



# In silico identification, characterization, and expression analysis of RNA recognition motif (RRM) containing RNA-binding proteins in *Aedes aegypti*

Melveetil Kishor Sumitha<sup>1</sup> · Mariapillai Kalimuthu<sup>1</sup> · Murali Aarthy<sup>1</sup> · Rajaiah Paramasivan<sup>1</sup> · Ashwani Kumar<sup>2</sup> · Bhavna Gupta<sup>1</sup>

Received: 14 January 2023 / Accepted: 4 September 2023 / Published online: 22 September 2023  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

RNA-binding proteins (RBPs) are the proteins that bind RNAs and regulate their functioning. RBPs in mosquitoes are gaining attention due to their ability to bind flaviviruses and regulate their replication and transmission. Despite their relevance, RBPs in mosquitoes are not explored much. In this study, we screened the whole genome of *Aedes aegypti*, the primary vector of several pathogenic viruses, and identified the proteins containing RNA recognition motif (RRM), the most abundant protein domain in eukaryotes. Using several in silico strategies, a total of 135 RRM-containing RBPs were identified in *Ae. aegypti*. The proteins were characterized based on their available annotations and the sequence similarity with *Drosophila melanogaster*. *Ae. aegypti* RRM-containing RBPs included serine/arginine-rich (SR) proteins, polyadenylate-binding proteins (PABP), heteronuclear ribonucleoproteins (hnRNP), small nuclear ribonucleoproteins (snRNP), splicing factors, eukaryotic initiation factors, transformers, and nucleolysins. Phylogenetic analysis revealed that the proteins and the domain organization are conserved among *Ae. aegypti*, *Bombyx mori*, and *Drosophila melanogaster*. However, the gene length and the intron-exon organization varied across the insect species. Expression analysis of the genes encoding RBPs using publicly available RNA sequencing data for different developmental time points of the mosquito life cycle starting from the ovary and eggs up to the adults revealed stage-specific expression with several genes preferentially expressed in early embryonic stages and blood-fed female ovaries. This is the first database for the *Ae. aegypti* RBPs that can serve as the reference base for future investigations. Stage-specific genes can be further explored to determine their role in mosquito growth and development with a focus on developing novel mosquito control strategies.

**Keywords** *Aedes aegypti* · RNA-binding proteins · RNA recognition motif (RRM) · Mosquito growth and development · Stage-specific expression

## Introduction

RNA-binding proteins (RBPs) are the proteins that bind RNAs and regulate their life cycle starting from synthesis to decay. RNAs after transcription are associated with RBPs.

RBPs and other cis-regulatory elements regulate RNA processing, transportation, localization, stability, modifications, translation, and degradation (Glisovic et al. 2008). Due to their role in several aspects of RNA biology, Diaz-Munoz and Turner in 2018 have aptly described RBPs as “readers, writers, editors, and erasers of the transcriptome” where “RNA writers” include the proteins involved in splicing, capping, and polyadenylation, “readers” are the RBPs involved in subcellular localization and translation, “editors” constitute methyltransferases and deaminases, and “erasers” include the destabilizing factors and nucleases that cause RNA instability and degradation (Díaz-Muñoz and Turner 2018). The regulation of all these aspects after transcription is known as post-transcriptional gene regulation (PTGR), and RBPs play a major role in PTGR.

Handling Editor: Una Ryan

✉ Bhavna Gupta  
bhavnaguptaster@gmail.com

<sup>1</sup> ICMR-Vector Control Research Centre (VCRC),  
Field Station, 4, Sarojini Street Chinna Chokkikulam,  
Madurai 625002, India

<sup>2</sup> ICMR-Vector Control Research Centre (VCRC), Puducherry,  
India

RBPs make up a significant part of the proteome in eukaryotes. They regulate cell differentiation and homeostasis during developmental stages, environmental challenges, and diseases/infections (Wurth 2012; Brinegar and Cooper 2016; Li et al. 2020). Several RBPs have been associated with different types of cancers and genetic disorders (Wurth 2012; Li et al. 2020). RBPs also play a role in the growth and development of plants (Muleya and Marondedze 2020), animals (Kerner et al. 2011), bacteria (Holmqvist and Vogel 2018), and insects (Norvell et al. 1999; Yoon et al. 2018). Interestingly, many RBPs have been found to regulate early embryo development in *Drosophila*. For example, mutations in the *Squid* gene disrupted Gurken-dependent dorsoventral patterning during oogenesis in *Drosophila* resulting in female sterility (Norvell et al. 1999; Gamberi et al. 2006). Smaug, another RBP, is involved in maternal to zygotic transition (Benoit et al. 2009). Localization and translational regulation of maternal transcripts during oogenesis in *Drosophila* rely on RBPs (Bansal et al. 2020). Staufen, a double-stranded RBP, has been identified as a key player in RNAi in certain Coleopteran insects (Yoon et al. 2018). However, RBPs in mosquitoes are not much explored except in reports where mosquito RBPs were found to control the replication of flaviviruses (Diosa-Toro et al. 2020; Yeh et al. 2022). RNA viruses use host RBPs for their replication and survival, and such RBPs can act as antivirals. For example, AeStaufen reduced genomic and sub-genomic flaviviral RNA copies in several mosquito tissues, including the salivary glands indicating its role in dengue transmission (Yeh et al. 2022). Taschuk et al. (2020) described the antiviral role of DDX56, Dead box helicase in *Drosophila* and human cell lines (Taschuk et al. 2020). Thus, the identification and characterization of mosquito RBPome would be helpful for such studies.

RBPs are generally identified by the presence of different RNA-binding structural motifs such as RNA recognition motif (RRM), the zinc fingers, K homology (KH), and others. These motifs locate and interact with their specific RNA targets. Among these motifs, RRM is one of the most abundant and well-characterized RNA-binding motifs in eukaryotes (Venter et al. 2001). RRM is found in all forms of life, including bacteria and viruses. RRM-containing proteins are involved in several post-transcriptional aspects like alternative splicing, translation, degradation, and processing of precursor ribosomal RNA (pre-rRNA). RRM-containing proteins form a prominent class of RBPs performing important biological functions.

In this study, we screened *Ae. aegypti* genome and identified the RRM-containing proteins using various in silico approaches. We also carried out expression analysis of the genes encoding the RBPs to elucidate their association with mosquito growth and developmental stages using stage-specific transcriptome data (Akbari et al. 2013).

## Material and methods

### Identification and characterization of RRM-containing proteins in *Ae. aegypti*

To search for RRM-containing proteins in *Ae. aegypti* genome, we used multiple in silico approaches. First, “RRM and *Ae. aegypti*” was used as the keyword in GenBank to retrieve proteins containing RRM. Second, 134 RRM-containing proteins of *D. melanogaster* (Gamberi et al. 2006) were used as a protein query against *Ae. aegypti* in BLASTp search. Third, the UniProt entries for *Ae. aegypti* for RRM-containing proteins were retrieved. Fourth, RRM profiles from PROSITE for IDs PDOC00030, PDOC51472, and PDOC51939 were used as a query for BLASTp against *Ae. aegypti* database with default parameters. The proteins obtained from each strategy were carefully examined to remove duplicates, and a list of unique proteins was obtained. This list was further used in the BLASTp search against *Ae. aegypti* protein database to retrieve proteins that might have been missed from earlier searches.

