SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

ISSN 1301-4048 | e-ISSN 2147-835X | Period Bimonthly | Founded: 1997 | Publisher Sakarya University | http://www.saujs.sakarya.edu.tr/

Title: Identification and Characterization of the CPP Gene Family in the Genome of Aedes aegypti L. (Yellow Fever Mosquito) (Diptera: Culicidae)

Authors: Murat TURAN

Recieved: 4.08.2023

Accepted: 3.10.2023

Article Type: Research Article

Volume: 27
Issue: 6
Month: December
Year: 2023
Pages: 1174-1184

How to cite
Murat TURAN; (2023), Identification and Characterization of the CPP Gene Family in the Genome of Aedes aegypti L. (Yellow Fever Mosquito) (Diptera: Culicidae). Sakarya University Journal of Science, 27(6), 1174-1184, DOI: 10.16984/saufenbilder.1338063
Access link
https://dergipark.org.tr/en/pub/saufenbilder/issue/80994/1338063

Sakarya University Journal of Science 27(6), 1174-1184, 2023

Identification and Characterization of the CPP Gene Family in the Genome of *Aedes aegypti* L. (Yellow Fever Mosquito) (Diptera: Culicidae)

Murat TURAN *[1]

Abstract

Aedes aegypti is an important vector organism responsible for carrying numerous arboviral pathogens and serious diseases, including yellow fever, Zika, Chikungunya, and Dengue fever. The CPP gene family, one of the crucial molecular defense systems, plays a significant role in the regulation of growth and development by controlling the production of proteins. In this study, a comprehensive genome analysis of the CPP gene family in *Ae. aegypti* was conducted. Each gene was thoroughly characterized, gene structures were examined, and conserved motifs were investigated. Additionally, the properties of these proteins were comprehensively analyzed. Expression analyses were performed to reveal the effects of CPP genes on development by calculating Reads Per Kilobase Million (RPKM) values. The findings emphasize the importance of CPP genes in controlling arboviral pathogens and understanding general stress responses in insects. The information derived from this research could contribute to the development of more effective intervention strategies for *Ae. aegypti* and other vector carriers to cope with stress. In conclusion, the systematic analysis of the CPP gene family in the *Ae. aegypti* genome is a crucial step in the management and development of effective disease prevention strategies for this species. Moreover, this study provides a significant foundation for future functional genomics research in understanding the structure and function of CPP genes.

Keywords: Aedes aegypti, CPP gene family, Yellow fever mosquito, characterization

1. INTRODUCTION

Cell cycle progression, metabolic balance, physiological regulations, and environmental responses are fundamental biological functions that are regulated in organisms through transcription factors (TFs), which control gene expression at the mRNA transcript level [1-4]. The CPP gene family, Cysteine-rich polycomb-like protein (CPP) transcription factors, is referred to as Tesmin/TSO1-like CXC (TCX) and represents a small but valuable group of TFs critical in the regulation of gene expression [1-3, 5]. These CPP genes are plant-derived and can be defined as CXCX4CX3YCXCX6CX3CXCX2C, containing one or two Cys-rich regions within this structure. These regions are called CXC domains [1-3, 6]. Both the CXC domains and the sequences that bind to them are universally found in organisms carrying the

* Corresponding author: m.turan@erzurum.edu.tr (M. TURAN)
[1] Erzurum Technical University, Türkiye
ORCID: https://orcid.org/0000-0003-2900-1755

entire CPP gene [1, 2]. These conserved regions confer the ability to bind to DNA [3, 6]. The preservation of CXC domains, particularly in plants and animals, has led to more extensive research in these species [6].

Studies on various plant species have shown that CPP genes can have diverse functions and contribute to the regulation of gene expression. These genes can perform different cellular tasks in plants, ranging from cell division to developmental processes [1, 4]. Several studies have reported the involvement of CPP genes in stress responses [7]. Studies have indicated that CPP genes potentially have a vital role in plant responses to abiotic stress factors like drought, heat, cold, and salinity. When plants are exposed to such stress conditions, the expression levels of CPP genes undergo significant changes, suggesting that the CPP gene family might play a crucial role in stress tolerance and adaptation mechanisms in plants [5, 6, 8].

As an example, specific genes in soybeans have been discovered to upregulate their expression in response to high-temperature stress. Similarly, in maize, CPP gene expression was significantly enhanced when the plants were exposed to heat, cold, and drought stresses. These findings indicate the potential involvement of CPP genes in facilitating plant responses to diverse abiotic stressors and highlight their importance in stress adaptation mechanisms [5]. Despite receiving less attention compared to other transcription factor families, CPP transcription factors are present in many species [9].

