Genome-wide analysis of various paralogs of aminopeptidase N (APN) gene in *Anopheles gambiae* (Diptera: Culicidae)

Renu Jakhar¹, Ritu Gill², Surender Kumar Gakhar*¹

¹Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak, India
²Centre for Biotechnology, Maharshi Dayanand University, Rohtak, India

**ABSTRACT**

Malaria continues to be a life threatening infectious disease throughout the tropical region of the world. aminopeptidase N1 (APN1) is one of the best choices for developing new Malaria Transmission-blocking vaccines. In this study an attempt has been made to overview genome-wide identification of APN genes in *Anopheles gambiae*. A total of eighteen *A. gambiae* APN sequences were found that contain conserved HEXXH and GAMEN signature sequences, indicate that large numbers of APN isomers present in mosquitoes. Multiple APN paralogs exist as a gene cluster may propose that huge synthesis of APNs is required for rapid digestion of peptides over a brief period. Gene structure study shows high sequence variations among them. Protein–protein interactions show that APN1 is highly connected protein, supporting their role as hub with other five types of APNs involved in glutathione metabolism, act as hub protein and disruption of one of these proteins may affect the whole pathway.

**1. INTRODUCTION**

The global health and socioeconomic burden imposed by the mosquitoes is well established, with an estimated 219 million malaria cases reported worldwide in 2017 (WHO Report, 2017). In addition to malaria, other vector-borne diseases are responsible and pose a significant threat to humanity, especially in tropical and subtropical countries [1]. The African continent contributes maximum to the total global malaria burden because of the presence of the *Anopheles gambiae* complex among different *Anopheles* vector, with over 400 million people living under the risk of infection. The use of chemical insecticides for decades has led to the development of resistance in the vector, making malaria control a challenging task, especially for the highly populated countries like India [2]. Also, the unavailability of an effective vaccine has pushed scientists worldwide to look for novel mechanisms for malaria control [3]. In recent years, transmission-blocking interventions (TBI) have emerged as a potential approach to replace our chemical toolkit and enhance our malaria control capabilities. The TBI has an additional advantage over conventional control methods because it avoids the selective resistance pressure within mosquitoes, as TBI relies on their survival instead of mortality [4].

A variety of proteins from *Plasmodium falciparum* has been previously tested for transmission blocking [5]. However, recent discoveries suggesting the use of multiple mosquito midgut molecules by *P. falciparum* has diverted the attention of the scientific community toward enzymes such as aminopeptidase N (APN) [5]. APNs are membrane-bound ubiquitous exopeptidase targeting single amino acids at N-terminal site of the polypeptide chain and have been identified in the midgut epithelial cells of various insect species [6]. Their role as cry toxin-binding proteins in insects was established [7,8] and in previous studies shown that till date more than 20 types of APN were reported in insects [9]. The alanyl aminopeptidase N is the leading malarial TBV immunogen [10]. A midgut specific protein, reported to play an important role in ookinete invasion of *Plasmodium* in the *A. gambiae*. Previous studies have shown that APN1 acts as a receptor. It is a glycosyl-phosphatidylinositol-anchored protein for attachment of *Plasmodium* parasite and later helps in sexual growth [11–13].

© 2019 Renu Jakhar, et al. This is an open access article distributed under the terms of the Creative Commons Attribution License -NonCommercial-ShareAlike Unported License (http://creativecommons.org/licenses/by-nc-sa/3.0/).
Rosenfeld and Vanderberg [6] validated the potential use of APN for malaria transmission blocking in A. gambiae by recording a significantly reduced number of oocysts using rabbit polyclonal antibodies against APN in infected mice [14]. Identified potent epitope from AgAPN1 is highly immunogenic. Similar results were observed in A. stephensi [15], suggesting possible multispecies blocking strategies. Also, antibodies against APN1 were responsible for 73% blockage in A. gambiae and 67% in A. stephensi [13]. Based on the crystal structure of the near-full length APN1, a structure-guided construct has been expressed in Escherichia coli and revealed B cells epitopes as transmission-blocking antigens [16]. Studies on the APN1 have also been identified and characterized from A. stephensi as a candidate for Transmission Blocking Vaccine [17]. Also, by applying omics technologies to understand mosquito biology and evolution, these help to explore the diversity and the variations among mosquitoes [18]. To identify protein-coding genes of A. gambiae based on its genomic sequence, a deep proteomic analysis using high-resolution Fourier transform mass spectrometry for both precursor and fragment ions was carried out [19]. Two isoforms of aminopeptidase N was also identified in Aedes aegypti larval midgut [20]. Five APN genes APN1 [13] and APN2, APN3, APN4, and APN5 [21] have been characterized in A. gambiae, but there is no understanding of molecular evolution in-paralogs. The availability of whole genome sequence of A. gambiae provides the opportunity to fill the gap in knowledge. Therefore, the current study was designed to analyze the organized structure of the paralogous gene structure in A. gambiae.

2. MATERIALS AND METHODS

2.1 Retrieval and analysis of A. gambiae APN paralogous gene sequences

The genomic DNA sequences, cDNA sequences, and amino acid sequences of various paralogous APN were obtained from Vectorbase database (https://www.vectorbase.org) and NCBI for comparison between these paralogous. The functional annotation of genes, number of nucleotides, gene products, orientation of gene and its position on the chromosome and amino acids sequences were analyzed in all paralogs. Distribution of intron and exon and intron phase patterns was obtained by online Gene Structure Display Server [22].

