The G12 gene which encodes for cell surface protein thought to be located on the microvillar membranes of the midgut epithelial cells; was studied in *Aedesaegypti* and named as *AEG12*. Due to its presence on the surface of microvilli it might play a role in receptor-ligand interactions when malarial parasite invades the mosquito midgut. The transcripts were reported significantly up regulated at 12 h after feeding on blood infected with *Plasmodium gallinaecum* (Shao *et al*. 2005).The analyzed amino acid sequence of *AEG12* suggested that it might be associated with digestive function but exact function is still not known due to lack of sequence similarities with other sequences present in all major databases. The similar protein arbitrarily named *ANG12* has also been studied in African malaria mosquito vector *Anopheles gambiae* (Diptera: Culicidae) sharing high degree of sequence similarity with *AEG12* pointing towards common function of both genes (Morlais *et al*. 2003). The G12 gene is similar to cockroach allergen, Bla g1 and Per a1 but in contrast these allergens contain multiple tandem amino acid repeats which shows that mosquitoes appeared later in evolution than cockroaches in which repeat allergen is conserved (Pomes *et al*. 1998). Nolan *et al*. 2011 used upstream region of *An. gambiae* G12 gene for driving transgene expression in *Anopheles stephensi* (Diptera: Culicidae) that was able to attack pre-sporogonic stages of *Plasmodium* parasite making vector refractory to malaria. However, *ANG12* gene was found massively induced at 24 hours post blood meal (PBM). This difference in expression profile might be due to difference in transciptional machinery between *Ae. aegypti* (Diptera: Culicidae) and *An. gambiae* (Diptera: Culicidae).

In India, *An. culicifacies* (Diptera: Culicidae) is the important vector of malaria responsible for 60-70 % of malaria cases throughout the country. *An. culicifacies* taxon exist as a complex of five isomorphic sibling species which are designated as A, B, C, D and E. Species A, B, C,D and E are the predominant vectors of malaria while species B has low vectorial capacity (Goswami *et al*. 2006). The blood inducible G12 gene has yet not been characterized in *An. culicifacies* in-spite of the fact that it is the principal vector in India. This manuscript describes the cloning and expression of *An. culicifacies*G12 gene.

MATERIALS AND METHODS

Mosquitoes

An. culicifacies A, mosquitoes were reared in an insectary constantly maintained at 28 ± 2º C and 70%-80% relative humidity with simulated dawn and dusk machine adjusted with a photoperiod of 14 h light and 10 hour dark. Adult mosquitoes were allowed to feed on 1% glucose soaked cotton pads as source of nourishment. For ovarian development four to five day old females were fed on rabbit blood. On the 3[rd]day post blood feeding females were allowed to lay eggs in water filled plastic

bowls lined with filter paper. After hatching larvae were reared in large enamel trays supplied with cat chow as food. The pupae were transferred to small plastic cups filled with water and kept in cloth cages for emergence to adult mosquitoes (Kumar *et al.* 2014).

DNA Isolation

Genomic DNA extraction involved homogenizing individual *An. culicifacies* female mosquito in bender buffer (0.1 M NaCl, 0.2 M sucrose, 0.1 M Tris, pH 9.0, 0.05 M EDTA, pH 8.0 0.5% SDS). The lysate was then treated with Rnase at 37ºC for 1 hr for RNA contamination removal. The sample was then treated with proteinase K at 50ºC overnight. DNA was then purified by standard phenol chloroform method (Sambrook & Russel 2001). The pellet was washed with 70% ethanol, air dried and redissolved in TE buffer and stored at -20ºC until use.

Polymerase Chain Reaction amplification and sequencing

Degenerate oligonucleotide primers corresponding to region of homology between *An. gambiae and Ae. aegypti* G12 gene was used for PCR. The primers and PCR amplification conditions used are mentioned in Table 1. Sequencing was performed with an ABI3730xl 96 capillary analyzer.

Table 1. List of successful Primers used in the study.

Sr. no.	Primer name	5'-3' sequence	Amplification conditions
1.	ACG12F2 ACG12R2	GGACAAATAGCATCGCCTATC CTGGACTTCTTGGTCGGTAAC	96°C,30s, 94°C,45s,64°C,30s,68°C,2 min, 68°C, 10 min, 35cycles
2.	ACG12F3 ACG12R3	ACGGACGACTTTGACGATTT CACTTCAGCATCTTCCAGCA	96°C,30s,94°C,45s,64°C,30s,68°C,1.3min, 68°C, 10 min, 35cycles

Dissection of mosquitoes and RNA extraction

Midguts to be used for RNA isolation were dissected at 0 h,2 h, 6 h, 12 h, 18 h, 24 h, 48 h, after blood feeding, and midguts from unfed females were dissected in DEPC treated water. For each experiment, a set of 20 midguts were homogenized in 200 ul of TriReagent (Sigma) and RNA was subsequently extracted, and stored at -80ºC until used for cDNA synthesis. Quantification of RNA was performed using ND2000c Nanodrop. cDNA was synthesised using Qiagen (QuantiTect Reverse Transcription Kit) according to manufacturer protocol using 1 ug of total RNA. Qrt-pcr was carried out using SYBR Green dye on Applied Biosystem Step One Real Time PCR system.

