PUBLIC HEALTH
OpenJournal
Openventio PUBLISHERS

Original Research Article

Codon Usage Analysis of Aminnopeptidase N Gene in Order Diptera

Renu Jakhar[1], SanjayKumar[2], Neelam Sehrawa[3], Deepika[4], Mohini Asija[5], S.K Gakha[6]

[1]Multi-Disciplinary Research Unit, PGIMS, Rohtak, Haryana, India
[2]Department of Pathology Nodal Officer, Multi-disciplinary Research Unit, PGIMS, Rohtak, Haryana, India
[3]Department of Genetics, Maharishi Dayanand University, Rohtak, Haryana, India
[4]Department of Medicine, PGIMS, Rohtak, Haryana, India
[5]Department of Health Research, Ministry of Health and Family Welfare, New Delhi, India
[6]Center for Biotechnology, Maharishi Dayanand University, Rohtak, Haryana, India

Corresponding Author
Renu Jakhar
Email: renujakhar22@gmail.com

Article Information:
Received: 14 December 2025 | Revised: 19 January 2026 | Accepted: 16 February 2026 | Published: March 22, 2026

Cite this article:
Renu Jakhar, SanjayKumar, Neelam Sehrawa, Deepika, Mohini Asija S.K Gakha(2026) Codon Usage Analysis of Aminnopeptidase N Gene in Order Diptera *Public Health Open Journal*. 11(1):203–223.
https://doi.org/10.17140/PHOJ.11.01.203

Abstract
Aminopeptidase N (APN) has been identified as Cry toxin-binding proteins in various insects and also responsible for digestion of peptides. An attempt has been made on the APN orthologous to overview the genome-wide analysis of APN gene in Order Diptera. The entire ORF of APN gene sequences belonging to mosquito and non-mosquito species in Order Diptera have been retrieved from NCBI and analysed by various bioinformatics tools. The goal of this study is to perform comparative analysis of APN nucleotide sequences, nucleotide composition, codon usage bias pattern, amino acid composition bias, pattern of conservation among amino acid residues, and phylogenetic analysis among insect species belonging to the order Diptera. Our results may provide useful insights on the patterns of codon usage bias that facilitate a better understanding of the structure and evolution of gene coding sequences of these insects.

Keywords: *Anopheles gambiae, Aminopeptidase N, Paralogous gene, Genome, Multigene, Evolution, Aedes ageypti, APN genes, Malaria, Dengue.*

Introduction

Mosquito-borne diseases continues to remain a life threatening infectious disease throughout the tropical region of the world. The global health and socioeconomic burden imposed by the mosquitoes is well established, with an estimated 282 million malaria cases and 14 million deaths reported worldwide in 2024 (WHO Report, 2025). It is a leading infectious disease of humans worldwide. It kills around 610 000 deaths over the world every year that account for a large number of death cases worldwide. In 2024, more than 14.6 million dengue cases were reported to WHO, with over 9,000 dengue-related deaths globally. The midgut is the site where the parasite first interacts with mosquito tissue and further developmental process takes place. The mosquito's midgut represents one of the most challenging environments for survival of parasites and thus midgut is a striking site for novel target malaria control strategies. The work reported suggests that genetic modification of mosquito vectorial capacity is feasible and represents a major step toward the goal of containing the spread of malaria. Strong and ubiquitous promoters of midgut could be used to drive the expression of effector genes that hinders transmission by either killing or interfering with parasite development. The alanyl Aminopeptidase N is the leading malarial TBV immunogen [1]. A midgut-specific protein, Aminopeptidase N 1 (APN1) is glycosylphospotidyl inositol anchored protein reported to play an important role in ookinete invasion of Plasmodium in the An. gambiae. The Anopheline Alanyl APN is originally isolated from the apical brush border microvilli fraction of An. gambiae midguts [2]. It is a ubiquitous enzyme which is found in a wide range of organisms from insects to mammals. APN belongs to a group of membrane-bound zinc enzymes [3] Dipteran Aminopeptidase N genes were observed to be highly biased genes among all the orders of class Insecta as revealed through nucleotide composition. In this study, an attempt has been made to overview genome-wide identification of APN genes in Diptera. The study covered comparative analysis of APN nucleotide sequences, nucleotide composition, codon usage bias pattern, amino acid composition analysis among insects to better understand the evolution of APN and its biological significance. A comparative sequence analysis was carried out to gain better understanding of the evolution of APN gene. The analysis of complete gene and CDS sequences of Aminopeptidase N was performed. The ORF of APNs belonging to Diptera order of class Insecta have been retrieved from various databases such as NCBI, Uniprot KB, vectorbase and flybase database etc and have been analysed by applying various bioinformatics tools. All the sequences show significant similarity with each other as revealed after performing BLAST. Although mosquitoes APN genes have diverged considerably from one another, the complete sequence can be aligned along their entire length. The nucleotide composition analysis was performed using Genomics GC% Content Calculate. The Amino acids composition of all known Aminopeptidase N proteins revealed mosquito-specific amino acid usage. The extent of codon bias in mosquito Aminopeptidase N genes based on the Relative Synonymous Codon Usage values was also determined. Gene structure study shows high sequence variations among them. In addition, nucleotide sequence alignment, intron analysis and nucleotide composition was carried out. The genus wise multiple sequence alignment of APN gene was performed using CLUSTAL W. Sequence comparisons of the genes showed that, except in conserved regions, there are many nonsynonymous substitutions, suggesting that the APN genes are highly evolving genes. Phylogenetic analyses were also carried out to establish the phylogenetic relationships among different orthologs present in other organisms. Protein-protein interactions show that APN is highly connected protein, supporting their role as hub with glutathione metabolism, Glutathione acts as insecticide resistance, so disrupting of APN protein or any other protein, this protein network may block the whole metabolic pathway of Glutathione. Models of proteins were also studied to further expand the knowledge of functioning of the gene.

MATERIALS AND METHODS

Retrieval and analysis of Mosquitoes APN1 orthologous gene sequences: The APN1 gene sequences of 15 different mosquitoes and 15 non mosquitoes have been retrieved from Vectorbase, Flybase, UniProt KB and NCBI databases for comparison. By using these databases, the functional annotation of genes, number of nucleotides, gene products, gene orientation and position, chromosome mapping and amino acids sequences were analysed in all orthologous.

Sequence analysis: BLAST searches were carried out to find sequences with similarity in the databases. Sequences for alignment and phylogenetic analysis were retrieved from the Genbank. Multiple sequence alignments were performed using the Bioedit tool. MEGA 7 software was used to analyse and establish phylogenetic relationships among the mosquitoes APN1 genes [4].

Conserved domains identification for APN proteins: The conserved domains were identified in all orthologous APN sequences by various domain search tools

like NCBI CDD (Conserved domain database) [5], Pfam [6] and InterProScan [7] software, which in turn by searching significantly close orthologous family members. The predicted cleavage sites signal peptides were analysed by using Signal IP algorithm. Glycosylation and phosphorylation sites were predicted by using the Prosite tool [8]. Possible transmembrane domains in APN proteins were also found by using the TMHMM [9]. The Molecular Weight (MW) and Isoelectric point (pI) were calculated by Compute pI/MW [10]. Multiple sequence alignments were performed with the BIOEDIT program with all paralogous sequences.

Nucleotide Composition Analysis: General nucleotide composition (A, T, C and G %) and nucleotide composition at first, second and third position of each codon were analysed for dipteran insect Aminopeptidase N CDS using MEGA6 and Genomics % GC content calculator. The AT3 and GC3 indices refer to the third position of synonymous codon and it helps to depict the biasness of the gene [11].

Synonymous Codon Usage: Relative synonymous codon usage (RSCU) is calculated as the ratio of the observed frequency of a codon to the frequency expected if all synonymous codons of a particular amino acid are used equally [12]. RSCU values were analysed using MEGA and CAIcal software The effective number of codons used by a gene (ENC) is generally used to measure the bias of synonymous codons [13]. Codon Adaptive Index (CAI) is a measure of relative adaptiveness of codon usage of a gene towards the codon usage of highly expressed genes. CAIcal software was used to calculate ENC and CAI values [14].

Multiple sequence alignment and Phylogenetic analysis: Order wise multiple sequence alignment was performed with MEGA 6 and BIOEDIT programmes in all insect APN. This helps to depict conserved characteristic features among these insect species. Phylogenetic tree were generated by MEGA6 using Bootstrap consensus Neighbour-joining Method [15-16]. These trees were constructed using APN protein sequences to infer the evolutionary distances among insects.

RESULTS

Nucleotide Sequence Analysis: To understand the evolutionary changes of APN gene among Diptera of insects, the complete coding sequences of fifteen mosquitoes and non-mosquitoes from Diptera orders have been obtained from the NCBI/ Genbank and UniProtKB databases. Further, In silico comparative analysis was performed. All orthologous sequences were compared and analysed for conserved sites using software called MEGA6.06 [14]. The multiple sequence alignments were carried out to detect the conserved sequences [17]. Dipteran APN from different classes shows divergence from one another, moreover, APN genes in mosquitoes reveal strong similarities to each other (Table 1 and 2). However, the nucleotide coding sequence was found to have different numbers and positions of exons (Table 1 and 2). Most of the mosquitoes APN gene possess almost similar ORF size, within a range of 3-4 kb except Aedes aegypti. AUG and TAA (mostly) are used as a start and stop codon, respectively. Nucleotide sequence alignments of dipteran APNs have shown the presence of limited conservation at few short coding regions.

Nucleotide Composition of APN: Nucleotide Composition analysis of mosquitoes APN coding regions shows that the Anopheles APN genes were found to have higher GC content as compared to Aedes and Culex APN genes. The analysis shows GC content varies significantly among different species. Among Anopheles, An. gambiae has the highest GC content of (54.4%), while among Aedes and Culex, Culex has more GC (50.4%) content. The Aedes surprisingly has lower GC content, lower than 50% and measured with highest AT content of 51.5% among all mosquitoes. The data is shown Table 3 for Anopheles, Aedes and Culex respectively. The % GC and AT content of An. gambiae APN gene was found to be 56.1 and 43.9 respectively, while the values for GC1s, GC2s and GC3s are 55%, 43% and 70% respectively. The frequency of codons ending with G/C is higher than A/T in all mosquitoes. Table 3 shows the distribution of overall GC content of APN gene in various mosquitoes. The nucleotide composition analysis of non-mosquitoes was also performed and found that the APN gene belonging to Drosophila were found to be GC rich except some other species (Table 4). Dipterans APN gene was found to be GC rich and as compared to 1st and 2nd codon position high GC content was observed at 3rd codon position. This clearly indicates that the Diptera APN gene is a biased gene (Table 3 4). The overall nucleotide composition and composition at the third codon position in APN gene suggests that compositional constraint might be influencing the codon usage pattern of these species.

Amino Acids Composition: As APN proteins are reported to play an important role in digestion of blood. The deduced amino acids composition of insect APN proteins was determined. The amino acids composition of mosquito APN is shown in Table 5. Three amino acids residues; Alanine, leucine and threonine were found to be very rich in this protein, showing that these residues may be in high biological demand during digestion of blood. There was no other significant difference found in the amino acids composition of

Table 1. List of genomic sequences of APN gene of various mosquitos (Diptera).

S No	Species	Accession no.	Length (bp)	Start codon	Stop codon	ORF (bp)	Exon No.	AA	Mol. Wt (kDa)	PI	Signal peptide	Zn binding site	Peptidase_M1 domain	ERAP1_C domain	Threonine rich region
1	Anopheles gambiae	AGAP004809	3588	ATG	TAA	3063	5	1020	113	5.0	1-19	H366EYAH370	292-519	597-905	887-933
2	Anopheles arabiensis	AARA016470	3586	ATG	TAA	3063	5	1020	113	5.1	1-19	H366EYAH370	292-519	597-905	947-993
3	Anopheles merus	AMEM002547	3599	ATG	TAA	3063	5	1020	113	5.1	1-19	H366EYAH370	292-519	597-905	947-993
4	Anopheles farauti	AFAF015666	3704	ATG	TAA	3078	5	1025	114	5.0	1-25	H366EYAH370	292-519	597-904	949 - 998
5	Anopheles quadrianulatus	AQUA016895	3586	ATG	TAA	3063	5	1020	113	5.0	1-19	H366EYAH370	292-519	597-905	947 - 993
6	Anopheles sinensis	ASIC009153	3688	ATG	TAA	3093	5	1030	114	4.9	1-20	H365ELAH369	291-518	596-902	907-1006
7	Anopheles atroparvus	AATE011993	3872	ATG	TAA	3072	5	1025	113	4.9	1-20	H361EYAH365	287-514	592-892	942-997
8	Anopheles darling	ADAC006959	3612	ATG	TAG	3060	5	1019	113	5.1	1-23	H360EYAH364	286-513	591-894	939-1003
9	Anopheles maculatus	AMAM007648	3805	ATG	TGA	2775	5	924	104	5.0	-	H356EYAH360	282-509	587-892	-
10	Anopheles albimanus	AALB015678	3591	ATG	TAA	3048	5	1015	113	5.1	1-23	H359EYAH363	285-512	590-893	889-999
11	Anopheles culicifacies	MK033514	3517	ATG	TAA	3084	5	1027	114	5.0	1-25	H372EYAH376	280-525	603-912	956-1001
12	Aedesalbopictus	AALF017287	3322	ATG	TGA	3003	5	1000	112	4.9	1-23	H359EYAH363	285-512	590-903	946-974
13	Aedesageypti	AAEL012778	2957	ATG	TAA	2697	4	898	112	4.8	1-22	H359EYAH363	285-512	590-898	-
14	Culex quanqueasciatus	CPIJ001048	3305	ATG	TAA	3036	5	1011	113	4.9	1-20	H299EYAH303	285-512	590-912	866 - 925

Table 2. List of genomic sequences of APN gene of various non-mosquitos (Diptera)

Sr. no	Order	Family	Species	Accession No.	Length	ORF	Exons	Introns	AA	Mol. Wt.	Signal Peptide	Mature peptide	Start codon	Stop codon	Peptidase_M1	ERAP1_C
1.	Diptera	Drosophilidae	Drosophila ananassae	FBtr0122736	3274	2886	5	4	961	108	1-20	21-961	ATG	TAG	298 – 510	589 – 890
			Drosophila mojavensis	FBgn0147011	3037	2859	4	3	952	107	1-20	22-952	ATG	TAA	291 – 505	585 – 902
			Drosophila virilis	FBgn0210169	3143	2850	4	3	949	109	1-22	23-949	ATG	TAG	288 – 485	582 – 902
			Drosophila erecta	FBgn0109703	3180	2862	4	3	953	106	1-24	25-953	ATG	TAG	294 – 508	587 – 888
			Drosophila melanogaster	FBgn0051233	3185	2859	4	3	952	106	1-25	26-952	ATG	TAA	293 – 507	586 – 904
			Drosophila persimilis	FBgn0161378	3092	2877	4	3	958	107	1-25	26-958	ATG	TAG	294 – 505	585 – 888
			Drosophila grimshawi	FBgn0126717	3348	2880	5	4	959	109	1-20	21-959	ATG	TAA	297 – 508	588 – 897
			Drosophila pseudoobscura	FBgn0076210	3089	2877	4	3	958	107	1-22	23-958	ATG	TAG	294 – 505	585 – 887
			Drosophila yakuba	FBgn0228109	3504	2883	5	4	960	108	1-22	23-960	ATG	TAA	299 – 510	591 – 888
			Drosophila sechellia	FBgn0181222	3575	2900	5	4	966	108	1-22	23-966	ATG	TAA	303 – 515	595 – 893
			Drosophila simulans	FBgn0192352	3566	2886	5	4	961	108	1-22	23-961	ATG	TAA	298 – 510	590 – 888
			Drosophila willistoni	FBtr0243538	3065	2871	4	3	956	108	1-21	22-956	ATG	TAA	293 – 505	584 – 887
			Musca domestica	MDOA004623	8955	2346	5	4	781	89	-	-	GGC	TAA	106 – 330	409 – 705
		Muscoidea	Stomoxys calcitrans	SCAU003701	6181	2871	5	4	956	108	1-25	25-956	ATG	TAG	283-507	586-904
		Psychodidae	*Lutzomyia longipalpis*	LLOJ009515	4700	2808	6	5	934	105	1-21	22-934	ATG	TGA	270 – 497	578 – 879

mosquito APN. Furthermore, other insects belonging to Diptera also have more of the alanine and leucine amino acids (Table 5). In some non-mosquitoes like Drosophila grimshawi a significant decline in the number of alanine residues was noticed with an increase in arginine and serine residues. The Leucine residues are present in highest frequency in the amino acid composition of protein of each species (accounting for 9% of

Table 3. Nucleotide composition of APN gene of mosquitoes (Diptera)

