Characterization of STAT Immune Signaling Pathway and ABC Transporters in Mosquitoes

THESIS

Submitted in partial fulfillment
Of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by

KULDEEP GUPTA

Under the Supervision of

Prof. Lalita Gupta

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
PILANI (RAJASTHAN) INDIA

2016

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, RAJASTHAN, INDIA

CERTIFICATE

This is to certify that the thesis entitled **"Characterization of STAT Immune Signaling Pathway and ABC Transporters in Mosquitoes"** submitted by **Kuldeep Gupta** ID No **2008PHXF024P** for award of Ph.D. of the institute embodies original work done by him under my supervision.

Signature of the Supervisor

Prof. LALITA GUPTA
Associate Professor
Department of Biological Sciences,
BITS-Pilani, Pilani Campus

Date:

Dedicated to My Beloved Family

TABLE OF CONTENT

Acknowledgements

*This thesis arose in part out of years of research that has been done since I came to **Prof. Lalita Gupta's** research group. There are many people who I would sincerely like to thank for their support, help and encouragement during my time as a graduate student. This thesis would not have been possible without the guidance, help and support of my supervisor, family, colleagues and friends. It is a pleasure to convey my gratitude to all of them in my humble acknowledgment.*

*Firstly, I would like to extend my deepest gratitude and special appreciation to my esteemed supervisor **Dr. Lalita Gupta**, Associate Professor, Department of Biological Sciences, BITS-Pilani, for giving me the opportunity to join her research group. It is my proud privilege to work under her, whose adroit supervision, expert guidance, genius ideas, never compromising attitude and sympathetic nature; imbibed the strength in me to work hard throughout this endeavour.*

*I wish to express my gratitude towards **Dr. Sanjeev Kumar** Associate Professor Department of Biological Sciences and Associate Dean SRCD for his inspiring suggestions, valuable guidance and advice during the course of my research work. Indeed, the words at my command are inadequate to express my gratifications to Sanjeev Sir who helped me in surging this path of research. He has supported me in every aspect of the scientific process and experiments in forming hypotheses, designing experiments, analyzing results, writing and submitting papers for publications.*

*I am grateful to**, Prof. V.S. Rao** Acting Vice-Chancellor, BITS, **Prof. B.N. Jain and Prof. L.K. Maheshwari** (ex Vice-Chancellors, BITS) and **Prof. A.K. Sarkar** Director**,** BITS-Pilani campus, **Prof. G. Raghurama** (Ex-Director)**,** BITS-Pilani for providing us supportive atmosphere, excellent research facilities in & around the region, exposure to the scientific world and a platform to rise. I express my thanks to **Prof S.K. Verma**, Dean, ARD, BITS-Pilani and **Prof. A.K. Das** (ex-Dean, ARD) for their motivating words throughout my period of stay in Pilani. I never forget to appreciate the support endow with endurance by **Dr. Shibasish Chowdhury** (Ex-Head, Department of Biological Sciences) and **Dr. R.N. Saha** (Ex-Deputy Director) during the tough time of my PhD tenure. They steering my research with acceleration and energized me to continue in academic research. I express my deepest respects and regards for them.*

*I am extremely thankful to **Dr. Rajesh Mehrotra,** Head, Department of Biological Sciences, and **Dr. Jitendra Panwar** (DRC Convenor, Department of Biological Sciences) for their constructive comments during my work.*

*I sincerely thank to my thesis Doctoral Advisory Committee members, **Dr. Sanjeev Kumar** and **Dr. Vishal Saxena** for the valuable advice and intellectual guidance. I am thankful for their constructive suggestions and untiring support during every progressive step in the research work.*

*It has been wonderful to be a part of the Molecular Parasitology and Vector Biology group for the last five years, a group of interesting, supportive and highly knowledgeable people. I owe my special deepest gratitude to **Dr. Shilpi Garg, Dr. Deepak Pakalapati, Dr. Amit Kumar Subudhi, Mr. P.A. Boopathi and Rini** for the enthusiasm and support laid by them during all stages of my work, which boosted me to complete my thesis. **Miss. Mithilesh, Isha, Tania**, **Ramandeep, Ranita, Parik**, **Tripti, Pinky** and **Mr.Vikas** have all*

been incredibly patient and provided plenty of laughs along the way. I also thanks to them for their advice and willingness to share their bright thoughts with me. I would not forget the name of my LOP and RP student names ***Swati, Yamna, Naveen, Ashish, Ankita, Vijay, Divya, Mamta*** *and* ***Anjali****. I feel very lucky to be a part of this team and I hope to keep in touch with everyone.*