The list of proteins obtained was analyzed using SMART and Conserved Domain (CD) databases to confirm the presence of RRM. The domain organization for each protein was retrieved from the SMART database. The chromosomal locations, protein length, number of transcripts, and protein IDs were obtained from the VectorBase (<https://vectorbase.org/>). Functional annotation was done using Blast2GO (Conesa et al. 2005), and the descriptions for each protein were based on available annotations in Genbank/VectorBase or sequence similarity with *D. melanogaster* (> 30% identity).

### Phylogenetic analysis, chromosomal mapping, and gene structure

Phylogenetic analysis of full-length amino acid sequences of *Ae. aegypti* RBPs was aligned using CLUSTAL W, and neighbour-joining tree with 500 bootstraps was constructed using MEGA7.0 (Kumar et al. 2016). The most similar proteins in *D. melanogaster* and *Bombyx mori* were retrieved from VectorBase (<https://vectorbase.org/vectorbase/app>) and Ensembl (<https://metazoa.ensembl.org/index.html>), respectively. Annotations were added to the tree using the iTOL server (Letunic and Bork 2021). Chromosomal mapping of the genes on *Ae. aegypti* chromosomes was done using Mapchart 2.32 (Voorrips 2002). Intron-exon gene structure schematics were generated using GSDS an online server (Hu et al. 2015).

### Expression profiling of the genes encoding RRM-containing proteins

RNA sequencing data from 41 different developmental time points throughout the *Ae. aegypti* life cycle from eggs up to

the adult stage were obtained from Sequence Read Archive (SRA) a public repository database, with accession number SRP026319 (Akbari et al. 2013). Transcriptomic data were mapped to the *Ae. aegypti* AagL5.0 reference genome using HISAT2 (Kim et al. 2019) to identify their genomic positions. The reads mapped to each gene were assembled and quantified using Cufflinks (Trapnell et al. 2012). The read count was quantified as the number of fragments per kilobase of exons per million mapped fragments (FPKM). The FPKM for *Ae. aegypti* RRM-containing RBPs were used for clustering and heatmap generation using Morpheus, (<https://software.broadinstitute.org/morpheus>).

## Results and discussion

### Overview of RBPs in *Ae. aegypti*

Using multiple in silico strategies, *Ae. aegypti* genome was screened, and 135 RRM-containing RBPs were identified (Supplementary Table 1) which is comparable to the number of RBPs in *D. melanogaster* ( $n = 134$ ) (Gamberi et al. 2006; Sysoev et al. 2016) and *B. mori* ( $n = 123$ ) (Wang and Zhou 2009). Orthologs of 133 RBPs were identified in *D. melanogaster* (Supplementary Table 1). Annotations were taken as available in Genbank or VectorBase, and the remaining unannotated genes were putatively characterized based on sequence similarity with *D. melanogaster*. One hundred twenty-six proteins could be described (Supplementary Table 1), while nine RBPs could not be assigned any description. Six of these nine RBPs contained multiple RRMs: AAEL022693 and AAEL012243 had six RRM copies; AAEL019864 and AAEL023907 contained five copies; AAEL022113 and AAEL004699 had four copies. Three of the nine RBPs contained a single RRM (AAEL022876, AAEL025927, AAEL019879).

RRM is often found as a single copy or in multiple copies and is also commonly found in combination with other motifs (Maris et al. 2005; SenGupta 2013; Loerch and Kielkopf 2015). Out of 135 RBPs, 41 (30%) had a single RRM, and 52 proteins had multiple RRMs (37%). The remaining 42 RBPs contained RRM along with other motifs (33%). Zinc finger (Znf) was the most common as earlier seen in many other organisms (Maris et al. 2005; Mahalingam and Walling 2020). Znf is a small protein motif with multiple finger-like protrusions that make tandem contacts with their target molecule. Many superfamilies of Znf have binding affinities to DNA/RNA/proteins based on their sequence and structure. For example, CCHHs are known to bind double-stranded DNA/RNA, and CCCH and CCHC types bind single-stranded RNA (Summers 1991; Iuchi 2001; Michel et al. 2003; Wang et al. 2021). A total of 13 Znf motifs including five C3H1, three C2HC, two C2H2,

one ring finger, and three RBZ were observed in 11 proteins. The ring finger is involved in protein-protein interaction, and the Znf-RBZ is related to the ubiquitin-binding function (Lorick et al. 1999).

Other motifs included PWI, KH, and G-patch, which are also known for their RNA-binding affinities (Szymczynna et al. 2003; Dong et al. 2004; Valverde et al. 2008; Aksaas et al. 2011; Zhang et al. 2016). Apart from this, motifs like SURP, SAP, LA, SPOC, RPR, HAT, Lsm\_interact, methyltransferase, muHD, and CID are involved in transcriptional and translational processes. For example, SURP is found in splicing regulatory proteins (Kuwasako et al. 2006). SAP motif has been identified in proteins involved in DNA repair, RNA processing, and apoptotic chromatin degradation (Aravind and Koonin 2000). Three proteins contained the LA motif. LA motif is involved in the maturation of RNA polymerase III transcripts (Alfano et al. 2004; Dong et al. 2004). LA motif also recognizes mRNAs with a 5'-terminal oligopyrimidine motif (5'TOP) important for protein synthesis (Pellizzoni et al. 1997). One protein had RRM in combination with the cleavage stimulation factor (CSTF) which is known for 3'-end cleavage and polyadenylation of pre-mRNAs (Mandel et al. 2008). HAT, a helical repeat motif, is involved in RNA-binding and RNA remodeling (Hammani et al. 2012). Nuclear transport-like factor (NTF2) was identified in one *Ae. aegypti* RBP. NTF2 functions as a cytosolic factor for nuclear import and also interacts with nuclear pore complex protein (Paschal and Gerace 1995). The ubiquitin-binding-associated (UBA) motif was found in one of the proteins. UBA mediates ubiquitination (Hurley et al. 2006). The motifs having specific functions may modulate the binding affinity, specificity, and versatility of RBPs. They may also be responsible for locating and interacting with multiple RBPs.

RRMs are also commonly seen in multiple copies within a protein. Fifty-two proteins contained multiple RRMs ranging from two to seven. Among proteins with multiple RRM copies, AAEL021950 contained the maximum number of seven RRMs, and four RBPs, namely, AAEL004075, AAEL012243, AAEL020196, and AAEL022693, contained six RRMs. Some of the RBPs are known to have a specific number of copies. For example, a single RRM is found in spliceosomal U1 70 kDa protein and two in U2 auxiliary factor 65 kDa subunit, three RRM copies are generally seen in ELAV protein, and the polyadenylate binding proteins (PABPs) are known to have four RRM copies (SenGupta 2013). Although the function of each RRM among multiple copies is unknown, they may increase specificity by binding to the long stretch of RNA sequence and functioning cooperatively (Maris et al. 2005). They may also have different specificities and thus diversify the biological functions of the protein. Moreover, it is also possible that all of the RRMs may not be functional or may not be able to bind RNAs (SenGupta 2013).