They are absent in prokaryotes, fungi, and fungi-like organisms, but widespread in plants and animals. The observed responses of CPP genes to various abiotic stress conditions in different plant species suggest that the CPP gene family has likely played a significant role in the evolutionary processes of plants. The ability of these genes to modulate their expression in the face of environmental challenges indicates their potential importance in enabling plants to adapt and survive in diverse and changing environments throughout their evolutionary history [1-3, 5, 6].

Arbovirus outbreaks, such as Dengue, Chikungunya, and Zika, are transmitted by *Aedes* mosquitoes and are widely spread through human mobility and trade [10-12]. *Aedes aegypti* L. (Diptera: Culicidae) is also a vector for zoonotic diseases like Rift Valley fever (RVF), which is considered a serious emerging zoonotic disease [12]. These diseases infect a total of 50-100 million people annually, while more than 2.5 billion people live in areas where these diseases can be transmitted [11, 12]. *Ae. aegypti* is a mosquito species commonly found in tropical and subtropical regions across the globe, spanning South America, Southeast Asia, and Africa. Its presence in these areas presents a persistent and substantial public health challenge for numerous countries. [11]. Dengue fever is a significant viral disease transmitted to humans by *Ae. aegypti* and *Aedes albopictus* Skuse (Diptera: Culicidae) mosquitoes [13-15].

Ae. aegypti mosquitoes can carry all four dengue virus serotypes and play a crucial role in the transmission of the disease. Dengue outbreaks lead to significant health problems and economic burdens in affected regions [16]. Dengue outbreaks currently affect more than 100 countries, spanning across Africa, the Americas, the Indian subcontinent, Southeast Asia, the Eastern Mediterranean, and the Western Pacific regions [13].

According to the estimates of the World Health Organization (WHO), there are around 50 million cases of dengue infections each year, leading to approximately 500.000 hospitalizations and 22.000 dengue-related deaths worldwide [13]. The global population at risk of contracting this disease exceeds 2.5 billion people [13]. *Aedes* mosquitoes are closely associated with human habitats and lay their eggs in water containers around houses. The accumulation of plastic and

rubber provides additional breeding sites, contributing to increased population density in urban areas [11]. Dengue continues to pose a significant threat to global health [12]. The role of mosquitoes as vectors facilitates the spread of the disease [17]. Therefore, developing effective interventions against dengue outbreaks and controlling the mosquito population are crucial to prevent the spread of the disease. Understanding the genes of *Aedes* mosquitoes and their functions could help develop more effective strategies for disease control and prevention.

In this study, a total of 11 different CPP genes have been identified from species including *Ae. aegypti*, *Anopheles arabiensis* Patton (Diptera: Culicidae), *Anopheles albimanus* Wiedemann (Diptera: Culicidae), *Anopheles funestus* Giles (Diptera: Culicidae), *Anopheles stephensi* Liston (Diptera: Culicidae), *Culex quinquefasciatus* Say (Diptera: Culicidae), *Drosophila ananassae* Doleschall (Diptera: Culicidae), and *Drosophila melanogaster* Meigen (Diptera: Culicidae). Each CPP gene has been fully characterized in these species, with their gene structures and conserved motifs investigated.

Additionally, the properties and subcellular locations of these proteins were comprehensively analyzed. Furthermore, our study involved the phylogenetic comparison of CPP genes in *Ae. aegypti* with those in *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*, *D. ananassae*, and *D. melanogaster*, using various bioinformatics tools. RPKM (Reads Per Kilobase Million) values were also calculated from the examination of the PRJNA419241 project at the National Center for Biotechnology Information (NCBI) to reveal the effects of CPP genes on development. This research presents the first comprehensive characterization of CPP genes in *Ae. aegypti*, adding unique value to the efforts in understanding the genetic and molecular biology of this species. The results obtained can serve as a valuable reference for cloning and functional analysis of CPP genes in this species. Furthermore, it provides valuable insights into the control of arbovirus pathogens and understanding general stress responses in insects.

2. MATERIAL AND METHODS

2.1. Determination of CPP Genes

In this study, accession number was determined using the InterPro database (https://www.ebi.ac.uk/interpro/). Accession number of the CPP (PF03638) gene family were identified in the genomes of the following species: *Ae. aegypti* (Aaeg) [18], *A. albimanus* (Aali) [19], *A. arabiensis* (Aara) [20], *A. funestus* (Afun) [21], *A. stephensi* (Aste), *Cx. quinquefasciatus* (Cqui) [22], *D. ananassae* (Dana) [23], and *D. melanogaster* (Dmen) [24]. In the conducted study, a total of 11 genes belonging to the relevant gene families were identified in the InsectBase 2.0 database (http://v2.insect-genome.com/). Protein and CDS sequences of these genes were retrieved. The SMART database (Simple Modular Architecture Research Tool) (http://smart.embl-heidelberg.de/), was utilized to identify the existence of CPP proteins and to explore potential extra regions within their sequences [25].