Sno.	Species	A+T %			G+C%			AT	GC
1.	*Aedes ageypti*	47.0	61.4	46.2	53.2	38.5	53.6	51.5	48.5
2.	*Culex quanquefasciatus*	47.4	59.2	41.6	52.3	40.9	58.1	49.6	50.4
3.	*Anopheles darlingi*	44.3	56.1	30.4	56.0	43.9	69.9	43.4	56.6
4.	*Anopheles culicifacies*	45.0	55.4	29.6	55.1	44.9	70.1	43.4	56.7
5.	*Anopheles arabiensis*	44.5	55.9	30.9	55.4	43.8	68.8	44.0	56.0
6.	*Anopheles gambiae*	44.5	56.9	30.7	55.4	43.6	69.5	43.9	56.1
7.	*Anopheles merus*	44.5	56.0	32.3	55.6	43.6	68.1	44.3	55.7
8.	*Aedes albopictus*	46.9	58.3	40.8	53.3	41.1	59.6	48.6	51.3
9.	*Anopheles albimanus*	42.2	56.6	22.0	57.8	43.5	77.9	40.4	59.7
10.	*Anopheles atroparvus*	43.6	55.6	20.4	56.8	44.4	80.0	39.5	60.4
11.	*Anopheles farauti*	41.8	56.7	18.3	58.0	43.7	81.5	38.9	61.1
12.	*Anopheles maculatus*	44.4	56.7	33.0	56.0	43.0	67.0	44.6	55.3
13.	*Anopheles quadriannulatus*	44.6	55.8	31.4	55.6	43.9	68.1	44.2	55.8
14.	*Anopheles sinensis*	43.6	56.0	25.5	56.0	44.3	74.6	41.6	58.3
15.	*Anopheles stephensai*	43.1	55.2	23.0	56.4	45.3	77.5	40.3	59.8

Table 4. Nucleotide composition of APN gene of non-mosquitoes (Diptera)

Sno.	Species	A+T %			G+C%			AT	GC
1.	Drosophila melanogaster	46.6	61.5	30.9	53.4	38.4	68.8	46.5	53.6
2.	*Drosophila mojavensis*	44.3	60.6	34.9	55.9	38.1	65.1	46.5	53.4
3.	*Drosophila virilis*	45.9	60.7	38.9	53.8	39.2	61.5	48.5	51.5
4.	Drosophila erecta	45.6	60.8	25.0	53.9	39.8	74.8	44.0	56.0
5.	Drosophila ananassae	49.7	58.8	30.8	48.0	40.2	68.9	53.2	53.2
6.	Drosophila persimilis	48.1	59.3	24.8	52.2	51.7	75.3	56.1	56.1
7.	Drosophila grimshawi	61.3	47.9	50.1	39.0	40.7	49.5	46.7	46.7
8.	Drosophila pseudoobscura	48.4	59.7	25.0	52.0	42.2	75.2	55.9	55.9
9.	Drosophila yakuba	51.8	57.9	29.0	55.9	39.0	71.0	54.0	54.0
10.	Drosophila sechellia	45.3	60.6	34.9	48.9	42.7	65.5	53.4	53.4
11.	*Drosophila simulans*	50.3	56.9	32.9	49.0	-	67.1	47.1	52.5
12.	*Drosophila willistoni*	51.3	59.8	54.8	45.7	39.8	45.1	55.4	44.7
13.	*Musca domestica*	50.6	62.7	45.6	50.4	37.2	54.6	54.1	45.8
14.	*Stomoxys calcitrans*	54.4	60.3	40.3	49.4	38.8	59.4	50.5	49.6
15.	*Lutzomyia longipalpis*	51.0	60.8	53.4	49.4	39.6	46.7	54.7	45.2

total amino acids) whereas cysteine residues are least abundant (1%) in the protein of insects (Table 6).

Synonymous Codon Usage: Codon bias is the non-uniform use of synonymous codons which encode the same amino acid. Some codons are more frequently used than others in several organisms, particularly in highly expressed genes. The spectacular diversity of insects makes them a suitable candidate for analysing the codon usage bias. To understand the pattern of non-random usage of synonymous codons in the insects, RSCU of individual codons were compared between mosquitoes and non-mosquitoes groups of species. RSCU> 1 represents codons that are used more frequently than expected whereas <1 represents codons which are used less frequently. The overall RSCU values for the 59 codons in Anopheles indicated that C and G occurred more frequently at the third codon position (as shown in Table 7). These are TTC, CTG, ATC, GTG, TCC, TCG, CCG, ACC, ACG, GCC, GCG, TAC, CAG, AAC, AAG, GAG, TGC, CGC and GGC. Among these frequent codons, the two most frequently used codon having highest values are CTG/Leu and CCG/Pro and amino acids such as asparagine, lysine, glutamine, phenylalanine, and tyrosine are encoded by one of the preferential synonymous codon. Ae. aegypti and Cx. quinquefasciatus APN genes also exhibit high synonymous codon usage bias (Table 7). The number of frequently used codons and rarely used codons varies among species, for example there were three amino acids in Ae. aegypti were GAG is preferred over GAA for glutamic acid and CCG is preferred over CCA for proline. The APN gene of mosquitoes showed a high synonymous codon usage bias in comparison to other insects, preferentially using one or two synonymous codons (Table 7 8). Three codons (CGC/Arg, ACC/Thr and UCC/Ser) are used frequently in Diptera. Furthermore some codons are found that are used frequently in each species of Diptera order of insects (Table 7 8). Specific codon such as CUA/Leu and GGG/Gly are found that are avoided in each species of each order. Within

Table 5. Amino acids composition of APN in mosquitoes (Diptera)

S no	Species	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
1.	*Aedes aegypti*	8.35	0.55	5.34	6.3	3.45	4.9	1.78	5.79	4.12	9.24	2.00	6.12	4.23	3.78	4.45	5.67	8.12	7.57	2.33	5.79
2.	*Culex quinquefasciatus*	9.59	0.59	5.24	6.42	3.65	4.94	2.07	6.13	3.36	9.69	2.07	6.13	3.65	3.06	4.45	6.23	9.59	6.33	1.87	4.84
3.	*Anopheles darlingi*	9.61	0.78	5.78	5.69	3.33	5.20	2.45	4.80	1.96	10.6	2.25	4.71	4.31	3.63	5.49	5.78	10.9	5.88	1.76	4.80
4.	*Anopheles culicifacies*	10.6	1.07	4.96	5.74	3.99	4.96	1.46	4.47	2.23	9.44	2.72	5.06	4.18	3.21	5.55	6.62	9.93	7.10	1.94	4.67
5.	*Anopheles arabiensis*	10.2	0.68	5.09	5.49	3.72	5.98	1.66	4.50	2.64	9.80	2.64	5.39	3.82	3.62	5.09	6.17	10.0	6.66	1.76	4.90
6.	*Anopheles gambiae*	10.2	0.68	5.09	5.58	3.72	5.98	1.76	4.50	2.64	10.0	2.64	5.29	3.72	3.62	4.90	6.17	10.0	6.56	1.76	4.80
7.	*Anopheles merus*	9.90	0.68	5.29	5.58	3.72	5.98	1.66	4.60	2.74	9.60	2.64	5.19	3.82	3.62	5.19	6.37	9.90	6.86	1.76	4.80
8.	*Aedes albopictus*	9	1	5.3	6	3.9	5	2	6	3	9	2	5.7	4	4	4	7	9	7	2	5
9.	*Anopheles albimanus*	10.5	0.59	5.51	5.91	3.25	5.32	2.26	4.63	2.56	10.1	2.16	4.23	4.03	3.94	5.32	5.91	9.95	6.79	1.77	5.12
10.	*Anopheles atroparvus*	10.8	0.78	5.36	5.46	3.60	6.24	1.75	5.26	2.04	9.65	2.34	4.97	3.90	3.31	5.36	5.56	10.0	7.12	1.75	4.58
11.	*Anopheles farauti*	10.8	0.58	5.65	5.36	4.09	4.87	1.95	4.39	1.85	9.75	2.24	4.68	4.09	3.41	5.75	5.56	10.2	8.19	1.75	4.68
12.	*Anopheles maculatus*	10.2	0.64	4.76	7.90	4.11	5.30	1.94	5.08	2.16	8.87	2.16	4.97	4.11	3.46	6.70	5.95	8.00	6.16	1.83	5.51
13.	*Anopheles quadriannulatus*	10	0.68	5.29	5.58	3.72	5.88	1.66	4.41	2.64	9.70	2.64	5.19	3.72	3.52	5.19	6.47	10.1	6.86	1.76	4.80
14.	*Anopheles sinensis*	9.90	0.58	4.75	6.40	3.39	5.82	1.45	4.27	2.33	10.1	2.03	5.24	4.07	3.00	5.43	5.53	11.1	7.76	1.84	4.75
15.	*Anopheles stephensi*	10.6	0.68	5.20	5.29	3.82	5.59	1.57	4.71	1.96	10.1	1.96	4.90	3.72	3.33	5.59	6.77	10.5	6.96	1.76	4.80

Table 6. Amino acids composition of APN in non-mosquito Diptera

S no	Species	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
1.	*Drosophila melanogaster*	8.19	0.73	5.46	6.09	4.62	5.14	2.10	4.09	3.67	10.8	1.36	6.51	3.67	4.62	3.46	7.14	8.29	7.24	1.68	5.04
2.	*Drosophila mojavensis*	9.13	0.73	5.77	5.98	4.30	4.72	2.20	4.30	3.67	10.9	1.68	6.09	3.88	4.93	4.20	6.93	7.56	6.19	1.78	4.93
3.	*Drosophila virilis*	9.16	0.73	5.58	6.63	4.53	4.53	2.31	4.53	3.79	10.7	1.47	6.53	3.58	4.32	3.89	6.63	8.74	5.90	1.79	4.53
4.	*Drosophila erecta*	8.49	0.62	5.24	5.35	4.93	5.56	2.09	3.67	3.67	10.2	1.36	6.71	3.67	4.93	3.56	7.66	7.97	7.66	1.67	4.82
5.	*Drosophila ananassae*	7.59	0.72	5.61	5.20	4.26	4.47	2.08	5.09	2.70	10.8	1.14	6.76	3.64	4.16	4.05	9.78	8.42	5.82	2.28	5.30
6.	*Drosophila persimilis*	8.49	0.73	5.24	6.19	5.03	4.72	1.99	4.82	3.25	11.2	1.36	7.24	3.77	5.03	3.46	6.92	7.13	6.71	1.78	4.82
7.	*Drosophila grimshawi*	5.64	3.87	0.73	1.35	2.92	3.45	3.55	7.21	3.13	10.0	4.81	3.97	7.84	3.76	10.5	10.0	8.15	4.07	2.61	2.19
8.	*Drosophila pseudoobscura*	8.35	0.62	6.15	4.69	4.48	4.90	1.77	4.59	3.02	11.4	1.04	6.88	3.13	5.11	3.86	9.39	8.03	5.01	2.40	5.01
9.	*Drosophila yakuba*	7.60	0.63	5.63	4.68	4.27	4.68	1.88	5.31	3.02	10.7	1.04	6.56	3.13	3.43	4.38	10.1	9.47	5.52	2.18	5.72
10.	*Drosophila sechellia*	9.13	0.73	5.77	5.98	4.30	4.72	2.20	4.30	3.67	10.9	1.68	6.09	3.88	4.93	4.20	6.93	7.56	6.19	1.78	4.93
11.	*Drosophila simulans*	8.22	0.62	5.93	4.37	4.37	4.47	1.76	5.41	3.01	10.5	1.04	6.55	3.32	3.53	4.37	9.78	9.46	5.09	2.39	5.72
12.	*Drosophila willistoni*	8.05	0.83	5.54	6.17	4.70	4.39	2.19	5.33	3.45	9.20	1.15	6.06	3.24	4.81	3.66	8.47	8.99	5.96	2.19	5.54
13.	*Musca domestica*	7.55	0.76	5.63	5.12	3.84	4.73	2.43	5.63	5.12	8.57	2.30	6.91	3.58	4.99	3.58	8.19	6.65	5.76	2.17	6.40
14.	*Stomoxys calcitrans*	8.47	0.62	5.64	5.64	4.18	5.33	2.82	5.43	3.87	8.68	2.40	7.32	4.28	4.18	4.39	6.38	7.21	5.75	2.09	5.23
15.	*Lutzomyia longipalpis*	7.81	0.85	6.42	4.92	4.81	5.56	1.71	6.95	2.03	9.20	2.78	6.10	3.85	3.64	4.49	6.20	8.77	5.78	2.03	5.99

the Insect, the preference shifts accomplished by high bias are seen in mosquitoes of Diptera and show preference shifts in three amino acids phenylalanine, histidine and threonine.

Effect of base composition on codon bias: Because of differences in mutational bias, the percentage of GC content varies greatly among different species, even in the species from the same order. An earlier study shows that genes in the Hymenoptera genome are generally localized in GC poor regions whereas among dipteral genomes such bias is not evident [18]. To determine the association of GC biasness with codon bias among different species, we move to the non-directional codon bias measures effective number of codon (ENC), which depends on nucleotide composition of genes [13]. An ENC value varies from 20 to 61, when ENC=20 means that only one codon is used for each amino acid (extreme codon bias) and ENC =61 means all codons are equally likely to code the amino acid (no biased usage of codon). The dipterans and particularly the mosquitoes were found to have the highest codon usage bias with an average ENC value of 45 and also with the highest average GC content at third codon position; 77.9 (Table 9). The greater the extent of codon preference in a gene, the lesser is the corresponding ENC value. The ENC values range from 40 to 56 in mosquitoes. Lower ENC value means more bias (Table 9 10). The average for the non-mosquito species was 47.4. It was found that there was a positive correlation between GC content and the degree of synonymous codon usage bias measured by ENC. The highest synonymous codon usage bias was found in An. albimanus APN with an ENC of 40.4 and GC3s is 78% (9-10). CAI is a directional measure of codon usage bias similar to relative codon bias score unlike ENC. Thus comparison between CAI and ENC provides a good quantitative assessment between nucleotide composition and codon bias selection. This gene shows increase in CAI value with decrease in ENC value in Diptera. CAI is used to calculate the level of gene expression. The average CAI value ranges from 0.5 to 0.7, a level close to 1.0, which might give a high expression level of genes according to the index definition. Drosophila willistoni is found to have lower codon biases with highest ENC value.

Table 7. Relative Synonymous Codon Usage of APN gene in mosquitoes (Diptera)