## Chromosomal mapping and functional enrichment

RBPs were mapped onto the *Ae. aegypti* chromosomes, and 135 RBP encoding genes were found evenly distributed across the length of all the three chromosomes (Fig. 1). The largest chromosome (chr\_2) contained a maximum of 53 genes; chromosome 3 (chr\_3) had 45, and the smallest chromosome (chr\_1) contained 37, the lowest number of genes. The length of proteins varied from 91 to 5680 amino acids including 43 small proteins (< 320 amino acids) and 32 large proteins (> 700 amino acids). The majority (75%) of the proteins that had only one exon were small proteins. The number of predicted transcripts for each protein varied from one to 29, and 56% (76/135) of the total genes had only one transcript.

Gene ontology (GO) analysis with Blast2GO categorized RBPs based on their putative functions. As expected, RNA-binding was the most enriched term among molecular functions, and in biological processes, *Ae. aegypti* RBPs were found to be associated with gene expression regulation, nucleic acid and protein metabolism, macromolecule biosynthesis, cellular component assembly, protein-containing complex, macromolecule modification, and ribonucleoprotein complex. This indicates the involvement of RBPs in several biological processes as seen in several other organisms (Fig. 2).

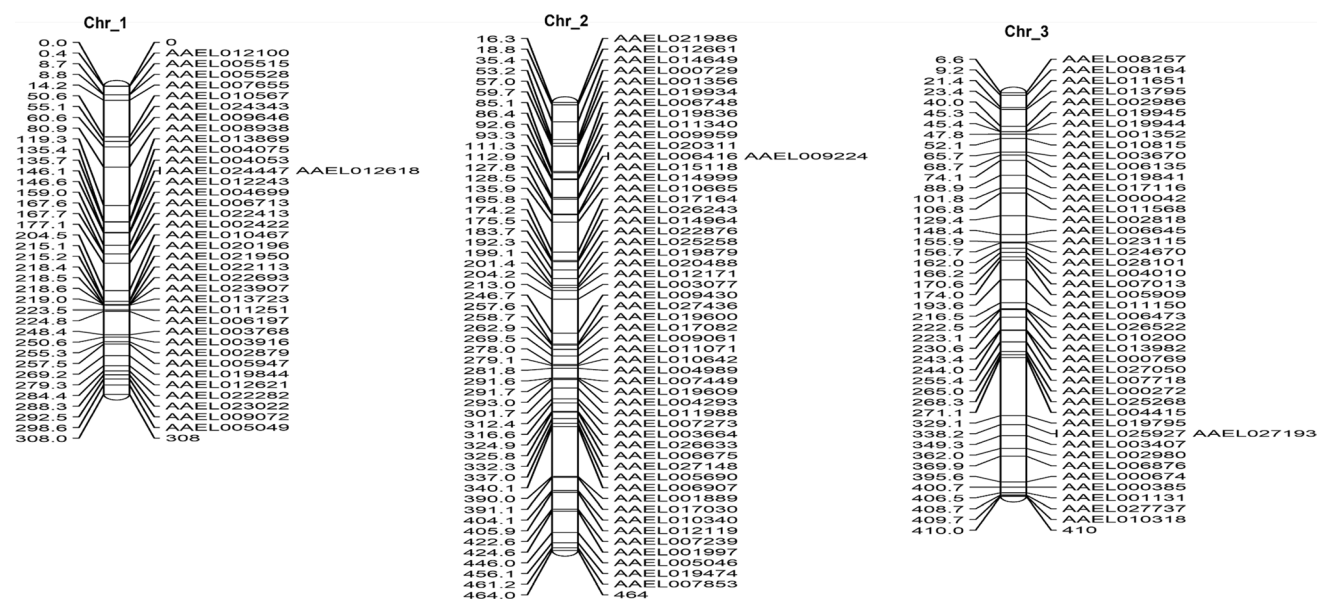
## Expression analysis of RBP genes using RNA sequencing data

Since developmental stage-specific transcriptomes of *Ae. aegypti* were available in a public database for several time

points throughout its life cycle, starting from early embryo up to adult mosquitoes (Akbari et al. 2013), it was possible to explore the stage-specific expression of RBPs in *Ae. aegypti*. From the analysis of RNA seq data, we retrieved FPKM values for 135 RBP encoding genes at 41 developmental time points. All the genes had > 1.0 FPKM in more than one stage and thus were considered expressed. Hierarchical clustering of the gene expression data indicated that many genes exhibited preferentially high expression in the embryonic stages and in blood-fed ovaries (Fig. 3). A subset of RBPs were found highly expressed in early embryonic stages, and interestingly, those RBPs were also highly expressed in blood-fed ovaries (Fig. 3). This indicates that these RBPs might be maternally inherited and play essential roles in embryonic development. Some of these genes included SR proteins, cytoplasmic polyadenylation-binding protein, La protein, gawky protein, boule, and among others. Further characterization of these genes might explain molecular basis of embryogenesis in mosquitoes. Two RBPs (AAEL010665 and AAEL013869) were also identified in the *Ae. aegypti* egg shell proteome analysis (Marinotti et al. 2014) suggesting their possible role in egg development. The detailed characterization of these genes is warranted to identify the candidates for developing new mosquito growth regulators or insecticides.

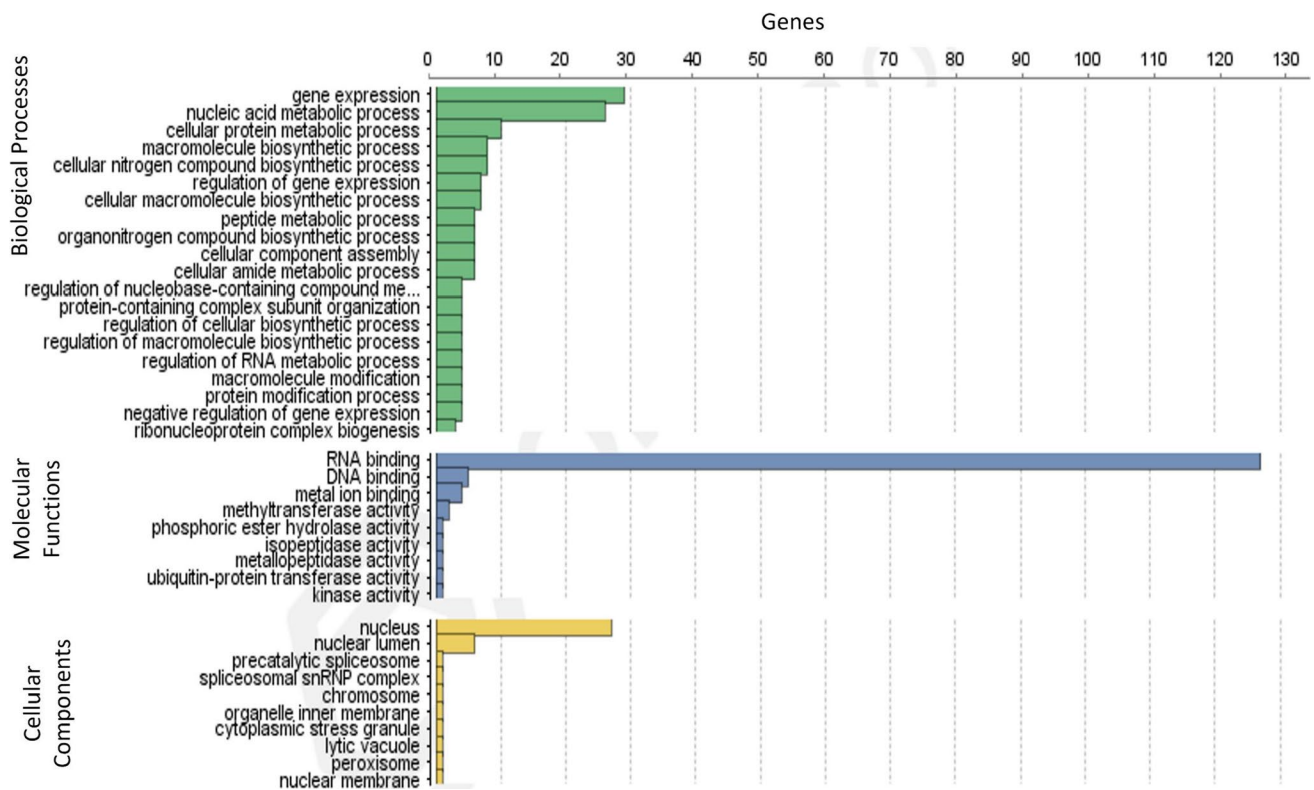
## Phylogenetic relationships, gene structure, and domain organization