I also express my thanks to my seniors ***Dr. Manav Kapoor, Dr. Shipra Mehra, Dr. Shaifali Gurjar and Dr. Pradeep Kumar*** *for their enthusiasm and kind assistance in various ways in facilitating my research. I would also like to acknowledge* ***Dr. Satish Kumar Pandey, Dr. Dilip Kumar Pandey, Dr. K.P. Chandrasekhar, Dr. Bhupendra Mishra****,* ***Dr. Ankur Zindal, Dr. Vadiraj Kurdekar*** *and* ***Mr. Gagandeep Singh Saggu, Mr. Arpit Bhargava,*** *who made my stay in Pilani, a pleasant experience. I express my sincere thanks to my Colleagues* ***Isha, Zarna, Rini, Purva, Senthil, Garima, Gurpreet, Panchsheela, Parva, Rajnish, Divya, Vidushi, Shraddha, Sandeep, Vikram, Vandana, Jyothi, Shubhra, Dhilip, Akansha, Leena, Chetna, Zaiba*** *for their companionship and much enriched help during all times whenever needed.*

I express my thanks to ***Dr. Sushil Yadav*** *(Veterinarian In-charge) and other BITS laboratory facility staff,* ***Mr. Parmeshwarji****,* ***Mr. Subhashji, Mr. Vishal, Mr. Mukesh, Mr. Manoj*** *and* ***Mr. Shyam*** *those also took care of my laboratory requirement and animals during the studies. I would like thank to non-teaching staff* ***Mr. Kamlesh, Mr. Naresh, Mr. Mahendra, Mr. Virendra, Mr. Kuldeep and Mr. Mahipal*** *and all the students of Biological Sciences Department for the encouragement provided by them in each stage of the research work.*

I am also thankful to ***University Grant Commission****, New Delhi, India, and ARD, BITS-Pilani for providing the financial assistance throughout the entire research work****.***

I would be failing in my duties if I forget those pretty animals that, I sacrificed for my research work. I pay homage to those experimental animals that have sacrificed their lives in making my endeavour successful. May god grant them eternal peace! I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

This list is incomplete without paying high regards to my parents ***Mr. Hari Narayan Gupta*** *and* ***Mrs. Chandra Prabha Gupta.*** *I know the amount of trouble and pain they have faced in bringing me to this stage of life. They remained in uncomfortable zone to provide me comfortable space. Their constant encouragement and inspiration has always provided me strengths much needed to achieve this tough goal. I would like to express my heartfelt thanks to my elder brother* ***Mr. Sumit Kumar Gupta*** *and sisters* ***Mrs. Chayanika Gupta*** *who's teaching and continuous support have boost me to achieve all the success in my life. The moral support extended by my sister-in law* ***Mrs. Sulekha Gupta*** *and brother-in law* ***Mr. Aditya Narayan Agrahari*** *is highly appreciated. Last but not least, I would like to dedicate this work to my beloved family especially my parents and wife* ***Priti Gupta*** *who have given me an enormous amount of support and for their continuous belief in me. At last but not least, I am great thankful to God with whose mercy I reached through my endeavour so far.*

Date: *Kuldeep Gupta*

LIST OF TABLES

LIST OF FIGURES

LIST OF USED ABBREVIATIONS

Ae.	*Aedes*
An.	*Anopheles*
aa	Amino acid
ABC	ATP Binding Cassette
Ae	*Aedes aegypti*
Ag	*Anopheles gambiae*
AGCC	*Anopheles* Genome Cluster Committee
AMP	Anti-Microbial Peptide
As	*Anopheles stephensi*
BL	Basal Lamina
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
CO- I & II	Cytochrome Oxidase- I & II
ct	Threshold cycle number in a PCR
Δct	Difference in ct values
DDT	Dichloro-Diphenyl-Trichloroethane
dsRNA	double stranded RNA
DVS	Dominant Vector Species
EMBL	European Molecular Biology Laboratory
GM	Genetically Modified
HPX	Heme-Peroxidase
ITS-2	Internal Transcribed Spacer-2
JAK	Janus Kinase
JIR	JAK Interacting Region
KIR	Kinase Inhibitory Region
MDR	Multi-Drug Resistance
MEGA	Molecular Evolutionary Genetics Analysis
ML	Maximum Likelihood
NBD	Nucleotide Binding Domain
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NJ	Neighbour Joining
NO	Nitric Oxide