2.2. Sequence Alignment and Phylogenetic Analyses

The protein sequences of the members of the CPP gene family in the genomes of *Ae. aegypti*, *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*, *D. ananassae* and *D. melanogaster* species were aligned using the Multiple Sequence Alignment by CLUSTALW tool (https://www.genome.jp/tools-bin/clustalw) [26]. Subsequently, the aligned protein sequences were used to construct a phylogenetic tree using the ITOL (Interactive Tree of Life) tool (https://itol.embl.de/) [27] Also, for the sequence analysis of conserved regions in these species, the BioEdit 7.7.1 [28] program was used for visualization purposes. In addition, the TBTools v1.123

[29] program was employed specifically for sequence logo analysis.

2.3. Characteristics and Subcellular Localizations of CPP Proteins

The amino acid count, molecular weight (kDa), theoretical isoelectric point (pI), amino acid composition, stability, aliphatic index, and GRAVY (Grand Average of Hydropathy) of CPP proteins in *Ae. aegypti*, *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*, *D. ananassae*, and *D. melanogaster* were calculated using the "ProtParam Tool" (https://web.expasy.org/protparam/) [30]. Additionally, DeepLoc (https://services.healthtech.dtu.dk/services/DeepLoc-2.0/) was used to predict the subcellular localization of CPP genes [31].

2.4. Structure of CPP Genes

In order to better understand the evolutionary changes in the gene structures of CPP genes found in the genomes of *Ae. aegypti*, *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*, *D. ananassae*, and *D. melanogaster*, and their potential effects, the exon and intron distributions were analyzed using the Gene Structure Display Server (GSDS) v2.0 database (http://gsds.gao-lab.org/) [32]. During the analysis process using the GSDS database, the gene sequences were retrieved from the InsectBase database in ".gff3" format files, and the phylogenetic tree outputs were obtained from the CLUSTALW tool in ".nwk" (Newick) format files.

2.5. Determination of Conserved Motifs

The conserved motifs present in the protein sequences encoded by CPP genes were discovered using the "Multiple EM for Motif Elicitation (MEME)" tool (https://meme-suite.org/meme/index.html) [33]. The analysis considered a width range for identified motifs from 2 to 50 amino acids, with a maximum limit of 10 motifs being determined. The motif regions, which varied in length from 6 to 300 amino acids, were identified with a flexible distribution, allowing for any number of repetitions.

2.6. Expression Analysis of *Ae. aegypti* CPP Genes

The profiles of CPP genes in *Ae. aegypti* were examined using Illumina RNA-seq data obtained from the Sequence Read Archive (SRA) database at NCBI (https://www.ncbi.nlm.nih.gov/sra). The RNA-Seq data for *Ae. aegypti* from 17 different developmental stages (L1 Larvae, L2 Larvae, L3 Larvae Male, L3 Larvae Female, L4 Larvae Male, L4 Larvae Female, Early Pupae Male, Early Pupae Female, Mid Pupae Male, Mid Pupae Female, Late Pupae Male, Late Pupae Female, Male Accessory Gland, Carcass Male, Carcass Female, Testes and Ovaries) were obtained from the NCBI Sequence Read Archive (SRA) database using the BioProject PRJNA419241 [18]. To normalize gene expression values, the Reads Per Kilobase per Million mapped reads (RPKM) algorithm was employed, which calculates the number of reads per kilobase per million mapped reads for each exon model [34]. The RPKM values were converted to log2 scale, and a circular heatmap was generated using the SRPLOT (http://www.bioinformatics.com.cn/en) algorithm.

3. RESULTS AND DISCUSSION

3.1. Determination of CPP Genes

A total of 2 genes from *Ae. aegypti*, 1 gene from *A. arabiensis*, 1 gene from *A. albimanus*, 1 gene from *A. funestus*, 1 gene from *A. stephensi*, 1 gene from *Cx. quinquefasciatus*, 2 genes from *D. ananassae*, and 2 genes from *D. melanogaster* were identified. Subsequently, abbreviations for the species names were used to improve clarity in the article, and the genes were numbered according to their chromosomal sequences. In Figure 1, the CPP protein sequences of *Ae. aegypti*, *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*,

D. ananassae, and *D. melanogaster* were aligned, and the CXC domain region was determined. The amino acids more frequently present in these domain regions were visualized using a logo representation. It was observed that if a gene possesses two CXC domains, both regions contain the same domain motif, whereas if a gene has only one CXC domain, the second region contains this domain.