RRM-containing proteins belonged to a wide range of functional categories including serine arginine (SR)-rich proteins, heterogeneous nuclear RNP (hnRNPs), small nuclear



**Fig. 1** Chromosomal distribution of the genes encoding RNA-binding proteins containing RNA recognition motif (RRM) across three chromosomes (chr\_1, chr\_2, and chr\_3) of *Aedes aegypti*. The num-

bers on the left of each chromosome are the locations of the genes (in Mb), and the protein IDs are shown on the right side



**Fig. 2** Gene ontology classifications of the biological process, molecular function, and cellular localization of the genes encoding RNA-binding proteins containing RNA recognition motif (RRM) in *Aedes aegypti*. GO distribution for the top 20 levels has been shown

ribonucleoproteins (snRNPs), nucleolysins, polyadenylate-binding proteins (PABP), ELAV-like proteins, and eukaryotic translation initiation factors (Table 1). We analyzed the gene structure, domain organization, and phylogenetics of some of the major *Ae. aegypti* RBP classes and compared with *D. melanogaster* and *B. mori*.

### Spliceosome-associated proteins

Pre-mRNA splicing needs core spliceosomes, cis-sequence elements, and several RBPs (Fredericks et al. 2015; Vuong et al. 2016). Among RBPs, snRNPs form the core components of the spliceosomes, and SR and hnRNPs proteins are the major splicing factors. Thirty-six proteins with a putative role in splicing were identified, including 12 SR proteins, eight hnRNPs, three pre-mRNA splicing factors, eight snRNPs, and five other splicing factors (Supplementary Table 1). Brooks and co-workers (Brooks et al. 2015) reported 56 *Drosophila* RBPs involved in splicing. A BLAST search using 56 proteins against *Ae. aegypti* genome identified 37 proteins containing RRM. These 37 proteins belong to different categories including six SR proteins, two pre-mRNA splicing factors, one each from hnRNP, PABP, and ELAV along with two nucleolysins, and two translation

initiation factors. This is possible, as many RBPs are known to perform multiple functions. However, the functional characterization of *Ae. aegypti* may provide accurate information.

### Serine arginine-rich (SR) proteins

A total of 12 SR proteins were identified in *Ae. aegypti*. SR proteins contain SR repeat regions at the C terminal and RRM at the N terminal. Nine proteins had a single RRM and five had two RRMs. One of the proteins contained Zn along with RRM, and another protein had an RPR motif which helps in protein-protein interaction. SR proteins are known to be a part of spliceosome assembly and regulate the alternate splicing (Jeong 2017), while RRM in the protein binds to the splicing enhancers to regulate processing events and the SR-rich region binds to the other proteins or RNAs.

We analyzed 12 *Ae. aegypti* putative SR proteins along with ten SR proteins from *D. melanogaster* and nine from *B. mori*. Phylogenetic analysis revealed that SR proteins were highly conserved among the three insect species. *Ae. aegypti* proteins clustered closely with their orthologs from both *D. melanogaster* and *B. mori* (Fig. 4). Protein length and domain organization were conserved (Matthews et al. 2018), while gene length and the intron-exon structure were

**Fig. 3** Expression profiles of the genes encoding RNA-binding proteins containing RNA recognition motif (RRM) in *Aedes aegypti* across 41 developmental time points starting from eggs up to adult mosquitoes. These developmental stages include specific time intervals of embryonic development, larval stages, sex-biased expression of male versus female pupae and carcass, blood-fed versus non-blood-fed ovary and female carcass, and testes-specific expression. Colors from white to black represent increasing gene expression. The genes showing high expression in the early embryonic stages and the blood-fed ovaries are shown within brown box

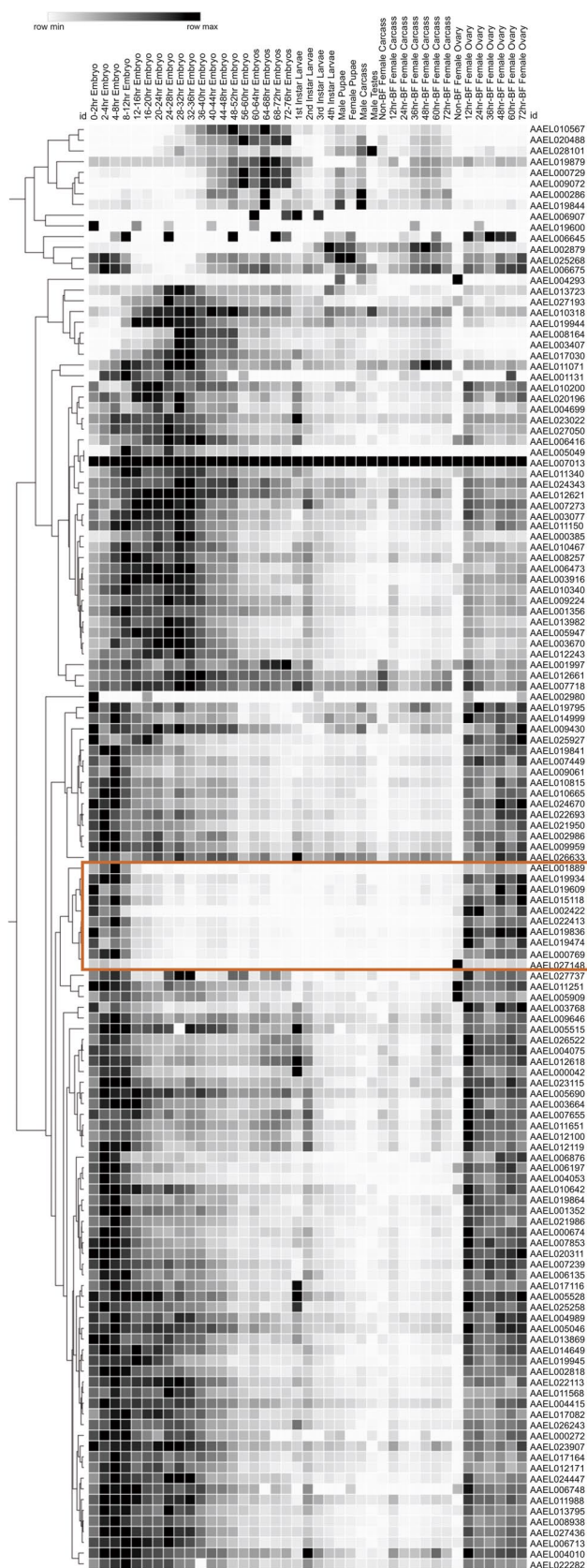
highly variable across the species. The number of introns and their organization also varied among the species. These findings are in corroboration with the findings of genomic comparisons between *Ae. aegypti*, *Anopheles gambiae*, and *D. melanogaster* (Nene et al. 2007) which observed that *Ae. aegypti* genes are generally longer and contain longer introns. This suggested that the genes have evolved across the species but they are highly conserved at the protein level indicating the biological importance of these proteins and the conserved functional roles played by them.