NOS	Nitric Oxide Synthetase
NTCR	N-Terminal Conserved Region
NVBDCP	National Vector Borne Disease Control Programme
P.	*Plasmodium*
Pb	*Plasmodium berghei*
PCR	Polymerase Chain Reaction
Pf	*Plasmodium falciparum*
PIAS	Protein Inhibitor of Activated STAT
PM	Peritrophic Matrix
PO	Phenol Oxidase
PPO	Pro-Phenol Oxidase
PRR	Pattern Recognition Receptor molecule
PTGS	Post-Transcriptional Gene Silencing
PTPs	Protein Tyrosine Phosphatases
qPCR	quantitative real time PCR
RING	Really Interesting New Gene
RISC	RNA-Induced Silencing Complex
RNAi	RNA interference
S7	Ribosomal protein Sub-unit 7
SAP	Scaffold Attachment Protein-A/B
SBP	Substrate-Binding Proteins
siRNA	Small interfering RNA
SOCS	Suppressor Of Cytokine Signaling
SP	Serine Protease
STAM	Signaling Transducing Adaptor Molecules
STAT	Signal Transducer and Activator of Transcription
SUMO`	Small Ubiquitin-like Modifier
StIP	STAT Interacting proteins
SUR	Sulphonyl Urea Receptors
TEP-1	Thio-Ester like Protein-1
TMD	Transmembrane Domain
TLR	Toll-Like Receptors
WGS	Whole Genome Shotgun sequences
WHO	World Health Organization

ABSTRACT

Mosquitoes are the vectors of several important human infectious diseases. *Anopheles* mosquitoes are responsible for spreading malaria whereas *Aedes* species are for arbovirus. To cope up with these infections mosquitoes possesses strong immune responses. Most of the time, both the mosquitoes are present in one particular location and able to transmit diseases. *Anopheles* and *Aedes* mosquitoes are present in different region of Rajasthan, India. We collected the larvae of both mosquitoes from the Pilani region of Rajasthan and characterized at a morphological and molecular level as *Anopheles stephensi* and *Aedes aegypti*. Further to understand their immunity against pathogens, we characterized STAT signaling pathway in *An. stephensi*. Further, we compared this pathway and its downstream genes with other genome sequenced *Anopheles* species. ABC transporters are also play a role in insect immunity, therefore we identified and annotate ABC transporter in *A. aegypti* mosquitoes.

STAT is an evolutionarily conserved transcription factor of JAK-STAT signaling pathway that is involved in immunity, growth and development of organisms. Unlike to *Drosophila* and *A. aegypti*; *An. gambiae* contains two STAT genes; conserved *Ag*STAT-A and retro-duplicated, intron-less *Ag*STAT-B. To identify whether other *Anopheles* species also contain two STATs, genomic data of different *Anopheles* were used for annotation. Comparative genome analysis has revealed that out of 18 *Anopheles*, 5 species do not contain retro-duplicated STAT-B. *An. sinensis*, *An. albimanus*, *An. darlingi*, *An. dirus* and *An. farauti* comprises only conserved STAT-A gene. Interestingly, 13 species that contain both STATs are mutually diverged and showed sequence variability in protein motifs. In taxonomical classification, only *Anopheles* species have two STATs in class Insecta.

The suppressors of cytokine signaling5 (SOCS5) and protein inhibitor of activated STAT (PIAS) genes play an imperative role in regulation of STAT pathway. Existence of two STATs in *Anopheles* mosquito provides the high chances for complication in the functioning of this pathway. This study has sought to understand the subsequent evolution and diversification through the examination of SOCS5 and PIAS gene in 18 *Anopheles* species. We have annotated SOCS5 gene and retrieved the highly conserved SH2 and SOCS box domain in all *Anopheles* species. The N-terminal region was found to be diverged however, 92 amino acid long region showed the conservation in sequence which could be the putative orthologs of JAK interacting region of higher organism.