2.2. Conserved domains identification for APN proteins

The conserved domains were identified in all paralogous APN sequences by various domain search tools like NCBI CDD (Conserved domain database) [23], Pfam [24], and InterProScan [25] software, which in turn by searching significantly close orthologous family members. The predicted cleavage sites signal peptides were analyzed by using Signal IP algorithm. Glycosylation and phosphorylation sites were predicted by using the Prosite tool [26]. Possible transmembrane domains in APN proteins were also found by using the TMHMM [27]. The Molecular Weight (MW) and Isoelectric point (pI) were calculated by Compute pI/MW [28]. Multiple sequence alignments were performed with BIOEDIT program with all paralogous sequences. Generation of sequence logo was done by Skylign tool to explore conservation of the sequence of motifs in APN paralogs of A. gambiae.

2.3. Protein–protein interaction (PPI) Network construction

The PPIs for 18 APNs protein were examined from STRING v10.5database [29]. PPIs networks were generated based on some interaction sources, i.e., experiments, text mining, co-occurrence, co-expression, databases, gene fusion, and neighborhood. PPIs that have at least a confidence score between medium ranges of 0.4–0.7 were selected for generation of networks. Topological parameters are also evaluated, i.e., the number of edges, number of nodes, and average node degree, etc. Network visualization was also carried out by Cytoscape v 3.5.1 [30]. Cyto-Hubba [31], a java plugin for Cytoscape software, was used to construct the APN hub of mined PPI network.

2.4. Phylogenetic analysis

Eighteen APN gene sequences retrieved from A. gambiae were compared and analyzed for conserved sites using MEGA6.06 [32]. Generation of the phylogenetic tree was done by using Bootstrap consensus neighbor-joining method for APN paralogous gene sequences.

2.5. Expression profile of the A. gambiae APN

The expression patterns were determined separately for all paralogous genes from MozAtlas (http://mozatlas.gen.cam.ac.uk) [33]. The specific expressions of the individual paralogous genes were analyzed in different tissues (midgut, salivary gland, fat body, ovary, and hemolymph, etc) of A. gambiae.

3. RESULTS AND DISCUSSION

3.1 Comparison of paralogous APN genes in A. gambiae

A total of 18 A. gambiae APN gene sequences were retrieved from the PEST genome on Vectorbase [34] (Table 1). Furthermore, the A. gambiae APN cDNA and protein sequences were searched with NCBI. Data consist of various types of APN, CDS length, protein length, orientation, and chromosomal location, and the number of exons for APN genes was extracted out for further analysis. Each of the 18 paralogous APN genes contains a variable number of exons and introns (Table 2). The lengths of the intron regions varied from 0 to 9 among these 18 APN genes (Fig. 1). APN paralog with accession no. AGAP0013001 had the maximum number of introns while AGAP003926 is intronless. The result indicates a large number of APN isomers are present in A. gambiae mosquitoes.

3.2. Genomic localization of APNs

The APNs were distributed over two chromosomes (X and chromosome 2) of A. gambiae. No APN genes are found to be localized on the third chromosome and at telomere positions. Most of these sequences were found in a cluster on the 2R arm of the chromosome (Table 1). The previous studies identified four types of APN, i.e., APN2, APN3, APN4, and APN5 had been found as a gene cluster located on chromosome 2R [10]. APN2, APN3, and APN4 genes along with four other APNs were found organized...
Table 1: Detail of paralogous APN genes of *A. gambiae* retrieved from the Vectorbase and NCBI database. The table covers genomic organization, vectorbase ID, gene orientation, chromosomal location, type of APN, number of nucleotides, exon-intron, ORF region, number of amino acids, Molecular weight, transmembrane domain, isoelectric point and signal peptide.

