Identification and characterization of long noncoding RNAs and their association with acquisition of blood meal in *Culex quinquefasciatus*

Azali Azlan, Mardani Abdul Halim, Faisal Mohamad and Ghows Azzam

School of Biological Sciences, Universiti Sains Malaysia, Penang, Malaysia

Abstract The Southern house mosquito, *Culex quinquefasciatus* (*Cx. quinquefasciatus*) is an important vector that transmit multiple diseases including West Nile encephalitis, Japanese encephalitis, St. Louis encephalitis and lymphatic filariasis. Long noncoding RNAs (lncRNAs) involve in many biological processes such as development, infection, and virus–host interaction. However, there is no systematic identification and characterization of lncRNAs in *Cx. quinquefasciatus*. Here, we report the first lncRNA identification in *Cx. quinquefasciatus*. By using 31 public RNA-seq datasets, a total of 4763 novel lncRNA transcripts were identified, of which 3591, 569, and 603 were intergenic, intronic, and antisense respectively. Examination of genomic features revealed that *Cx. quinquefasciatus* shared similar characteristics with other species such as short in length, low GC content, low sequence conservation, and low coding potential. Furthermore, compared to protein-coding genes, *Cx. quinquefasciatus* lncRNAs had lower expression values, and tended to be expressed in temporally specific fashion. In addition, weighted correlation network and functional annotation analyses showed that lncRNAs may have roles in blood meal acquisition of adult female *Cx. quinquefasciatus* mosquitoes. This study presents the first systematic identification and analysis of *Cx. quinquefasciatus* lncRNAs and their association with blood feeding. Results generated from this study will facilitate future investigation on the function of *Cx. quinquefasciatus* lncRNAs.

Key words *Culex quinquefasciatus*; blood-feeding; long noncoding RNA; transcriptome

Introduction

The Southern house mosquito, *Culex quinquefasciatus* (*Cx. quinquefasciatus*), is an important member of the *Cx. pipiens* complex that spreads multiple types of diseases such as West Nile encephalitis, Japanese encephalitis, St. Louis encephalitis, and lymphatic filariasis (Arensburger *et al*., 2010; Nchoutpouen *et al*., 2019). *Cx. pipiens* complex is composed of six members namely, *Cx. molestus* Forskll, *Cx. pipiens* Linneaus, *Cx. quinquefasciatus* Say, *Cx. pallen* Coquille, *Cx. globocoxi-...
ability of *Cx. quinquefasciatus* to carry many harmful diseases is due to its blood-feeding behavior. Newly eclosed female mosquitoes need a minimum of 48 h before they can feed on blood, and subsequently mate (Reid *et al*., 2015). For example, cytochrome P450s, proteases, and odorant-binding proteins were upregulated during blood-feeding period (Reid *et al*., 2015). Another study performed RNA-seq on antennae of blood-fed and nonblood-fed females, and found that chemosensory genes, odorant-binding proteins, ionotropic and odorant receptors were among the genes that were differentially expressed (Taparia *et al*., 2017). Besides blood meal, RNA-seq was also used to identify genes that may be involved upon insecticide exposure and resistance. Taken together, the utilization of RNA-seq approach enables the researcher to observe the differences in *Cx. quinquefasciatus* gene expression level at a genomic scale under different physiological conditions (Reid *et al*., 2015; Taparia *et al*., 2017).

Even though protein-coding genes have always been the focus in transcriptomic studies, several reports showed that noncoding RNAs (ncRNAs) are implicated to play major roles in mosquitoes (Gu *et al*., 2013; Etebari *et al*., 2016, 2017; Miesen *et al*., 2016; Göertz *et al*., 2019; Rückert *et al*., 2019; Azlan *et al*., 2019b). Regulatory ncRNAs include small RNAs and long noncoding RNA (lncRNAs). Small RNA consists of micro-RNAs (miRNAs), short-interfering RNAs (siRNAs), and PIWI-interacting RNA (piRNAs), all of which regulate gene expression at the posttranscriptional level (Coller & Parker, 2005; Brennecke *et al*., 2007; Djuranovic *et al*., 2012; Azlan *et al*., 2016). lncRNAs, on the other hand, are RNA species of more than 200 nt in size that do not encode amino acids (Clark & Mattick, 2011; Ulitsky & Bartel, 2013). Due to next-generation sequencing and bioinformatics, identification of lncRNAs has been made possible in many organisms including mosquito such as Aedes aegypti (*Ae. aegypti*) (Etebari *et al*., 2016; Azlan *et al*., 2019b).