Amino Acid	**Codons**	Ae. aegypti	Cu. quinquefasciatus	An. darlingi	An. culicifacies	An. arabiensis	An. gambiae	An. merus	Ae. albopictus	An. albimanus	An. atroparvus	An. farauti	An. maculatus	An. quadriannulatus	An. sinensis	An. stephensi
Phe (F)	TTT	0.77	0.54	0.29	0.2	0.47	0.37	0.42	0.56	0.24	0.22	0.29	0.26	0.47	0.34	0.21
	TTC	1.23	1.46	1.71	1.8	1.53	1.63	1.58	1.44	1.76	1.78	1.71	1.74	1.53	1.66	1.79
Leu (L)	TTA	0.14	0.12	0.11	0.31	0.18	0.29	0.18	0.07	0.06	0.06	0.06	0.29	0.18	0.06	0.06
	TTG	1.16	1.35	1.21	0.74	0.84	0.82	0.73	1.4	0.93	0.91	0.78	0.88	0.67	1.2	0.58
	CTT	1.23	1.04	0.5	0.49	0.42	0.47	0.43	0.63	0.06	0.48	0.06	0.8	0.42	0.23	0.47
	CTC	0.8	0.86	0.94	0.74	1.02	0.99	0.98	1.12	1.34	0.97	1.38	0.88	0.97	1.03	1.05
	CTA	0.29	0.37	0.61	0.49	0.66	0.52	0.67	0.42	0.52	0.12	0.3	0.8	0.61	0.4	0.41
	CTG	2.39	2.27	2.64	3.22	2.88	2.91	3	2.37	3.09	3.45	3.42	2.34	3.15	3.09	3.44
Ile (I)	ATT	1.5	1.16	0.49	0.59	0.85	0.85	0.96	1.36	0.26	0.72	0.8	0.64	0.87	0.41	0.69
	ATC	1.21	1.74	2.27	1.96	1.89	1.89	1.79	1.42	2.74	2.11	2.13	1.98	1.93	2.39	2.19
	ATA	0.29	0.1	0.24	0.46	0.26	0.26	0.26	0.22	0	0.17	0.07	0.38	0.2	0.2	0.13
Met (M)	AUG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Val (V)	GTT	1.35	1.5	0.53	0.22	0.59	0.66	0.74	1.08	0.64	0.55	0.43	0.56	0.57	0.8	0.34
	GTC	0.88	0.75	1.4	1.1	0.88	0.9	0.8	1.03	0.87	1.1	0.95	1.12	0.86	1.1	0.85
	GTA	0.47	0.63	0.33	1.1	0.47	0.48	0.46	0.59	0.41	0.38	0.38	0.91	0.51	0.6	0.51
	GTG	1.29	1.13	1.73	1.59	2.06	1.97	2	1.3	2.09	1.97	2.24	1.4	2.06	1.5	2.31
Ser (S)	TCT	0.47	1.24	0.71	0.44	0.48	0.38	0.46	0.79	0.1	0.11	0.11	0.22	0.36	0.21	0.35
	TCC	1.41	1.24	0.51	1.24	1.33	1.43	1.57	1.5	1	1.79	1.89	1.31	1.36	1.47	1.48
	TCA	0.35	0.38	0.31	0.44	0.29	0.19	0.09	0.71	0.2	0.11	0	0.44	0.18	0.21	0.17
	TCG	1.53	0.86	2.24	1.94	2	2	1.94	0.79	2.3	2.53	1.89	1.75	2.09	1.89	1.74
	AGT	1.88	0.86	1.02	0.79	1.05	1.05	1.02	1.06	0.7	0.32	0.42	1.53	1	0.84	0.96
	AGC	0.35	1.43	1.22	1.15	0.86	0.95	0.92	1.15	1.7	1.16	1.68	0.76	1	1.37	1.3
Pro (P)	CCT	0.95	0.32	0.36	0.19	0.41	0.42	0.41	0.76	0.29	0.4	0.1	0.53	0.42	0.29	0.32
	CCC	0.95	0.76	0.55	1.02	0.72	0.74	0.82	1.05	0.49	1	1.14	0.84	0.74	0.76	0.95
	CCA	1.37	1.3	0.64	0.28	1.03	0.95	1.03	1.05	0.59	0.3	0.19	0.74	0.95	0.57	0.42
	CCG	0.74	1.62	2.45	2.51	1.85	1.89	1.74	1.14	2.63	2.3	2.57	1.89	1.89	2.38	2.32
Thr (T)	ACT	0.93	0.66	0.5	0.47	0.55	0.66	0.59	0.75	0.4	0.54	0.27	0.43	0.65	0.52	0.41
	ACC	1.53	1.73	1.39	1.22	1.53	1.4	1.35	1.58	1.31	1.79	1.75	1.51	1.38	1.57	1.38
	ACA	0.6	0.41	0.57	0.39	0.27	0.27	0.32	0.4	0.36	0.23	0.27	0.22	0.35	0.28	0.07
	ACG	0.93	1.2	1.54	1.92	1.65	1.67	1.74	1.27	1.94	1.44	1.71	1.84	1.62	1.63	2.13
Ala (A)	GCT	1.07	1.03	0.69	0.59	0.76	0.72	0.79	0.72	0.49	0.29	0.29	0.59	0.75	0.55	0.29
	GCC	1.12	1.48	1.51	1.36	1.18	1.22	1.11	1.57	1.64	1.77	1.44	1.52	1.18	1.88	1.76
	GCA	1.07	0.78	0.82	0.77	0.95	0.88	0.99	1.11	0.34	0.36	0.4	0.72	0.98	0.59	0.62
	GCG	0.75	0.7	0.98	1.28	1.1	1.18	1.11	0.6	1.53	1.59	1.87	1.18	1.1	0.98	1.32
Tyr (Y)	TAT	0.58	0.73	0.37	0.58	0.44	0.45	0.45	0.72	0.35	0.3	0.25	0.59	0.49	0.37	0.24
	TAC	1.42	1.27	1.63	1.42	1.56	1.55	1.55	1.28	1.65	1.7	1.75	1.41	1.51	1.63	1.76
STOP	TAA(*)	3	3	0	3	3	3	3	0	3	3	3	0	3	3	3
	TAG(*)	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
His (H)	CAT	1.25	0.86	0.64	1.2	1.18	1	1.18	0.7	0.7	0.56	0.5	1	1.18	0.4	1.25
	CAC	0.75	1.14	1.36	0.8	0.82	1	0.82	1.3	1.3	1.44	1.5	1	0.82	1.6	0.75
Gln (Q)	CAA	0.94	0.65	0.59	0.42	0.27	0.27	0.32	1.1	0.35	0.24	0.23	0.31	0.28	0.32	0.18
	CAG	1.06	1.35	1.41	1.58	1.73	1.73	1.68	0.9	1.65	1.76	1.77	1.69	1.72	1.68	1.82
Asn (N)	AAT	0.91	0.55	0.67	0.5	0.65	0.63	0.68	0.67	0.33	0.24	0.17	0.74	0.64	0.41	0.32
	AAC	1.09	1.45	1.33	1.5	1.35	1.37	1.32	1.33	1.67	1.76	1.83	1.26	1.36	1.59	1.68
Lys (K)	AAA	0.97	0.94	0.3	0.52	0.3	0.3	0.43	0.63	0.31	0.19	0.32	0.6	0.44	0.25	0.4
	AAG	1.03	1.06	1.7	1.48	1.7	1.7	1.57	1.38	1.69	1.81	1.68	1.4	1.56	1.75	1.6
Asp (D)	GAT	1.21	0.98	1.15	1.06	1.04	1	1	1.06	1.14	0.44	0.93	0.91	1	0.73	0.72
	GAC	0.79	1.02	0.85	0.94	0.96	1	1	0.94	0.86	1.56	1.07	1.09	1	1.27	1.28
Glu (E)	GAA	1.19	1.48	0.55	0.85	0.64	0.7	0.67	1.11	0.53	0.68	0.73	0.82	0.77	0.73	0.74
	GAG	0.81	0.52	1.45	1.15	1.36	1.3	1.33	0.89	1.47	1.32	1.27	1.18	1.23	1.27	1.26
Cys (C)	TGT	0.8	1	0.5	1.27	0.86	0.86	0.86	0.8	0.33	0	0.67	0.67	0.86	0.33	0.86
	TGC	1.2	1	1.5	0.73	1.14	1.14	1.14	1.2	1.67	2	1.33	1.33	1.14	1.67	1.14

Amino Acids Sequence Alignment: Multiple sequence alignment of APN protein sequences were performed separately to know the conservation patterns within an order. The complete sequences of amino acids of APN belonging to Diptera order have been analysed. Sequence analysis of APN indicates all are evolved from common ancestors. The amino acids sequence alignment of is shown in Fig.1 2. The areas of highly conserved regions for amino acid were clearly observed. From the MSA it is clear that residues GAMEN, HEXXH and NEGFA are highly conserved, showing that these conserved sites are present in almost all the aligned sequences. All these three motifs are found conserved in Anopheles at similar positions. These repeats were also found in other mosquito species and showed homology to other insects. Glycosylation sites are also found conserved in all mosquitoes. Four cysteine residues are also found conserved in all insects. GPI prediction program exhibited the presence of GPI anchor signal sequence at the C-terminus. Such a sequence was found to be identical with An. gambiae and An. stephensi. Furthermore analysis of the C-terminal part of all insect APNs revealed a high degree of conservation of tyrosine residues.

Phylogenetic Analysis: Coding regions of AcAPN1 were examined for Phylogenetic analysis to study the evolutionary patterns among mosquitoes. This analysis includes 15 selected mosquitoes for the APN1 gene. The evolutionary pattern on phylogenetic tree showed that Anopheles, Aedes and Culex APN1 gene may have originated from common ancestors and revealed close similarity to each other. The relationship among all three genus of mosquito is more strikingly represented in two clades: Anopheles and Aedes/ Culex, this show Aedes and Culex shows close similarity than Anopheles. The evolutionary relationship of APN1 protein derived from different Anopheles was also evaluated, among this Anopheles group An. culicifacies found to be placed nearer to An. stephensi and An. gambiae, to these which is in agreement with the peptide sequences of these genera as all contains GAMEN catalytic domain (Fig. 3). The tree topology shows that the Anopheles APN1 shares orthologs from other mosquitoes such as Aedes and Culex. Phylogenetic tree based on APN protein of fifteen species of Diptera order of Non mosquitoes were analysed separately to know the evolutionary relationship of organisms within an order. Phylogenetic tree also supported the grouping of species into two major clades. Phylogenetic analysis of APN indicates that all are evolved from common ancestors (Fig. 3 4).

DISCUSSION

Conserved DNA sequences of APN gene have been retrieved from 30 mosquitoes and non-mosquito species from NCBI, GenBank etc and has been analysed with various bioinformatic tools. The APN gene sequences consist of a start site, ATG followed by TAA (in most species) as termination codon site. The nucleotide sequences of all APN insects were found to have different numbers and positions of exons. The ORF region of APN gene almost falls in between 2-3 kb. The present investigation highlights comparative studies on nucleotide composition and codon bias between 30 species belonging to order Diptera of class Insecta. As synonymous is not uniform during the translational process, the identification of the codon usage pattern is important to understand the translational selection of codons in protein coding genes in various insects' species. The overall RSCU values in mosquitoes indicated that C/G occurred more frequently at the 3rd codon position whereas in Hymenoptera, Lepidoptera and Hemiptera insect orders most frequent codon ends with A/T at the 3rd position (Sharma et al., 2014). In Anopheles and Culex species APN gene was found to be CG rich, whereas in Aedes AT richness was observed. Also for comparison, coding sequences of APN gene were taken from various insects to perform nucleotide composition analysis. All insect species were found to be AT rich except Diptera. The analysis shows that at 1st and 2nd codon positions, the A content was found higher than the T content except for the 3rd codon position. Dipterans APN gene was found to be GC rich and as compared to 1st and 2nd codon position high GC content was observed at 3rd codon position. These observations clearly indicate that the Diptera APN gene is a biased gene. The pattern of GC contents of the honeybee genome and its relationship with codon usage was studied [19]. This showed that genes located in GC poor regions showed much larger deviation in both codon usage bias and amino acid bias than genes located in GC rich regions. The codon context pattern in the honey bee genome indicates that T and G combination occurs more frequently for adjacent 1st and 3rd positions of codon pairs than A and C combinations [20]. It thus suggested that such a pattern of codon context bias might have evolved in order to minimize the effect of nucleotide mutations coding regions. Except methionine and tryptophan, the APN1 amino acid composition revealed mosquito-specific amino acid usage. This study shows the amino acid alanine is the amino acid that accounts for the greatest usage and cysteine accounts for the least usage in the genes of Diptera. The amino acid composition analysis of other orders of APN reveals the presence of thymine residues in high content

Amino Acid	**Codons**	Ae. aegypti	Cu. quinquefasciatus	An. darlingi	An. culicifacies	An. arabiensis	An. gambiae	An. merus	Ae. albopictus	An. albimanus	An. atroparvus	An. farauti	An. maculatus	An. quadriannulatus	An. sinensis	An. stephensi
STOP	TGA(*)	0	0	0	0	0	0	0	3	0	0	0	3	0	0	0
Trp (W)	TGG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Arg (R)	CGT	1.2	2.13	1.5	2	1.62	1.56	1.58	1.4	0.78	0.87	0.51	0.87	1.7	1.71	0.63
	CGC	1.35	1.33	2.68	2	2.42	2.4	2.49	1.12	2.56	2.84	2.24	1.94	2.38	2.14	2.74
	CGA	1.8	1.07	0.64	0.21	0.81	0.96	0.91	1.53	0.22	0.33	0.41	0.77	0.91	0.54	0.84
	CGG	0.6	0.4	0.75	1.26	0.69	0.72	0.57	0.7	2.11	1.31	2.34	0.87	0.57	0.96	1.16
	AGA	1.05	0.53	0.21	0.21	0.35	0.24	0.23	0.42	0.22	0.44	0.2	1.26	0.23	0.32	0.21
	AGG	0	0.53	0.21	0.32	0.12	0.12	0.23	0.84	0.11	0.22	0.31	0.29	0.23	0.32	0.42
Gly (G)	GGT	1	1.28	1.96	1.57	1.38	1.31	1.25	1.2	1.93	1.63	1.28	1.39	1.4	1.47	1.33
	GGC	1.18	0.64	0.98	1.18	1.44	1.57	1.57	0.8	1.26	1.5	1.6	1.63	1.33	1.13	1.61
	GGA	1.73	2.08	0.75	0.63	0.85	0.79	0.92	1.68	0.44	0.63	0.48	0.82	0.87	0.87	0.7
	GGG	0.09	0	0.3	0.63	0.33	0.33	0.26	0.32	0.37	0.25	0.64	0.16	0.4	0.53	0.35

[21]. The present study compares the codon usage patterns between different types of species from Diptera order of insects. As synonymous codon usage is not uniform during the translation process, the identification of the codon usage pattern is important to understand the translational selection of codons in protein coding genes among related species. A significant high synonymous codon usage bias was observed in mosquito APN gene belonging to the order Diptera, by predominantly using one or two optimal synonymous codons over others. Similar results observed in non-mosquitoes species of Diptera order. In this study, the overall RSCU values for the 59 codons in mosquitoes indicated as C and G (C is more) occurs more frequently at the 3rd codon position. These results indicate that codon usage pattern in the mosquitoes is mostly contributed by compositional constraints. This main factor affects the synonymous codon usage bias reported as in the case of D. melanogaster [22-23]. In Drosophila generally G and especially C are favoured at synonymous sites in biased genes. In Anopheles, the codon (CTG/Leu and CCG/Pro) are used more frequently. Amino acids like phenylalanine, tyrosine, glutamine, asparagine, lysine and cysteine are almost solely encoded by one synonymous codon. The codon bias of Aedes ageypti and Anopheles gambiae mosquito's genome was analysed [24]. Genome wide analysis of relative synonymous codon usage value of gene indicated that only a select set of codon was optimized in these mosquitoes. Optimized and rare codons were also identified in both species. Results support that high codon bias occurs in highly expressed genes. The comparison shows that although the pattern of codon usage of the Culicinae mosquitoes (Ae. ageypti and Cu. quinquefasciatus) is similar to the Anopheline mosquito, there was variation in the usage of individual codons [25]. Because half of the codons are more biased in An. gambiae than in Ae. aegypti and Cu. quinquefasciatus, while the other half of codon tends to be more biased in Culicinae than An. gambiae. Detailed analysis of codon bias among twelve species of Drosophila sequenced genome was done[26], It was found that one species, D. willistoni has unusually low codon usage bias pattern in the usage of synonymous codons. While G or C ending codons were preferentially used in other species, D. willistoni genes showed a shift toward A or T ending codon preference and high ENC value. Genome sequences of 22 insect species of order Diptera and Hymenoptera where was known [27]. The study revealed that in Diptera more frequent codon ends with G or C mostly at 3rd codon position whereas in Hymenoptera insect order, most frequent codon contain A or T at its 3rd position. Also in Lepidoptera, analysis of codon usage bias of Bombyx mori revealed that most of the preferentially codons end with A or T at the 3rd codon position [28]. Codon usage on two Hemiptera insect species namely, Bemisia tabaci and Homalodisca coagulata was studied [21]. They found that most frequent codons end with A or C at 3rd codon position. The above said studies suggest that the third position of codon seems to affect codon bias. The results of our analysis are in agreement with these results in similar manner in species of Diptera orders. This finding may be the result of compositional constraint that occurs in codon usage patterns in these insect species. The extent of codon bias in mosquitoes APN genes based on the RSCU value was

Table 10. Relative Synonymous Codon Usage (RSCU) of APN gene in non-mosquitoes (Diptera)