Sequence analysis

The ORF was deduced using NCBI ORF finder tool. The signal peptide was predicted using Signal P software (Petersen *et al*. 2011). Manual searching was done to identify TATA box and other transcription factors in upstream region of the sequence. BLAST searches were carried out to find sequences with similarity in the databases. Sequences for alignment and phlogenetic analysis were retrieved from the Genbank. Multiple sequence alignments were performed using Clustal W program. MEGA 6 software was used to analyse and establish phylogenetic relationship among the insect G12 genes with high similarity (Tamura *et al*. 2013). The 3D model of AcG12 protein sequence was obtained using Phyre2 server (Kelly *et al*. 2009). Conserved protein domains were identified using the Conserved domain database (CDD) (Marchler*et al*. 2015). The isoelectric point (pI) and molecular weight (MW) were calculated by Compute pI/MW (Bjellqvist *et al*. 1993). Transmembrane domain searches were performed by TMHMM software (Krogh *et al*. 2001). The potential for glycosylation sites and phosphorylation sites were obtained with the NetNGlyc 1.0 and NetPhos 2.0 program (Blom*et al*. 1999).

RESULTS

Isolation and structure of an *Anopheles culicifacies* G12 gene

Based on the conserved regions of G12 among various organisms, sets of primers were designed to isolate G12 gene of *An. culicifacies*. By this screening we were able to isolate G12 gene from *An. culicifacies* and was named *AcG12*. The resulting sequence was found consisting of 2753 bp (Genbank accession number KR011199) including a 500 bp upstream region followed by a630 bp open reading frame (ORF) and a 300 bp downstream region. The ORF was predicted to encode a 209 amino acids protein. Sequence analyses suggested that *AcG12* has 3 exons separated by 2 introns as outlined in Fig.1B.It is comparable in its general organization with the *An. gambiae* and *Ae. aegypti* G12 gene. The introns found in mosquito G12 gene varied in length. The first intron ranged from 60 bp (*Ae. aegypti*) to 80 bp (*An. gambiae*) and the second intron ranged from 67 bp (*Ae. aegypti*) to 75 bp (*An. gambiae*) (Fig. 1A).Sequence comparison by BLAST showed that this gene is similar to *An. gambiae* and *Ae. aegypti* G12 gene having 69% and 66% identity respectively.