| S. no | Paralogous gene sequence | NCBI Accession No. | Name of protein | Orientation | Chromosome location | Position | No. of nucleotides (bp) | Start codon | Stop codon | ORF (bp) | Transmembrane Helix | Signal peptide | Molecular Wt (KDa) | PI | No. of exons and Introns | No. of amino acids | Ref. |
|-------|-------------------------|-------------------|----------------|-------------|---------------------|----------|------------------------|------------|------------|----------|----------------------|----------------|------------------|-----|------------------------|------------------|-----|
| 1     | AGAP013001 XM_003436526 | APN Reverse       | 2R             | 42,09,339–42,10,400 | 3,802 | ATG | TAA | 3,216 | 1 (36–58) | - | 122.4 | 5.37 | 10 | 9 | 1,071 | [34] |
| 2     | AGAP013146 XM_003436520 | APN 5 Forward     | 2R             | 42,05,810–42,06,317 | 3,209 | ATG | TGA | 2,826 | 1 (917–939) | 1–25 | 5.49 | 106.9 | 7 | 6 | 941 | [34,21] |
| 3     | AGAP013188 XM_003436525 | APN 4 Forward     | 2R             | 442,04,916–442,09,276 | 3,169 | ATG | TAG | 2,808 | 2 (7–26, 890–912) | 1–23 | 5.48 | 106.5 | 6 | 5 | 935 | [34,21] |
| 4     | AGAP013255 XM_003436524 | APN 3 Forward     | 2R             | 42,08,785–42,09,324 | 3,166 | ATG | TGA | 2,808 | 2 (7–26, 911–933) | 1–23 | 5.34 | 106.4 | 6 | 5 | 935 | [34,21] |
| 5     | AGAP013393 XM_003436523 | APN2 Forward      | 2R             | 42,07,019–42,07,346 | 3,170 | ATG | TAG | 2,757 | 1–26, 911–933 | 1–23 | 5.41 | 104.2 | 6 | 5 | 920 | [34,21] |
| 6     | AGAP003692 XM_313476    | APN Reverse       | 2L             | 3,79,024–3,79,032 | 3,322 | ATG | TAA | 2,763 | 0 | 1–23 | 4.86 | 105.6 | 6 | 5 | 920 | [34,21] |
| 7     | AGAP003690 XM_318000    | APN 1 Reverse     | 2L             | 3,79,35,54–3,79,37,232 | 3,588 | ATG | TAA | 3,063 | 0 | 1–19 | 5.06 | 113.2 | 5 | 4 | 1020 | [34,13] |
| 8     | AGAP004808 XM_554544    | APN Reverse       | 2L             | 3,79,242–3,79,236 | 2,926 | ATG | TAA | 2,604 | 0 | 1–23 | 5.20 | 97.8 | 5 | 4 | 867 | [34] |
| 9     | AGAP004808 XM_554544    | APN Reverse       | 2L             | 3,79,242–3,79,167 | 2,926 | ATG | TAA | 2,604 | 0 | 1–23 | 5.20 | 97.8 | 5 | 4 | 867 | [34] |
| 10    | AGAP003692 XM_562862    | APN Forward       | 2R             | 4,02,35,80–4,02,35,81 | 3,322 | ATG | TAA | 2,763 | 0 | 1–23 | 4.86 | 105.6 | 6 | 5 | 920 | [34,21] |
| 11    | AGAP004860 XM_514331    | APN Reverse       | 2R             | 4,50,85,39–4,51,03,59 | 3,077 | ATG | TAA | 2,748 | 0 | 1–27 | 5.39 | 104.1 | 5 | 4 | 915 | [34] |
| 12    | AGAP000885 XM_316862    | APN Reverse       | X              | 16,72,033–16,73,074 | 3,782 | ATG | TAG | 2,862 | 1 | 7–29 | 5.41 | 108.9 | 4 | 3 | 955 | [34] |
| 13    | AGAP013150 XM_00343651 | APN 2 Reverse     | 2R             | 48,72,171–48,73,237 | 3,067 | ATG | TAA | 2,820 | 0 | 1–20 | 5.05 | 106.5 | 4 | 3 | 939 | [34] |
| 14    | AGAP013155 XM_00343591 | APN Forward       | 2R             | 11,83,0,120–11,83,1,167 | 2,984 | ATG | TAA | 2,889 | 0 | 1–19 | 5.48 | 107.5 | 5 | 2 | 962 | [34] |
| 15    | AGAP001881 XM_001689265 | APN Forward       | 2R             | 11,83,3,852–11,83,3,198 | 3,262 | ATG | TGA | 2,874 | 0 | 1–39 | 5.37 | 107.2 | 2 | 1 | 957 | [34] |
| 16    | AGAP012984 XM_00343591 | APN 1 Reverse     | 2R             | 11,82,6,70–11,82,6,118 | 2,925 | ATG | TAA | 2,847 | 1 | 12–34 | 5.73 | 107.7 | 2 | 1 | 948 | [34] |
| 17    | AGAP012757 XM_301714    | APN Forward       | Unknown        | 27,43,282–27,43,6,76 | 2,431 | ATG | TTG | 1,964 | 0 | 1–20 | 5.76 | 658(Partial) | 4 | 3 | 955 | [34] |
| 18    | AGAP003926 XM_318379    | APN Forward       | 2R             | 46,38,1,954–46,36,584 | 2,631 | ATG | TAA | 2,631 | 0 | 1–19 | 5.71 | 99.0 | 5 | 4 | 876 | [34] |
Table 2: Structure of APN genes from *A. gambiae* species. E and I with number represents corresponding exon and intron numbers of the gene. 5' and 3' UTR regions are excluded.