Amino Acid	Codons	D. ananassae	D. mojavensis	D. melanogaster	D. virilis	D. erecta	D. persimilis	D. grimshawi	D. pseudoobscura	D. yakuba	D. sechellia	D. simulans	D. willistoni	Musca domestica	Stomoxys calcitrans	Lutzomyia longipalpis
Phe (F)	TTT	0.59	0.54	0.59	1.02	0.64	0.54	1.14	0.47	0.54	0.54	0.71	0.84	0.53	0.75	0.67
	TTC	1.41	1.46	1.41	0.98	1.36	1.46	0.86	1.53	1.46	1.46	1.29	1.16	1.47	1.25	1.33
Leu (L)	TTA	0.12	0.12	0.12	0.24	0.06	0	1	0.07	0.06	0.18	0.12	0.41	0	0.14	0.42
	TTG	1.46	0.98	1.46	1.29	1.29	0.56	1.69	0.65	0.93	0.98	0.89	1.77	4.03	2.89	1.19
	CTT	0.7	0.4	0.7	0.12	0.43	0.5	0.56	0.38	0.47	0.4	0.65	0.75	0.45	0.58	1.33
	CTC	0.29	0.69	0.29	0.82	0.49	1.01	0.75	0.98	0.93	0.69	0.65	0.41	0.72	0.72	2.16
	CTA	0.64	0.75	0.64	0.71	0.12	0.17	0.94	0.44	0.35	0.75	0.59	1.43	0.18	0.43	0.35
	CTG	2.8	3.06	2.8	2.82	3.61	3.76	1.06	3.49	3.26	3.06	3.09	1.23	0.63	1.23	0.56
Ile (I)	ATT	1.69	1.17	1.69	1.4	1.29	0.91	0.78	0.82	0.94	1.17	0.75	1.29	1.5	1.38	1.43
	ATC	1.23	1.54	1.23	1.26	1.63	2.02	1.26	1.91	1.76	1.54	2.08	1.06	1.43	1.21	1.34
	ATA	0.08	0.29	0.08	0.35	0.09	0.07	0.96	0.27	0.29	0.29	0.17	0.65	0.07	0.4	0.23
Met (M)	AUG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Val (V)	GTT	0.58	0.61	0.58	0.86	0.66	0.44	0.82	0.25	0.53	0.61	0.82	1.33	1.6	1.02	1.78
	GTC	1.04	1.02	1.04	1.29	1.15	1.81	0.82	1.58	1.28	1.02	1.06	0.84	1.33	1.09	0.52
	GTA	0.29	0.34	0.29	0.21	0.05	0.38	0.92	0.25	0.15	0.34	0.33	0.7	0.36	0.51	0.3
	GTG	2.09	2.03	2.09	1.64	2.14	1.38	1.44	1.92	2.04	2.03	1.8	1.12	0.71	1.38	1.41
Ser (S)	TCT	0.26	0.82	0.26	0.76	0.08	0.36	1.13	0.4	0.12	0.82	0.32	0.67	0.47	0.69	0.72
	TCC	2.12	1.18	2.12	1.43	2.3	2	0.81	1.67	1.92	1.18	1.98	1.19	1.88	1.57	0.83
	TCA	0.18	1	0.18	0.29	0.25	0.09	1.13	0.27	0.31	1	0.19	1.33	0.47	0.69	1.24
	TCG	0.79	0.55	0.79	1.05	0.58	1.36	0.75	1.33	1.42	0.55	1.15	0.89	0.75	1.28	0.41
	AGT	0.35	0.55	0.35	0.38	0.41	0	1.13	0.73	0.56	0.55	0.77	0.96	1.03	1.08	1.86
	AGC	2.29	1.91	2.29	2.1	2.38	2.18	1.06	1.6	1.67	1.91	1.6	0.96	1.41	0.69	0.93
Pro (P)	CCT	0.46	0.65	0.46	1.29	0.34	0.44	1.28	0.4	0.4	0.65	0.75	0.9	0.71	0.78	0.78
	CCC	2.29	1.73	2.29	1.53	2.4	2.56	0.59	2.4	1.73	1.73	1.88	1.29	2.14	2.63	1.56
	CCA	0.69	0.65	0.69	0.71	0.57	0.56	1.55	0.67	0.65	0.65	0.75	1.42	0.71	0.49	1.56
	CCG	0.57	0.97	0.57	0.47	0.69	0.44	0.59	1.2	1.2	0.97	0.63	0.39	0.43	0.11	0.11
Thr (T)	ACT	0.76	1.22	0.76	0.48	0.47	0.76	0.87	0.52	0.75	1.22	0.84	1.16	0.85	0.64	0.59
	ACC	2.13	1.44	2.13	1.64	2.37	2	0.41	1.82	1.85	1.44	1.63	0.88	2.31	2.26	0.73
	ACA	0.51	0.61	0.51	0.92	0.37	0.47	1.54	0.57	0.4	0.61	0.44	1.16	0.46	0.64	1.9
	ACG	0.61	0.72	0.61	0.96	0.79	0.76	1.18	1.09	1.01	0.72	1.1	0.79	0.38	0.46	0.78
Ala (A)	GCT	0.82	0.87	0.82	0.84	0.69	0.81	0.25	0.99	0.87	0.87	0.96	1.25	1.22	0.64	1.21
	GCC	2.56	1.98	2.56	1.93	2.57	2.57	0.52	2.85	2.19	1.98	2.13	1.56	2.31	2.91	1.15
	GCA	0.31	0.51	0.31	0.46	0.2	0.44	1.7	0.3	0.44	0.51	0.56	1.04	0.34	0.3	1.42
	GCG	0.31	0.64	0.31	0.55	0.4	0.3	0.96	0.6	0.38	0.64	0.35	0.16	0.14	0.15	0.22
Tyr (Y)	TAT	0.83	1.11	0.83	1.44	0.74	0.43	1.05	0.75	0.62	1.11	0.76	1.43	1.4	1	0.82
	TAC	1.17	0.89	1.17	0.56	1.26	1.57	0.95	1.25	1.38	0.89	1.24	0.57	0.6	1	1.18
STOP	TAA(*)	1	0	1	0	0	0	0.96	0	0	0	0	0	3	0	0
	TAG(*)	0	0	0	0	0	3	0.75	0	0	0	0	0	0	3	0
His (H)	CAT	0.8	0.76	0.8	1.45	0.9	0.53	0.82	0.71	1	0.76	0.59	1.43	0.95	0.81	1.38
	CAC	1.2	1.24	1.2	0.55	1.1	1.47	1.18	1.29	1	1.24	1.41	0.57	1.05	1.19	0.63
Gln (Q)	CAA	0.41	0.34	0.41	0.54	0.34	0.25	1.06	0.45	0.42	0.34	0.47	0.83	1.23	1.1	1
	CAG	1.59	1.66	1.59	1.46	1.66	1.75	0.94	1.55	1.58	1.66	1.53	1.17	0.77	0.9	1
Asn (N)	AAT	0.68	0.97	0.68	1.13	0.72	0.49	1.16	0.7	0.63	0.97	0.76	1.31	1.07	1.2	1.26
	AAC	1.32	1.03	1.32	0.87	1.28	1.51	0.84	1.3	1.37	1.03	1.24	0.69	0.93	0.8	0.74
Lys (K)	AAA	0.4	0.74	0.4	0.39	0.29	0.19	1.2	0.28	0.28	0.74	0.34	1.03	1.1	0.92	0.84
	AAG	1.6	1.26	1.6	1.61	1.71	1.81	0.8	1.72	1.72	1.26	1.66	0.97	0.9	1.08	1.16
Asp (D)	GAT	1.04	1.13	1.04	1.32	0.8	1	1.14	0.98	1.41	1.13	1.3	1.32	1.27	1.07	1.53
	GAC	0.96	0.87	0.96	0.68	1.2	1	0.86	1.02	0.59	0.87	0.7	0.68	0.73	0.93	0.47
Glu (E)	GAA	0.52	0.46	0.52	0.48	0.27	0.34	1.38	0.36	0.67	0.46	0.57	1.25	1.7	1.07	1.26
	GAG	1.48	1.54	1.48	1.52	1.73	1.66	0.62	1.64	1.33	1.54	1.43	0.75	0.3	0.93	0.74
Cys (C)	TGT	0.57	0.29	0.57	0.86	0.67	0.57	0.59	0.67	0.29	0.67	0.75	1.33	0.67	0.33	0.86
	TGC	1.43	1.71	1.43	1.14	1.33	1.43	1.41	2	1.33	1.71	1.33	1.25	0.67	1.33	1

determined by the effective number of codons (ENC). The APN gene of mosquitoes has a codon bias ranging

Amino Acid	**Codons**	D. ananassae	D. mojavensis	D. melanogaster	D. virilis	D. erecta	D. persimilis	D. grimshawi	D. pseudoobscura	D. yakuba	D. sechellia	D. simulans	D. willistoni	Musca domestica	Stomoxys calcitrans	Lutzomyia longipalpis
STOP	TGA(*)	0	0	0	0	0	0	0	1.29	0	0	0	0	0	0	3
Trp (W)	TGG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Arg (R)	CGT	1.64	1.65	1.64	1.62	1.24	1.27	0.48	0.81	1.14	1.65	1.57	1.89	3	2.86	1.29
	CGC	2.91	3.45	2.91	2.43	3	4	0.89	2.11	2.71	3.45	1.86	1.89	1.93	2.43	1.86
	CGA	0.36	0.15	0.36	0.97	0.53	0	0.83	1.14	0.14	0.15	0.57	0.51	0	0.14	0.71
	CGG	0.18	0.6	0.18	0.81	0.18	0.36	0.83	0.16	0.71	0.6	0.71	0.34	0.21	0.14	0.14
	AGA	0	0	0	0	0	0.18	1.25	0.49	0.43	0	0.57	0.86	0.64	0.14	1
	AGG	0.91	0.15	0.91	0.16	1.06	0.18	1.72	1.3	0.86	0.15	0.71	0.51	0.21	0.29	1
Gly (G)	GGT	1.39	1.24	1.39	1.86	0.83	0.8	0.97	1.02	1.16	1.24	1.02	1.71	2.05	1.73	1.08
	GGC	0.98	2.04	0.98	1.67	1.58	2.31	1.21	1.79	1.96	2.04	1.77	0.86	1.73	1.65	0.77
	GGA	1.47	0.53	1.47	0.19	1.51	0.71	1.09	0.68	0.89	0.53	1.21	1.43	0.22	0.47	1.69
	GGG	0.16	0.18	0.16	0.28	0.08	0.18	0.73	0.51	0	0.18	0	0	0	0.16	0.46

Table 12. Effective number of synonymous codons and GC content of APN gene in mosquitoes (Diptera)

Species	CAI	ENC	GC3s	GC %
Culex quinquefasciatus	0.749	52.1	58.1	50.4
Aedes ageypti	0.771	54.5	53.6	48.5
Anopheles darling	0.783	45.4	69.9	56.6
Anopheles gambiae	0.527	56.2	69.4	47.8
Anopheles arabiensis	0.734	47.9	68.8	56
Anopheles merus	0.661	48.2	68.1	55.8
Anopheles culicifacies	0.735	46.2	70	56.6
Anopheles albimanus	0.737	40.4	77.9	59.7
Aedes albopictus	0.705	55.4	59.5	51.4

from ENC from 40.4 (An. albimanus) to 56.2 (An. gambiae), the larger the extent of codon preference in a gene shows synonymous codon usage, the smaller the corresponding ENC value. It was found that there was a positive correlation between GC content and degree of synonymous codon usage bias measured by ENC. The highest synonymous codon usage bias of An. albimanus APN with an ENC of 40.4 indicates the biased GC3s (77.9). This gene shows increase in CAI value with decrease in ENC value in Diptera. The average CAI value was found as 0.7, a level close to 1.0, which might give a high expression level of genes according to the index definition. Many highly expressed genes show high synonymous codon usage bias for fast and accurate translation. In conclusion, codon bias is most prominent in highly expressed genes for translational competency and accuracy [29-30]. The results revealed the similar pattern was observed from other highly expressed genes [31]. So, higher codon biasness is found in APN1 gene among various mosquitoes and it is clear that APN1 is one of the highly expressed gene plays important in blood digestion and parasite invasion processes. APN amino acid sequence analysis shows the presence of high amino acid conservation among specific areas in all insects, all aligned confidently along their entire lengths. These include M1 Peptidase and ERAP1-like C- terminal domains. The zinc ions play an important role in the catalysis mechanism of metalloproteases like aminopeptidase. Multiple alignment results of different APN sequences showed the highest similarity between the Dipteran APN sequences [32]. In this study multiple alignments showed three major conserved sites in almost all the aligned sequences Where HEXXH motif is also conserved, which is responsible for peptidase activity [33-34]. The zinc metalloproteases like Aminopeptidase N among various phyla were shown to bear a signature sequence motif HEXXH where they bind the zinc atom in tetrahedral fashion. The highly conserved HEYAH amino acid residues are found in all insects [35]. Here the histidine residue acts as ligands of zinc ion where GLU plays a catalytic role. The GAMEN domain signature was also found in all insects and

Table 13. Effective number of synonymous codons and GC content of APN in non-mosquitoes (Diptera)

Species	CAI	ENC	GC3s	GC %
Drosophila mojavensis	0.714	45.8	65.5	53.5
Drosophila ananassae	0.801	40.4	74.8	55.9
Drosophila virilis	0.789	48.0	61.5	51.5
Drosophila erecta	0.790	40.8	74.8	56.0
Drosophila sechellia	0.796	42.1	70.8	54.2
Drosophila melanogaster	0.782	43.7	68.8	53.5
Drosophila persimilis	0.768	41.5	75.3	56.1
Drosophila grimshawi	0.661	55.2	52.2	47.4
Drosophila pseudoobscura	0.718	42.0	75.2	55.9
Drosophila yakuba	0.739	45.1	70.9	54.0
Drosophila simulans	0.748	48.0	67.1	52.9
Drosophila willistoni	0.808	55.2	45.1	44.7
Musca domestica	0.815	42.9	54.6	45.8
Lutzomyia longipalpis	0.777	52.7	46.7	49.5
Stomoxys calcitrans	0.748	45.7	59.4	45.3