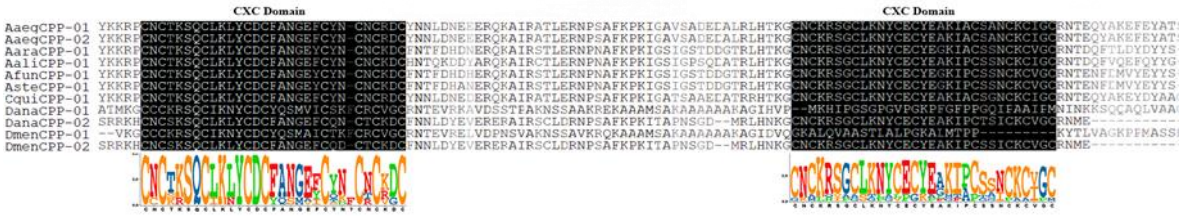

Figure 1 Protein sequences belonging to *Ae. aegypti*, *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*, *D. ananassae*, and *D. melanogaster* were aligned using ClustalW, and visualized using the Bioedit program. The CXC domains were highlighted in black

3.2. Phylogenetic Analysis of CPP Proteins

Prior to constructing the phylogenetic tree of the CPP family, sequence alignment was performed using CLUSTALW, and visualization was achieved using the ITOL tool. As a result of the analysis, CPP genes have been divided into 3 main groups: Group 1, Group 2, and Group 3. While Group 1 contains only 1 CXC domain, both Group 2 and Group 3 have 2 CXC domains (Figure 2).

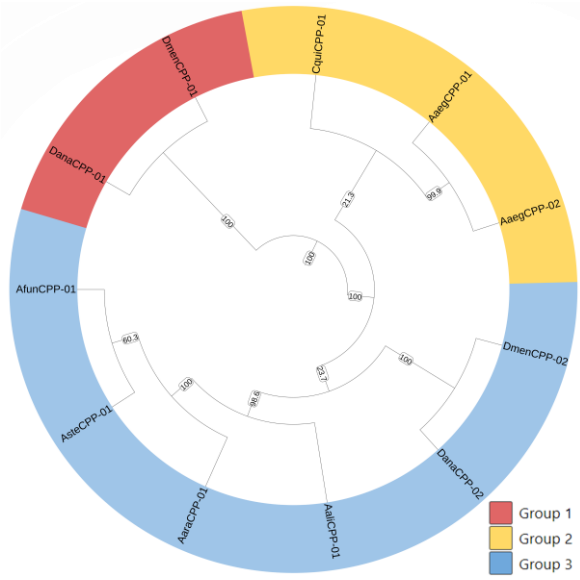

Figure 2 Phylogenetic tree illustrating the relationship of CPP family in *Ae. aegypti*, *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*, *D. ananassae* and *D. melanogaster*

3.3. Characteristics and Subcellular Localizations of CPP Proteins

CPP proteins with a single CXC domain (DanaCPP-01 and DmenCPP-01) have amino acid lengths in the range of 243-330 and molecular weights between 26 and 35 kDa. CPP proteins with two CXC domains have amino acid lengths in the range of 759-950 and molecular weights between 81 and 100 kDa. All proteins exhibit hydrophilic properties. AaliCPP-01 exhibits acidic properties, AaraCPP-01 shows neutral characteristics, while the others possess basic properties.

DanaCPP-02 was found to be stable, while the other proteins were unstable. When the aliphatic index and GRAVY values were considered together, it can be said that the proteins have hydrophobic characteristics in their interior structures and hydrophilic characteristics in their exterior structures (Table 1). The potential subcellular localizations of CPP genes were analyzed with DeepLoc, and it was determined that all CPP genes are located in the nucleus and contain a nuclear localization signal (Table 2).

Table 1 The properties of CPP proteins

Gene	Number of amino acids	Molecular weight (kDa)
AaegCPP-01	785	83.53
AaegCPP-02	785	83.48
AaliCPP-01	810	88.61
AaraCPP-01	751	80.70
AfunCPP-01	750	80.79
AsteCPP-01	759	80.61
CquiCPP-01	801	84.57
DanaCPP-01	330	35.24
DanaCPP-02	921	96.73
DmenCPP-01	243	26.18
DmenCPP-02	950	100.02
Gene	Theoretical pI	Stability
AaegCPP-01	8.89	Unstable
AaegCPP-02	8.85	Unstable
AaliCPP-01	6.23	Unstable
AaraCPP-01	7.43	Unstable
AfunCPP-01	8.2	Unstable
AsteCPP-01	8.39	Unstable
CquiCPP-01	8.75	Unstable
DanaCPP-01	9.67	Unstable
DanaCPP-02	9.71	Stable
DmenCPP-01	9.38	Unstable
DmenCPP-02	9.53	Unstable
Gene	Aliphatic index	GRAVY
AaegCPP-01	75.38	-0.348
AaegCPP-02	75.38	-0.346
AaliCPP-01	74.64	-0.532
AaraCPP-01	70.6	-0.49
AfunCPP-01	70.08	-0.476
AsteCPP-01	71.23	-0.398
CquiCPP-01	73.82	-0.343
DanaCPP-01	67.21	-0.452
DanaCPP-02	71.61	-0.469
DmenCPP-01	80.41	-0.229
DmenCPP-02	73.03	-0.423