Thorough analysis of N-terminal domain showed the repetition of several amino acids in this region *viz.* 8 histidine, 3 phenylalanine, 4 serine and variable alanine repeats. This *Anopheles* genus specific conservativeness of N-terminal region suggests its regulatory role in STAT pathway. The PIAS protein has been proposed to interact with many transcription factors involved in the immune system. PIAS proteins regulate transcription through several mechanisms, including blocking the DNA-binding activity of transcription factors, recruiting transcriptional co-repressors and promoting protein sumoylation. Annotation of PIAS gene in eighteen different *Anopheles* species revealed much conserved domains *viz.* SAP domain, PINIT domain and the MIZ/SP-RING zinc finger domain. Phylogenetic study of SOCS and PIAS genes reveal the same evolutionary pattern as it was followed in taxonomical hierarchy of *Anopheles* species.

Further, laboratory colonized *An. stephensi* mosquito was used to identify the STAT pathway genes; AsSTAT-A, AsSTAT-B, AsSOCS5 and AsPIAS has cloned and sequenced using specific primers. The expression of these five genes (AsSTAT-A, AsSTAT-B, AsPIAS, AsSOCS5 and AsNOS) in *An. stephensi* mosquito 24 hour after *Plasmodium berghei* infection was investigated. Quantitative real time PCR analysis revealed that AsSTAT-A and AsSTAT-B mRNA levels were remained constant after *P. berghei* infection. As these are the transcription factors therefore their requirement is to switch-on the pathway. In case of AsSOCS5, near about six times higher expression was observed in the infected midguts, whereas in the carcass samples it was three times highly express compare to controls. In addition to that AsPIAS and AsNOS expression was found significantly high in midgut but not in carcass compare to blood-fed control. This indicates that STAT pathway induced during ookinete invasion and in parallel the suppressor of this pathway is also active to counter balance the immune reactions.

Functional characterization through RNAi experiment envisaged a new paradigm in this mosquito species. We have individually silenced the STAT-A or STAT-B gene by injecting dsRNA of the gene through nanoinjection into hemolymph of adult female *An. stephensi*. After silencing of AsSTAT-A gene, we found there was no significant change in expression of downstream genes of the pathway. However, silencing of other AsSTAT-B resulted in the silencing of AsSTAT-A as well as reduced the expression of other downstream genes. There was a marked reduction of expression recorded of AsSTAT-A, AsSOCS5, AsPIAS and AsNOS gene after AsSTAT-B silencing. This

functional depiction clearly indicated that STAT-B regulates the basal level of STAT-A gene as well as other downstream genes.

In another section of thesis we characterized and identified the field collected *Aedes aegypti* species as well as excavate their ABC transporters gene family. We collected the *Aedes* larvae from Pilani region of Rajasthan and characterize them at morphological and molecular level. The molecular identification was done through nuclear internal transcribed spacer-2 (ITS-2) and mitochondrial markers cytochrome oxidase subunit-I (mtCOI). Lab adapted *Aedes* species were further compared with globally distributed *Aedes* species for their ITS-2 and COI sequences. A comprehensive multiple sequence alignment and phylogenetic analysis revealed that COI gene of *A. aegypti* showed extremely low genetic variability with one of the Indian isolate from Thirumala, Andhra Pradesh region. Though, in context of different geographical location, it indicates close similarity with Thailand and high variability from Madagascar population. On the other hand, ITS-2 illustrated highest identity with *A. aegypti* of Saudi Arabia whereas high divergence from Mayotte, France. These finding suggest that isolate from Rajasthan is similar to other Asian continent strains possibly due to the same origin.

The ABC transporter is the largest transporter gene family in all living organisms. In insects, ABC transporters have diverse physiological functions which include molecule transport, metabolism, insecticide resistance and immunity. Here, we analysed the genome of *Aedes aegypti* and identified different subfamilies of ABC transporter genes. Total 71 putative ABC genes identified which were classified into eight subfamilies (A-H). 25 ABC transporter genes were identified as full transporters lies in the subfamilies ABC-A, ABC-C and ABC-D and 46 were reported as half transporters. Subfamilies E to H contain only half transporters whereas subfamilies A to D contain both full and half transporters. We also found that *A. aegypti* ABC family is significantly larger than other known dipterans. Comparative genome analysis revealed that this increase is due to gene expansion within a single clade of subfamily ABC-C and ABC-G. Gene duplication was found very evident in ABC-A, ABC-C and ABC-G subfamilies, whereas gene numbers and structures are well conserved in ABC-D, ABC-E, ABC-F, and ABC-H subfamilies. Phylogenetic analysis of all *A. aegypti* ABC proteins exactly confirmed their position in that particular subfamily and showed the similarity with other subfamily members.