Previous study in *Ae. aegypti* showed that a subset of lncRNAs were highly expressed during early embryo (0–8 h), and they were also enriched in blood-fed ovary. This suggests that lncRNAs may be maternally inherited, and they may play critical roles in early embryonic development (Azlan *et al*., 2019b). Beside embryonic development, lncRNAs in mosquitoes are also involved in host–virus interaction. For example, knockdown of *Ae. aegypti* lncRNA (lincRNA-1317) resulted in an increase in virus replication (Etebari *et al*., 2016). Although these studies suggest the regulatory roles of lncRNAs, the specific mechanism of lncRNAs functions are still unknown. Up until now, there have been no lncRNA studies in *Cx. quinquefasciatus*. We believed that, similar to *Ae. aegypti*, lncRNAs may possess certain regulatory roles in *Cx. quinquefasciatus*. In order to dissect their functional roles, it is important to first systematically annotate lncRNAs in the genome of *Cx. quinquefasciatus*.

Here, we report the first annotation of lncRNAs in *Cx. quinquefasciatus* genome using 31 publicly available RNA-seq data. We applied stringent bioinformatics analysis to confidently predict a total of 4763 novel lncRNAs that correspond to 4037 loci in the genome. We then characterized the newly identified lncRNAs such as expression dynamics, genomic features, and sequence conservation. We also investigate the transcriptional response of *Cx. quinquefasciatus* lncRNAs upon blood meal acquisition. Furthermore, through our coexpression network analysis, a subset of *Cx. quinquefasciatus* lncRNAs may participate in the taking and processing of blood meal.

### Materials and methods

#### Preparation of public RNA-seq datasets

Thirty-one RNA-seq datasets were downloaded from NCBI SRA websites. List of accession number can be found in Table S1. Strand information and library selection methods used in all 31 RNA-seq public libraries can also be found in Table S1. FASTQC was used to check the quality of raw reads. Trimmomatic version 0.38 (Bolger *et al*., 2014) was used to clip adapters and low quality reads (<20 phred score).

#### Alignment of RNA-seq reads to reference genome and transcriptome assembly

Cleaned reads were individually aligned to *Cx. quinquefasciatus* genome (CpipJ2, VectorBase) using Hisat2 (Kim *et al*., 2015). The reads were mapped according to the strandedness of the libraries. Strandedness of each library was determined using infer_experiment.py in RSEQC package (Wang *et al*., 2012). BAM files generated by Hisat2 were used as input for transcriptome assembly using Stringtie version 1.0.1 (Pertea *et al*., 2015). Annotation file of *Cx. quinquefasciatus* in GTF format (VectorBase) was used to guide the assembly. We set the minimum size of transcripts to be assembled was 200 bp. All Stringtie output files were merged into a single unified transcriptome assembly using Stringtie merge option. The resulting unified transcriptome assembly (GTF format) was compared to the reference annotation using
Gffcompare (https://github.com/gpertea/gffcompare), and transcripts with class code “u,” “i,” and “x” were kept for lncRNA prediction analysis.

**lncRNA prediction**

Transcripts with class code “u,” “i,” and “x” were subjected to TransDecoder analysis to predict long ORF, which was set to be minimum of 100 amino acids. Transcripts that do not have long ORF were then analyzed for their coding potential using CPC2 (Kang et al., 2017), PLEK (Li et al., 2014), and CNCI (Sun et al., 2013). Only transcripts that were categorized as noncoding by all three softwares were retained for downstream analyses. We then used BLASTX to find sequence similarity of the potential lncRNA transcripts against Swissprot database. Transcripts having high sequence similarity (E-value < 1e-3) were discarded. We then removed transcripts that had no strand information and located within less than 2 kb scaffold-end range.

**Transcript quantification and expression**

Salmon version 0.10.1 (Patro et al., 2017) was used with default parameters to compute the expression of each transcript. We combined protein-coding and lncRNA transcripts together into a single FASTA file when running Salmon. TPM value from Salmon was used for downstream analysis.

**Validation of lncRNAs by reverse-transcribed PCR**

About 20 adult females of VCRU-lab strain *Cx. quinquefasciatus* were obtained from Vector Control Research Unit, Universiti Sains Malaysia. A pool of 20 mosquitoes were homogenized for RNA isolation and subsequent reverse transcription for cDNA synthesis. For RNA isolation, Rneasy Mini Kit (Qiagen, cat. no: 74104) was used according to manufacturer’s protocol. Reverse transcription was carried out using iScript Reverse Transcription Supermix (Bio-Rad Laboratories, California; cat. no. 1708840) according to manufacturer’s protocol. We then performed PCR using the cDNA as templates. We designed PCR primers that span exon–exon junction using Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers used in this study are listed in Table S2.