3.4. Structure of CPP Genes

When the intron/exon structures of homologous and paralogous genes exhibit similarity across different species, this similarity can be employed as an indicator to assess the proximity or evolutionary distance between the two species [6]. It has been observed that all genes contain multiple exon regions, and the number of exon regions varies depending on the number of CXC domains. Genes with 1 CXC domain have 2 exon regions, while those with 2 CXC domains have 4 exon regions (Figure 3). The protein structures align appropriately with the phylogenetic tree, indicating similarity among the proteins (Figure 2).

Table 2 The subcellular Localizations of CPP proteins

Gene	Probability	Predicted localizations
AaegCPP-01	0.8697	Nucleus
AaegCPP-02	0.8718	Nucleus
AaraCPP-01	0.9247	Nucleus
AaliCPP-01	0.8757	Nucleus
AfunCPP-01	0.9235	Nucleus
AsteCPP-01	0.919	Nucleus
CquiCPP-01	0.92	Nucleus
DanaCPP-01	0.8512	Nucleus
DanaCPP-02	0.9009	Nucleus
DmenCPP-01	0.8688	Nucleus
DmenCPP-02	0.902	Nucleus
Gene	Predicted signals	
AaegCPP-01	Nuclear localization signal	
AaegCPP-02	Nuclear localization signal	
AaegCPP-02	Nuclear localization signal	
AaraCPP-01	Nuclear localization signal	
AaliCPP-01	Nuclear localization signal, Nuclear export signal	
AfunCPP-01	Nuclear localization signal	
AsteCPP-01	Nuclear localization signal	
CquiCPP-01	Nuclear localization signal	
DanaCPP-01	Nuclear localization signal	
DanaCPP-02	Nuclear localization signal	
DmenCPP-01	Nuclear localization signal	
DmenCPP-02	Nuclear localization signal	

3.5. Conserved Motifs of CPP Genes

The conserved motifs of the CPP gene family were acquired through the utilization of MEME-Suite. During the analysis process, a

total of ten conserved motif patterns were successfully identified for further investigation and examination of the CPP gene family. The number of obtained motifs ranged from 3 to 8, with DanaCPP-01 and DmenCPP-01 having the fewest conserved motifs (3 motifs), and AaegCPP-01, AaegCPP-02, AaraCPP-01, AsteCPP-01, and AfunCPP-01 having the most conserved motifs (8 motifs). No motifs were found to be common among all genes. All motifs exhibited similarity in accordance with the phylogenetic tree, indicating that they are conserved and share similarities among the proteins of the CPP gene family (Figure 4).

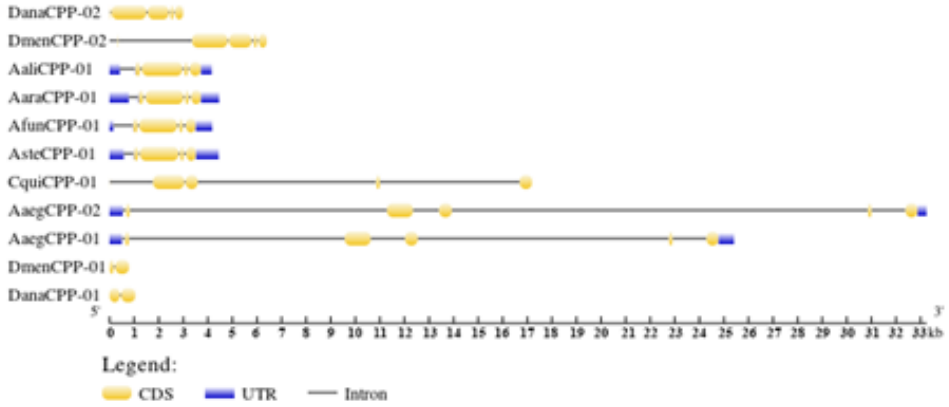

Figure 3 The phylogenetic relationships, length, and position of exons and introns in the CPP genes in *Ae. aegypti*, *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*, *D. ananassae* and *D. melanogaster*