The putative transcription start site starting with nucleotide A was found positioned at 32 bp upstream of the start codon ATG and was designated as +1. A unique TATA like sequence (TATAAAA) was found present at position -32 from transcription start site. The polyadenylation site was found present at 36 bp downstream of the stop codon.

```
cataggagaagaagatagttacagcaaaaaaaaaaatattcgggttttactactaaatgattcttagaaataaaatccgttttcattt    -380
tgaaagctagttgaatttgagtgggaaattgaatcacttatgaaatgcttaatatttcaatgaaaaagtataaatgaatgaagaata    -293
caattaatcattaaaacattatgccatgtcatggcgtgaaatctccatcaattttgattattttataatattttgattggttcggaagtca    -202
gtaaaaaaatgctggatttataattgataactgcgaggtatgttgtcaattactaacatgtccacttgaacatacgtaagtggcaaga    -114
taacattcaaaacaatgattaccgactttatcctgcttgttattcaaggcttaaatgtccataacaaaagaaaactagtaatataaaa    -26
gagaaagcgcatcaacattgaagtcatcagtagaaccataccaacaacacaagaaca    32
```

```
ATGAAACTCGCCGTTGTCGTGATCGCTTGCTTGGTAGCCACCGTCGCCTGTGCACCGA    90
 M  K  L  A  V  V  V  I  A  C  L  V  A  T  V  A  C  A  P (T)    20
```

```
CCCAACGTGCACTCACGGACGACTTTGACGATTTCGTAGCGTTACTTCCTTTGGAAGA    148
 Q  R  A  L  TD  D  F  D  D  F  V  A  L  L  P  L  E  D    39
```

```
TCTGCTGAACTTGGCCATGCGCTATCTCGTTACCGACCAAGAAGTCCAGCAAGCGTTG    206
 L  L  N  L  A  M  R  Y  L  V  TD  Q  E  V  Q  Q  A  L    58
```

```
CTTTATCTGCAAGGAGAAGAATTTGCTTCGGTTTGGGATCAGTTCTTCGCTCTTACGG    264
 L  Y  L  Q  G  E  E  F  A (S) V  W  D  Q  F  F  A  L  T  A    78
```

```
CAGTAAAAGATTTGTTGAACTATTTGGAGGCCGCTGGTGTGCCGGCATACGATTCATT    322
 V  K  D  L  L  N  Y  L  E  A  A  G  V  P  A (Y) D  S  L    97
```

```
GAATGCTGTCGCCGATTTCTTGGGCCTGACGCCTTTGAAACCAACCGTCCGAAGTTGT    380
 N  A  V  A  D  F  L  G  L  T  P  L  K  P  T  V  R (S) C    116
```

```
AAGTATACCCACAAAACATACATATACACACCCATACAAAATCAATTAgtgtgataccttca    442
 K (Y) T  H  K  T  Y  I  Y  T  P  I  Q  N  Q  L    132
```

```
ataatgacgtttcgagtagtgcgtacgggaggattgaacggactgcttgaagagGCTCTTGCCTTGTTGC    512
                                            A  L  A  L  L  P    138
```

```
CACATGAAGAACTGGAAGCCATGTTTGAAGAGAAGCTGAAGACTAGTCCCGAGTTCA    569
 H  E  E  L  E  A  M  F  E  E  K  L  K  T (S) P  E  F  K    157
```

```
AGGCTTTGTTCGAAAAACTCCAGAAGTTTGACCATAAGCAGCTGAGAGATCTATACG    626
 A  L  F  E  K  L  Q  K  F  D  H  K  Q  L  R  D  L (Y) E    176
```

```
AGgtaagtgtttactggaatcggccaaaagaaatgatttaatcgcacgtgactctttccgctcattgctcaacag    701
```

```
AATTCAGCTGAAGTACAAGGAATGATCCAAAAGCTGCGCGATCAAGGTGTTGATGTT    758
 N  S  A  E  V  Q  G  M  I  Q  K  L  R  D  Q  G  V  D  V    195
```

```
GATCACATAGTCCAGGTGGTAAAGGATTTCTTCGGATGGAACTAG    803
 D  H  I  V  Q  V  V  K  D  F  F  G  W  N*    209
```

```
aatgcttaggatttggcgtttggaaaaatagaagaataaataaaattataccaagtgaattgttagtcatattcattggataacaaaaatc
aaaaagaaaataaaacgaaaaaagcacaacattatgacgagaaagtttaaattatgtagagacaaagaagaaagagaaaatatatatt
acctctgattagtgaataatgaagaaataattaagaaattatattctgttggctatagaaaacttttttccggcatcatttttgtaatgattga
aagcatgtgtccattcagtttccacgt
```

Fig.1A Sequence of a genomic fragment containing the *AcG12 gene* (KR011199). The nucleotide coding sequence is shown above the putative translation product of the corresponding cDNA. The 5', 3'-UTRsand introns are in lowercase, while the two invariable nucleotides of the donor and acceptor sites are in bold characters. The putative TATA box, the translation initiation (ATG) and polyadenylation signal are underlined. Circles highlight potential N-linked phosphorylation sites. The asterisk shows the stop codon. Nucleotide and amino acid numbers are given on the right.
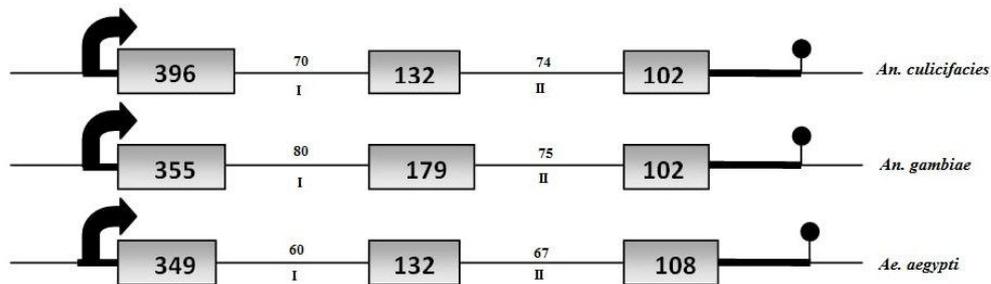
Fig.1.B structural comparison of the putative G12 genes from *An. culicifacies*, *An. gambiae* and *Ae. aegypti*. Shaded boxes represent exons and roman numerals refer to introns. Numbers express length in nucleotides. The transcription start point is shown by an arrow. Untranslated regions are represented in black and the polyadenylation site at the end of transcript is indicated by a dot on a vertical line.