| S. no | Gene ID   | E1    | I1    | E2    | I2    | E3    | I3    | E4    | I4    | E5    | I5    | E6    | I6    | E7    | I7    | E8    | I8    | E9    | I9    | E10   |
|-------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1     | AGAP003926| 2631  |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 2     | AGAP013155| 1417  | 85    | 1546  |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 3     | AGAP012984| 1043  | 78    | 1804  |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 4     | AGAP001881| 1037  | 388   | 1837  |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 5     | AGAP013150| 680   | 72    | 369   | 88    | 1651  | 87    | 186   |       |       |       |       |       |       |       |       |       |       |       |       |
| 6     | AGAP000885| 15    | 1240  | 200   | 77    | 2402  | 843   | 268   |       |       |       |       |       |       |       |       |       |       |       |       |
| 7     | AGAP004808| 713   | 105   | 502   | 63    | 239   | 76    | 655   | 78    | 495   |       |       |       |       |       |       |       |       |       |       |
| 8     | AGAP002508| 677   | 69    | 295   | 65    | 352   | 112   | 1255  | 76    | 595   |       |       |       |       |       |       |       |       |       |       |
| 9     | AGAP004809| 572   | 110   | 450   | 270   | 341   | 82    | 1328  | 63    | 526   |       |       |       |       |       |       |       |       |       |       |
| 10    | AGAP012757| 488   | 250   | 162   | 56    | 121   | 91    | 351   | 66    | 846   |       |       |       |       |       |       |       |       |       |       |
| 11    | AGAP004860| 491   | 79    | 649   | 69    | 239   | 92    | 1152  | 89    | 285   |       |       |       |       |       |       |       |       |       |       |
| 12    | AGAP003695| 488   | 147   | 147   | 75    | 133   | 74    | 354   | 80    | 1383  | 183   | 258   |       |       |       |       |       |       |       |       |
| 13    | AGAP003692| 488   | 250   | 147   | 71    | 133   | 83    | 462   | 139   | 1305  | 88    | 318   |       |       |       |       |       |       |       |       |
| 14    | AGAP001318| 677   | 80    | 1333  | 82    | 191   | 67    | 166   | 61    | 140   | 71    | 397   |       |       |       |       |       |       |       |       |
| 15    | AGAP013255| 677   | 80    | 1333  | 82    | 191   | 67    | 166   | 61    | 140   | 68    | 459   |       |       |       |       |       |       |       |       |
| 16    | AGAP013146| 689   | 85    | 1333  | 76    | 191   | 79    | 166   | 71    | 140   | 72    | 450   |       |       |       |       |       |       |       |       |
| 17    | AGAP013393| 677   | 80    | 533   | 64    | 737   | 82    | 191   | 67    | 166   | 61    | 140   | 71    | 397   |       |       |       |       |       |       |
| 18    | AGAP013001| 905   | 80    | 159   | 49    | 398   | 66    | 344   | 64    | 400   | 65    | 281   | 80    | 200   | 86    | 312   | 59    | 82    | 97    | 335   |

**Figure 1**: Representation of APN and their paralogous on X, 2L, and 2R arm of chromosomes. All APNs represented by their geneId and the arrow below this Id indicate forward and reverse orientations of APN gene. Integers at the right side of each bar locate arm name. The scale at the top is in megabases (Mb) and at the end of each arm size in Mb is also mentioned.
on a single locus in the forward orientation, suggesting their 
co-evolution or duplication from a single gene. These clusters 
may have a similar function or co-expression in the A. gambiae 
genome. APN1 are present on a different chromosome (2L) and 
are organized in the reverse orientation (Table 1). Also, most of 
the APN genes were located in forward orientation over the 2R arm 
of the chromosome. The chromosomal location and orientation of 
APNs are summarized in Figure 1 for easy readability. Clustering 
of the paralogous genes suggests facilitated functional regulation 
and have been reported in several organisms [35]. The gene 
conversion or indirect selective forces are the usual mechanisms 
for coincidental evolution and clustering of paralogs.

3.3. Conserved domain identification

Protein BLAST and MSA was performed using all these APN 
sequences for the known conserved regions. APN proteins 
characterized by the presence of three domains, aminopeptidase 
N like, N terminal (IPR042097), Pfam (PF01433) M1 Peptidase 
(IPR014782), and Pfam (11838) ERAP1-like C-terminal domain 
(IPR024571) are found in all APNs which is in accordance with 
the previous studies made in insects and other organisms [36,37]. 
These domains showed homology to Homo sapiens endoplasmic 
reticulum aminopeptidase 1 isoform b precursor and human 
leukotriene A4 hydrolase (PDB Id 2R59). CBB, Pfam, and InterPro 
also carried out for confirmation of results. GAMEN (300-350 aa), 
HEXXH (350-390 aa), and WWDV AWLNEGFA (360-400 aa) are 
highly conserved and they may act as catalytic and zinc-binding 
sites for peptides (Fig. 2A). The amino acid residues involved 
in the APN catalytic site of human APN include the amino acid 
motif GAMEN (residues 352-356) of the human APN [38]. This 
motif was completely conserved in Anopheles paralogous except 
for AGAP004808, which have AAMEN. The two H residues in

![Figure 2: (A) Amino acid sequence alignment of various paralogous AgAPNs. The conserved gluzincin APN motif represents in box from 320 to 340 aa residues. The zinc-binding conserved sites are shown in box from 335 to 375 aa residues. (B) The APNs HMM logo based on alignment of all paralogous APNs. (C) Showing number of glycosylation and phosphorylation sites in different types of APNs.](image)
HEXXH act as zinc ion binding sites and E residue is the catalytic site found in all APNs. APNs molecular weight varied from 97 kDa to 220 kDa. There was a slightly low variation in the PI values, ranging from 4.5 to 5.7. Sequence logo was generated to explore the sequence conservation of sequences of APN motifs in *A. gambiae* (Fig. 2B). Analysis of the predicted phosphorylation and glycosylation sites of APN sequences carried out by using the PROSITE program. It shows a maximum of 14 glycosylation are present in AGAP002508 (Fig. 2C). The previous studies show that three glycosylation sites are present in APN2, APN3, and APN4 [21] and one in APN1 [16]. Analysis of phosphorylation sites in these sequences shows that the number lies between 20 and 35. AGAP013001 has the maximum number, and APN5 is found to have the minimum number of phosphorylation sites.