Chapter 1

Introduction and Review of literature

Part of this chapter has been published in this paper: Kajla M, Gupta K, Gupta L, Kumar S. A fine-tuned management between physiology and immunity maintains the gut microbiota in insects. **Biochem Physiol.** 2015, 4(4):182.

1.1 Mosquito and their life-cycle

Mosquitoes are the vectors of several important human infectious diseases including malaria, lymphatic filariasis and arbovirus, transmitted by the few species of *Anopheles*, *Culex* and *Aedes* respectively (Krzywinski and Besansky 2003). There are more than 3500 species of mosquitoes distributed throughout the world in 34 genera; mostly belongs to genus *Anopheles*, *Culex* and *Aedes* (Dong et al., 2004; Ghosh et al., 2013). Mosquitoes are insects belong to the order Diptera, the "True Flies". Like other dipterans, they have two wings with scales, their mouthparts form a long piercing sucking structure called proboscis (Robinson, 1939). Males differ from females by having feathery antennae and proboscis, only suitable for sucking and feed on nectar, plant sap, flowers, overripe fruits throughout their life-span (Wahid et al., 2003). The mosquitoes have a holometabolous type of development, having four distinct stages in their life cycle: egg, larva, pupa, and adult (Truman and Riddiford, 1999). In **Figure 1.1** each of these stages can be easily recognized by their special appearance. Moreover, these specific key morphological features are the basis for their taxonomical categorization (Clements, 1992).

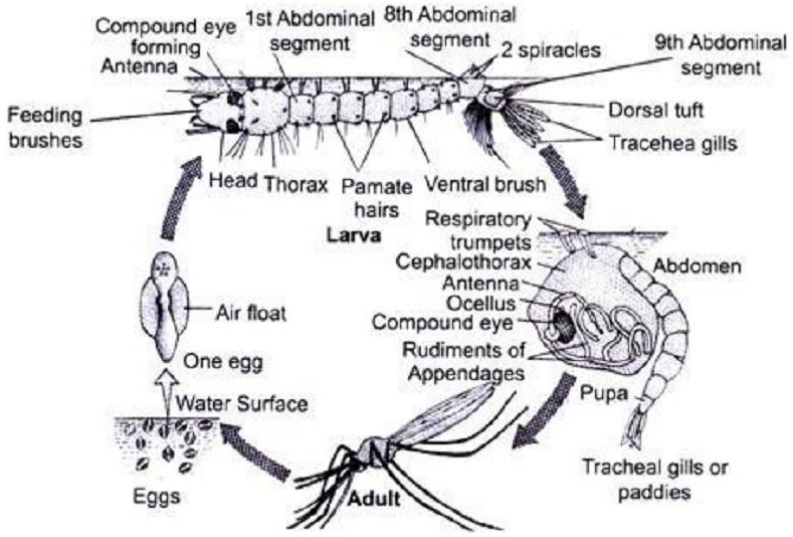


Figure 1.1 Life-cycle of *Anopheles* mosquitoes comprises three life stages in aquatic environment (*viz.* eggs, larvae, pupae) and adults live only in aerial mode of life. Figure has been adopted from http://www.biology-resources.com/insects.

Egg: *Anopheles* and *Aedes*, as well as many other genera, lay their eggs singly but *Culex* genus mainly laid their eggs in the form of raft (Raikhel et al., 2002). Each raft may contain

100-200 or more eggs, loosely cemented together. A female *Anopheles* mosquito usually laid 50-60 eggs at a time, 3-4 days after the blood-meal. They float on the surface of the water through the definite flap present on the edge (Malhotra et al., 2000). *Culex* and *Anopheles* lay their eggs on the water surface while many *Aedes* species lay their eggs on damp soil that will be flooded by water (Farnesi et al., 2015). Most eggs hatch into larvae within 48 hours under the availability of constant temperature (Impoinvil et al., 2007).

Larva: After egg hatching, there are four larval stages (L-I, L-II, L-III and L-IV). Each stage has the larval time period of about 24 to 26 hours depend upon optimum temperature. The larva shed (molt) their skins four times, growing larger after each molt (Jarośík et al., 2004). Most larvae have siphon tubes for breathing and hang upside down from the water surface. *Anopheles* larvae have very tiny siphon and lie parallel to the water surface to get a supply of oxygen through a breathing opening (Mereta et al., 2013). *Culex* and *Aedes* larvae show distinctive swimming style, so they are known as "wrigglers". They have a distinctive tube for breathing which extends from the end of their body. All the instars are voracious eaters, taking of microscopic size food materials into the buccal cavity by instant vibration of its feeding brushes (Merritt et al., 1992).