Tissue and Stage specific expression

Expression profile of the *AcG12* was investigated by using quantitative real time PCR. RNA was isolated from a pool of 20 female midguts kept exclusively on a sugar diet or a similar pool collected at different time intervals after blood feeding. Tissue specificity was also assessed by extracting RNA from salivary glands, ovaries and carcasses. Very low expression was detected in guts dissected from sugar fed females. Peak expression of gene was recorded 24hours after the blood meal (Fig. 2) and thereafter transcripts level dropped continuously. No expression was detected in salivary glands, ovaries and carcasses indicating its strict midgut specificity and blood meal associated inducibility (Fig. 3).
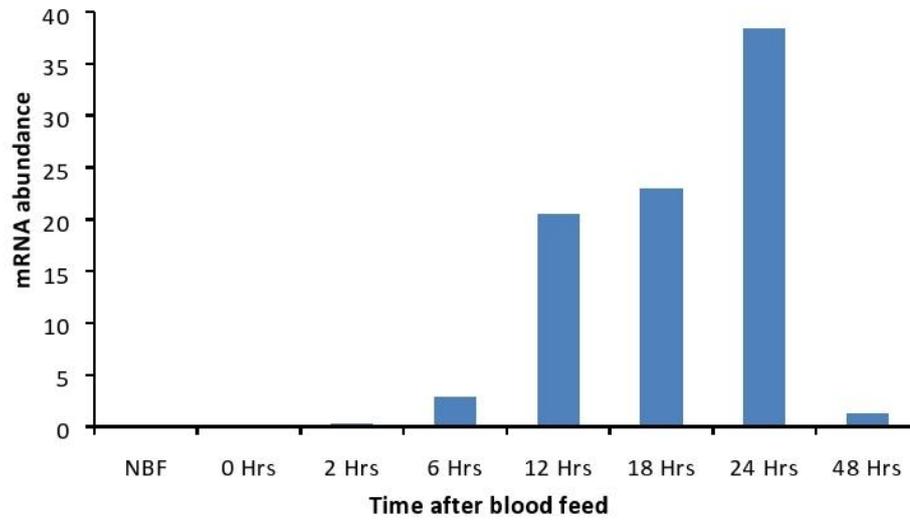
Fig 2: The expression of G12 in *An.culicifacies* female midgut at different time points after blood feeding using Real-Time PCR.
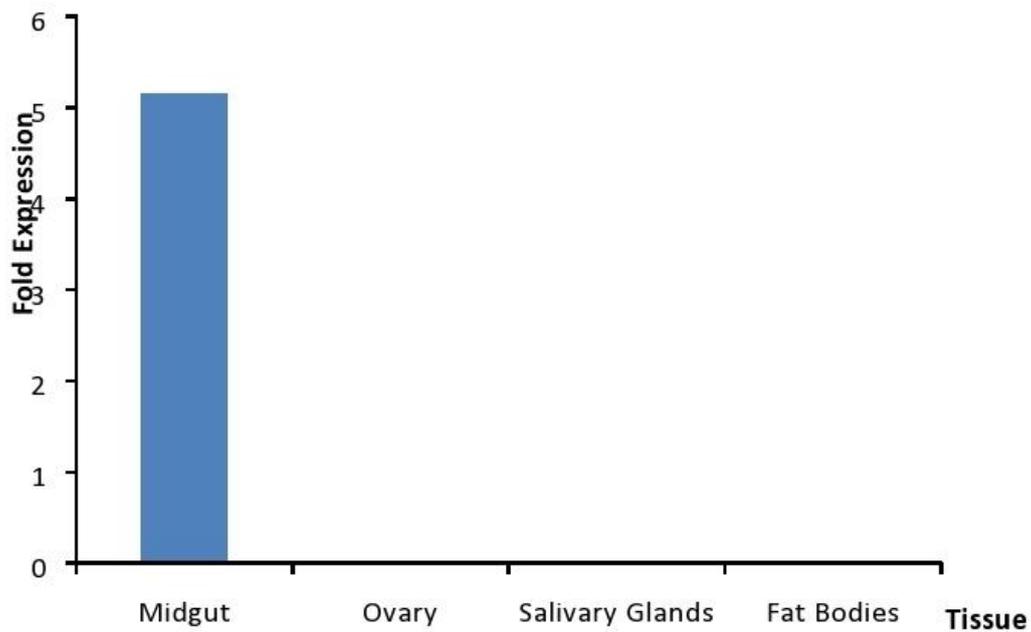

Fig.3: Tissue specificity of the *AcG12* expression. The mRNA was isolated from midguts, ovaries, salivary glands and fatbody and expression was analyzed using Real-Time PCR.

Analysis of predicted amino acid sequence

Analysis of deduced polypeptide predicted a protein of molecular mass of 23.7kDa having isoelectric point of 4.69. The first 17 amino acids of the amino terminal end were considered to form signal peptide and the remaining 191 amino acid constitute the mature protein. Search for the transmembrane domains and glycosylations sites revealed presence of no such sites in the predicted protein. But the search for phosphorylation sites revealed presence of 7 of such sites positioned at 20, 68,94, 115,118, 153 and175 which are thought to provide protection from proteolysis. *AcG12*was predicted having aninsect allergen repeat domain of 176 amino acids (Fig. 4; Pfam domain 06757 and Inter pro domain IPR010629) which is in accordance with the previous studies made in *An. gambiae* and *Ae. aegypti*. These domains showed 32% homology to cockroach allergens Bla g1 and Per a1 (Pomes *et al*.1998; Wu *et al*.1998);however the function of this domain is still not clear.
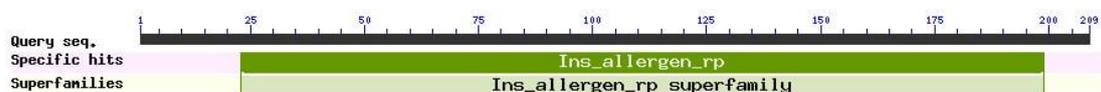