**Expression specificity**

We used JS divergence method to compute expression specificity of lncRNAs and protein-coding genes. MATLAB version R2018b was used to calculate JS score using the formula previously described (Cabili et al., 2011).

**Coexpression analysis**

WGCNA version 1.4.6 was used to perform coexpression analysis between lncRNAs and protein-coding genes (Zhang & Horvath, 2005; Langfelder & Horvath, 2008). In this analysis, we used seven RNA-seq data generated by Reid et al. (2015). Expression values generated from Salmon version 0.10.1 (Patro et al., 2017) were used as raw inputs. TPM of 1 in at least one in seven libraries was set as a minimum threshold for coexpression analysis. We computed variance of each lncRNA and protein-coding gene, and ranked them in descending order. We only picked top 5000 genes for coexpression analysis.

**Functional annotation**

g:Profiler was used for functional annotation and enrichment analysis (Reimand et al., 2007). g:SCS threshold was used for multiple testing correction.

**Results**

**Identification of Cx. quinquefasciatus lncRNAs**

In this study, we used 31 public RNA-seq datasets which were composed of approximately 849 million reads as raw inputs for lncRNA prediction. Public RNA-seq datasets used in the study were listed in Table S1. Pipeline for lncRNA prediction can be found in Figure 1. Each library (paired or single-end) was individually mapped against Cx. quinquefasciatus genome (CpipJ2, VectorBase) using splice-aware aligner tool, Hisat2 (Kim et al., 2015). Each alignment file was then used for transcriptome assembly using Stringtie, and all 31 Stringtie output files were merged using Stringtie merge option (Pertea et al., 2015). We then compared our merged transcriptome assembly with reference annotation (CpipJ2.4, VectorBase) using gffcompare (https://github.com/gpertea/gffcompare). A total of 64 689 transcripts that correspond to 37 532 loci were assembled by Stringtie. Out of 64 689, 45 238 of them were multi-exon transcripts. We managed to recover all Cx.
Fig. 1 IncRNA identification. Summary of IncRNA prediction pipeline.
**IncRNA in Cx. quinquefasciatus**

*quinquefasciatus* known transcripts (19,859) using Stringtie; hence, validating our transcriptome assembly. After transcriptome assembly, we then performed stringent transcript filtering steps (Fig. 1). Transcripts of less than 200 bp were discarded. We only chose transcripts with class code “i,” “u,” and “x,” all of which were given by gffcompare. Class code “i,” “u,” and “x” denote intronic, intergenic, and antisense to reference annotation respectively (https://github.com/gpertea/gffcompare). We discovered that a total of 11,467 transcripts were denoted as either “i,” “u,” or “x,” and all of them were at least 200 bp in length. We then used TransDecoder (Haas et al., 2013) to predict long open-reading frame (ORF) within these transcripts. Out of 11,467, 2,994 transcripts were discarded because they were predicted to have ORF of more than 100 amino acids.

The remaining 8,473 transcripts were then evaluated for coding potential using CPC2 (Kang et al., 2017), PLEK (Li et al., 2014), and CNCI (Sun et al., 2013), and those classified as “noncoding” by all three algorithms were retained for downstream analysis. A total of 7,413 transcripts were found to be predicted as noncoding by all three softwares. To avoid false positive prediction, we used BLASTX (E-value cut-off 10\(^{-3}\)) to find sequence similarity of those 7,413 transcripts with known proteins in Swissprot. The number of transcripts that have no significant BLASTX are 7,025. Finally, we removed unstranded transcripts, and transcripts located within less than 2 kb scaffold-end range on the same strand. By applying this pipeline, we identified a total of 4,763 novel lncRNA transcripts, which derived from 4,037 distinct loci in the genome. From these 4,763 lncRNA transcripts, 3,591 of them were intergenic, while the remaining 569 and 603 transcripts were intronic and antisense to reference gene respectively. List of lncRNAs *Cx. quinquefasciatus* and their corresponding genomic loci can be found in Table S3. We randomly selected 5 novel lncRNAs, and validated them through reverse-transcribed PCR (RT-PCR) using specific primers that span exon-exon junction (Fig. S1).