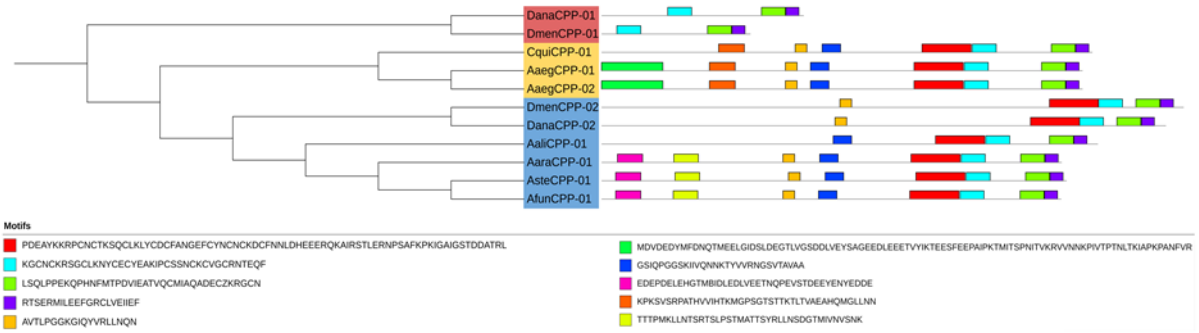

Figure 4 The conserved motifs of CPP proteins in *Ae. aegypti*, *A. arabiensis*, *A. albimanus*, *A. funestus*, *A. stephensi*, *Cx. quinquefasciatus*, *D. ananassae*, and *D. melanogaster*

3.6. Expression Analysis of *Ae. aegypti* CPP Genes

CPP genes play diverse roles in regulating cell division and the development of various tissues. They are involved in crucial biological processes that influence the growth, differentiation, and maintenance of tissues and organisms [1]. The expression of CPP genes is notably high in flowers, and their accumulation reaches peak levels during the development of ovules and microspores [3]. Additionally, a separate study has demonstrated that the TSO1 protein is widely distributed in flowers, with a particularly significant presence in developing microspores and eggs [6]. Moreover, a study focusing on Arabidopsis has revealed the elevated expression of CPP genes during the development of flower eggs and microspores, indicating their potential role in the development of reproductive organs [1]. Additionally, when investigating the expression of CPP genes in *Ae. aegypti* at various stages, the highest expression was observed in the testes. There was also an increased expression near the ovaries, while

larval stages exhibited a comparatively lower level of expression (Figure 5).

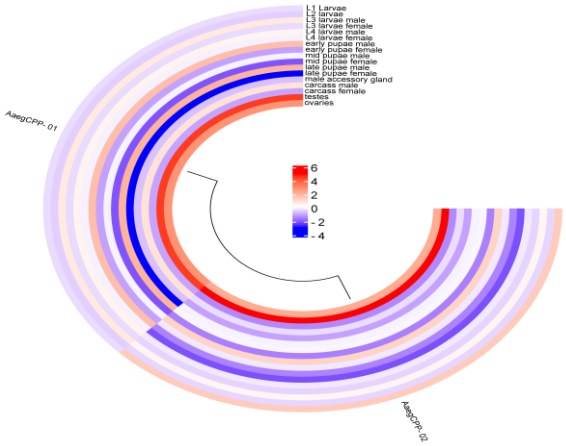

Figure 5 Heat map of differentially expressed CPP genes in *Ae. aegypti* from various developmental stages

4. CONCLUSION

The CPP gene family is a highly important gene family in plants, and although extensively studied in plant research, there has been limited research, particularly on vector organisms. This study represents the first investigation of the CPP gene family in *Ae. aegypti*, a mosquito species responsible for causing widespread human diseases and deaths. The research aims to analyze and characterize the CPP gene family systematically in *Ae. aegypti*, focusing on protein properties, structures, phylogenetic relationships, and expression levels. The findings highlight the significance of CPP genes and shed light on their association with insect adaptive strategies. By understanding the functions of CPP genes in *Ae. aegypti*'s development, stress response, and disease transmission in more depth, this research can significantly contribute to the management of this species and the development of effective disease prevention strategies. Additionally, the study provides valuable insights into *Ae. aegypti* and other vector carriers' abilities to cope with climate change, which could aid in developing more efficient intervention strategies to control these diseases.

Acknowledgments
The author would like to thank Dr. Emre İlhan for his contributions.

Funding
The author has no received any financial support for the research, authorship or publication of this study.

Authors' Contribution
The author contributed conceptualization, formal analysis, visualization, writing – original draft, writing – review & editing to the study.

The Declaration of Conflict of Interest/ Common Interest
No conflict of interest or common interest has been declared by the authors.

The Declaration of Ethics Committee Approval
This study does not require ethics committee permission or any special permission.