Fig.4A: Graphical summary of Insect allergen domain foundin *An. culicifacies*(*AcG12*) using Conserved domain database (CDD).
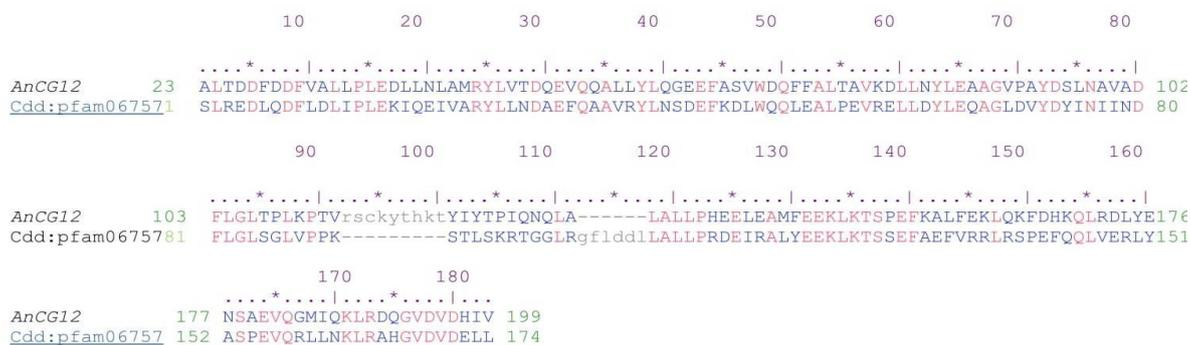

Fig.4B The comparsion of *AcG12* with superfamily member pfam06757 (a insect allergen related repeat). Identical residues are shown in red while mismatches are shown in blue.

The *AcG12* amino acid sequence was further compared with G12 sequences of other mosquitoes and cockroach allergen (Fig. 5). All of these sequences were found containing signal peptide indicating that the protein is secreted as zymogen and activated after cleavage releasing the mature protein. A unique feature of these amino acid sequences was the existence of allergen domain which strongly suggested an evolutionary relationship between mosquitoes and cockroaches. Based on homology modelling the three dimensional structure of protein was predicted using Phyre^2 program. The 85% of the residues were modeled with 100% confidence with major cockroach allergen bla g1 (Fig. 6).
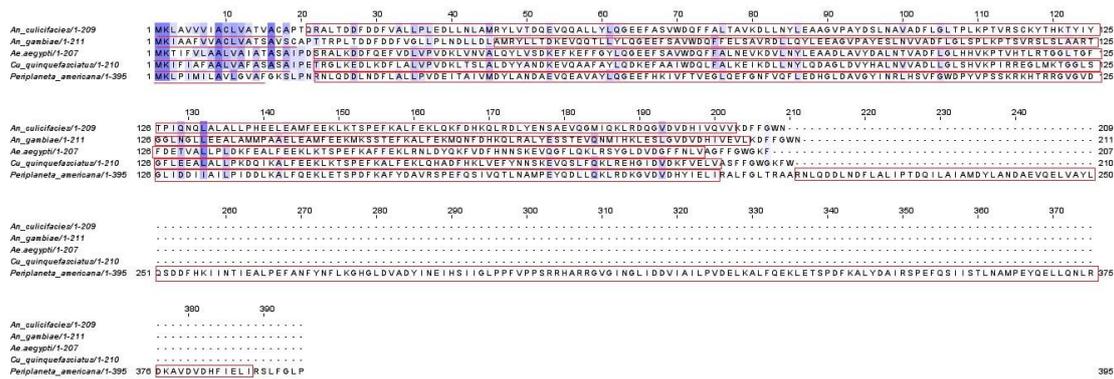