### Structural features of *Cx. quinquefasciatus* lncRNAs

Studies in other species showed that, when compared with protein-coding genes, lncRNAs displayed certain characteristics: shorter in size, lower GC content, low sequence conservation, and lower coding potential (Nam & Bartel, 2012; Young et al., 2012; Jenkins et al., 2015; Etebari et al., 2016; Wu et al., 2016; Liu et al., 2017; Akay et al., 2019; Azlan et al., 2019a, 2019b). We asked if this is also true for *Cx. quinquefasciatus* lncRNAs.

We discovered that, in general, *Cx. quinquefasciatus* lncRNAs were significantly shorter than protein-coding transcript (Kolmogorov–Smirnov test [KS test] \(P < 2.2 \times 10^{-16}\), Fig. 2A). Mean of lncRNAs and protein-coding transcripts were 699 bp and 1,377 bp, respectively. Meanwhile, lncRNAs had a median size of 463 whereas the median length protein-coding transcripts were 1,047 bp.

We then checked the GC content of lncRNAs and compared them with protein-coding transcripts, mRNAs. We also included other noncoding sequences sequences such as 3'UTR and 5'UTR. We found that, in general, noncoding sequences (lncRNAs, 3'UTR, and 5'UTR) have slightly lower GC content than mRNA (Fig. 2B). Median GC content of lncRNA, 3'UTR, 5'UTR, and mRNA were 42%, 35%, 45%, and 56%, respectively. In addition, mean GC content of noncoding sequences were lower than mRNA (mRNA: 55%, lncRNA: 41%, 3'UTR: 44%, 5'UTR: 36%). Beside GC content, we also examined coding potential of noncoding sequences and compared them with mRNA transcripts. CPC2 software showed that mRNA had the highest median and mean coding probability score (mean: 0.65 and median: 0.672). Median score for lncRNA, 5'UTR, and 3'UTR were 0.02, 0.004, and 0.007, respectively. Meanwhile, average probability of lncRNA, 5'UTR, and 3'UTR were 0.047, 0.023, and 0.027, respectively.

Previous studies reported that lncRNAs have low sequence conservation (Etebari et al., 2016; Wu et al., 2016; Azlan et al., 2019b). Here, we assessed sequence conservation of the newly predicted *Cx. quinquefasciatus* lncRNAs using BLASTN against other insect genomes such as *Ae. aegypti*, *Aedes albopictus* (*Ae. albopictus*), *Anopheles gambiae* (*An. gambiae*), and *Drosophila melanogaster* (*D. melanogaster*). To determine the level of nucleotide similarity, we used bit score from BLASTN algorithm (Etebari et al., 2016; Azlan et al., 2019b). Similar to previous reports (Etebari et al., 2016; Wu et al., 2016; Azlan et al., 2019b), *Cx. quinquefasciatus* lncRNAs, in general, had lower sequence conservation than protein-coding transcripts (Fig. 3A).

### Cx. quinquefasciatus lncRNAs are lowly expressed and temporally specific

To assess the dynamics of lncRNA expression across different time points, we analyzed transcriptome data by Reid et al. (2015), which generated RNA-seq libraries of female mosquitoes *Cx. quinquefasciatus* at 2, 12, 24, 36, 48, 60, and 72 h posteclosion (Reid et al., 2015). We reanalyzed this time-course transcriptome data to compute
the expression level of *Cx. quinquefasciatus* lncRNAs. We discovered that, across all time points, the expression level of lncRNA was lower than protein-coding genes (KS test $P < 2.2e-16$, Fig. 4A). RNA-seq libraries generated by Reid et al. (2015) were poly-A selected; hence, the lncRNA expression pattern observed (Fig. 4A) was exclusively for polyadenylated and RNAP II expressed lncRNA transcripts. It was reported that certain lncRNAs can also be expressed by RNAP III (Massone et al., 2012; Bhat et al., 2016).

Previous studies reported that lncRNA in various organisms displayed a more temporally and tissue-specific pattern in expression (Cabili et al., 2011; Wu et al., 2016; Azlan et al., 2019a, 2019b). We then checked if *Cx. quinquefasciatus* lncRNAs have the same characteristics. To investigate expression specificity of lncRNAs, we employed Jensen–Shannon (JS) score, that ranges between zero to one (Cabili et al., 2011). This specificity score quantifies the similarity in expression value across different time points. Genes having JS score of one

---

**Fig. 2** Features of *Cx. quinquefasciatus* lncRNAs. (A) Distribution of transcript length. (B) GC content. (C) Coding potential from CPC2 algorithm.