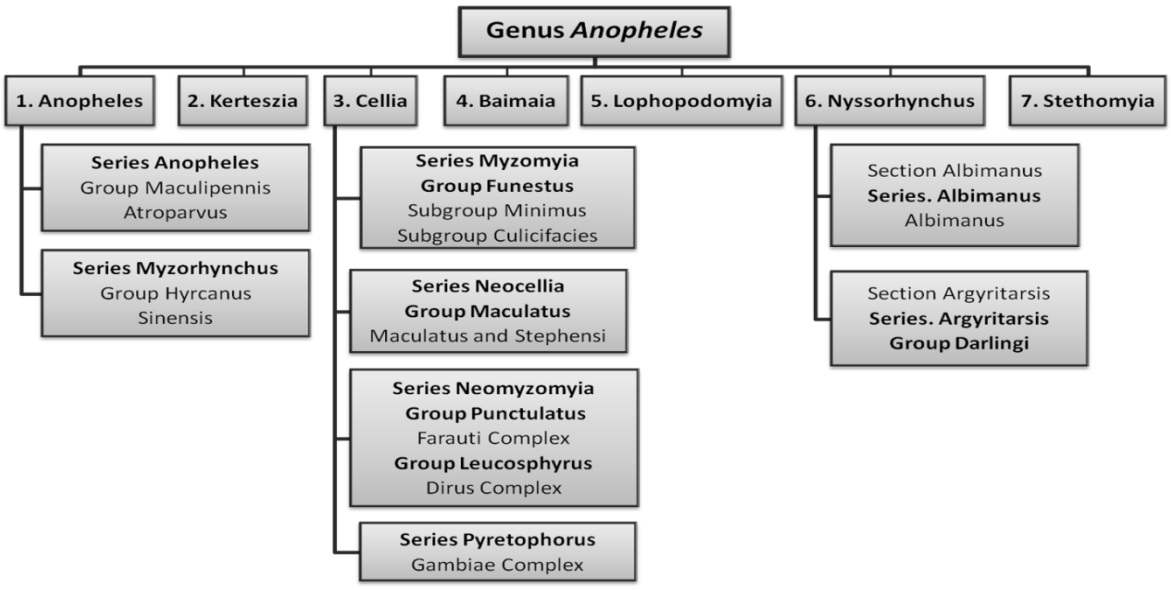
Pupa: The 4th instar larva at the end of its stage give rises to a comma shaped pupa. Pupae do not feed but are very active, respiring through its pair of breathing trumpets. The pupal stage is a resting, non-feeding stage of development. When metamorphosis is complete, the pupal skin splits and the adult mosquito emerges.

Adult: The newly emerged adult rests on the surface of the water for a short time to allow itself to dry and all its body parts to harden. The wings have to spread out and dry properly before it can fly. Male adult mosquitoes solely feed on nectar and do not bite humans while the female mosquito after mating requires a blood meal for their ovarian development and vitellogenesis. The entire cycle is completed in 10-14 days. How long each stage lasts depends on both temperature and species characteristics (Christiansen-Jucht et al., 2014).

1.2 Taxonomy and zoogeography of *Anopheles* mosquitoes

Anopheles mosquitoes are the vectors of several important human infectious diseases and found on all continents, with the exception of Antarctica. Genus *Anopheles* represents more

than 500 species that are often organized in sibling species complexes. Human malaria is solitary transmitted by the *Anopheles* mosquitoes, but not all species within the genus or even all members of each vector species (sibling species or vector complex) are efficient malaria vectors. This suggests an underlying genetic/genomic plasticity that results in variation of key traits determining vectorial capacity within the genus. Genus *Anopheles* currently includes 465 formally named species that are disproportionately divided between seven subgenera: *Anopheles*, *Baimaia*, *Cellia*, *Kerteszia*, *Lophopodomyia*, *Nyssorhynchus* and *Stethomyia* [Mosquito Taxonomic Inventory. http://mosquito-taxonomic-inventory.info/ (accessed on 1 September 2014)]. The current sub generic classification of *Anopheles* is primarily based on the number and positions of specialized setae on the gonocoxites of the male genitalia (Christophers, 1915). Three of the subgenera, *Anopheles*, *Cellia*, and *Nyssorhynchus*, include the major species that transmit malaria parasites whose genome sequence recently been submitted (Neafsey et al., 2013). The informal categories used in the classification of *Anopheles* include Sections, Series, Groups, Subgroups and Complexes. Since many of the primary malaria vectors belong to species complex (**Flowchart 1**). Their non ranked classification was compiled according to the Bulletin of Entomological Research (Harbach, 2004). Accurate phylogenetic reconstruction and demographical distribution of species are necessary for understanding their evolutionary relationship. Eighteen *Anopheles* species whose genomes have been sequenced shown in **Table 1.1** These species belong to different subgenus, series and spread in various part of the world.