showed homology to each other. Many workers have determined hydrolysis and catalytic activities of APNs. Above said site GAMEN found as Bastatin binding site in APN of E. coli [36]. Third motif WWDNL found as a peptide binding site in human APN and it was found that this motif is also important in tumor angiogenesis and binding with peptides [37]. Also, these conserved sites are reported for Cry toxin protein binding region on APN [38]. The entire above said three motifs are found to be conserved in APN of M. sexta and H. armigera the similar sites were predicted by Conseq and Consurf servers [39]. APN protein of all the insects encoding amino acids sequence within 900-1000 amino acids and weights of APN protein was found in between 90 to 100 kDa. Another important finding is that APN follow secretary pathway, as amino terminal is encoded by signal sequences. Moreover, in the present investigation, GPI prediction program exhibited the presence of GPI anchor signal sequence at the C-terminus. Such sequence was found to be more conserved in dipterans and lepidopterans. Besides, the amino acid sequence also contained four Cysteine residues. Cys residues are important in the structural integrity of proteins by forming disulphide bridges, which in turn helps to keep the molecules intact and to maintain the conformation of the active sites. These residues help in enhancing the stability of the proteins [40]. These residues are highly conserved among APN molecules of higher mammals (rat, rabbit, pig and human), was detected [41].The post translational modification i.e. phosphorylation was found at serine, threonine and tyrosine residues conferring protection from proteolytic degradation. Protein-protein interaction results show that APN protein connected to glutathione synthase may involve in metabolic pathways for metabolism of Glutathione. Glutathione makes mosquito insecticide resistance, so disrupting of APN protein or any other protein, this protein network may block the whole metabolic pathway of Glutathione [42]. Amino acid sequence analysis of the Dipterian APN gene shows that they all can be aligned confidently along their entire lengths. Specific areas of relatively high amino acid conservation were observed. The ORF's range among all APN was found from 2 to 3.5 kb, encoding a 900-1000 amino acids sequence. All APN protein molecular weights were found to be above 90 kDa. Three important domains namely Aminopeptidase N like, M1 Peptidase and ERAP1-like C- terminal domain were identified in all APNs. All APNs are found to have four conserved Cysteine residues. Signal peptide sequence of 15-30 amino acids was present at N terminal end and GPI anchored region (hydrophobic amino acid residues) at C terminal end of APN protein. The APN analysis revealed the three conserved regions, gluzicin APN motif GAMEN, second one is Zinc- binding/ gluzicin motif HEXXH, which is a part of a typical catalytic site for the majority zinc dependent metallopeptidase and is required for enzymatic function, the third one is zinc binding ligand which is conserved in the sequence motif NEXFA. Highly conserved region is present in domain I, about 64 amino acid residues believed to be important for Cry 1 toxin binding. The Phylogram was constructed using the mosquitoes APN sequences and similar sequences of non-mosquitoes origin retrieved from the NCBI database. Undoubtedly, the ancestry species of the gene should be more closely related to sequences from a similar genus. In the present investigations the An. gambiae and An. stephensi seems to be the closest relatives of the An. culicifacies [43]. Out of the non-mosquitoes species, with no surprise are closely related to each other as they are the latest to diversify on the path of the evolutionary journey. Phylogenetic tree

```
                                    10        20        30        40        50        60        70        80        90       100
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Anopheles_gambiae                   -------MVRLLLLVSLGLCCLLASADRPPYKSSGIYEEEPQLPGASDLSGGVASDEIGVAPAQAVDERYRLPTTSIPIHYDLHLRTEIHRNERTFTGTV
Anopheles_arabiensis                -------MVRSLLLVSLGLCCLLASADRPPYKSSGIYEEEPQLPGASDLSGGVASDEIGVAPAQAVDERYRLPTTSIPIHYDLHLRTEIHRNERTFTGTV
Anopheles_merus                     -------MVRSLLLVSLGLCCLLASADRPPYKSSGIYEEEPQLPGASDLSGGVASDEIGVVPAQPVDERYRLPKTSIPIHYDLHLRTEIHRNERTFTGTV
Anopheles_quadriannulatus           -------MVRSLLLVSLGLCCLLASADRPPYKSSGIYEEEPQLPGASDLSGGVASDEIGVAPAQAVDERYRLPTTSIPIHYDLHLRTEIHRNERTFTGTV
Anopheles_maculatus                 ------------MYLFEDNRFARAERPPFKSSSGLYEDEPQLAASEQEGDGT---FVGIVPAQPVDEAYRLPKTSIPIHYDLHLTTEIHRNERQFSGTV
Anopheles_stephensai                ------------------------------------------------------------------LPKTSVPIHYDLHLTTEIHRQLVLHTRGL
Anopheles_culicifacies              MVRFWILVSLGLCCLVAPNDRFVRAEVRPPFKTGGIYEDEPQLLLDESERSDASVEQGAVVPAQAVDERYRLPKTSVPIHYDLHLRTEIHRNERTFTGSV
Anopheles_farauti                   -MVRSLILATLGLWCLLGPANVFTRAERPPFKASGLYDQEPELAVTDP------DTVAVAVPVQAVDERYRLPKTSVPLHYDLHLRTEIHRNERTFQGSV
Anopheles_atroparvus                ------MVRALIVVSIGVCCLLLAGAERPPFKSRGIHGED------ELAEAGAASEVTGVAPSQAVDETYRLPKTSVPVHYDLHLRTEIHINERTFDGSV
Anopheles_sinensis                  ------MVRALSVVSIGLCWLLLAGAERPPFRNVGINGDDE--LLGRPVEAEAASEVTGVAPVQAVDETYRLPKTSVPLHYDLHLRTEIHRNERTFTGSV
Anopheles_darlingi                  ---MVRPLVIVSIGLWCLLGAGELLAERPPFKTDAIHAEEEE---------PVPSAEQAIVPAQDVDESYRLPQTSIPIHYAVHLTTNIHRNERTFTGTV
Aedes_albopictus                    ---MVRPLVIVSIGLWCLLGAGELLAERPPYKTDTVHAEEE----------PVVSAEQAFVPAQDVDESYRLPKASVPIHYDVHLTTNIHRNDRTFTGTV
Aedes_ageypti                       ---------MIVLKWSVVLWGLTLATYG-TVAERPLSDVSDNVDPPQIAPAFAVE-ESEIIPLQEVDESYRLPKTSYPTHYELRLRTEVHTGNRQFDGTV
Anopheles_albimanus                 ---------MIALKWSVVFLGLALTAFGTAAAERPLSGEFNVDDPPQIAPAFAAE-ESEIVPLQAVDETYRLPKTSYPTHYDIKLHTEVHTGERQFQGTV
Culex_quanquefasciatus              -----------MIAPKSILLCVLAVAGAALAADRPLSKSADVDPPLLIAPEIEAFGDDTAVPFQEIDEAYRLPTTITPTRYDVLLRTEVHTGDRAFEGTV
Clustal Consensus                                                                                                ** :  * :* : * *::*     .   :

                                   110       120       130       140       150       160       170       180       190       200
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Anopheles_gambiae                   GIQLQVVQATDKLVMHNRGLVMSSAKVSSLPNGVTGAPTLIGDVQYSTDTTFEHITFTSPTILQPGT-YLLEVAFQGRLATNDDGFYVSSYVADNGERRY
Anopheles_arabiensis                GIQLQVVQATDKLVMHNRGLVMSSAKVSSLPNGVTGAPTLIGDVQYSTDTTFEHITFTSPTILQPGT-YLLEVAFQGRLATNDDGFYVSSYVADNGERRY
Anopheles_merus                     GIQLQVVQATDKLVMHNRGLVMSSAKVSSLPNGVTGAPTLIGDVQYSTDTTFEHITFTSPTILQPGT-YLLEVAFQGRLATNDDGFYVSSYVADNGERRY
Anopheles_quadriannulatus           GIQLQVVQATDKLVMHNRGLVMSSAKVSSLPNGVTGAPTLIGDVQYSTDTTFEHITFTSPTILQPGT-YLLEVAFQGRLATNDDGFYVSSYVADNGERRY
Anopheles_maculatus                 AIQLQVLQATDQLVLHNRRLEIASAKISSLPNGVNGAPTLIGDATYSTNTTIEHITFTCPGILQPGF-YWLEVTFEGILATNDDGFYVSSYVADSGERRY
Anopheles_stephensai                VISSARISSLPNERQFSGSVAIQLQVLETTDNGVTGAPVLIGDATYSTNTTIEHITFTSPNILQPGF-FWLEVSFVGQLATNDDGFYVSSYVADSGERRY
Anopheles_culicifacies              AIQLQVLEATSTLVLHNRALSIASAKLSSLPNGVTGTPTPIGDATFSTDTTFEHITFTSPMLLQPGY-YLLEVSFEGRLATNDDGFYVSSYVADSGERRY
Anopheles_farauti                   GIQLQVLEVTDRLVLHNRNLVITSAKLSALPDGVTGAPQLIGDVTYATDATFEHITFTSPSVLQPGF-YLLEVVYEGRLATNDDGFYVSSYVNDRDERVY
Anopheles_atroparvus                GITLQVLEATNQLVLHNRALSISTAKISSLPNGVDGAPTLIGDVTNSYDTRVEHLTFTSPVVLQPGY-YLLEVTFQGRLTTNDDGFYVSSYVNDLGERRY
Anopheles_sinensis                  AIKLQVLETTDKLVLHNRGLSILTAKVSSLPNGVEGAPTLIGDATFSLDTRVEHLTFTSPTILQTGF-YLLEVTFEGRLTTNDDGFYVSSYVTDAGERRY
Anopheles_darlingi                  AIKLQVVEPTDNLVLHQLGLQLSTAKISSLPAGIDGPPTLIGDVTFATDTRVAHLTLTSPSVLLAGY-YLLEIDYSGRLSTDDDGFYVSSYVADNGERRY
Aedes_albopictus                    AIKLQVIEPTDKLVLHQRGLQLSAARISSLLAGIDGPATPIGDVTFATETQRAHLTLTAAMVLQPGY-YQVEIDYSGRLSTDDDGFYVSSYVADNGERRY
Aedes_ageypti                       AIHLNVVEATNAIVVHYRSLTIQNAKLAFIPT-PEADPQQLNDPTWTYDAKVEQLSFNSETLLNPGS-YILTVEYNGRLSNSEDGFYISSYVNKDGVTKY
Anopheles_albimanus                 AIDLNVVEATNSIVVHHRSLTIQSAKLVLVPT-AGE-PQELNDPTWTYDTRVEQLSFNSVNLLSPGS-YILTVTYAGRLSNNQDGFYMSSYVDSQNRTVY
Culex_quanquefasciatus              DIYFTVLEATNEIVIHQRDLAIQSYVLYSVG----ATDVELAAPAMTYNALYNKVTFTSGTLLTAGENYKLVIRFNGKLANDEDGFYISSYVDDNGVTKY
Clustal Consensus                    *    :.       .   : :    :             :     : ::   ::::..  :* .*  : : : : * *:..:****:**** . .   *

                                   210       220       230       240       250       260       270       280       290       300
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Anopheles_gambiae                   LATTQFESTSARMAFPCYDEPGLKATFTVSITHSLSYKAISNMPQKTTTDIETDMRTTFFEKTPAMSTYLLAFVVSDFQLRLSGAQRVYVRPNAFNEATF
Anopheles_arabiensis                LATTQFESTSARMAFPCYDEPGLKATFTVSITHSLNYKAISNMPQKTTTDIETDMRTTYFEKTPAMSTYLLAFVVSDFQFRVSGPQRVYVRPNAFNEAAF
Anopheles_merus                     LATTQFESTSARMAFPCYDEPGLKATFTVSITHSRSYKAISNMPQKTTTDIETDMRTTDFEKTPAMSTYLLAFVVSDFLFRVSGTQRVYVRPNAFNEAAF
Anopheles_quadriannulatus           LATTQFESTSARMAFPCYDEPGLKATFTVSITHSRSYKAISNMPQKTTTDIETDMRTTDFEKTPAMSTYLLAFVVSDFLFRVSGTQRVYVRPNAFNEATF
Anopheles_maculatus                 LATTQFESTSARMAFPCYDEPALKATFTVSITHSLTYTAISNMPENRRSDADG-MRTTVFEKTPIMSSYLLAFVVSDFQYRIATPQRVYVRPNAFQQASF
Anopheles_stephensai                LATTQFESTSARMAFPCYDEPALKATFTVSLTHSITYSAISNMRENRRSDADG-MRTTVFEIDGTQRVYVRPNAFQQATFAKTPIMSSYLLAFVVSDFLY
Anopheles_culicifacies              LATTQFESTSARMAFPCYDEPALKATFTVSITHSATYSAISNMPQQETFNEDA-MRTTVFQRTPLMSTYLLAFVVSDFQYRINGPQRVYVRPNAIQEATF
Anopheles_farauti                   LATTQFESTSARMAFPCYDEPALKATFTVSITHINAYSAISNMPVASTVDLADSMRTTAFQRTPLMSTYLLAFVVSDFQFRISGTQRVFVRPNAFNEASF
Anopheles_atroparvus                LATTQFESTSARMAFPCYDEPALKATFTVSITHHSSYFAISNMPVTGIIIVDGDMRTTRFNQSPAMSTYLLAFVVSDFLFRTSGTQRVYVRPNAFNEAGF
Anopheles_sinensis                  LATTQFESTSARMAFPCYDEPALKATFTVSITHSRSYFAISNMPVNTLTDVEPDMRTTLFNPTPAMSSYLLAFVVSDFLFRQSGTQRVYARPNAFDEIPF
Anopheles_darlingi                  LATTQFESTSARRAFPCYDEPALKATFTITITHGSLYGAISNMPEDTR-TVNGEMATTKFETTPLMSTYLIAFVVSDFLRRGDNQHRVFARPNAFQELPF
Aedes_albopictus                    LATTQFESTSARRAFPCYDEPALKATFTITITHGALYGAISNTEQVSR-IVSGESATTKFKTTPLMSSYLIAFVVSDFLNRGSDKHRVFARPNAFQELPF
Aedes_ageypti                       LATTQFESTSARMAFPCYDEPGLKATFALWITHDVLYTANSNMPYTSTIDG--DIRVTQFEVTPKMSTYLLAFVVSDFQRLGTLEHSVYARPNAIDEVVF
Anopheles_albimanus                 LATTQFESTSARMAFPCYDEPALKATFALQITHSRFYTANSNMEYDFDIDG--DYRISKFKVTPIMSTYLLAFVVSDFAYKGPPTHRVHARVNAINEVDF
Culex_quanquefasciatus              LATTQFESTSAREAFPCFDEPNLKARFTIKILHHGSYSARSNMGVAKEEDIGNGYKITTFGITPLMSTYLLAFVVSDFESIGTDTQRVYVRSNALDEATF
Clustal Consensus                   ************ ****:*** *** *:: : *   * * **              : *        *: . ...:          .    ...:  :

                                   310       320       330       340       350       360       370       380       390       400
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Anopheles_gambiae                   ALEAGVKILKVLDDHLGIPYDTYMPKLDQIAIPDFAAGAMENWGLVTYR-EQALLFNPAVSTYRGKTNVATTIAHEYAHQWFGNLVSPEWWEYIWLNEGF
Anopheles_arabiensis                ALEAGVKILKVLDDHLGIPYDTYMPKLDQIAIPDFAAGAMENWGLVTYR-EQALLFNPAVSTYRGKTNVATTIAHEYAHQWFGNLVSPEWWEYIWLNEGF
Anopheles_merus                     ALEAGVKILKVLDDHLGIPYDTYMPKLDQIAIPDFAAGAMENWGLVTYR-EQALLFNPTVSTYRGKTNVATTIAHEYAHQWFGNLVSPEWWEYIWLNEGF
Anopheles_quadriannulatus           ALEAGVKILKVLDDHLGIPYDTYMPKLDQIAIPDFAAGAMENWGLVTYR-EQALLFNPAVSTYRGKTNVATTIAHEYAHQWFGNLVSPEWWEYIWLNEGF
Anopheles_maculatus                 ALEAGVKILEVLDEHLGIPYSTYMPQLYQIAVPDFAAGAMENWGLVTYR-EQALLFDPAVSTYRTKTNIATTIAHEYAHQWFGDLVTPEWWEYIWLNEGF
Anopheles_stephensai                RLEAGVKILKVLDEHLGIPYSTYMPTLNQIAVPDFAAGAMENWGLVTYKREQLLLFDPAVSTYRTKTNIATTIAHEYAHQWFGDLVTPEWWEYIWLNEGF
Anopheles_culicifacies              ALEAGVKILKALDDHLDIPYSAYMPKLFQIAVPDFAAGAMENWGLVTYR-EQLLLFNSAVSTYRTKTNIVTTVAHEYAHQWFGDLVTPEWWEYIWLNEGF
Anopheles_farauti                   ALEAGVKILAALDEHLGIAYGAHMPKLDQIAVPDFAAGAMENWGLVTYR-EELLLFKPAANTYRTKTNIATTVAHEYAHQWFGDLVTPEWWEYIWLNEGF
Anopheles_atroparvus                ALEAGVKILKVLDDYLGIDYAAQMPKLDQIAVPDFAAGAMENWGLVTYR-EQLLLFNPDVSTYRTKTNIATTIAHEYAHQWFGDLVTPEWWEYIWLNEGF
Anopheles_sinensis                  ALEAGVKILKALDDYLDIPYSSVMPKLDQIAVPDFAAGAMENWGLVTYR-EQALLFNPTVSTYRTKVNVATTIAHELAHQWFGDLVSPEWWEYIWLNEGF
Anopheles_darlingi                  ALEASRLILQVLDDHLGVPYTNYMEKIDQVALPDFAAGAMENWGLVTYR-EQLLLFNPEVNSYRTKTNIATTISHEYAHQWFGDLVTPQWWEYIWLNEGF
Aedes_albopictus                    ALEASKLILPVLDEHLYVPYMSYMEKIDQVALPDFAAGAMENWGLVTYR-EQLLLFNPEVNTYRTKTNIATTIAHEYAHQWFGDLVTPQWWEYIWLNEGF
Aedes_ageypti                       AVEAGQKILEKLDAHLGIGYYDHMPQMKQFAIPDFAAGAMENWGLVTYR-EQYLLFNPELSTYRTKTNIATVIAHEYAHQWFGNLVSPEWWEYIWLNEGF
Anopheles_albimanus                 AVEAGQIILEKLDTHLGIGYYDHMPTMKQFAIPDFAAGAMENWGLVTYR-EQYLLFNQELSTYRTKTNIATVIAHEYAHQWFGNLVSPEWWEYIWLNEGF
Culex_quanquefasciatus              AQEAGEKILAALDSHLKMKYSEQMPEMKQFAIPDFAAGAMENWGLVTYR-EQYLLFNPALSTYRTKTNIATIIAHEYAHQWFGNLVSPEWWEYIWLNEGF
Clustal Consensus                     **.  **  ** :* : *   *  : *.*:****************: *: ***.   .:** *.*:.* ::** ******:**:*:***********
```