### Small nuclear ribonucleoproteins (snRNPs)

Removal of introns from pre-mRNA (splicing) is a crucial step in eukaryotes' gene expression. This is carried out by a ribonucleoprotein complex called spliceosome which is consisted of snRNPs and various other proteins. Seven snRNPs containing the RRM were identified in *Ae. aegypti*. snRNP is a complex of proteins and snRNA. snRNPs associated with specific U-rich snRNAs are the core components of the spliceosomes involved in splicing (Solymosy and Pollák 1993). The most abundant and well-characterized snRNPs include U1, U2, U4, U5, and U6 which form the major components of spliceosomes (Solymosy and Pollák 1993), and the minor spliceosomes contain U11, U12, U4atc, U5, and U6atac (Will and Lührmann 2011; Matera and Wang 2014). snRNP in *Ae. aegypti* included three U1 snRNPs, one U11/12, two proteins with U2 auxiliary factor, and one with U2-associated SURP motif-containing protein. Phylogenetic analysis of these snRNPs along with *D. melanogaster* and *B. mori* revealed a similar pattern as observed in SR proteins. The proteins and the domain organization were highly conserved, but variations were observed in gene length and intron-exon organization (Fig. 5).

### Heterogeneous nuclear ribonucleoproteins (hnRNPs)

RRM is also found in several hnRNPs. Many hnRNPs are known to regulate alternative splicing and protein components of snRNPs. hnRNPs associate with nascent RNAs

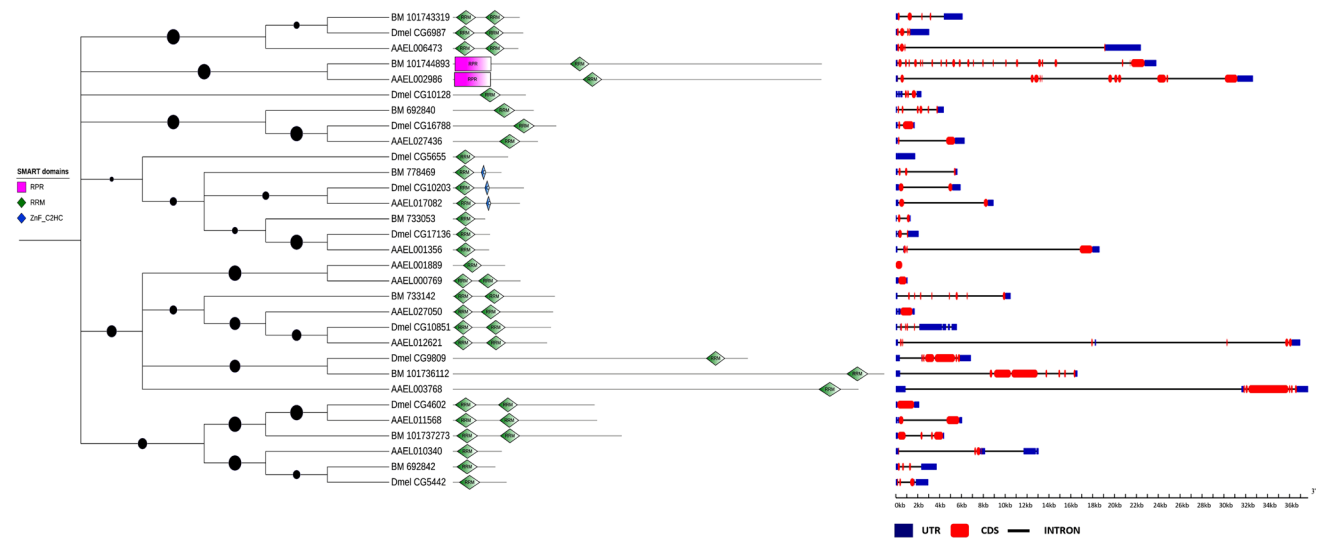


**Table 1** Summary of the functional classes of proteins and the number of genes in each class

Sl. No.	Functional class	Number of genes
1	Serine/arginine rich proteins	12
2	Zinc finger proteins	6
3	Small nuclear ribonuclear proteins	8
4	Pre-mRNA splicing factors	3
5	Heterogeneous nuclear ribonuclear proteins	8
6	Poly(A)-binding protein	5
7	Cytoplasmic polyadenylation element-binding proteins (CPEB)	2
8	ELAV-like proteins	5
9	Eukaryotic translation initiation factor	4
10	Nucleolysin	1
11	Peptidyl-propyl cis-trans isomerase	2
12	Musashi-like proteins	2
13	Splicing proteins	5
14	Transformer proteins	4
15	La proteins	3
16	SRA stem-loop interacting RBPs	3
17	Sex lethal	1
18	Other characterized RBPs	53
19	Unknown RBPs	8

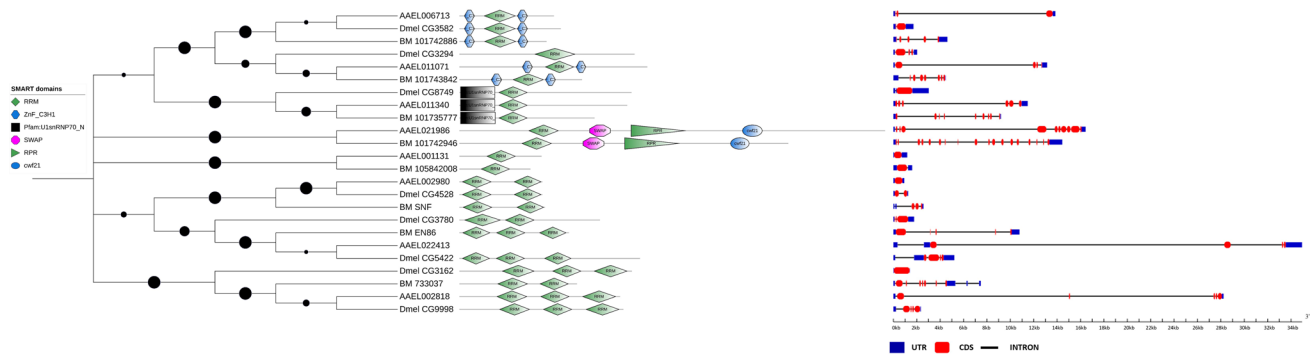
and lead to their export, translation, localization, and stability (Krecic and Swanson 1999; Geuens et al. 2016). A total of eight hnRNPs were identified, five of which encoded two RRM and three proteins had three RRM.