Fig.5: Multiple sequence alignment of *AcG12* and the four most similar sequences, as determined by an NCBI BLAST search. Predicted peptide signal sequence is underlined and insect allergen domains are boxed. Identical residues are shown in dark blue colour while light colour depicts similarity. The NCBI accession numbers for each of the sequences are as follows: *An. gambiae* (Z22925), *Ae. Aegypti* microvillar membrane protein (AY050565), *Culexquinquefasciatus* (XM_001863553), *Periplanetaamericana* (U69957).
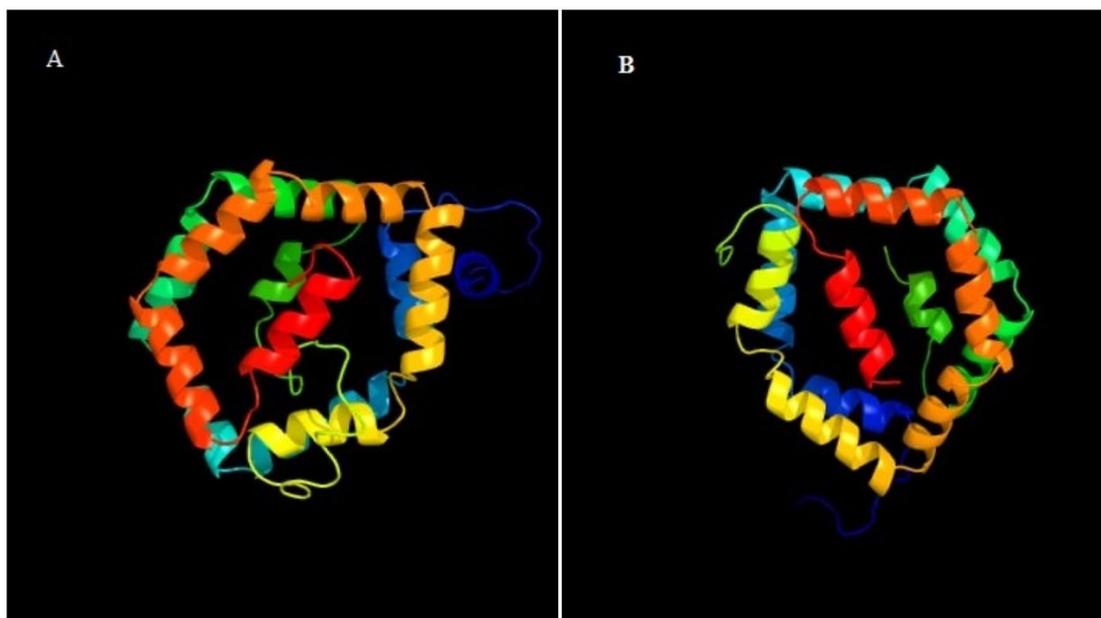
Fig.6: Predicted 3D structure of putative *An. culicifacies* G12 and comparison with *Blattellagermanica* allergen (bla g1).

Phylogenetic analysis

Phylogenetic analysis using coding region of *AcG12* was examined for evolutionary pattern among mosquitoes. The *AcG12* sequence was searched against the NCBI database to find similar sequences. The relationship among these retrieved sequences is more strikingly represented in Fig. 7 where *AcG12* was found most closely related to *An. gambiae* G12 and both seem to be originated from common ancestor. Interestingly *AcG12*, *ANG12*, *AEG12* and G12 genes of *Culex quinquefasciatus* (Diptera:culicidae) showed close similarity to *Periplaneta americana* allergen which is in agreement with the peptide sequences of these genera as all contains insect allergen domain.
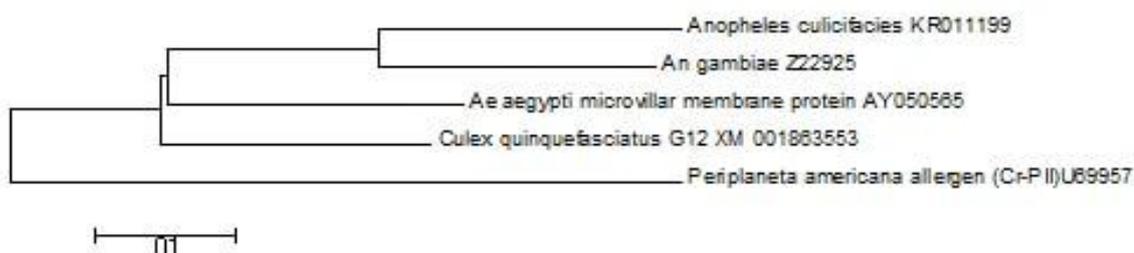
Fig.7: Phylogram showing the relationship between *AcG12* and other similar sequences. The evolutionary history was analyzed using the neighbor-joining method in MEGA 6 after aligning with Clustal W.

DISCUSSION

In the present study G12 gene was isolated from *An. culicifacies* which was found inducing in midgut in response to blood meal. The study revealed that the coding sequence of *An. culicifacies* G12 gene (*AcG12*) shares 69% and 66% identity to *An. gambiae* and *Ae. aegypti* G12 genes. The *AcG12* was found comprised of three exons separated by two small introns. The size of second exon was perfectly conserved in *AcG12* and *AEG1*2 whereas length of third exon was found to be constant in *AcG12* and *ANG12*. The intron boundaries have conserved splice sites however the length of introns varies in all three mosquito species. Length of 5'-UTR and 3'-UTR, region also varies in all the three species of mosquitoes. The variation in size and position of exons and introns indicate the divergence of G12 gene among mosquitoes.