The Declaration of Research and Publication Ethics
The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] U. Ullah, Z. A. Buttar, A. Shalmani, I. Muhammad, A. Ud-Din, H. Ali, "Genome-wide identification and expression analysis of CPP-like gene family in *Triticum aestivum* L. under different hormone and stress conditions", Open Life Sciences, vol. 17, pp. 544-562, 2022.

[2] L. Zhang, H. K. Zhao, Y. M. Wang, C. P. Yuan, Y. Y. Zhang, H. Y. Li, X. F. Yan, Q. Y. Li, Y. S. Dong, "Genome-wide identification and expression analysis of the CPP-like gene family in soybean", Genetics and Molecular Research, vol. 14, no.1, pp. 1260-1268, 2015.

[3] T. Lu, Y. Dou, C. Zhang, "Fuzzy clustering of CPP family in plants with evolution and interaction analyses", BMC Bioinformatics, vol. 14, no. Suppl 13, S10, 2013.

[4] Z. Yang, S. Gu, X. Wang, W. Li, Z. Tang, C. Xu, "Molecular Evolution of the CPP-like Gene Family in Plants: Insights from Comparative Genomics of *Arabidopsis* and Rice", Journal of Molecular Evolution, vol. 67, pp. 266-277, 2008.

[5] Y. Zhou, L. Hu, S. Ye, L. Jiang, S. Liu, "Genome-wide identification and characterization of cysteine-rich polycomb-like protein (CPP) family genes in cucumber (*Cucumis sativus*) and their roles in stress responses", Biologia, vol. 73, pp. 425-435, 2018.

[6] A. Rakhimzhanova, A. G. Kasapoğlu, A. Sapakova, E. İlhan, R. Zharmukhametova, M. Turan, L. Zekenova, S. Muslu, L. Kazhygeldiyeva, M. Aydın, A. Çiltaş, "Expression analysis and characterization of the CPP gene family of Melatonin-treated common bean cultivars under different abiotic stresses," South African Journal of Botany, vol. 160, pp. 282-294, 2023.

[7] M. Li, F. Wang, J. Ma, H. Liu, H. Ye, P. Zhao, J. Wang, "Comprehensive Evolutionary Analysis of CPP Genes in *Brassica napus* L. and Its Two Diploid Progenitors Revealing the Potential Molecular Basis of Allopolyploid Adaptive Advantage Under Salt Stress", Frontiers in Plant Science, vol. 13, 873071, 2022.

[8] Y. Sun, X. Jia, D. Chen, Q. Fu, J. Chen, W. Yang, H. Yang, X. Xu, "Genome-Wide Identification and Expression Analysis of Cysteine-Rich Polycomb-like Protein (CPP) Gene Family in Tomato", International Journal of Molecular Sciences, vol. 24, 5762, 2023.

[9] X. Y. Song, Y. Y. Zhang, F. C. Wu, L. Zhang, "Genome-wide analysis of the maize (*Zea may* L.) CPP-like gene family and expression profiling under abiotic stress", Genetics and Molecular Research, vol. 15, no. 3, gmr.15038023, 2016.

[10] F. Ding, J. Fua, D. Jiang, M. Hao, G. Lina, "Mapping the spatial distribution of *Aedes aegypti* and *Aedes albopictus*", Acta Tropica, vol. 178, pp. 155–162, 2018.

[11] J. Vontas, E. Kioulos, N. Pavlidi, E. Morou, A. della Torre, H. Ranson, "Insecticide resistance in the major dengue vectors *Aedes albopictus* and *Aedes aegypti*", Pesticide Biochemistry and Physiology, vol. 104, pp. 126–131, 2012.

[12] C. Lowenberger, "Innate immune response of *Aedes aegypti*", Insect Biochemistry and Molecular Biology, vol. 31, pp. 219–229, 2001.

[13] M. I. Salazar, J. H. Richardson, I. Sánchez-Vargas, K. E. Olson, B. J. Beaty, "Dengue virus type 2: replication and tropisms in orally infected *Aedes aegypti* mosquitoes", BMC Microbiology, vol. 7, 9, 2007.

[14] B. Kamgang, S. Marcombe, F. Chandre, E. Nchoutpouen, P. Nwane, J. Etang, V. Corbel, C. Paupy, "Insecticide susceptibility of *Aedes aegypti* and

Aedes albopictus in Central Africa", Parasites & Vectors, vol. 4, 79, 2011.

[15] D. J. Gubler, G. G. Clark, "Community involvement in the control of *Aedes aegypti*", Acta Tropica, vol. 61, pp. 169-179, 1996.

[16] K. Kamimura, I. T. Matsuse, H. Takahashi, J. Komukai, T. Fukuda, K. Suzuki, M. Aratani, Y. Shirai, M. Mogi, "Effect of temperature on the development of *Aedes aegypti* and *Aedes albopictus*", Medical Entomology and Zoology, vol. 53, no. 1, pp. 53-58, 2002.