### 3.4. Protein-protein interaction (PPI)

By using the STRING database, we mined the PPIs for 18 APNs protein based on the experiment, protein homolog, database evidence, text mining, co-expression, and co-occurrence available at STRING database. The search includes all these parameters to explore the PPI data thoroughly. The extracted analysis ranged from average confidence scores (CI-0.4 ≤ S ≤0.7). PPI data revealed that five APNs, APN1, AGAP000885, AGAP013001, AGAP003695, and AGAP002508 possessed more than four interacting partners in the network and these APNs were associated with each other (Fig. 3A). This group of proteins comprising nine more proteins and were found to having a strong connection with each member having 10–11 interacting partner. Also, enrichment and topological information of APNs system input were also analyzed by STRING. The statistics of the network revealed that this intercome comprised of 55 edges and 11 nodes. The average local clustering coefficient and node degree of the network were found to be 0.982 and 9.82. The majority of these *A. gambiae* system proteins were significantly enriched mainly in the biological process associated with glutathione synthase protein involved in glutathione metabolism. In Cytoscape plugin, cytohubba was used to explore the hub among above said five APNs from STRING TA system PPI network (Fig. 3B). PPI network determined some highly connected nodes that correspond with specific biological functions called Hub proteins. These hub proteins show evolutionary conservation than non-hubs. So, removal of one of protein from this hub may lead

![Figure 3](#)
to disruption of the network, and thus this can be considered as attractive blocking targets to reduce malaria. APN4 (AGAP013188) shows 11 nodes and 19 edges associated with cadherin protein (experimentally determined) and also co-expressed with alkaline phosphatases (Fig. 3C). The average local clustering coefficient and node degree of the network were found to be 0.855 and 3.45. On the other hand, remaining 12 types of APNs are connected separately with different types of interacting partners, i.e., alkaline phosphatase by 10 node and 11 edges, these are not showing any type of network between them (Fig. 3D). In Lepidoptera, APNs, along with alkaline phosphatases, play a major role in inserting Cry1A toxin oligomers into the cell membrane [21]. In A. aegypti, it is reported that alkaline phosphatase acts as a functional receptor for Cry 11Aa toxin [39]. Anopheles gambiae APN2 also reported to have a receptor for Cry 11Ba [40,41]. So, PPIs network with APN2 and other APNs with alkaline phosphatase play a role in Cry toxin binding.

3.5. Targeting and pathway study

This study showed that different types of APN proteins contained 900 amino acids to 1,060, at the N-terminal region, where 22–25 aa acts as a signal peptide. However, two types of APNs are without signal peptide (Table 1). Experimental evidence shows that most of the APNs were localized to the microvilli in the posterior midgut. Subcellular localization prediction suggests that all APNs were localized in the plasma membrane, so signal peptide is required for Membrane trafficking from Endoplasmic reticulum (ER) to Golgi transport, a forward secretory pathway in the form of Exosomal proteins of microglial cells. KEGG results show that five types of APN proteins including APN1, AGAP000885, AGAP013001, AGAP003695, and AGAP002508 act as a hub connected to glutathione synthase (1271707 in Fig. 3) may involve together in metabolic pathways for metabolism of Glutathione (Fig. 4). APNs of mosquitoes are attached to

![Figure 4: APN Metabolic pathway produced by KEGG, red color denotes the site where APN (APN1, AGAP000885, AGAP013001, AGAP003695, and AGAP002508) proteins are utilized to carry out glutathione metabolism.](image-url)
the midgut surface where they are preferentially localized in cholesterol-rich membrane rafts. Recent evidence suggests that lipid rafts may be an essential component of ookinete invasion of the midgut epithelium. Mosquito-based TBI antigens Anopheles alanyl aminopeptidase N (AnAPN1) [13] known ookinete-interacting proteins were found associated with apical midgut-microvilli detergent-resistant membranes (DRM), which are enriched in lipid rafts, it is reported that other paralogous APNs AGAP003695, AGAP003926, AGAP013146, AGAP012757, AGAP002692, AGAP001881, AGAP013393, AGAP013188, AGAP013255, and AGAP04860 were also present in DRM [42]. APNs of Lepidoptera are anchored to the midgut surface where they are preferentially confined in cholesterol-rich membrane rafts. A complex set of events that require lipid raft integrity, lateral mobility, and partitioning on midgut surfaces has been shown to be essential for Cry1a toxin insertion and pore formation in Heliothis virescens and Manduca sexta [43].