However, the upstream region of *AcG12* did not show much similarity. The position of unusual TATA like sequence, TATAAAA from start codon was found to be -65 in *An. culicifacies*, -117 in *An. gambiae*. Its position relative to start codon and absence of any other similar sequences nearby clearly indicate that it is the most probable site for the TFIID binding domain.

Another important finding is that G12 protein is targeted to the secretory pathway by amino terminal signal peptide that is encoded by signal sequences of 17 amino acids. Thus the residues after cleaving at Alanine 18 secrete the protein in active form. The propeptidase domain was found present in *An. gambiae*, *Ae. aegypti*, *Culex* and *Periplanta americana* but varies in length indicating that all are secreted as zymogen. The post translational modification i.e phosphorylation was found at serine and threonine residues conferring protection from proteolytic degradation.

The unique feature of the *AcG12* protein is that it contain insect allergen related repeat domain. These repeats were also found in other mosquito species and showed homology to major allergens identified in *Periplanetaamericana* and to a nitrile specific protein (*PrNSP*) from the midgut of *Pierisrapae* (Lepidoptera:Pieridae*)*. *PrNSP* helps in converting toxic compound like isothiocyanate into less toxic compound such as nitriles (Wittstock *et al*. 2004). Previous studies established that allergen sequences contain multiple tandem repeats of 100 amino acid residues but *AcG12*, *ANG12* and *AEG12* show no evidence of these repeats and sequence conservation among these species were poor indicating that mosquitoes had degenerated from cockroaches in which repeat allergen sequences was conserved (Pomes *et al*. 1998).

The three dimensional structure of *AcG12* was predicted and over all topology was found to be similar to *Blatellagermanica* (Blattodea: Blattellidae)bla g1 which clearly indicates an evolutionary relationship between mosquitoes and cockroaches. qRT-PCR analysis showed that the G12 transcripts were elevated to high level in the midgut after a blood meal, and its maximal expression was observed at 24h after a

blood meal. This expression profile resembles the dynamics of *ANG12* in which expression increases several fold after blood feeding (Nolan *et al*. 2011). However, these findings are in contrast to *Ae. aegypti* where rapid up regulation occured at 12h following feeding with blood infected with *P. gallinaceum* compared to uninfected blood. This might be due to parasite infection which elicits the immune response. (Morlais *et al*. 2003)

The lack of protein sequence similarity with other organisms makes it difficult to predict a function for *AcG12*. Distension in mosquito midgut was found immediately after blood ingestion causing shortening and loss of microvilli (Freyvogel and Stäubli 1965). Rapid blood induced expression suggests that *AcG12* might play a role in digestion and reorganization of the midgut epithelial membranes.

The phylogeny of the gene confirmed that it is genetically closer to *An. gambiae* G12 gene and both seemed to be diverged from common ancestor. The evolutionary pattern on phylogenetic tree showed that *Anopheles*, *Aedes* and *Culex* G12 gene may have originated from common ancestor and revealed close similarity to cockroach allergen.

The high level of similarity with *ANG12* and *AEG12* suggests that the gene in three species serve the same function. The deduced amino acid sequence of *AcG12* depicts that the protein follows a secretory pathway, and given its rapid blood induction and female midgut specific expression; the protein may have a function in the digestion of the blood.

CONCLUSION

In the present study we have done characterization of *An. culicifacies* A G12 gene which is expressed in response to blood meal and is engaged in blood meal digestion. Development of genetic transformation system in mosquitoes including *Anopheles* proposed a scheme in which wild population of mosquito is replaced with innocuous strain that can no longer transmit a pathogen. In particular, upstream regulatory elements of *AcG12* can be used for expressing genes that interfere with parasite development in transgenic mosquitoes. For exploring full potential of regulatory element further transgenic related studies are needed in *An. culicifacies* which is an important vector of malaria in rural and peri-urban areas.

REFERENCES

[1]   Abraham, E.G., Donnelly-Doman, M., Fujioka, H., Ghosh, A., Moreira, L. and Jacobs-Lorena, M., 2005, "Driving midgut-specific expression and secretion of a foreign protein in transgenic mosquitoes with AgAper1 regulatory elements," Insect Mol. Biol., 14, pp. 271-279.

[2]   Bjellqvist, B., Hughes, G.J., Pasquali, Ch., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S. and Hochstrasser, D.F., 1993,"*The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences,*" Electrophoresis., 14, pp. 1023-1031.

[3]   Blom, N., Gammeltoft, S. and Brunak, S., 1999, "Sequence and structure based prediction of eukaryotic protein phosphorylation sites," J. Mol. Biol.,294(5), pp. 1351-1362.

[4]   Freyvogel, T.A. and Stäubli, W., 1965, "The Formation of the Peritrophic Membrane in Culicidae," Acta Trop., XXII, 2, pp. 118-141.