**Fig. 3** Sequence conservation of *Cx. quinquefasciatus* lncRNAs. (A) Boxplot of similarity bit score using BLASTN algorithm. (B) The Venn diagram shows the number of *Cx. quinquefasciatus* with E-value below 1e-5 in other species using BLASTN algorithm.
indicate the extreme case in which it is only expressed in one time point, whereas those having a score close to zero are ubiquitously expressed in all time points with relatively similar value of expression (Cabili et al., 2011). Similar to previous reports, the expression of *Cx. quinquefasciatus* lncRNAs was more temporally specific than protein-coding genes (KS test $P < 2.2e^{-16}$, Fig. 4B). We found that 58% of lncRNA transcripts had perfect specificity score of one, whereas only 15% of protein-coding transcripts had JS score of one.

**Cx. quinquefasciatus** lncRNAs have potential roles in blood meal acquisition

To investigate the possible involvement of lncRNA in blood meal acquisition, we analyzed transcriptome data by Reid *et al.* (2015), which aimed to identify genes, especially protein-coding genes, that might be crucial for taking and processing the blood meal in adult female *Cx. quinquefasciatus* (Reid *et al.*, 2015). They generated paired-end RNA-seq libraries of seven different time points after eclosion (2, 12, 24, 36, 48, 60, and 72 h) of adult females in order to examine temporal gene expression of posteclosion and prior to blood-feeding (Reid *et al.*, 2015). They showed that adult females *Cx. quinquefasciatus* need at least 48 h posteclosion before they can take their first blood meal (Reid *et al.*, 2015). Therefore, prior to 48 h, it was hypothesized that genes necessary for blood meal acquisition would be differentially expressed, and after 48 h, the differentially expressed genes would be important for processing the blood meal. We hypothesized that, besides protein-coding genes, lncRNAs in *Cx. quinquefasciatus* would also play regulatory roles in blood-feeding process.

Weighted Gene Correlation Network Analysis (WGCNA) has always been used to predict potential roles of lncRNAs in many species (Zhang & Horvath, 2005; Langfelder & Horvath, 2008; Wu *et al.*, 2016; Azlan *et al.*, 2019a). WGCNA can be used to search for clusters or modules of highly correlated genes, which in this case, correlation between lncRNAs and protein-coding genes (Zhang & Horvath, 2005; Langfelder & Horvath, 2008). To predict the potential roles of lncRNAs in blood meal acquisition, we used RNA-seq data from Reid *et al.* (2015), and performed WGCNA analysis on lncRNAs and protein-coding genes (Reid *et al.*, 2015). Using WGCNA, we searched for significant correlation between lncRNAs and a set of annotated protein-coding genes in *Cx. quinquefasciatus*. Then, the protein-coding genes that were significantly correlated with lncRNAs were passed through g:Profiler for functional enrichment analysis. WGCNA analysis yielded a total of 21 modules, and 7 of them exhibited strong and statistically significant correlations with lncRNAs.
Our functional enrichment analysis using g:Profiler revealed that numerous GO terms were significantly associated with each module (Table S4). Several GO terms that were enriched in this study such as regulation of
metabolic processes, cofactor binding, oxidoreductase activity, and enzyme regulator activity (Table S4). Two modules were found to be significantly associated with 2 h posteclosion (Table 1). Functional annotation using g:Profiler of these two modules revealed that lncRNAs at 2 h had potential roles in diverse processes such as compound metabolic pathways, biosynthesis, catabolic process, and proteolysis (Table S4).

Previous studies reported that genes encoding vitellogenin, chymotrypsin, and serine proteases were differentially expressed during blood-feeding in blood-sucking mosquitoes such as Cx. quinquefasciatus (Reid et al., 2015), An. sinensis (Chen et al., 2017), Ae. aegypti (Bonizzoni et al., 2011; Sim et al., 2012), and An. gambiae (Dana et al., 2005; Rinker et al., 2013). In this study, several Cx. quinquefasciatus lncRNAs were found to be significantly correlated with genes that encode vitellogenin (CPIJ003531), chymotrypsin (CPIJ003915), and serine proteases (CPIJ004675). At 12 h posteclosion, several genes such as salivary proteins and cytochromes were found to be differentially expressed (Reid et al., 2015). We next asked if the same set of genes with similar functions could be found in our magenta module that was highly associated with 12 h of genes with similar functions could be found in our genome using a total of 31 relatively high-depth published transcriptomes.