3.6. Expression profiles of APNs at various tissue level

Data are available at MozAtlas (http://mozatlas.gen.cam.ac.uk) on the expression of paralogous APN genes concerning different target tissues, i.e., midgut, malpighian tubules, head, testis, ovary, carcass, and salivary glands. These are collected from EST library and it shows that different types of APN genes were expressed in various tissues [33]. AGAP001881 and AGAP013155 only expressed in midgut tissue. From APN1 to APN5, AGAP012745, AGAP003692, and AGAP012757 showed their highest expression level in the midgut with some expression in the malpighian tubules (Table 3). However, AGAP003695 showed lower expression in the midgut and much higher expression in the malpighian tubules. APN5 showed expression in the midgut, carcass, and the malpighian tubules with very small expression in the head, testis, and ovary. In fact, AGAP013001 shows their expression in carcass and head. AGAP000885 expressed in carcass only. AGAP004808 shows most of their expression in the testis and less expression in the midgut. However, AGAP002508 is expressed in the midgut and the ovary. Also, the expression level of AGAP04860 was observed in the midgut, malpighian tubules, and salivary gland with less expression in the head and the testis. Among all these APNs, most of their expressions were seen in the midgut. It was interesting to note that most of APNs were not expressed in the salivary gland, head, ovarian and testis tissue.

3.7. Phylogenetic analysis

The phylogenetic construction revealed that each branch of tree evolves independently from the other, showing different paralogous APN gene sequences (Fig. 5). Phylogenetic analysis of A. gambiae genome showed a distinct organization of APN gene family. The phylogenetic tree opens two major clusters, including 15 APNs, which are in one group and three APNs in the second cluster. A clade formed by APN2, APN3, and APN4 indicates that these genes are closely evolved [21, 41 and 45]. APN1 is out of the above-said clade and showed that the evolution of these genes occurs independently from their common ancestors. However, APN3 acts as a sister-group to APN2 and APN4, with 100% bootstrap support (Fig. 5). APN5 branched off next, clustering with APN4, APN3, and APN2. Among these five APN genes, APN2, APN3, APN4, and APN5 formed separate single clusters revealing their similar evolution as compared to APN1 genes which formed a different group.

Table 3: Expression pattern of A. gambiae APN in different tissues based on data obtained from MozAtlas.

| Gene ID          | Carcass | Head | Salivary gland | Midgut | Malpighian | Ovary | Testis | Unknown |
|------------------|---------|------|----------------|--------|------------|-------|--------|---------|
| AGAP004809 (APN1) | 0       | 0    | 0.7%           | 93.8%  | 9%         | 0     | 0      | 0       |
| AGAP013393 (APN2) | 0.1%    | 0    | 0.8%           | 91.5%  | 7.5%       | 0     | 0.1%   | 0       |
| AGAP013255 (APN3) | 0       | 0    | 0              | 94%    | 6%         | 0     | 0      | 0       |
| AGAP013146 (APN4) | 4.8%    | 0.2% | 0.5%           | 73%    | 20%        | 0.2%  | 0.1%   | 1.2%    |
| AGAP013188 (APN5) | 0.1%    | 0    | 0.8%           | 91.5%  | 7.5%       | 0     | 0.1%   | 0       |
| AGAP012745        | 0       | 0    | 0              | 94%    | 6%         | 0     | 0      | 0       |
| AGAP001881        | 0       | 0    | 0              | 100%   | 0          | 0     | 0      | 0       |
| AGAP012984        | NA      | NA   | NA             | NA     | NA         | NA    | NA     | NA      |
| AGAP013150        | 0       | 0    | 0              | 14%    | 0          | 0     | 0      | 86%     |
| AGAP013155        | 0       | 0    | 0              | 100%   | 0          | 0     | 0      | 0       |
| AGAP003695        | 0       | 0    | 1.1%           | 29.3%  | 69.6%      | 0     | 0      | 0       |
| AGAP003692        | 0       | 0    | 1.1%           | 93%    | 5.9%       | 0     | 0.2%   | 0       |
| AGAP012757        | 0       | 0    | 0              | 66%    | 34%        | 0     | 0      | 0       |
| AGAP013001        | 58%     | 42%  | 0              | 0      | 0          | 0     | 0      | 0       |
| AGAP000885        | 100     | 0    | 0              | 0      | 0          | 0     | 0      | 0       |
| AGAP002508        | 0       | 0    | 0              | 97%    | 0          | 0     | 3%     | 0       |
| AGAP004808        | 0       | 0    | 0              | 23.5%  | 0          | 0     | 76.5%  | 0       |
| AGAP004860        | 1.3%    | 0    | 3.2%           | 80.2%  | 14%        | 0     | 0.9%   | 0.2%    |
4. CONCLUSION

Through genome-wide analysis of the APN paralogous genes in *A. gambiae*, covered insilico study on gene structure, location on chromosome, protein properties, functional motifs analysis, protein-protein interaction study, metabolic pathway, and phylogeny. Previous studies are limited to the function of only five types of APNs (APN1-5). Further functional characterization and expression analysis of other APNs will be necessary to get more thoughtful understanding of the evolution of the *A. gambiae* APN gene families. TBVs promise a more efficient way to malaria control. The previous studies have identified an epitope from *A. gambiae* APN1 is highly immunogenic. But, there is a weakness due to improper folding of the protein in addition to transmission irrelevant epitopes dilute the production of functional antibodies. In this study, some proteins are identified as hub proteins and may express together during ookinete invasion. So, targeting these hub APNs or other protein from a protein-protein interaction network may also be able to induce more antibody response that significantly inhibits parasite development and may be more effective to control malaria.

ACKNOWLEDGMENTS

The author Renu Jakhar highly acknowledges the financial assistance and support provided by DBT-IPLS, Govt. of India. The authors also would like to thank Arvind Sharma at the University of Nevada, Reno (USA) for valuable comments and suggestions.

