. doi: 10.1093/nar/gky1141

[81] Vargas-Chavez C, Longo Pendy NM, Nsango SE, et al. Transposable element variants and their potential adaptive impact in urban populations of the malaria vector Anopheles coluzzii. Genome Res. 2022;32(1):189–202. doi: 10.1101/gr.275761.121

[82] 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. doi: 10.1186/s12864-015-1507-3

[83] Carissimo G, Eiglmeier K, Reveillaud J, et al. Identification and characterization of two novel RNA viruses from anopheles gambiae species complex mosquitoes. PLoS One. 2016;11(5):e0153881. doi: 10.1371/journal.pone.0153881

[84] Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25. doi: 10.1186/gb-2009-10-3-r25

[85] Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2). doi: 10.1093/gigascience/giab008

[86] Fu L, Niu B, Zhu Z, et al. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28 (23):3150–3152. doi: 10.1093/bioinformatics/bts565

[87] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–930. doi: 10.1093/bioinformatics/btt656

[88] Antoniewski C. Computing siRNA and piRNA overlap signatures. Methods Mol Biol. 2014;1173:135–146. doi: 10.1007/978-1-4939-0931-5_12

[89] Ge SX, Jung D, Yao R, et al. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36 (8):2628–2629. doi: 10.1093/bioinformatics/btz931

[90] Kanehisa M, Furumichi M, Sato Y, et al. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–d592. doi: 10.1093/nar/gkac963

[91] Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576–589. doi: 10.1016/j.molcel.2010.05.004

[92] Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics. 2009;25 (14):1754–1760. doi: 10.1093/bioinformatics/btp324

[93] Uhrig S, Klein H, Valencia A. PingPongPro: a tool for the detection of piRNA-mediated transposon-silencing in small RNA-Seq data. Bioinformatics. 2019;35(2):335–336. doi: 10.1093/bioinformatics/bty578

[94] Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8): e1003118. doi: 10.1371/journal.pcbi.1003118

[95] Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–192. doi: 10.1093/bib/bbs017

[96] Cassone BJ, Kay RG, Daugherty MP, et al. Comparative transcriptomics of malaria mosquito testes: function, evolution, and linkage. G3 Bethesda. 2017;7(4):1127–1136. doi: 10.1534/g3.117.040089

[97] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12 (4):357–360. doi: 10.1038/nmeth.3317

[98] Liao Y, Smyth GK, Shi W. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41(10):e108. doi: 10.1093/nar/gkt214

[99] Ryazansky S, Radion E, Mironova A, et al. Natural variation of piRNA expression affects immunity to transposable elements. PloS Genet. 2017;13(4):e1006731. doi: 10.1371/journal.pgen.1006731