Discussion

In this study, we performed genome-wide identification and characterization of lncRNAs in Cx. quinquefasciatus genome using a total of 31 relatively high-depth publicly available RNA-seq libraries. We presented a set of 4763 novel lncRNAs transcripts which derived from 4037 loci in Cx. quinquefasciatus genome. RNA-seq libraries used here were generated from many tissues and developmental stages, making our prediction pipeline to be relatively robust and comprehensive. We applied stringent filtering steps in our pipeline to reduce false positive and false negative prediction of noncoding transcripts (Nam & Bartel, 2012; Etebari et al., 2016; Wu et al., 2016; Azlan et al., 2019a, 2019b). The filtering steps include using more than one coding potential assessment algorithms and removing transcripts with no strand information (Fig. 1). Taken together, lncRNA prediction pipeline used in this study is comparably equivalent to previous studies in other organisms, and it is relatively robust and comprehensive.

Analysis of genomic features revealed that lncRNAs identified in our study showed known characteristics of lncRNAs from other species. The characteristics are shorter in size, low sequence conservation, and lower GC content. The majority of Cx. quinquefasciatus lncRNAs shared high degree of sequence similarity with Ae. albopictus and Ae. aegypti, followed by An. gambiae and D. melanogaster. This lncRNA sequence similarity pattern concurs with previous phylogenetic relationship study on known mosquitoes which showed that Ae. albopictus and Ae. aegypti are more evolutionarily closer to Cx. quinquefasciatus than An. gambiae (Chu et al., 2018). A total of 10 Cx. quinquefasciatus lncRNAs shared high sequence similarity with all insects (Fig. 3B), suggesting that they are evolutionarily conserved and may play vital roles in insect development or basic cellular functions.

Investigation on the expression across different time points posteclosion revealed that, compared to protein-coding genes, Cx. quinquefasciatus lncRNAs were more temporally specific and have lower expression level. This high temporal specificity indicates that lncRNAs have a smaller time window of expression that protein-coding genes. Even though lncRNAs have lower expression that protein-coding genes, their temporal specificity in expression suggests that they putatively execute specific biological functions at specific time point in development. Furthermore, coexpression analysis by WGCNA revealed that Cx. quinquefasciatus lncRNAs were significantly correlated with specific time-point of preblood feeding immediately after eclosion, suggesting that lncRNAs may possess potential roles in blood meal acquisition by adult female mosquitoes. This blood meal acquisition by adult female is physiologically important as it is not only required for reproduction and mating, but it serves as a gateway for pathogens, making Cx. quinquefasciatus one of the most harmful mosquito vectors on Earth.

In this study, several GO terms such as regulation of metabolic processes, cofactor binding, oxidoreductase activity, and enzyme regulator activity were significantly enriched during blood meal acquisition in Cx. quinquefasciatus (Table S4). Similar GO terms were also
reported to be significantly enriched in the studies of transcriptome changes in *Anopheles sinensis* (*An. sinensis*), and *Ae. aegypti* during pre- and postblood feeding (Bonizzi et al., 2011; Chen et al., 2017). *Cx. quinquefasciatus* lncRNAs were also found to be significantly correlated with genes encoding vitellogenin (CPIJ003531), chymotrypsin (CPIJ003915), and serine proteases (CPIJ004675), all of which were found to be involved in blood meal in mosquitoes (Dana et al., 2005; Bonizzi et al., 2011; Sim et al., 2012; Rinker et al., 2013; Reid et al., 2015; Chen et al., 2017). It was previously reported that blood meal acquisition stimulated the vitellogenesis in *Ae. aegypti* female mosquitoes (Valzania et al., 2019). Oogenesis in *Ae. aegypti* was halted before blood feeding, and the process resumed to produce mature eggs during vitellogenic stage after blood meal acquisition (Valzania et al., 2019). Besides, chymotrypsin and serine proteases were found to be important for digesting proteins in blood meal in the midgut of female mosquitoes (Bian et al., 2008). Taken together, WGCNA and g:Profiler analysis revealed that *Cx. quinquefasciatus* lncRNAs were significantly correlated with protein-coding genes which were involved in blood feeding. Therefore, results generated in this study suggest an important participation of noncoding genes in blood meal acquisition in *Cx. quinquefasciatus*.

However, the data presented here is simply descriptive, and it does not provide empirical evidence via experimentation on the specific mechanisms of lncRNA functions in blood meal acquisition. We believed that results generated in this study can be a starting point for dissecting the mechanism of *Cx. quinquefasciatus* lncRNAs functions. Further studies are required to investigate the mechanism of lncRNA actions inside cells during blood feeding in *Cx. quinquefasciatus*. In conclusion, this research provides the first annotation and characterization of *Cx. quinquefasciatus* lncRNAs, and their putative roles in blood-feeding. We believed that data presented here will be a valuable resource for genetics and molecular studies of ncRNAs in *Cx. quinquefasciatus*.