AUTHOR’S CONTRIBUTION

Renu Jakhar conducted the study, performed in silico analysis and wrote the manuscript. S.K. Gakhar planned the study. Ritu Gill revised the manuscript.

CONFLICT OF INTERESTS

There is no conflict of interests.

REFERENCES

1. Singh N, Shukla M, Chand G, Barde PV, Singh MP. Vector-borne diseases in central India, with reference to malaria, filaria, dengue and chikungunya. WHO South-East Asia J Public Health 2014;3(1):28.
2. Raghavendra K, Velamuri PS, Verma V, Elamathi N, Barik TK, Bhatt RM et al. Temporo-spatial distribution of insecticide-resistance in Indian malaria vectors in the last quarter-century: need for regular resistance monitoring and management. J Vector Borne Dis 2017;54(2):111.
3. Piemmarini P, Esquivel C, Denton J. Malpighian tubules as novel targets for mosquito control. Int J Environ Res Public Health 2017;14(2):111.
4. Wangdi K, Gatto ML, Kelly GC, Banwell C, Dev V, Clements AC. Malaria elimination in India and regional implications. Lancet Infect Dis 2016;16(10):e214–24.
5. Bokharaei H, Raz A, Zakeri S, Djadid ND. 3'-RACE Amplification of aminopeptidase N gene from *Anopheles stephensi* applicable in transmission blocking vaccines. Avicenna J Med Biotechnol 2012;4(3):131–41.
6. Rosenfeld A, Vanderberg JP. *Plasmodium berghei*: induction of aminopeptidase in malaria resistant strain of *Anopheles gambiae*. Exp Parasitol 1999;93(2):101–4.
7. Gill SS, Cowles. EA, Francis V. Identification, isolation, and cloning of a *Bacillus thuringiensis* Cry1Ac toxin-binding protein from the midgut of the lepidopteran insect *Heliothis virescens*. J Biol Chem 1995;270:27277–82.
8. Knight PJ, Knowles. BH, Ellar. DJ. Molecular cloning of an insect aminopeptidase N that serves as a receptor for *Bacillus thuringiensis* Cry1Ac toxin. J Biol Chem 1995;270:17765–70.
9. Chen, Aimanova, S.Pan, S.S.Gill. Identification and characterization of *Aedes aegypti* aminopeptidase N as a putative receptor of *Bacillus thuringiensis* Cry11A toxin. Insect Biochem Mol Biol 2009;39:688–96.
10. Dinglasan RR, Jacobs-Lorena M. Flipping the paradigm on malaria transmission blocking vaccines. Trends Parasitol 2008;24:364–70.
11. Dinglasan RR, Fields I, Shahabuddin M, Azad AF, Sacci JB Jr. Monoclonal antibody MG96 completely blocks Plasmodium yoelii development in *Anopheles stephensi*. Infect Immun 2003;71(12):6995–7001.
12. Dinglasan RR, Valenzuela JG, Azad AF. Sugar epitopes as potential universal disease transmission blocking targets. Insect Biochem Mol Biol 2005;35(1):1–10.

Figure 5: Showing phylogenetic analysis of different types of APNs with their Accession No.
13. Dinglasan RR, Kalume DE, Kanzok SM, Ghosh AK, Muratova O, Pandey A. Disruption of Plasmodium falciparum development by antibodies against a conserved mosquito Midgut antigen. Proc Natl Acad Sci USA 2007;104(33):13461–6.

14. Mathias DK, Flieskatt JL, Arnimstead JS, Bethony JM, Abdul-Majid KB, McMillan A, et al. Expression, immunogenicity, histopathology, and potency of a mosquito-based malaria transmission-blocking recombinant vaccine. Infect Immun 2012;80(4):1606–14.

15. Armistead JS, Morlais I, Mathias DK, Jardim JG, Joy J, Fridman A, et al. Antibodies to a single, conserved epitope in Anopheles APN1 inhibit universal transmission of falciparum and vivax malaria. Infect Immun 2013;82(2):818–29.

16. Atkinson SC, Armimstead JS, Mathias DK, Sandeu MM, Tao D, Borhani-Dizaji N, et al. Structural analysis of Anopheles midgut aminopeptidase N reveals a novel malaria transmission-blocking vaccine B-cell epitope. Nat Struct Mol Biol 2015;22(7):532.

17. Shabani S, Karimi A, Mahboudi F, Poursarif A, Djadid ND, Motalleb G, et al. Molecular identification and characterization of aminopeptidase N1 from Anopheles stephensi: A candidate for transmission blocking vaccines. Gene Rep 2016;5:157–62.

18. Ruzzante L, Reijnders MJM, Waterhouse RM. Of genes and genomes: mosquito evolution and diversity. Trends Parasitol 2019;35(1):32–51.

19. Chaerkady R, Kelkar DS, Muthusamy B, Kandasamy K, Dvivedi SB, Sahasrabuddhe NA, et al. A proteogenomic analysis of Anopheles gambiae using high-resolution Fourier transform mass spectrometry. Genome Res 2011;21(11):1872–81.

20. Pootanakit K, Angsuthanasombat C, Panyim S. Identification of two isoforms of aminopeptidase N in Aedes aegypti larval midgut. J BiochemMol Biol 2003;36(5):508–13.