Flowchart 1. Taxonomical categorization of genus *Anopheles*

Table 1.1 Taxonomic and geographical information of *Anopheles* species whose genome has been sequenced (Neafsey et al., 2015; Sinka et al., 2010a; Sinka et al., 2010b; Sinka et al., 2011)

Mosquito Name	**Infra specific name (Strain)**	**Subgenus**	**Series**	**Group**	**Region**
An. gambiae	Pest	Cellia	Pyretophorus	gambiae comp.	Ethiopian
An. arabiensis	DONG5_A	Cellia	Pyretophorus	gambiae comp.	Ethiopian
An. quadriannulatus	QUAD4_A	Cellia	Pyretophorus	gambiae comp.	Ethiopian
An. melas	CM1001059_A	Cellia	Pyretophorus	gambiae comp.	Ethiopian
An. merus	MAF	Cellia	Pyretophorus	gambiae comp.	Ethiopian
An. christyi	ACHKN1017	Cellia	Pyretophorus		Ethiopian
An. epiroticus	epiroticus2	Cellia	Pyretophorus		Oriental
An. maculatus	maculatus3	Cellia	Neocellia	Maculatus	Oriental
An. stephensi	Walter Reed	Cellia	Neocellia	Maculatus	Oriental
An. culicifacies	species A-37-1	Cellia	Myzomyia	Funestus	Oriental
An. minimus	MINIMUS1	Cellia	Myzomyia	Funestus	Oriental
An. funestus	FUMOZ	Cellia	Myzomyia	Funestus	Ethiopian
An. dirus	WRAIR2	Cellia	Neomyzomyia	Leucosphyrus	Oriental
An. farauti	FAR1	Cellia	Neomyzomyia	Punctulatus	Australasian
An. atroparvus	EBRO	Anopheles	Anopheles	Maculipennis	Palaearctic
An. sinensis	AS2	Anopheles	Anopheles	Hyrcanus	Palaearctic
An. albimanus	ALBI9_A	Nyssorhynchus	Albimanus	Albimanus	Nearctic
An. darlingi	A_darlingi	Nyssorhynchus	Argyritarsis	Darlingi	Neotropical

Approximately 60 species of *Anopheles* mosquitoes have the capacity to transmit malaria parasites and 41 are dominant vector species (DVS) (Cohuet et al., 2010; Hay et al., 2010). Each individual *Anopheles* mosquito species was categorized in Wallace's zoogeographical provinces. All human malaria causing vector species have been shown in **Figure 1.2** (Manguin et al., 2008). Among these 60 *Anopheles* species, 18 vector species whose genome sequences are available has been shown in the map legend (◄). Out of eighteen, 2 are non-vector species for malaria present in Ethiopian region.

The specific vector and parasite relationship determines the dynamicity of parasite development and malaria transmission among humans (Sinka et al., 2012). It has been hypothesized that the molecular mechanism of malaria vector species varies with pathogen due to disproportional susceptibility (Molina-Cruz and Barillas-Mury, 2014). Recently, a landmark study was published by Molina-Cruz et al., 2015 in which they have proposed the lock-key theory; which implies that with different haplotypes of Pfs47 gene (gametocyte surface protein, P47) of *Plasmodium* strains, the malaria vector species of different world regions behave differently. Why only certain species of anopheline are malaria vectors which determines the vector competence remains an enigma and warrants its elucidation.