[17] S. Leta, T. J. Beyene, E. M. De Clercq, K. Amenu, M. U. G. Kraemer, C. W. Revie, "Global risk mapping for major diseases transmitted by *Aedes aegypti* and *Aedes albopictus*", International Journal of Infectious Diseases, vol. 67, pp. 25–35, 2018.

[18] B. J. Matthews, O. Dudchenko, S. B. Kingan, S. Koren, I. Antoshechkin, J. E. Crawford, W. J. Glassford, M. Herre, S. N. Redmond, N. H. Rose, G. D. Weedall, Y. Wu, S. S. Batra, C. A. Brito-Sierra, S. D. Buckingham, C. L. Campbell, S. Chan, E. Cox, B. R. Evans, T. Fansiri, et al., "Improved reference genome of *Aedes aegypti* informs arbovirus vector control", Nature, vol. 563, no. 7732, pp. 501-507, 2018.

[19] A. Compton, J. Liang, C. Chen, V. Lukyanchikova, Y. Qi, M. Potters, R. Settlage, D. Miller, S. Deschamps, C. Mao, V. Llaca, I. V. Sharakhov, Z. Tu, "The Beginning of the End: A Chromosomal Assembly of the New World Malaria Mosquito Ends with a Novel Telomere", G3 (Bethesda), vol. 10, no. 10, pp. 3811-3819, 2020.

[20] A. Zamyatin, P. Avdeyev, J. Liang, A. Sharma, C. Chen, V. Lukyanchikova, N. Alexeev, Z. Tu, M. A. Alekseyev, "Chromosome-level genome assemblies of the malaria vectors *Anopheles coluzzii* and *Anopheles arabiensis*", Gigascience, vol. 10, no. 3, giab017, 2021.

[21] J. Ghurye, S. Koren, S. T. Small, S. Redmond, P. Howell, A. M. Phillipy, N. J. Besansky, "A chromosome-scale assembly of the major African malaria vector *Anopheles funestus*", Gigascience., vol. 8, no. 6, giz063, 2019.

[22] O. Dudchenko, S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, A. P. Aiden, E. L. Aiden, "De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds", Science, vol. 356, no. 6333, pp. 92-95, 2017.

[23] E. S. Tvedte, M. Gasser, B. C. Sparklin, J., Michalski, C. E. Hjelmen, J. S. Johnston, X. Zhao, R. Bromley, L. J. Tallon, L. Sadzewicz, D. A. Rasko, J.C. Dunning Hotopp, "Comparison of long read sequencing technologies in interrogating bacteria and fly genomes", G3 (Bethesda), vol. 11, no. 6, jkab083, 2021.

[24] M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, et al., "The genome sequence of *Drosophila melanogaster*", Science, vol. 287, no. 5461, pp. 2185-2195, 2000.

[25] I. Letunic, S. Khedkar, P. Bork, "SMART: recent updates, new developments and status in 2020",

Nucleic Acids Research, vol. 49, pp. D458–D460, 2021.

[26] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, D. G. Higgins, "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools", Nucleic Acids Research, vol. 25, pp. 4876-4882, 1997.

[27] I. Letunic, P. Bork, "Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation", Nucleic Acids Research, vol. 49, pp. W293–W296, 2021.

[28] T. A. Hall, "BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT", Nucleic Acids Symposium Series, vol. 41, pp. 95-98, 1999.

[29] C. Chen, H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He, R. Xia, "TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data", Molecular Plant, vol. 13, no. 8, pp. 1194-1202, 2020.

[30] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, "Protein Identification and Analysis Tools on the Expasy Server," In John M. Walker (Ed): The Proteomics Protocols Handbook, Humana Press, pp. 571-607, 2005.

[31] V. Thumuluri, J. J. A. Armenteros, A. R. Johansen, H. Nielsen, O. Winther, "DeepLoc 2.0: multi-label subcellular localization prediction using protein language models", Nucleic Acids Research, vol. 50, no. W1, pp. W228–W234, 2022.

[32] B. Hu, J. Jin, A. Y. Guo, H. Zhang, J. Luo, G. Gao, "GSDS 2.0: an upgraded gene feature visualization server," Bioinformatics, vol. 31, pp. 1296-1297, 2015.

[33] T. L. Bailey, N. Williams, C. Misleh, W. W. Li, "MEME: discovering and analyzing DNA and protein sequence motifs," Nucleic Acids Research, vol. 34, pp. W369-W373, 2006.

[34] A. Mortazavi, B. A. Williams, K. Mccue, L. Schaeffer, B. Wold, "Mapping and quantifying mammalian transcriptomes by RNA-Seq," Nature Methods, vol. 5, pp. 621-628, 2008.