based on APN protein of non-mosquitoes were analysed separately to know the evolutionary relationship of organisms. Phylogenetic analysis of APN indicates that all are evolved from common ancestors. Phyloge-

```
                                  410       420       430       440       450       460       470       480       490       500
                                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Anopheles_gambiae                 ATLYEFYALDMAYPGQEYWELFNQQVIQYAMGQDGQASTRPMNWNAATPGEISALFDRVAYDKSGSVLNMMRHVLGDDNWKAGLKAYLTDRALQGAVDEQ
Anopheles_arabiensis              ATLYEFYALDMAYPGQEYWELFNQQVIQYAMGQDGQASTRPMNWNAATPGEISALFDRVAYDKSGSVLNMMRHVLGDDNWKAGLKAYLTDRALQGAVDEQ
Anopheles_merus                   ATLYEFYALDMAYPGQEYWELFNQQVIQYAMGQDGQASTRPMNWNAATPGEISALFDRVAYDKSGSVLNMMRHVLGDDNWKAGLKAYLSDRALQGAVDEQ
Anopheles_quadriannulatus         ATLYEFYALDMAYPGQEYWELFNQQVIQYAMGQDGQASTRPMNWNAATPGEISALFDRVAYDKSGSVLNMMRHVLGDDNWKAGLKAYLTDRALQGAVDEQ
Anopheles_maculatus               ATLYEFYAAHLAYPEAEYWELFNPQVIQAAMVPDGLATTRPMNWNAATPGEISALFDRVAYPKSGSVLNMMRHVLGEENWKAGLNAYLTDRALSVAVDEQ
Anopheles_stephensai              ATLYEFYAAHLAYPEAEYWELFNTQVIQAAMVPDGLASTRPMNWNAATPGEISALFDRVAYG---SVLNMMRHVLGEENWKAGLNTYLNDRALSVAVDEQ
Anopheles_culicifacies            RTLYEFYAAHLAYPEMEYWELFNTQVIQAAMVPDALATTRPMNWNAATPGEISALFDRVAYPKSGSVLNMMRHVLGEENWRSGLKVYLEDRALATAVDEQ
Anopheles_farauti                 ATLYEFYATHLAYPEQEYWELFNLQVIHAAMVPDALVSTRPMNWNAATPGEISALFDRVAYPKSGSVLNMMRHVLGDDNWRAGLRTYLTDRTYKSANDEQ
Anopheles_atroparvus              ATLYEYYATHLAYPEHEYWELFNTQVIQAAMVPDGLASTRPMNWNAATPGEISALFDRVAYPKSGSVLNMMRHVLGDDNWRAGLKIYLEARSLKEANDEQ
Anopheles_sinensis                ATLYEYYATDLAYPGMEYWELFNTQVIQAAMVPDGQASTRPMNWNAATPGEIAALFDRVAYPKSGSVLNMMRHVLGDDNWRAGLKIYLEDRALKEANDEQ
Anopheles_darlingi                ATLYEYLATDLAYPGMEYWELFNTQVIQAAMVPDGLASTRPMNYNAATPGEISALFDRVAYPKSGSVLNMMRHVLGDDNWRAGLRTYLLDRALNKATDED
Aedes_albopictus                  ATLYEYLATDMAYPGMEYWELFNTQVIQAAMVPDGLASTRPMNYNAATPAEISALFDRVAYPKSGSVLNMMRHVLGDDNWREGLRIYLLDRALKGATDED
Aedes_ageypti                     ATLYEYYATHLAYPEVGYWELFNTQVIQAAMVPDGLETTRPMTWNADTPRSIASLFDRVAYPKSGSVLNMMRNVLGEINWTAGLKAYLTARQFDGANADH
Anopheles_albimanus               ATLYEYYATHMAYPEVGYWELFNTQVIQAAMVPDGLDSTRPMTWNADTPASIGSLFDRVAYPKSGSVLNMMRNVLGEDNWTAGLKAYLTARQFDTANSDH
Culex_quanquefasciatus            ATLYEYLATHEAYPNMEYWELFNLQVIQAAMVPDGLESTRPMTWNADSPQSIARLFDRVAYPKSGSVLNMMRNIVGDENWATGLRLYLQARQYKGANVEH
Clustal Consensus                  ****: * . ***   ****** ***: **  *.  :****.:** :* .*. *******    *******:::*: **  **. **  *    *  :.

                                  510       520       530       540       550       560       570       580       590       600
                                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Anopheles_gambiae                 LYAGLQSAIEGKGVLPNGVTVAQIMRTWTNEAGYPVLNVRRSYDTGDVIISQERFYNDRKVPNTNIWMIPYNYVHQAKADFNEFDDFQWLATKAARIETT
Anopheles_arabiensis              LYAGLQSAIEGKGVLPNGVTVAQIMRTWTNEAGYPVLNVRRSYDTGDVIISQERFYNDRKVPNTNIWMIPYNYVHQAKADFNEFDDFQWLATKAARIETT
Anopheles_merus                   LYAGLQSAIEGKGVLPNGVTVGQIMRTWTNEAGYPVLNVRRSYDTGDVIISQERFYNDRKVPNTNIWMIPYNYVHQAKADFNEFDDFQWLATKAARIETT
Anopheles_quadriannulatus         LYAGLQSAIEGKGVLPNGVTVSQIMRTWTNEAGYPVLNVRRSYDTGDVIISQERFYNDRKVPNTNIWMIPYNYVHQAKADFNEFDDFQWLATKAARIETT
Anopheles_maculatus               LYEGLQSAIEGKGVLPNGVTVAQIMRTWTNEAGYPVLTVRRSYDTGDVIISQERFYSDRKVPNTNIWMIPYNYVQQARADFNEFDDFEWLATKAARIETT
Anopheles_stephensai              LYTALQSAIEGKDVLPNGVTVAQIMRTWANEAGYPVLSVRRSYDTGDVIISQERFYSDRKVPNTNVWMIPYNYVQQARADFNEFDDFQWLATKAARIATT
Anopheles_culicifacies            LYTSLQSVIEGKNVLPGDVTVAQIMRTWANEAGYPVLNVRRSYDTGDVIISQERFYSDRKVPNTNVWMIPYNFVQQARADFNEFDDFQWLATKAARISTT
Anopheles_farauti                 LYQGLQSAIEGKGVLPGGVTVAQIMRTWTTEAGYPVLNVRRSYDTGDVIISQERFYSDRKIPNANVWMIPYNFVQRARADFNEFDDFQWLATKAARIETG
Anopheles_atroparvus              LYAGLQAAIEGKGVLPAGVTVAQIMRTWTTEAGYPVLNVRRSYDTGDVIISQERFYSDRKVPNSNIWMIPYNMVQQSRADFNEFDDFAWLATKAARLQTS
Anopheles_sinensis                LYRGLQSAIEGKNVLPAGVTVAQIMRTWTTEAGYPVLNVRRSYDTGDVIISQERFYSDRKVPNENVWMIPYNMVNQARADFNEFDDFEWLASKAARIQTT
Anopheles_darlingi                LYRGLQSAIEGKGVLPDGVTVGQIMRTWTNEAGYPVLNVRRSYDTGEIIISQERFYSDRKIPNTNIWMIPYNYVHQAVADFNEFDDFKWLATKADRFTTT
Aedes_albopictus                  LYNALQQAIEGKGVLPAGVTVGQIMRTWTNEAGYPVLNVRRSYDTGEIIISQERFYSDRKIPNANIWMIPYNYVHQAVADFNEFDDFKWLATKADRFMTT
Aedes_ageypti                     LYVGLQSAIQGNNVLPEGVTVKDIMDTWANEKGYPVLSVRRTYETGDIIISQERFISDRKVPNTNVWMIPYNYVHQSKADFDDLSTFSWLSTKAARINTE
Anopheles_albimanus               LYTGLQSAIQGKNVLPDGVTVKDIMESWATEKGYPVLSVRRTYETGSVIISQERFISDRKVPNTNIWMIPYNFVRQSRADFDDLSTFSWLSTKAARISTD
Culex_quanquefasciatus            LHEGLQAAIEGKNILPDGMTIKQVMDSWTTEKGYPVLNVRRSYDTGDVIISQERFISDRKVPNTNVWMLPYNYVNQSVADFHDTNTYSWLTTKAARISTD
Clustal Consensus                 *: .** .*:*:.:** .:*: ::* :*:.* *****.***:*:**.:******* .***:** *:**:*** *.:: ***.: . : **::** *: *

                                  610       620       630       640       650       660       670       680       690       700
                                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Anopheles_gambiae                 VPANEWIVFNKQQVGYYRVNYDEHNWELITNALHENWASIHRLNRAQLIDDAYWLARSGRLDLRVALRFMTYLRNEREYAPWTAANVALTYFNNRLRGTA
Anopheles_arabiensis              VPANEWIVFNKQQVGYYRVNYDERNWELITNALHENWASIHRLNRAQLIDDAYWLARSGRLDLRVALRFMTYLRNEREYAPWTAANVALTYFNNRLRGTA
Anopheles_merus                   VPANEWIVFNKQQVGYYRVNYDERNWELITNALHENWASIHRLNRAQLIDDAYWLARSGRLDLRVALRFMTYLRNEREYAPWTAANVALTYFNNRLRGTA
Anopheles_quadriannulatus         VPANEWIVFNKQQVGYYRVNYDERNWELITNALHENWASIHRLNRAQLIDDAYWLARSGRLDLRVALRFMTYLRNEREYAPWTAANVALTYFNNRLRGTA
Anopheles_maculatus               VPATEWIIFNKQQVGYYRVNYDDHNWELITNALIDNFAAVHRLNRAQLIDDAYWLARSGRLDLRVALRLMTYLRGEREYAPWTAANVALSYFNSRLRGTE
Anopheles_stephensai              VPATEWLIFNKQQVGYYRVNYDDRNWELITNALIDNLAAVHRLNRAQLIDDAYWLARSGRLDLRVALRLMTYLRGEREYAPWAAANTALSYFNSRLRGTE
Anopheles_culicifacies            VPANEWIIFNKQQVGYCCVPYAVRNWKLITNAWWENCAGCIRLNRALLIDDANWLARSGRLDMRVALNLMTYLRGEREYAPWTAASVVLSYFNSRLRGTA
Anopheles_farauti                 VPADEWIIFNKQHVGYYRVNYDDHNWELITQALFDNYASIHRLNRAQLIDDAYWLARSGRLDMRVTLRLMTYLRNEREYAPWTAANVVLTYFNSRLRGTA
Anopheles_atroparvus              VPATEWIVCNKQQVGYYRVNYDDRNWELITDALIANWASIHRLNRAQLIDDAYWFARSGRLDMRITLRLMTYLRNEEEYAPWTAANVALTYFNNRLRGTA
Anopheles_sinensis                VPENEWIIFNKQQVGYYRVNYDERNWELITNALIANWASVHRLNRAQLIDDAYWLARSGRLDMRITLRLLTYLRHEEEYAPWTAANVVLSYFNSRLRGTA
Anopheles_darlingi                VPAHEWIIFNKQEVGYYRVNYDDHNWELITNALHECFGCIHRLNRAQLIDDAYWLARSGRLDLRIALRLMTYLHNEREYSPWAAADVALSYFNSRLRGTE
Aedes_albopictus                  VPAHEWIIFNKQEVGYYRVNYDDHNWELITNALHASYGSVHRLNRAQLIDDAYWLARSGRLDLRIALRLMAYLQHEREYSPWASADVALSYFNSRLRGTE
Aedes_ageypti                     VPANEWIIFNKQQVGYYRVNYDANNWELITNALINNLSSIDRLNRAQLIDDAYWLARSGRLDIEVLMKLLTYLKNETEYAPWTAANNVLSYFNGKLRGTP
Anopheles_albimanus               VSANEWIIFNKQQVGYYRVNYDDVNWELITNALINNVDSIHRLNRAQLIDDAYWLARSGRLDLEILMKLLTYLPNETEYAPWTAANNVLSYFNGKLRGTP
Culex_quanquefasciatus            VPDHLWIIFNIQQVGFYRVNYDERNWELITNSLITNWASVNRLNRAQLIDDAFWLARSERLDLEIFMKLLTYLKDEREYAPWTAASNALSYFNGKLRGTN
Clustal Consensus                 *.   *:: * *.**:  * *   **:***::         ***** ***** *:*** ***:.: :.:::**  * **:**::*. .*:***.:****

                                  710       720       730       740       750       760       770       780       790       800
                                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Anopheles_gambiae                 EYHNFLIFVDALIEDIYSLLTIDAVSPDDTLLHKYLVQTISTWACSMGYTDCLMKTAALLKAEASGTGP-AVHPDIASVTYCYGMRSALESEFQYLYRKM
Anopheles_arabiensis              EYHNFLIFVDALIEDIYSLLTIDAVSPDDTLLHKYLVQTISTWACSMGYTDCLMKTAALLKAEASGTGP-AVHPDIASVTYCYGMRSALESEFQYLYRKM
Anopheles_merus                   EYHNFLIFVDALIEDIYSLLTIDAVSPDDTLLHKYLVQTISTWACSMGYTDCLMKTAALLKAEASGTGP-AVHPDIASVTYCYGMRSAAESEFQYLYRKM
Anopheles_quadriannulatus         EYHNFLVFVDALIEDIYSLLTIDAVSPDDTLLHKYLVQTISTWACSMGYTDCLMKTAALLKAEASGTGP-AVHPDIASVTYCYGMRSALESEFQYLYRKM
Anopheles_maculatus               NYHDFLVFASALIENIYELLTIDVVSPNDTLLHKYLVQTISTWACSLGYKDCLERTGALLRAEAAGTGP-AVHPDIATVTYCYGMHTTGETEYQYLYRKM
Anopheles_stephensai              NYHDFILFVDALIEELYGLLTIDVVSPNDSLLHKYLVQTITSWACSLGYKDCLERTGALLRAGASGTGP-AVHPDIATVTYCYGMRTTGAAEYQYLYRKM
Anopheles_culicifacies            AYHDFLLFVDALIENIYGMLTIDVVSPNDTLLHKYLVQTISTWACTIGYKDCAEKTAALLKAEVASAGQTPVHPDIASVTYCYGMRSANDAEYQYFYRKM
Anopheles_farauti                 AYHDFLVFVDALIENIYGLLAIDAVAEEDTLLHKYLVQTISTWACASGYTDCLTKTRALLTAEATSTTGFTVHPDIASVTYCYGMRSVGETEFQYLYRKM
Anopheles_atroparvus              QYHDFLQFVDALIENIYGKLTIDAVADDETLLHKYLKQTISTWACMIGYTDCLTRTAALLSREALGTGP-AVHPDIATVTYCYGMRNIGEPEFQYIYRKM
Anopheles_sinensis                QYHDFLQFVDALIENVYSKLKIDEVQENETLLHKYLKQTISTWACTIGYTDCLTRTGALLRGEALNTGP-AVPADIATVTYCYGMRNADEAVYQYIYGKL
Anopheles_darlingi                AYEHFTTFVHSLTEPLYSLLNIDAVDAQDTLLHKYLVQTISSWACRLDHPDCLARTKTLLTSEAS--GRAMVHADLSTVTYCYGMRTAGEQEYQYVYRKM
Aedes_albopictus                  AYEHFSTFVQSLVVPLYSLLNIEAVDPQDSLLHKYLVQTISSWACRLDHPDCLARTKALLATEAS--GGAKVHPDLATVTYCYGMRTAGEQEYQYVYRKM
Aedes_ageypti                     AYKDFTTMVDHLIKKVYKTLDVTAVSDTEPLLHKYLKQTISTWACLIGNEDCLKRTKEALQKEVTEGIP--VHPDVATVVYCNGLRTADVAEYQYLYKRI
Anopheles_albimanus               DYENFTTLVNHLIRKVYATLEVDSVSESEPLLHKYLKQTISTWACMVGNEDCLKRTKAALQKEVTEGVK--VHPDIATVVYCQGLRTADAAEYQHLYKRI
Culex_quanquefasciatus            EYHDFLIMVHRLIKYVYSTLSVATVSETESMLHKYLKQTITNWACIVGDLECLRLTKEALENEAHNGVM--VHPDVSSVVYCHGLRDADENVFVYLYNRI
Clustal Consensus                  *..*  :.  *   :*  * :  *   :.:***** ***:.***  .  :*   *   *   .       * .*:::*.** *::      : :.* ::
```

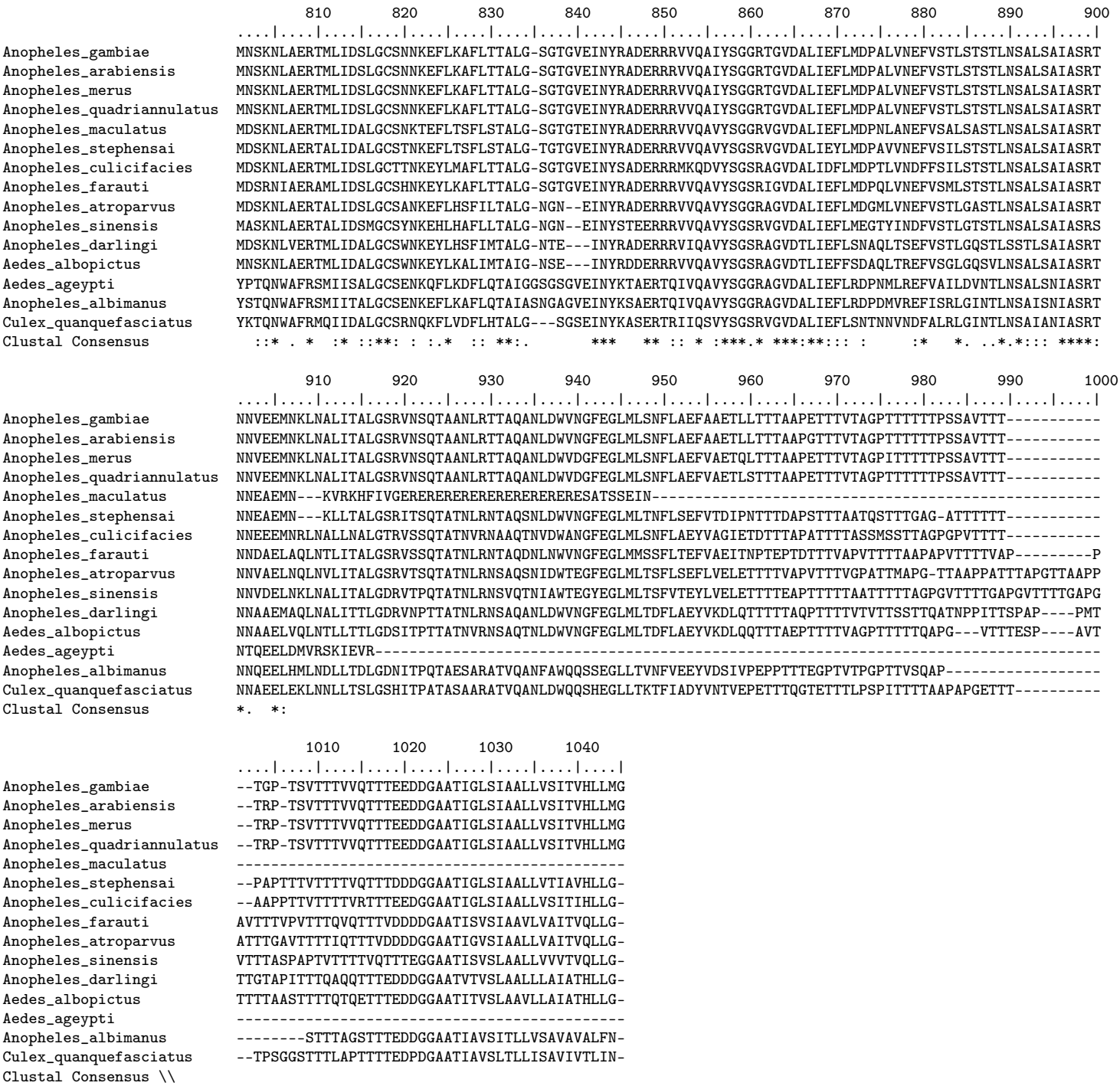


Figure 1. Amino acid sequence alignment of APN in mosquitoes (Diptera). Signal peptide was underlined at N terminal region. Yellow represents GAMEN motif, Blue represents HEXXH, Zn binding region, green represents NEGFA conserved regions, red represents conserved Cys residues, royal blue colour represents glycosylation sites and pink represents GPI anchored region at C terminal end.

netic tree also supported the grouping of each order of non-mosquitoes into two major clades. In Particular, in Diptera two major clades belonged to Anopheline and Culicinae. The phylogenetic relationship followed by aligning the amino acid sequences from various insect species belonging to six distinct orders on one scale showed that the Lepidoptera and Diptera APN sequences situated into a single cluster [44]. The accumulation of similar sequences from other insects species reveal closer relative of the mosquitoes. Coleoptera and Hymenoptera were found in another single cluster. However, Hemiptera are situated in separate clades.