[5]   Goswami, G., Singh, O.P., Nanda, N., Raghavendra, K., Gakhar, S.K. and Subbarao, S.K., 2006, " Identification of all members of the *Anopheles culicifacies* complex using allele-specific polymerase chain reaction assays," Am J. Trop. Med. Hyg., 75, pp. 454– 460.

[6]   Ito, J., Ghosh, A., Moreira, L.A., Wimmer, E.A. and Jacobs-Lorena, M., 2002, "Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite," Nature.,417, pp. 452-455.

[7]   Kelly, L.A. and Sternberg, M.J.E, 2009, "Protein structure prediction on the web: a case study using the Phyre server," Nat. Protoc.,4, pp. 363–371.

[8]   Krogh, A., Larsson B., von Heijne, G. and E.L.L. Sonnhammer, 2001, "Predicting transmembrane protein topology with Hidden Markov model: Application to complete genomes," J. Mol. Biol.,305 (3), pp. 567-580.

[9]   Kumar, A., Sharma, A., Sharma, R. and Gakhar, S. K., 2014, "Identification, characterization and analysis of expression of gene encoding carboxypeptidaseA in *Anopheles culicifacies* A (Diptera: culicidae)," Acta Trop., 139**,** pp. 123–30.

[10]   Lombardo, F., Lycett, G.J., Lanfrancotti, A., Coluzzi, M. and Arca, B., 2009, "Analysis of apyrase 59 upstream region validates improved *Anopheles gambiae* transformation technique," BMC Res. Notes.,2: 24.

[11]   Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C. and Bryant, S.H., 2015, "CDD: NCBI's conserved domain database," Nucleic Acids Res., 28: 43.

[12]   Morlais, I., Mori, A., Schneider J.R. and Severson, D.W., 2003, "A targeted approach to the identification of candidate genes determining susceptibility to

Plasmodium gallinaceum in *Aedesaegypti,*" Mol. Genet. Genomics.,269, pp. 753–764.

[13] Nolan, T., Petris, E., Muller H.M., Cronin, A., Catteruccia, F. and Crisanti, A., 2011, "Analysis of Two Novel Midgut-Specific Promoters Driving Transgene Expression in *Anopheles stephensi* Mosquitoes," PLoS ONE.,6, pp. 1-8.

[14] Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011, "SignalP 4.0: discriminating signal peptides from transmembrane regions," Nat. Methods., 8, pp. 785–786.

[15] Pomes, A., Melen, E., Vailes, L.D., Retief, J.D., Arruda, L.K. and Chapman, M.D., 1998, "Novel allergen structures with tandem amino acid repeats derived from German and American cockroach," J. Biol. Chem.,273, pp. 30801–30807.

[16] Sambrook, J. and Russell, D.W., 2001, "Molecular Cloning: A Laboratory Manual," third ed. CSHL Press, New York.

[17] Shao, L., Devenport, M., Fujioka, H., Ghosh, A. and Jacobs-Lorena, M., 2005, "Identification and characterization of a novel peritrophic matrix protein, Ae-Aper50, and the microvillar membrane protein, AEG12, from the mosquito, *Aedesaegypti,*" Insect Biochem. Mol. Biol.,35, pp. 947–59.

[18] Shen, Z., Dimopoulos, G., Kafatos, F.C. and Jacobs-Lorena, M., 1999, "A Cell Surface Mucin Specifically Expressed in the Midgut of the Malaria Mosquito *Anopheles gambiae,*" Proc Natl. Acad. Sci USA., 96, pp. 5610–5615.

[19] Tamura, K.,Stecher, G., Peterson, D., Filipski, A. andKumar, S.,2013, "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0," Mol. Biol. Evol.,30 (12), pp. 2725–2729.

[20] Wittstock, U., Agerbirk, N., Stauber, E.J., Olsen, C.E., Hippler, M., Mitchell-Olds, T., Gershenzon, J. and Vogel, H., 2004, "Successful herbivore attack due to metabolic diversion of a plant chemical defense," Proc Natl. Acad. Sci USA.,101, pp. 4859–4864.

[21] World Health Organisation., 2015, World Malaria Report.

[22] Wu, C.H., Wang, N.M., Lee, M.F., Kao, C.Y. and Luo, S.F., 1998, "Cloning of the American cockroach Cr-PII allergens: evidence for the existence of cross-reactive allergens between species," J. Allergy., 101, pp. 832-840.

[23] Yoshida, S. and Watanabe, H., 2006, "Robust salivary gland specific transgene expression in *Anopheles stephensi* mosquito," Insect Mol. Biol.,15, pp. 403–410.

[24] Zieler, H., Nawrocki, J.P. and Shahabuddin, M., 1999, *"Plasmodium gallinaceum* ookinetes adhere specifically to the midgut epithelium of *Aedesaegypti* by interaction with a carbohydrate ligand," J. Exp. Biol.,202, pp. 485–495.