**Acknowledgments**

We would like to thank all our collaborators and colleagues for the discussion and the work conducted in this lab. We would especially like to thank the Vector Control Research Unit, Universiti Sains Malaysia for providing the mosquito samples. This study was funded by the Universiti Sains Malaysia Research University Grant (1001/PBIOLOGI/8011604) and MOSTI Science-Fund (305/PBIOLOGI/613238).

© 2020 Institute of Zoology, Chinese Academy of Sciences, 28, 917–928
Chu, H., Li, C., Guo, X., Zhang, H., Luo, P., Wu, Z. et al. (2018) The phylogenetic relationships of known mosquito (Diptera: Culicidae) mitogenomes. Mitochondrial DNA Part A, 29, 31–35.

Clark, M.B. and Mattick, J.S. (2011) Long noncoding RNAs in cell biology. Seminars in Cell & Developmental Biology, 22, 366–376.

Coller, J. and Parker, R. (2005) General translational repression by activators of mRNA decapping. Cell, 122, 875–886.

Dana, A.N., Hong, Y.S., Kern, M.K., Hillenmeyer, M.E., Harker, B.W., Lobo, N.F. et al. (2005) Gene expression patterns associated with blood-feeding in the malaria mosquito Anopheles gambiae. BMC Genomics, 6, 5.

Djuranovic, S., Nahvi, A. and Green, R. (2012) miRNA-mediated gene silencing by translational repression followed by mRNA deadenlyation and decay. Science, 336, 237–240.

Etebari, K., Asad, S., Zhang, G. and Asgari, S. (2016) Identification of Aedes aegypti long Intergenic non-coding RNAs and their association with Wolbachia and Dengue virus infection. PLoS Neglected Tropical Diseases, 10, e0005069.

Etebari, K., Hegde, S., Saldaña, M.A., Widen, S.G., Wood, T.G., Asgari, S. et al. (2017) Global transcriptome analysis of Aedes aegypti mosquitoes in response to Zika virus infection. mSphere, 2, e00456-17.

Farajollahi, A., Fonseca, D.M., Kramer, L.D. and Marm Kilpatrick, A. (2011) “Bird biting” mosquitoes and human disease: A review of the role of Culex pipiens complex mosquitoes in epidemiology. Infection, Genetics and Evolution, 11, 1577–1585.

Göertz, G., Miesen, P., Overheul, G., van Rij, R., van Oers, M. and Pijlman, G. (2019) Mosquito small RNA responses to Wolbachia infection profiles of pre blood-fed adult females immediately following Wolbachia infection. PLoS Neglected Tropical Diseases, 13, e0007229.

Göransson, M., Johansson, I., Persson, H., Olofsson, J. and Durkan, C. (2015) De novo transcript sequence reconstruction from RNA-seq reads. Nature Biotechnology, 33, 290–295.

Göransson, M., Persson, H., Olofsson, J. and Durkan, C. (2017) Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nature Methods, 14, 417–419.

Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559.

Li, A., Zhang, J. and Zhou, Z. (2014) PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on improved k-mer scheme. BMC Bioinformatics, 15, 311.

Liu, F., Guo, D., Yuan, Z., Chen, C., Xiao, H. (2017) Genome-wide identification of long non-coding RNA genes and their association with insecticide resistance and metamorphosis in diamondback moth, Plutella xylostella. Scientific Report, 7, 1–11.

Massone, S., Ciarlo, E., Vella, S., Nizzari, M., Florio, T., Russo, C. et al. (2012) NDM29, a RNA polymerase III-dependent non-coding RNA, promotes amyloidogenic processing of APP and amyloid β secretion. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1823, 1170–1177.

Miesen, P., Ivens, A., Buck, A.H. and van Rij R.P. (2016) Small RNA profiling in dengue virus 2-infected aedes mosquito cells reveals viral piRNAs and novel host miRNAs. PLOS Neglected Tropical Diseases, 10, e0004452.

Nam, J.W. and Bartel, D.P. (2012) Long noncoding RNAs in C. elegans. Genome Research, 22, 2529–2540.

Nchoutpouen, O., Talipouo, A., Djiamou-Djonkam, L., Kopya, E., Ngadjeu, C.S. et al. (2019) Culex species diversity, susceptibility to insecticides and role as potential vector of Lymphatic filariasis in the city of Yaoundé, Cameroon. PLoS Neglected Tropical Diseases, 13, e0007229.