Conclusion

In silico analysis of the APN gene among six orders of class Insecta was also carried to understand more about APN genes and their evolution. A comparative

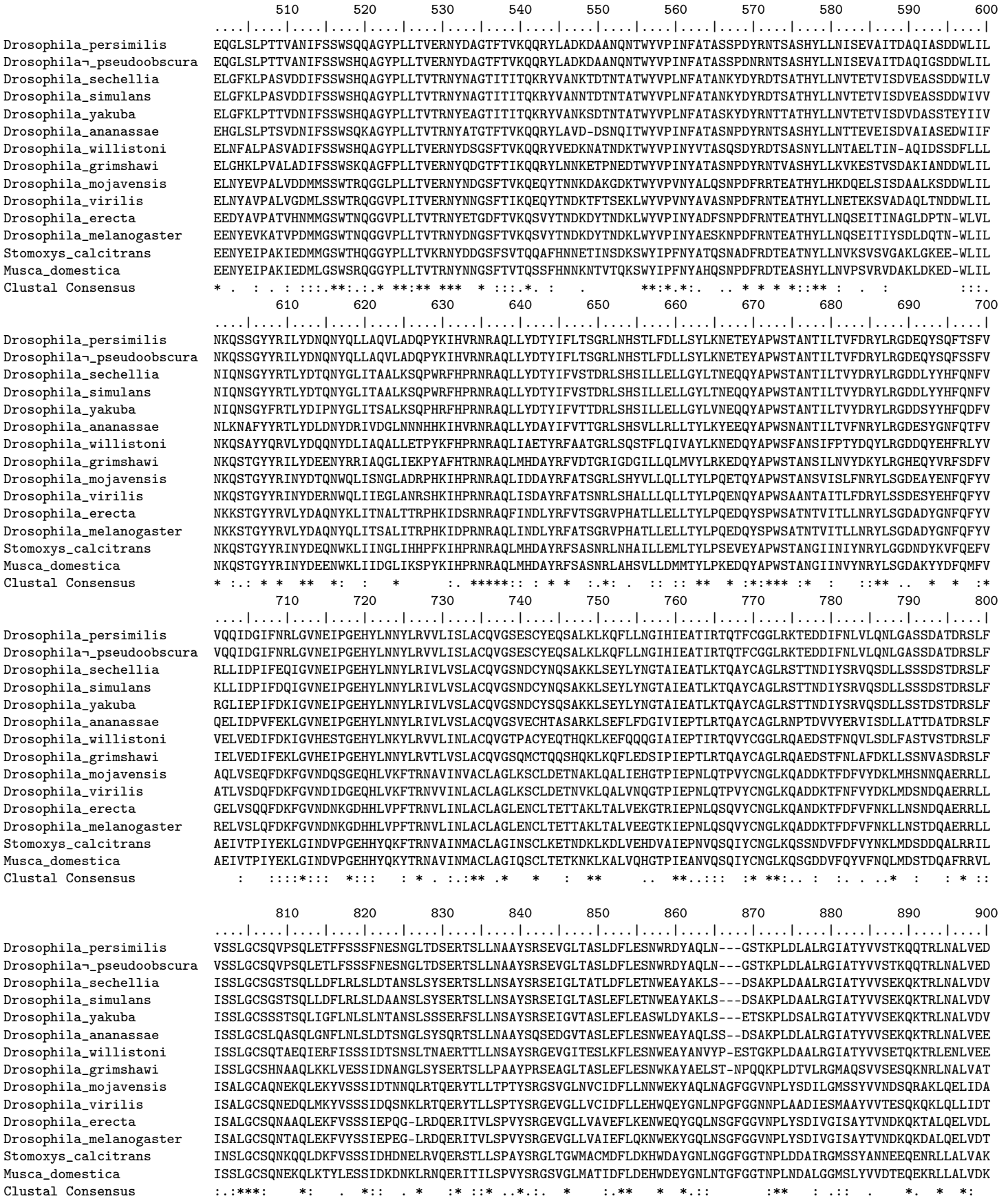
```
                                    10        20        30        40        50        60        70        80        90       100
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Drosophila_persimilis               --MAMWLLNSSLLLLLLAAVSLQARLILPQ---------DDGQLETKEQFFATPRDDIDEGNYRLPNNTAPESYKVELWTNVHSGARDFNGSVQIDLQVL
Drosophila¬_pseudoobscura           --MAMWLLNSSLLLLLLAAVSLQARLILPQ---------DDGQLETKGQFFATPRDDIDEGNYRLPNNTAPESYKVELWTNVHSGARDFNGSVQIDLQVL
Drosophila_sechellia                --MARWLLNSFSLILLLAVGSLQARLILPLSPDDGPRDQGKMETKGAALFAAPRVDDNSAGNYRLPNTTEPESYNVELWTNVHNGDTQFTGTVNIDIRVL
Drosophila_simulans                 --MARWLLNSFSLILLLAVGSLQARLILPLSPDD-----GQLETKGAALFAAPRVDDNSAGNYRLPNTTEPESYNVELWTNVHNGDTQFTGTVNIDIRVL
Drosophila_yakuba                   --MARWLLNSFSLILLLAVGSLQARLILPLSLDD-----GQLETKGAALYATPRVDDNSAGNYRLPNTTEPESYNVELWTNVHTGDDQFNGIVNIDIRVL
Drosophila_ananassae                ----MAMWLLKSCILFLAVASSQARLIPPLSLDIED---VQLETKGTYSLISPRADDNSEGNYRLPNNTAPEAYRVELWTNVHNGTREFNGSVQIDIQVL
Drosophila_willistoni               --MAKWLLNSCLLLLLGAVSLQAGLIIPHTDVEQIET--------KGYQVITPRVDDNSPDNYRLPNNSIPVSYNVELWTNVHEGDTNFNGTVKIDIQIV
Drosophila_grimshawi                ---MMRWLITSGLLLLLAASLQAGLLPPPIDNESQ----LETKGYQQVSLNTPREDDNNDENYRLPNNTVPIAYDVELWTQVDTGIREFKGIVKIDLQVQ
Drosophila_mojavensis               ----MARVFTPILILLLVLATTHAAVVRQFPPFER------LQQQLIENPNLRAAQRADEQTYRLPNNTEPLEYNIDLTTNVHDGTTEFQGTVTITLKVL
Drosophila_virilis                  ---MARVLVTILALLLVVATTAHAAVVRNFPPFEQ------LQQLEN--PNSRLAPRADEDTYRLPNNTEPVEYTVELHTNVHTGDTAFSGTVAISINVL
Drosophila_erecta                   MSRVQVGGMSLFLVLLLAISATQAAVVRQYPVFDN------LQHLEN-PMQRSVVPFADEDNYRLPNDTIPSHYAVSLSTNVHTGDTVFNGTVAITLSVS
Drosophila_melanogaster             MSCVQVGGMSVFLVLLLAISATQAAVVRQYPILDH------LQHLEN-PMQRLVVPLADEDNYRLPYDTIPSHYAVSLSTNVHTGDTVFNGTVAITLSVL
Stomoxys_calcitrans                 ------------------------------------------------------------NTTKPTHYDIDLSTMVHEGTRAFVGTVKIHVTFE
Musca_domestica                     ----------------------------------------------------------------------------------------------
Clustal Consensus

                                   110       120       130       140       150       160       170       180       190       200
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Drosophila_persimilis               ELSSSITLHYRQTSNFTASIISRD-EVAQAIELDVATDTVREFLTLTPKLATVTFAANTNWTLSIDYNGTLRGDMGGFYISSYTDDDSTVHYLATTQFES
Drosophila¬_pseudoobscura           ELSSSITLHYRQTSNFTASIISRD-EVAQAIELDVATDTDREFLTLTPKLATVTFAANTNWTLSIDYNGTLRGDMGGFYISSYTDDDSTVHYLATTQFES
Drosophila_sechellia                SETSNITLHYRQTSNFEATIISRDGATPTAIPLTVTPELQREFLVLTRTTAGEAFGANTNWTITINYTGIHRSDMGGFYISSYTDDDGKQHFLATTQFES
Drosophila_simulans                 SETSNITLHYRQTSNFEATIISRDVATPTAIPLTVTPELQREFLVLTRTTAGEAFGANTNWTITINYTGIHRSDMGGFYISSYTDDDGKQHFLATTQFES
Drosophila_yakuba                   NETSNITLHYRQTSNFEASIISRDVVTPTSIPLTVTTELQREFLVLTRTTAGEAFGANTNWTITIKYNGTHRTDMGGFYLSSYTDDDGKQHFLATTQFES
Drosophila_ananassae                EPTTEIKLHFLQTSDHVATIISRDQTDPTAIPLTVANETLRQFLILTRPDSTEPFEANTNWTLSITYNGTLRSDMGGFYISSYTGVDNQVHYLATTQFES
Drosophila_willistoni               EASVNITLHARQTSNFVATIIDRDNATATEINLEVDASDTTREFLVLSAEG-LTFDPNTNWTLTIKYTGLLRLDMGGFYMSSYTEN-NTVHYLATTQFES
Drosophila_grimshawi                TASNTITLHAKETENYKATLKSKDNAASSSIPLKWN-VGTREFLQLTPVDTTLTFEKDSNWTLTIEYEGSLRTDNGGFYLSTYKDDDDATHYLATTQFES
Drosophila_mojavensis               ETSSDIVVHARQLAILGATIRKSDAAESTDVALDIVEDKEREFLHLS-ALG-LTFEEDSIWLLTIKYTGNLRLDNGGFYLSRYTDETGATKYLATTQFES
Drosophila_virilis                  ETSTIIVVHARQLENFTATIRKD--GETTDTALEITHDTDREFLKLS-AKD-LTFEEGTSWLLTIKYEGNLRTDNGGFYLSKYTDENGATKYLATTQFES
Drosophila_erecta                   STTSKIVVHARQLENFTASIIQQGVTGAVAQDLVYEYEAEREFLTFS-KTG-LSFPEDTTWILTIKYQGHLRKDNGGFYLSTYTDEQGNTKYLATTQFES
Drosophila_melanogaster             NTTTKIVVHARQLENFTASIIQQGVTEAVAQELVYEYEAEREFLTFS-KTG-LTFPEDTTWILTINYQGHLRTDNGGFYLSTYTDEEGNTKYLATTQFES
Stomoxys_calcitrans                 EATTSIVLHARQLIITAATIQFG---TTAPEDLEPTYEEEREFLTLT-RKTNTPFPKDSQWVVTLDYNGQLRLDNGGFYLSTYNDANGNVRHLATTQFES
Musca_domestica                     ------------------------------------------------------------------------------GNERYLAITQFES
Clustal Consensus                                                                                                  .  :.** *****

                                   310       320       330       340       350       360       370       380       390       400
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Drosophila_persimilis               VESSLAGYFGVPFALPKLDQAGIPDFSAGAMENWGLATYREQYMWWNKQNSTINQKTNIANIIGHEYAHMWFGDLVSIKWWTYLWLKEGFATLFSYESND
Drosophila¬_pseudoobscura           VESSLAGYFGVPFALPKLDQAGIPNFSAGAMENWGLATYREQYMWWNKQNSTINQKTNIANIIGHEYAHMWFGDLVSIKWWTYLWLKEGFATLFSYESND
Drosophila_sechellia                VESSLASYFGVPFALPKLDQAGIPDFSAGAMENWGLATYRERYMWWNKQNSTINLKTNIANIIGHEYAHMWFGDLVSVKWWTYLWLKEGFATLFSYESND
Drosophila_simulans                 VESSLASYFGVPFALPKLDQAGIPDFSAGAMENWGLATYREQYMWWNKQNSTINLKTNIANIIGHEYAHMWFGDLVSVKWWTYLWLKEGFATLFSYESND
Drosophila_yakuba                   VEASLAGYFGVPFALPKLDQAGIPDFSAGAMENWGLATYREQYMWWNEQNSTINLKTNIANIIGHEYAHMWFGDLVSVKWWTYLWLKEGFATLFSYESND
Drosophila_ananassae                VESSLASYFGVPFPLPKLDQAGIPDFSAGAMENWGLATYREQYMWWTKENSTINLKTNIANIIGHEYAHMWFGDLVSIQWWTYLWLKEGFATLFSYESND
Drosophila_willistoni               VESSLASYFGVPFALPKLDQAGIPDFSAGAMENWGLATYREQYMWWTKEESTINLKTNIANIIGHEYTHMWFGDLVSVKWWTYLWQKEGFATLFSYESND
Drosophila_grimshawi                VEASLAKYYDVPFPLPKLDQAGIPQFSAGAMENWGIATYREQYMWWTKEKSTINLKTSIANIIGHEYAHMWFGDLVSLKWWTYLWLKEGFATLFSYESND
Drosophila_mojavensis               VEKRLSGYFGVPFALPKLDQAAIPDFAAGAMENWGLATYREEYLLYNVENSTTSTQTNIATIEAHEDAHMWFGDLVAIEWWSYLWLKEGFATLFENLAVN
Drosophila_virilis                  VEKRLSGYFGVPFALPKLDQAAIPDFAAGAMENWGLATYREEYLLYNIENSTTNTQTNIATIEAHEDAHMWFGDLVAIEWWSYLWLKEGFATLFENLAVH
Drosophila_erecta                   VEKRLSGYFGVPYTLPKLDQAGIPDFAAGAMENWGLATYREEYLLYNTENSTTSTQTNIATIEAHEDAHMWFGDLVAIEWWSFLWLKEGFATLFENLAVD
Drosophila_melanogaster             VEKRLSGYFDVPFALPKLDQAGIPDFAAGAMENWGLATYREEYLLYNTENSTTSTQTNIATIEAHEDAHMWFGDLVAIEWWSFLWLKEGFATLFENLAVD
Stomoxys_calcitrans                 ILERLAEYYGVPFALPKIDQAAIPDFAAGAMENWGLATYREEYMLYNKDHSTITTQTNIANIIAHEYAHQWFGNYVAIQWWTYLWLKEGFATLFSYKATD
Musca_domestica                     ILERLAEYYNVDFALPKIDQAAIPDFAAGAMENWGLATYREEYMLYNKDTSAVTTQTNIASIIAHEYCHQWFGNYVAVHWWTYLWLKEGFATLFSYKAVD
Clustal Consensus                   :   *: *:.* :.***:***.**:*:********:*****.*: :. : *: . :*.**.* .**  * ***: *::.**::** ********.  : .

                                   410       420       430       440       450       460       470       480       490       500
                                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Drosophila_persimilis               IAFPEWNTYQIYHVNDYNSALLNDAAGAATPMTHYVQTPSEISSRYNTFSYAKPASVLFMFKNAWTDSVFRLGLNKYLTKNQFNSCDEWDLFESLQEAAN
Drosophila¬_pseudoobscura           IAFPEWNTYQIYHVNDYNSALLNDAAGSATPMTHYVQTPSEISNRYNTFSYAKPASVLFMFKNAWTDSVFRLGLNKYLTKNQYNSCDEWDLFESLQEAAN
Drosophila_sechellia                IAFPLWDTYQIFHVSDYNSALINDALASAVPMTHYVQTPTEISNRYNTFSYAKPASVLYMFKNAWTDKVFRTGLNKYLTKNKYTSCDEWDLFASFQESAD
Drosophila_simulans                 IAFPLWDTYQIFHVSDYNSALINDALASAVPMTHYVQTPTEISNRYNTFSYAKPASVLYMFKNAWTDKVFRTGLNKYLTKNKYTSCDEWDLFASFQESAD
Drosophila_yakuba                   IAFPLWDTYQIYHVNDYNSALLNDALASAVPMTHYVQTPSEISNRYNTFSYAKPASVLYMFKNAWSDKVFRTGLNKYLTKNQYTSCDEWDLFASFQESAD
Drosophila_ananassae                IAFPEWDTYQIYHVNDYNSALLNDASTSAVPMTHYVQTPTEISGRYNTFSYAKPSAVLYMFKNAWTDRVFRAGLNKYLTKNQYTSCDEWDLFAALQESAD
Drosophila_willistoni               IAFPEWDTYQIFHVADYHSALMNDAGENAVPMTHYVQTPNEISNRYNTYSYAKPASVLYTFKNAWTDKVFRLGVHKYLDQHKFTSCDEWDLFASFQEAAD
Drosophila_grimshawi                IAFPEWDTYQIFHVSDYNSALINDASISTP-MTHYVQTPREISGRYGTISYAKPASVLYMFKNAWTDDVFRRGLNLYLTQFAYTSADEWDLFSSLQKAAD
Drosophila_mojavensis               LAYPEWDIFQTFHTGSYQSAMVSDGNPSVRPMTYYVEKPSDISLLYDSIAYAKAGSVLDMWRHALTNTVFQQGLHNYLVEKAYSAANETDLFNAIALAAR
Drosophila_virilis                  LAYPEWDIFQTFHAGSYQSALIADANPNVRPMTTFVEKPSEISQLYDSIAYAKAGSVLDMWRHALTNTVFQQGLHNYLVTKAYSAANETDLFDAISLAAR
Drosophila_erecta                   LAYPEWDIFQTFHAGSYQSALVNDASANARPMSHFVQKPSEIALLYDSVSYAKAGSVLDMWRHALTNTVFQRGLHNYLVDNQFTAANPSKLFDGIAKAAV
Drosophila_melanogaster             LAYPEWDIFQTFHAGSYQSALVNDASANARPMSHFVQKPSEIALLYDSVSYAKAGSVLDMWRHALTNTVFQRGLHNYLVDNQFTAANPSKLFEAIAKAAV
Stomoxys_calcitrans                 AAYPEWGIWQQFHTDDYQGALNNDAGVSPRPMTHYVQKPQEISRLYDTVSYAKAGSVLNMWNHALTDTVFQRGLHNYLTANNFSAAVEDKLFTAIQIASK
Musca_domestica                     AAYPEWGVWQQFQTDDYQSALTNDASDYPRPMTHYVQKPSEIAMSYGTVTYAKAGSVLNMWNHALTDKVFQRGLHNYLDANQYSSAVEEQLFLAIQTAAK
Clustal Consensus                    *:* *. :* ::. .*:.*:  *.      *: :*:.* :*:  *.: :***..:**  :.:* :: **: *:: **    :.:.   .** .:  ::
```