All the proteins are clustered in a phylogenetic tree as per their domain organization (Fig. 6). Two of the *Ae. aegypti* hnRNPs (AAEL005515 and AAEL005049) were clustered with *Drosophila* squid protein. Squid is the most abundant hnRNP protein which is expected to bind most cellular RNAs. *Squid* plays an important role in dorsoventral axis formation during oogenesis by localizing *Gurken*RNA (Norvell et al. 1999; Steinhauer and Kalderon 2005; Cáceres and Nilson 2009). Mutants of Squid have shown its importance in oogenesis and embryo development. For example, homozygous mutants laid severely dorsal eggs that could not mature into adulthood (Kelley 1993; Matunis et al. 1994); complete deletion of *squid* genes lead to lethality (Matunis et al. 1994), while deletion in the first exon of the *squid* resulted in larval death (Matunis et al. 1994). *Ae. aegypti* proteins AAEL005515 and AAEL005049 have not been characterized. The amino acid sequences of both proteins are > 75% identical to the *Drosophila* squid protein (Supplementary Figure 1). However, the gene structure and the overall gene length were different (Fig. 6). Protein-protein interaction analysis with STRING database (Szklarczyk et al. 2021) using default parameters revealed interactions between AAEL005515 and AAEL005049 as well as with the other nine proteins (Fig. 7). Seven of the nine proteins coded for RNA-binding proteins. In particular, AAEL005515 and AAEL005049 were found interacting with all other nine proteins indicating their importance in the network. The sequence conservation across the species indicates functional relevance. However, an in-depth analysis of these genes could provide greater insights and envision.



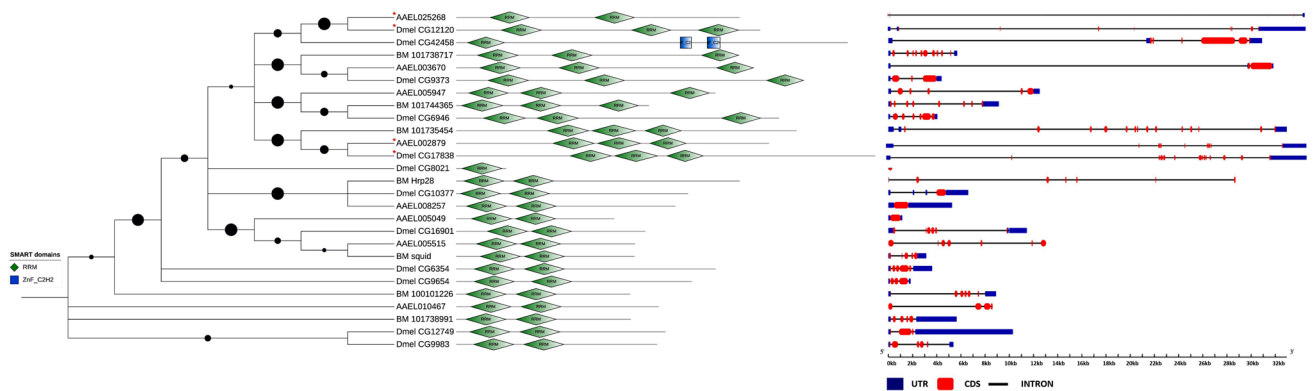
**Fig. 4** Phylogenetic tree of *Aedes aegypti*, *Drosophila melanogaster*, and *Bombyx mori* serine/arginine-rich proteins containing RNA recognition motif (RRM). The tree was constructed using the neighbour joining (NJ) method with 500 bootstraps. The domain organization

of each RBP and the intron-exon architecture of the genes encoding them are also shown. Bootstrap values are represented by the size of the red circle on each branch



**Fig. 5** Phylogenetic tree of *Aedes aegypti*, *Drosophila melanogaster*, and *Bombyx mori* small nuclear RNA nucleoproteins (snRNPs) containing RNA recognition motif (RRM). The tree was constructed using the neighbour-joining (NJ) method with 500 bootstraps. The

domain organization of each RBP and the intron-exon architecture of the genes encoding them are also shown. Bootstrap values are represented by the size of the red circle on each branch



**Fig. 6** Phylogenetic tree of *Aedes aegypti*, *Drosophila melanogaster*, and *Bombyx mori* heterogeneous nuclear ribonucleoproteins (hnRNPs) containing RNA recognition motif (RRM). The tree was constructed using the neighbour-joining (NJ) method with 500 boot-

straps. The domain organization of each RBP and the intron-exon architecture of the genes encoding them are also shown. Bootstrap values are represented by the size of the red circle on each branch

### Polyadenylate- binding proteins (PABPs)

PABPs, another RRM containing RBPs, are involved in protein synthesis, mRNA stability, and mRNA biogenesis (Bernstein et al. 1989; Kühn and Wahle 2004). As per published records, only one PABP has been predicted in *D. melanogaster* (Smith et al. 2014), three in humans, while 12 in *Arabidopsis thaliana* (Lorković and Barta 2002), five in rice (Belostotsky 2003) and 14 in barley (Mahalingam and Walling 2020). In *Ae. aegypti*, we identified five PABPs and two cytoplasmic polyadenylation element-binding (CPEB) proteins. Similar to other eukaryotes, four PABPs had multiple RRM domains ranging from two to six. However, only one of them contained a polyA domain with four RRM domains which is the well-characterized domain pair identified in PABPs. One protein had two RRM domains, and three had a single

RRM copy. Similar to plants, where PABPs with one or two RRM domains have been identified (Lorković and Barta 2002; Mahalingam and Walling 2020), however, the functional characterization of the proteins could shed light on their exact role and functional capabilities.

### Conclusion

This study has generated the first reference database for *Ae. aegypti* RBPs containing RRM domains and their expression analysis specific to different developmental time points of the mosquito life cycle. This database would be useful for future studies to select and characterize candidate proteins and elucidate their role in mosquito growth and development. The proteins having a significant impact on the mosquito life cycle can be targeted





**Fig. 7** The STRING protein-protein interaction network for *Aedes aegypti* protein AAEL005515. Colored lines between the proteins indicate interactions

to develop growth regulators or insecticides. This will also be a useful reference base for future studies to characterize the entire RBP repertoire of *Ae. aegypti* as well as other mosquito species. It can further be extended to identify and validate the genes experimentally through interactome capture from the different developmental stages of mosquitoes to identify candidate genes.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00436-023-07969-2>.

**Acknowledgements** The authors would like to thank the Indian Council of Medical Research (ICMR), New Delhi for intramural support. Melveetil Kishor Sumitha would like to thank ICMR for the Senior Research Fellowship and Madurai Kamaraj University (MKU) for supporting the research.

**Author contribution** Conceptualization, methodology, formal analysis and investigation, visualization, and writing the manuscript: Bhavna Gupta. Data retrieval and analysis, visualization, and writing the manuscript: Melveetil Kishor Sumitha. Data retrieval and visualization: Mariapillai Kalimuthu and Murali Aarthy. Manuscript review: Ashwani Kumar and Rajaiah Paramasivan.