21. Zhang R, Hua G, Andacht TM, Adang MJ. A 106-kDa aminopeptidase is a putative receptor for Bacillus thuringiensis Cry11Ba toxin in the mosquito Anopheles gambiae. Biochemistry 2008;47:11263–72.

22. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. OSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 2015;31(8):1296–7; doi: 10.1093/nar/gks1067.

23. Marchler BA, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 2016;45(D1):D200–3.

24. Schaeffer RD, Liao Y, Cheng H, Grishin NV. ECOD: new and continuing developments at PROSITE. Phylogenet Evol 2014;76:127–33.

25. Wang AH, Zhou D, Rini JM. The X-ray crystal structure of human aminopeptidase N reveals a novel dimmer and the basis for peptide processing. J Biol Chem 2012;287:36804–13.

26. Fernandez LE, Aimanova KG, Gill SS, Bravo A, Soberon M. A GPI-anchored alkaline phosphatase is a functional midgut receptor of Cry11Aa toxin in Aedes aegypti larvae. Biochem J 2006;394:77–84.

27. Zhang R, Hua G, Urbauer JL, Adang MJ. Synergistic and inhibitory effects of aminopeptidase peptides on Bacillus thuringiensis Cry11Ba toxicity in the mosquito Anopheles gambiae. Biochemistry 2010;49:8512–919.

28. Zhang R, Hua G, Urbauer JL, Adang MJ. Synergistic and inhibitory effects of aminopeptidase peptides on Bacillus thuringiensis Cry11Ba toxicity in the mosquito Anopheles gambiae. Biochemistry 2010;49:8512–919.

29. Zhang R, Hua G, Urbauer JL, Adang MJ. Synergistic and inhibitory effects of aminopeptidase peptides on Bacillus thuringiensis Cry11Ba toxicity in the mosquito Anopheles gambiae. Biochemistry 2010;49:8512–919.

30. Marlin EE, Hughes AL. Evolutionary diversification of aminopeptidase N in Lepidoptera by conserved clade-specific amino acid residues. Mol Phylogenet Evol 2014;76:127–33.

31. Dong CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY, cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;(4):S11.

32. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 2013;30:2725–9.

33. Baker DA, Nolan T, Fischer B, Pinder A, Crisanti A, Russell S. A comprehensive gene expression atlas of sex-and tissue-specificity in the malaria vector, Anopheles gambiae. BMC Genomics 2011;(1):296.

34. Mengin E, Louis C, Holt RA, Birney E, Collins FH. The Anopheles gambiae genome: an update. Trends Parasitol 2004;20(2):49–52.

35. Deng XJ. Gene expression divergence and evolutionary analysis of the drosomycin gene family in Drosophila melanogaster. J Biomed Biotechnol 2009;Article ID 315423:9.

36. Agrawal N, Malhotra P, Bhattachar JK. Interaction of gene-cloned and insect cell-expressed aminopeptidase N of Spodoptera litura with insecticidal crystal protein CryIc. Appl Environ Microbiol 2002;68(9):4583–92.

37. Hughes AL. Evolutionary diversification of aminopeptidase N in Lepidoptera by conserved clade-specific amino acid residues. Mol Phylogenet Evol 2014;76:127–33.

38. Wang AH, Zhou D, Rini JM. The X-ray crystal structure of human aminopeptidase N reveals a novel dimmer and the basis for peptide processing. J Biol Chem 2012;287:36804–13.

39. Fernandez LE, Aimanova KG, Gill SS, Bravo A, Soberon M. A GPI-anchored alkaline phosphatase is a functional midgut receptor of Cry11Aa toxin in Aedes aegypti larvae. Biochem J 2006;394:77–84.

40. Zhang R, Hua G, Urbauer JL, Adang MJ. Synergistic and inhibitory effects of aminopeptidase peptides on Bacillus thuringiensis Cry11Ba toxicity in the mosquito Anopheles gambiae. Biochemistry 2010;49:8512–919.

41. Zhang R, Hua G, Urbauer JL, Adang MJ. Synergistic and inhibitory effects of aminopeptidase peptides on Bacillus thuringiensis Cry11Ba toxicity in the mosquito Anopheles gambiae. Biochemistry 2010;49:8512–919.

42. Zhang R, Hua G, Urbauer JL, Adang MJ. Synergistic and inhibitory effects of aminopeptidase peptides on Bacillus thuringiensis Cry11Ba toxicity in the mosquito Anopheles gambiae. Biochemistry 2010;49:8512–919.

43. Zhang R, Hua G, Urbauer JL, Adang MJ. Synergistic and inhibitory effects of aminopeptidase peptides on Bacillus thuringiensis Cry11Ba toxicity in the mosquito Anopheles gambiae. Biochemistry 2010;49:8512–919.

44. Hughes AL. Evolutionary diversification of aminopeptidase N in Lepidoptera by conserved clade-specific amino acid residues. Mol Phylogenet Evol 2014;76:127–33.

How to cite this article:
Jakhar, Gill, and Gakhar. Genome-wide analysis of various paralogs of aminopeptidase N (APN) gene in Anopheles gambiae (Diptera: Culicidae). J Appl Biol Biotech 2019;7(06):15–24. DOI: 10.7324/JABB.2019.70603.