Paradkar, P.N., Duchemin, J.B., Rodriguez-Andres, J., Trinidad, L. and Walker, P.J. (2015) Cullin4 is pro-viral during West Nile Virus infection of Culex mosquitoes. PLoS Pathogens, 11, e1005143.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nature Methods, 14, 417–419.

Peretea, M., Peretea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology, 33, 290–295.

Reid, W.R., Zhang, L., Liu, F. and Liu, N. (2012) The transcriptome profile of the mosquito Culex quinquefasciatus following permethrin selection. PLoS ONE, 7, e47163.

Reid, W.R., Zhang, L. and Liu, N. (2015) Temporal gene expression profiles of pre blood-fed adult females immediately following eclosion in the southern house mosquito Culex quinquefasciatus. International Journal Biological Sciences, 11, 1306–1313.

Reimand, J., Kull, M., Peterson, H., Hansen, J. and Vilo, J. (2007) g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Research, 35, W193–W200.
A. Azlan et al.

Rezende, T.M.T., Rezende, A.M., Luz Wallau, G., Santos Vasconcelos, C.R., de-Melo-Neto, O.P., Silva-Filha, M.H.N.L. et al. (2019) A differential transcriptional profile by Culex quinquefasciatus larvae resistant to Lysinibacillus sphaericus IAB59 highlights genes and pathways associated with the resistance phenotype. Parasites & Vectors, 12, 407.

Rinker, D.C., Pitts, R.J., Zhou, X., Suh, E., Rokas, A. and Zwiebel, L.J. (2013) Blood meal- induced changes to antennal transcriptome profiles reveal shifts in odor sensitivities in Anopheles gambiae. Proceeding of the National Academy of Sciences USA, 110, 8260–8265.

Rückert, C., Prasad, A.N., Garcia-Luna, S.M., Robison, A., Grubaugh, N.D., Weger-Lucarelli, J. et al. (2019) Small RNA responses of Culex mosquitoes and cell lines during acute and persistent virus infection. Insect Biochemistry and Molecular Biology, 109, 13–23.

Shin, D., Civana, A., Acevedo, C. and Smartt, C.T. (2014) Transcriptomics of differential vector competence: West Nile virus infection in two populations of Culex pipiens quinquefasciatus linked to ovary development. BMC Genomics, 15, 513.

Sim, S., Ramirez, J.L. and Dimopoulos, G. (2012) Dengue virus infection of the Aedes aegypti salivary gland and chemosensory apparatus induces genes that modulate infection and blood-feeding behavior. PLoS Pathogens, 8, e1002631.

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C. et al. (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Research, 41, e166–e166.

Taparia, T., Ignell, R. and Hill, S.R. (2017) Blood meal induced regulation of the chemosensory gene repertoire in the southern house mosquito. BMC Genomics, 18, 393.

Ulitsky, I. and Bartel, D.P. (2013) lincRNAs: Genomics, evolution, and mechanisms. Cell, 154, 26–46.

Valzania, L., Mattee, M.T., Strand, M.R. and Brown, M.R. (2019) Blood feeding activates the vitellogenic stage of oogenesis in the mosquito Aedes aegypti through inhibition of glycogen synthase kinase 3 by the insulin and TOR pathways. Developmental Biology, 454, 85–95.

Wang, L., Wang, S. and Li, W. (2012) RSeQC: quality control of RNA-seq experiments. Bioinformatics, 28, 2184–2185.

Wu, Y., Cheng, T., Liu, C., Liu, D., Zhang, Q., Long, R. et al. (2016) Systematic identification and characterization of long non-coding RNAs in the silkworm, Bombyx mori. PLoS ONE, 11, e0147147.

Young, R.S., Marques, A.C., Tibbit, C., Haerty, W., Bassett, A.R., Liu, J.L. et al. (2012) Identification and properties of 1,119 candidate lncRNA loci in the Drosophila melanogaster genome. Genome Biology and Evolution, 4, 427–442.

Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. Statistical Application in Genetics and Molecular Biology, 4, 17.

Zittra, C., Flechl, E., Kothmayer, M., Vitecek, S., Rossiter, H., Zechmeister, T. et al. (2016) Ecological characterization and molecular differentiation of Culex pippin complex taxa and Culex torrentium in eastern Austria. Parasites & Vectors, 9, 197.

Manuscript received January 30, 2020
Final version received June 20, 2020
Accepted June 22, 2020

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1 PCR products in 1.5% agarose gel.
Table S1 List of 31 publicly available RNA-seq libraries used in this study.
Table S2 Primers used in this study.
Table S3 List of lncRNAs in Cx. quinquefasciatus genome.
Table S4 Gene ontology and enrichment from g:Profiler.