**Data availability** All the data has been included in the manuscript.

## Declarations

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Conflict of interest** The authors declare no competing interests.

## References

- Akbari OS, Antoshechkin I, Amrhein H et al (2013) The developmental transcriptome of the mosquito *Aedes aegypti*, an invasive species and major arbovirus vector. *G3* 3:1493–1509. <https://doi.org/10.1534/g3.113.006742>
- Aksaas AK, Larsen ACV, Rogne M et al (2011) G-patch domain and KOW motifs-containing protein, GPKOW; a nuclear RNA-binding

- protein regulated by protein kinase A. *J Mol Signal* 6:10. <https://doi.org/10.1186/1750-2187-6-10>
- Alfano C, Sanfelice D, Babon J et al (2004) Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nat Struct Mol Biol* 11:323–329. <https://doi.org/10.1038/nsmb747>
- Aravind L, Koonin EV (2000) SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* 25:112–114. [https://doi.org/10.1016/S0968-0004\(99\)01537-6](https://doi.org/10.1016/S0968-0004(99)01537-6)
- Bansal P, Madlung J, Schaaf K et al (2020) An interaction network of RNA-binding proteins involved in *Drosophila* oogenesis. *Mol Cell Proteomics* 19:1485–1502. <https://doi.org/10.1074/mcp.RA119.001912>
- Belostotsky DA (2003) Unexpected complexity of poly(A)-binding protein gene families in flowering plants: three conserved lineages that are at least 200 million years old and possible auto- and cross-regulation. *Genetics* 163:311–319. <https://doi.org/10.1093/genetics/163.1.311>
- Benoit B, He CH, Zhang F et al (2009) An essential role for the RNA-binding protein Smaug during the *Drosophila* maternal-to-zygotic transition. *Development* 136:923–932. <https://doi.org/10.1242/dev.031815>
- Bernstein P, Peltz SW, Ross J (1989) The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. *Mol Cell Biol* 9:659–670. <https://doi.org/10.1128/mcb.9.2.659-670.1989>
- Brinegar AE, Cooper TA (2016) Roles for RNA-binding proteins in development and disease. *Brain Res* 1647:1–8. <https://doi.org/10.1016/j.brainres.2016.02.050>
- Brooks AN, Duff MO, May G et al (2015) Regulation of alternative splicing in *Drosophila* by 56 RNA binding proteins. *Genome Res* 25:1771–1780. <https://doi.org/10.1101/gr.192518.115>
- Cáceres L, Nilson LA (2009) Translational repression of *gurken* mRNA in the *Drosophila* oocyte requires the hnRNP Squid in the nurse cells. *Dev Biol* 326:327–334. <https://doi.org/10.1016/j.ydbio.2008.11.030>
- Conesa A, Götz S, García-Gómez JM et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Díaz-Muñoz MD, Turner M (2018) Uncovering the role of RNA-binding proteins in gene expression in the immune system. *Front Immunol* 9. <https://doi.org/10.3389/fimmu.2018.01094>
- Diosa-Toro M, Prasanthi KR, Bradrick SS, Garcia Blanco MA (2020) Role of RNA-binding proteins during the late stages of *Flavivirus* replication cycle. *Virol J* 17:1–14. <https://doi.org/10.1186/s12985-020-01329-7>
- Dong G, Chakshumathi G, Wolin SL, Reinisch KM (2004) Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch. *EMBO J* 23:1000–1007. <https://doi.org/10.1038/sj.emboj.7600115>
- Fredericks AM, Cygan KJ, Brown BA, Fairbrother WG (2015) RNA-binding proteins: splicing factors and disease. *Biomolecules* 5:893–909. <https://doi.org/10.3390/biom5020893>
- Gamberi C, Johnstone O, Lasko P (2006) *Drosophila* RNA binding proteins. *Int Rev Cytol* 248:43–139. [https://doi.org/10.1016/S0074-7696\(06\)48002-5](https://doi.org/10.1016/S0074-7696(06)48002-5)
- Geuens T, Bouhy D, Timmerman V (2016) The hnRNP family: insights into their role in health and disease. *Hum Genet* 135:851–867. <https://doi.org/10.1007/s00439-016-1683-5>
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* 582:1977–1986. <https://doi.org/10.1016/j.febslet.2008.03.004>
- Hammani K, Cook WB, Barkan A (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *Proc Natl Acad Sci U S A* 109:5651–5656. <https://doi.org/10.1073/pnas.1200318109>
- Holmqvist E, Vogel J (2018) RNA-binding proteins in bacteria. *Nat Rev Microbiol* 16:601–615. <https://doi.org/10.1038/s41579-018-0049-5>
- Hu B, Jin J, Guo AY et al (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296–1297. <https://doi.org/10.1093/bioinformatics/btu817>
- Hurley JH, Lee S, Prag G (2006) Ubiquitin-binding domains. *Biochem J* 399:361–372. <https://doi.org/10.1042/BJ20061138>
- Iuchi S (2001) Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci* 58:625–635. <https://doi.org/10.1007/PL00000885>
- Jeong S (2017) SR proteins: binders, regulators, and connectors of RNA. *Mol Cells* 40:1–9. <https://doi.org/10.14348/molcells.2017.2319>
- Kelley RL (1993) Initial organization of the *Drosophila* dorsoventral axis depends on an RNA-binding protein encoded by the *squid* gene. *Genes Dev* 7:948–960. <https://doi.org/10.1101/gad.7.6.948>
- Kerner P, Degnan SM, Marchand L et al (2011) Evolution of RNA-binding proteins in animals: insights from genome-wide analysis in the sponge *Amphimedon queenslandica*. *Mol Biol Evol* 28:2289–2303. <https://doi.org/10.1093/molbev/msr046>
- Kim D, Paggi JM, Park C et al (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37:907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Krecic AM, Swanson MS (1999) hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* 11:363–371. [https://doi.org/10.1016/S0955-0674\(99\)80051-9](https://doi.org/10.1016/S0955-0674(99)80051-9)
- Kühn U, Wahle E (2004) Structure and function of poly(A) binding proteins. *Biochim Biophys Acta Gene Struct Expr* 1678:67–84. <https://doi.org/10.1016/j.bbaexp.2004.03.008>
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Kuwasako K, He F, Inoue M et al (2006) Solution structures of the SURP domains and the subunit-assembly mechanism within the splicing factor SF3a complex in 17S U2 snRNP. *Structure* 14:1677–1689. <https://doi.org/10.1016/j.str.2006.09.009>
- Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>
- Li K, Guo ZW, Zhai XM et al (2020) RBPTD: a database of cancer-related RNA-binding proteins in humans. *Database* 2020:1–7. <https://doi.org/10.1093/database/baz156>
- Loerch S, Kielkopf CL (2015) Dividing and conquering the family of RNA recognition motifs: a representative case based on hnRNP L. *J Mol Biol* 427:2997–3000
- Lorick KL, Jensen JP, Fang S et al (1999) RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc Natl Acad Sci U S A* 96:11364–11369. <https://doi.org/10.1073/pnas.96.20.11364>
- Lorković ZJ, Barta A (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. *Nucleic Acids Res* 30:623–635. <https://doi.org/10.1093/nar/30.3.623>
- Mahalingam R, Walling JG (2020) Genomic survey of RNA recognition motif (RRM) containing RNA binding proteins from barley (*Hordeum vulgare* ssp. *vulgare*). *Genomics* 112:1829–1839. <https://doi.org/10.1016/j.ygeno.2019.10.016>
- Mandel CR, Bai Y, Tong L (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* 65:1099–1122. <https://doi.org/10.1007/s00018-007-7474-3>
- Marinotti O, Ngo T, Kojin BB et al (2014) Integrated proteomic and transcriptomic analysis of the *Aedes aegypti* eggshell. *BMC Dev Biol* 14:1–11. <https://doi.org/10.1186/1471-213X-14-15>
- Maris C, Dominguez C, Allain FHT (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate

- post-transcriptional gene expression. *FEBS J* 272:2118–2131. <https://doi.org/10.1111/j.1742-4658.2005.04653.x>
- Matera AG, Wang Z (2014) Erratum: a day in the life of the spliceosome (*Nature Reviews Molecular Cell Biology* (2014) 15 (108–122)). *Nat Rev Mol Cell Biol* 15:294. <https://doi.org/10.1038/nrm3778>
- Matthews BJ, Dudchenko O, Kingan SB et al (2018) Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature* 563:501–507. <https://doi.org/10.1038/s41586-018-0692-z>
- Matunis EL, Kelley R, Dreyfuss G (1994) Essential role for a heterogeneous nuclear ribonucleoprotein (hnRNP) in oogenesis: hrp40 is absent from the germ line in the dorsoventral mutant *squid*. *Proc Natl Acad Sci U S A* 91:2781–2784. <https://doi.org/10.1073/pnas.91.7.2781>
- Michel SLJ, Guerrero AL, Berg JM (2003) Selective RNA binding by a single CCCH zinc-binding domain from Nup475 (Tristetraprolin). *Biochemistry* 42:4626–4630. <https://doi.org/10.1021/bi034073h>
- Muleya V, Marondedze C (2020) Functional roles of RNA-binding proteins in plant signaling. *Life* 10:1–8. <https://doi.org/10.3390/life10110288>
- Nene V, Wortman JR, Lawson D et al (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316:1718–1723. <https://doi.org/10.1126/science.1138878>
- Norvell A, Kelley RL, Wehr K, Schüpbach T (1999) Specific isoforms of Squid, a *Drosophila* hnRNP, perform distinct roles in Gurken localization during oogenesis. *Genes Dev* 13:864–876. <https://doi.org/10.1101/gad.13.7.864>
- Paschal BM, Gerace L (1995) Identification of NTF2, a cytosolic factor for nuclear import that interacts with nuclear pore complex protein p62. *J Cell Biol* 129:925–937. <https://doi.org/10.1083/jcb.129.4.925>
- Pellizzoni L, Lotti F, Maras B, Pierandrei-Amaldi P (1997) Cellular nucleic acid binding protein binds a conserved region of the 5' UTR of *Xenopus laevis* ribosomal protein mRNAs. *J Mol Biol* 267:264–275. <https://doi.org/10.1006/jmbi.1996.0888>
- SenGupta D (2013) RNA-binding domains in proteins. In: Brenner's encyclopedia of genetics: Second Edition. Elsevier Inc., pp 274–276
- Smith RWP, Blee TKP, Gray NK (2014) Poly(A)-binding proteins are required for diverse biological processes in metazoans. *Biochem Soc Trans* 42:1229–1237. <https://doi.org/10.1042/BST20140111>
- Solymosy F, Pollák T (1993) Uridylate-rich small nuclear RNAs (UsnRNAs), their genes and pseudogenes, and UsnRNPs in plants: structure and function. A comparative approach. *CRC Crit Rev Plant Sci* 12:275–369. <https://doi.org/10.1080/07352689309701904>
- Steinhauer J, Kalderon D (2005) The RNA-binding protein Squid is required for the establishment of anteroposterior polarity in the *Drosophila* oocyte. *Development* 132:5515–5525. <https://doi.org/10.1242/dev.02159>
- Summers MF (1991) Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance. *J Cell Biochem* 45:41–48. <https://doi.org/10.1002/jcb.240450110>
- Sysoev VO, Fischer B, Frese CK et al (2016) Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in *Drosophila*. *Nat Commun* 7. <https://doi.org/10.1038/ncomm512128>
- Szklarczyk D, Gable AL, Nastou KC et al (2021) The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 49:D605–D612. <https://doi.org/10.1093/nar/gkaa1074>
- Szymczyna BR, Bowman J, McCracken S et al (2003) Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. *Genes Dev* 17:461–475. <https://doi.org/10.1101/gad.1060403>
- Taschuk F, Tapescu I, Moy RH, Cherry S (2020) Ddx56 binds to Chikungunya virus rna to control infection. *MBio* 11:1–16. <https://doi.org/10.1128/mBio.02623-20>
- Trapnell C, Roberts A, Goff L et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562–578. <https://doi.org/10.1038/nprot.2012.016>
- Valverde R, Edwards L, Regan L (2008) Structure and function of KH domains. *FEBS J* 275:2712–2726. <https://doi.org/10.1111/j.1742-4658.2008.06411.x>
- Venter JC, Adams MD, Myers EW et al (2001) Celera\_genoma. *Science* 291:1–49
- Voorrips RE (2002) Mapchart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93:77–78. <https://doi.org/10.1093/jhered/93.1.77>
- Vuong CK, Black DL, Zheng S (2016) The neurogenetics of alternative splicing. *Nat Rev Neurosci* 17:265–281. <https://doi.org/10.1038/nrn.2016.27>
- Wang L-L, Zhou Z-Y (2009) RNA recognition motif (RRM)-containing proteins in *Bombyx mori*. *African J Biotechnol* 8:1121–1126
- Wang Y, Yu Y, Pang Y et al (2021) The distinct roles of zinc finger CCHC-type (ZCCHC) superfamily proteins in the regulation of RNA metabolism. *RNA Biol* 18:2107–2126. <https://doi.org/10.1080/15476286.2021.1909320>
- Will CL, Lührmann R (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3. <https://doi.org/10.1101/cshperspect.a003707>
- Wurth L (2012) Versatility of RNA-binding proteins in cancer. *Comp Funct Genomics* 2012. <https://doi.org/10.1155/2012/178525>
- Yeh S-C, Diosa-Toro M, Tan W-L et al (2022) Characterization of dengue virus 3' UTR RNA binding proteins in mosquitoes reveals that AeStaufen reduces subgenomic flaviviral RNA in saliva. *PLoS Pathog* 18:1–46
- Yoon JS, Mogilicherla K, Gurusamy D et al (2018) Double-stranded RNA binding protein, Staufen, is required for the initiation of RNAi in coleopteran insects. *Proc Natl Acad Sci U S A* 115:8334–8339. <https://doi.org/10.1073/pnas.1809381115>
- Zhang Y, Rataj K, Simpson GG, Tong L (2016) Crystal structure of the SPOC domain of the *Arabidopsis* flowering regulator FPA. *PLoS One* 11:1–13. <https://doi.org/10.1371/journal.pone.0160694>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.