sequence analysis was carried out to gain a better understanding of the evolution of APN gene. The conservation of residues was considerable and widespread throughout the coding region. However, the high level of divergence was observed in the size and positions of introns. This could be ascribed to the presence of repetitive DNA elements. Great sequence and structural diversity suggests that APN genes are highly evolving with selective functional constraints. APN sequences were found that contain conserved HEXXH and GAMEN



signature sequences in multiple APN orthologs. The APN gene belonging to different orders was found to be AT rich except Diptera. Dipterans APN gene was found to be GC rich and at third codon position high GC content was observed. The presence of high GC content in Diptera clearly indicates that dipteran APN gene is a biased gene. The Leucine residues are present in highest frequency in the amino acid composition of protein of

```
                                  910        920        930        940        950        960        970
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Drosophila_persimilis    VKTNGPEYVNDDLSTVVDSRVKANLDWLAVHRDPLISWLAAYNTVKDNGSPALTTSLATILASALTLLLWQLF--
Drosophila¬_pseudoobscura VKTNGPEYVNDDLSTDIDSRVKDNLDWLAVHREPLISWLAAYNTVKDNGSPALTTSLATILASALTLLLWQLF--
Drosophila_sechellia     VKAEGPSYINDDLSTTIDSRITFNFAWLEVNRDPVLNWIEDS--FRENGSPSLTSSITTILVSAMALLLFRLY--
Drosophila_simulans      VKAEGPSYINDDLSTTIDSRITSNFAWLEVNRDPVLNWIEDS--FRENGSPSLTSSITTILVSAMALLLFRLY--
Drosophila_yakuba        VKAEGPSYVNDDLSDTIDSRITSNFAWLEVNRDPVLNWIEDS--LRENGSPSLTASITTILVSAMALLLFRL---
Drosophila_ananassae     VQSKENGYLQDDLSSTVNSRIQSNFDWLDTNRDPVISWVNSH--ISGSGSSGLTVSLATVLISAMALLLSKLY--
Drosophila_willistoni    VKAHGPAYVNDDLSEIITARIDANFAWLNEHRDSVINWLVSYR---TSGSSALTTSLTTFMAAALTLLLCKVF--
Drosophila_grimshawi     INGNGIEHVNDDIDTAVNRNIQANFDWLNVHRDPVINWLIAYR----SGTAPLSASIFATLTALITLFLCKFS--
Drosophila_mojavensis    VK--DSDYVPATLQASVDTNVEANMAWLTSNRGPIISWVTSFR----NGSPALTASVLAMALCLFVAMFF-----
Drosophila_virilis       VK--GSEFVSDTLQDSVDTTIAANMEWLSTNREPIISWVTGFR----NGSPAMTASVLALALCFLAAKLF-----
Drosophila_erecta        VK--NSEYVPSTLQASVDSNVNANYAWLASNRDPLMSWVAGFR----SGSSALSASIVAILACLLAAKLF-----
Drosophila_melanogaster  VK--GSEYVPSTLQDSVDSNVNANYAWLASNRDPLMTWVAGFR----SGSSALNASILAILACLLAAKLF-----
Stomoxys_calcitrans      VK--NSVHASANLEEVVEANFKTNRDWLKINRDPLMSWLISYR----SNGGATLAISTLSMVS--VILLVVSRLF
Musca_domestica          VK--SSEHVSSSLEASVKYNIQSNFDWLKSNRDPLINWFQSFN----IKSGASAVISSSSMAISTLLLLVISRLF
Clustal Consensus        ::     .    :.  :   .  *  **  :* .:::.*.                          :
```

Figure 2. Amino acids sequence alignment of APN in non-mosquitoes (Diptera). Signal peptide was underlined at N terminal region. Yellow represents GAMEN motif, Blue represents HEXXH , Zn binding region, green represents NEGFA conserved regions, red represents conserved Cys residues.

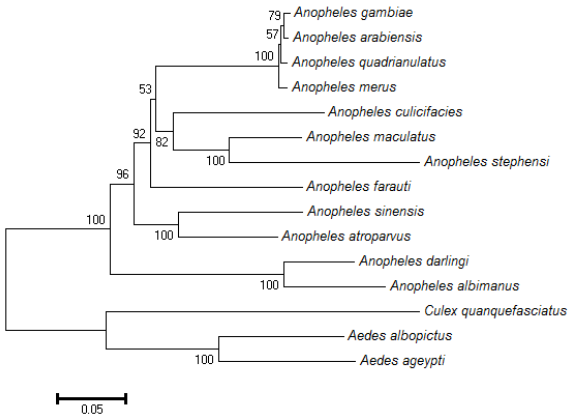


Figure 3. Phylogenetic analysis of APN Protein of mosquitoes (Diptera).

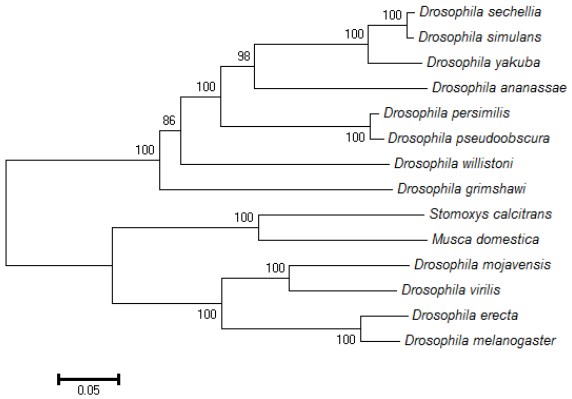


Figure 4. Phylogenetic analysis of APN protein of non-mosquitoes (Diptera).

each species, accounting for 9% of total amino acids and cysteine residues are least abundant, 1% in the APN protein of insects. A significant high synonymous codon usage bias has been shown by mosquito APN genes belonging to the order Diptera (preferentially using one or two synonymous codons) as compared to other species.

References

1. Dinglasan R R, Kalume D E, Kanzok S M, Ghosh A K, Muratova O, Pandey A. (2007). Disruption of Plasmodium falciparum development by antibodies against a conserved mosquito Midgut antigen. Proc Natl Acad Sci USA. 104 (33):13461-13466.
2. Zhang K B and Adang M J. (2009). Anopheles gambiae alkaline phosphatase is a functional receptor of Bacillus thuringiensis jegathesan Cry11Ba toxin. Biochemistry.48: 9785–9793.
3. Rosenfeld A, Vanderberg J P. (1999). Plasmodium berghei: induction of aminopeptidase in malaria resistant strain of Anopheles gambiae. Exp Parasitol. 93(2):101-104.
4. Tamura K, Stecher G, Peterson D, Filipski A and Kumar S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution. 30(12): 2725-2729.
5. Marchler B A, Bo Y, Han L, He J, Lanczycki C J, Lu S. (2014). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic acids research. 45(D1): D200-3.
6. Schaeffer R D, Liao Y, Cheng H, Grishin N V. (2017). ECOD: new developments in the evolutionary classification of domains. Nucleic Acid Research. 45(D1): D296-D302.
7. Finn R D, Attwood T K, Babbitt P C, Bateman A, Bork P, Bridge A J. (2016). InterPro in 2017 beyond protein family and domain annotations. Nucleic acids research. 45(D1): D190-9.
8. Sigrist C J, De Castro E, Cerutti L, Cuche B A, Hulo N, Bridge A, Bougueleret L and Xenarios I. (2012). New and continuing developments at PROSITE. Nucleic acids research. 41(D1): D344-D347.
9. Gabor E T, Istvan S. (2001). The HMMTOP transmembrane topology prediction server. Bioinfor-

matics. 9(17): 849-850.
10. Bjellqvist B, Hughes G J, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S and Hochstrasser D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis. 14(1): 1023-1031.
11. Richard J E, Lin K, Tan T. (2000). A functional significance for codon third bases. Gene, 245: 291-298.
12. Sharp P M, Tuohy T M F, Mosurski K R. (1986). Codon usage in yeast cluster-analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14: 5125-5143.
13. Wright F. (1990). The effective number of codons used in a gene. Gene. 87: 23-29.
14. Puigbo P, Bravo I G, Garcia-Vallve S. (2008). CAIcal: a combined set of tools to assess codon usage adaptation. Biol. Dir. 3: 38.
15. Saitou N, Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.
16. Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evol. 39: 783-791.
17. Thompson J D, Higgins D G, Gibson T J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22): 4673-80.
18. Jorgensen F G, Schierup M H, Clark A G. (2007). Heterogeneity in regional GC content and differential usage of codons and amino acids in GC poor and GC-rich regions of the genome of Apis mellifera. Molecular Biology and Evolution. 24: 611–961.
19. Wright F. (1990). The effective number of codons used in a gene. Gene. 87: 23-29.
20. Cutler R W, Chantawannakul P. (2007). The effect of local nucleotides on synonymous codon usage in the honeybee (Apis mellifera L.) genome. Journal of Molecular Evolution. 64: 637–645.
21. Sharma J, Chakraborty S, Uddin A. (2014). Codon usage bias in two hemipteran insect species: Bemisia tabaci and Homalodisca coagulata. Advances in Biology. http://dx.doi.org/10.1155/2014/145465.
22. Powell J R and Moriyama E N. (1997). Evolution of codon usage bias in Drosophila. Proc. Nad. Acad. Sci. U. S. A. 94: 7784-7790.
23. Shields D C, Sharp P M, Higgins D G and Wright F. (1988). "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. 5: 704-716.
24. Behura S K, & Severson D W. (2011). Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in Aedes aegypti and Anopheles gambiae. Insect molecular biology. 20(2):177-187.
25. Rodriguez O, Singh B K, Severson D W, Behura S K. (2012). Translational selection of genes coding for perfectly conserved proteins among three mosquito vectors. Infection, Genetics and Evolution. 12, 1535–1542.
26. Vicario S, Moriyama EN, Powell JR. (2007). Codon usage in twelve species of Drosophila. BMC evolutionary biology.7(1):226.
27. Behura S K, Severson D W. (2012). Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. PLoS One. 7(8): e43111.
28. Jia X, Liu S, Zheng H, Li B, Qi Q, Wei L, Zhao T, He J. and Sun J. (2015). Non-uniqueness of factors constraint on the codon usage in Bombyx mori. BMC genomics. 16(1): 356.
29. Hershberg R, Petrov D A. (2008). Selection on codon bias. Annu Rev Genet. 42: 287–299.
30. Drummond D A, Bloom J D, Adami C, Wilke C O, Arnold F H. (2005). Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 10240: 14338–14343.
31. Ikemura T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2:13-34.
32. Rajagopal R, Agrawal N, Selvapandiyan A, Sivakumar S, Ahmad S, Bhatnagar R K. (2003). Recombinantly expressed isoenzymic aminopeptidases from Helicoverpa armigera (American cotton bollworm) midgut display differential interaction with closely related Bacillus thuringiensis insecticidal proteins. Biochemical Journal. 370(3): 971-978.
33. Blomster M, Wetterholm A, Mueller M J, Haeggström J Z. (1995). Evidence for a catalytic role of tyrosine 383 in the peptidase reaction of leukotriene A4 hydrolase. European journal of biochemistry. 231(3): 528-534.
34. Rudberg P C, Tholander F, Andberg M, Thunnissen M M, Haeggstrom J Z. (2004). Leukotriene A4 hydrolase identification of a common carboxylate recognition site for the epoxide hydrolase and aminopeptidase substrates. Journal of Biological Chemistry. 279(26): 27376-27382.
35. Hughes A L. (2014). Evolutionary diversification of aminopeptidase N in Lepidoptera by

conserved clade-specific amino acid residues. Molecular phylogenetics and evolution. 76: 127-133.

36. Ito K, Nakajima Y, Onohara Y, Takeo M, Nakashima K, Matsubara F , Yoshimoto T. (2006). Crystal structure of aminopeptidase N (proteobacteria alanyl aminopeptidase) from Escherichia coli and conformational change of methionine 260.
37. Pasqualini R, Koivunen E, Ruoslahti E. (1995). A peptide isolated from phage display libraries is a structural and functional mimic of an RGD-binding site on integrins. The Journal of cell biology. 130(5): 1189-1196.
38. Atsumi S, Mizuno E, Hara H, Nakanishi K, Kitami M, Miura N, Tabunoki H, Watanabe A and Sato R. (2005). Location of the Bombyx mori aminopeptidase N type 1 binding site on Bacillus thuringiensis Cry1Aa toxin. Appl. Environ. Microbiol. 71(7): 3966-3977.
39. Singh A, Sivaprasad C V. (2009). Functional interpretation of APN receptor from M. sexta using a molecular model. Bioinformation, 3(8): 321.
40. Martelli P L, Fariselli P, Malaguti L, Casadio R. (2002). Prediction of the disulfide bonding state of cysteines in proteins at 88% accuracy. Protein Science. 11(11): 2735-2739.
41. Agrawal N, Malhotra P and Bhatnagar R K. (2002). Interaction of gene-cloned and insect cell-expressed aminopeptidase N of Spodoptera litura with insecticidal crystal protein Cry1C. Appl. Environ. Microbiol. 68 (9): 4583-4592.
42. Jakhar, Renu, Ritu Gill, and Surender Kumar Gakhar. (2019). Genome-wide analysis of various paralogs of aminopeptidase N (APN) gene in Anopheles gambiae (Diptera: Culicidae). Journal of Applied Biology and Biotechnology: 15-24.
43. Jakhar R, Gakhar SK. (2021). Identification and Characterization of Aminopeptidase N 1 Gene of the Indian Malaria Vector Anopheles culicifacies (Diptera: Culicidae). Journal of Medical Entomology. 58(3):1476-81.
44. Rajagopal R, Agrawal N, Selvapandiyan A, Sivakumar S, Ahmad S, , Bhatnagar R K. (2003). Recombinantly expressed isoenzymic aminopeptidases from Helicoverpa armigera (American cotton bollworm) midgut display differential interaction with fp closely related Bacillus thuringiensis insecticidal proteins. Biochemical Journal. 370(3): 971-8.