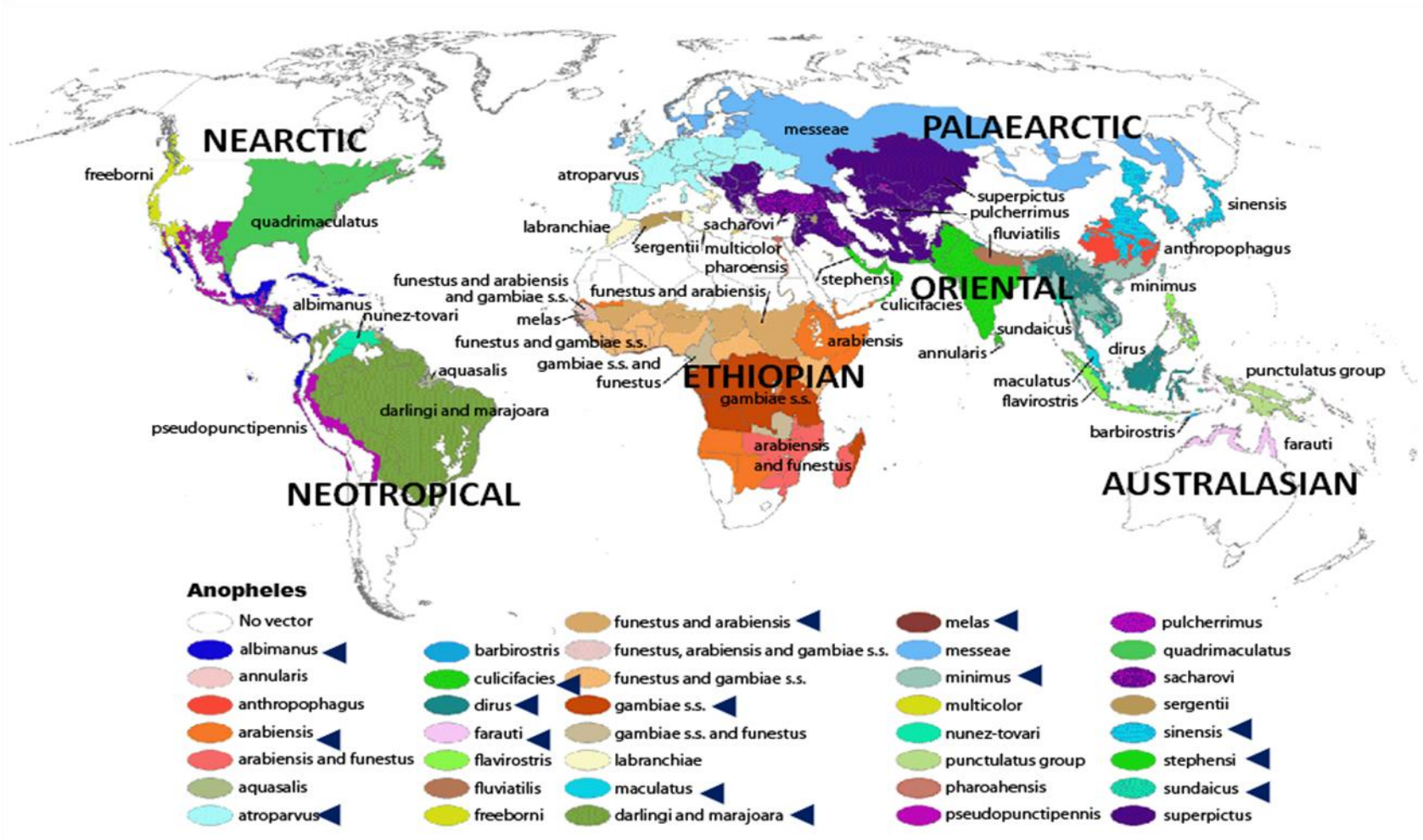


Figure 1.2 Zoogeographical provinces on world map showing different endemic species of genus *Anopheles*. The map presented here provides particular species location information and highlights the existence of a greater number of vector species in that realm. Maps clearly illustrate the spatial extent of a species' distribution (Sinka et al., 2010a; Sinka et al., 2010b; Sinka et al., 2011). Symbol (◄) is showing those *Anopheles* species whose genome assemblies have been submitted to the NCBI (Neafsey et al., 2013).

1.3 Mosquito immune system

Mosquitoes are encountered by several pathogens throughout their lifetime. To fight against the infections, they have strong innate immunity but lack adaptive immunity (Vilmos and Kurucz, 1998). Mosquito innate immune system can be divided into two main classes: cellular immunity and humoral immunity (Hillyer, 2010). The cellular response include: cell-mediated phagocytosis and encapsulation by hemocytes and pericardial cells (Castillo et al., 2006; Hillyer et al., 2003a and 2003b). Three primary mechanisms employed by the mosquitoes to kill the pathogens *viz.* phagocytosis, melanization and lysis (**Figure 1.3**).

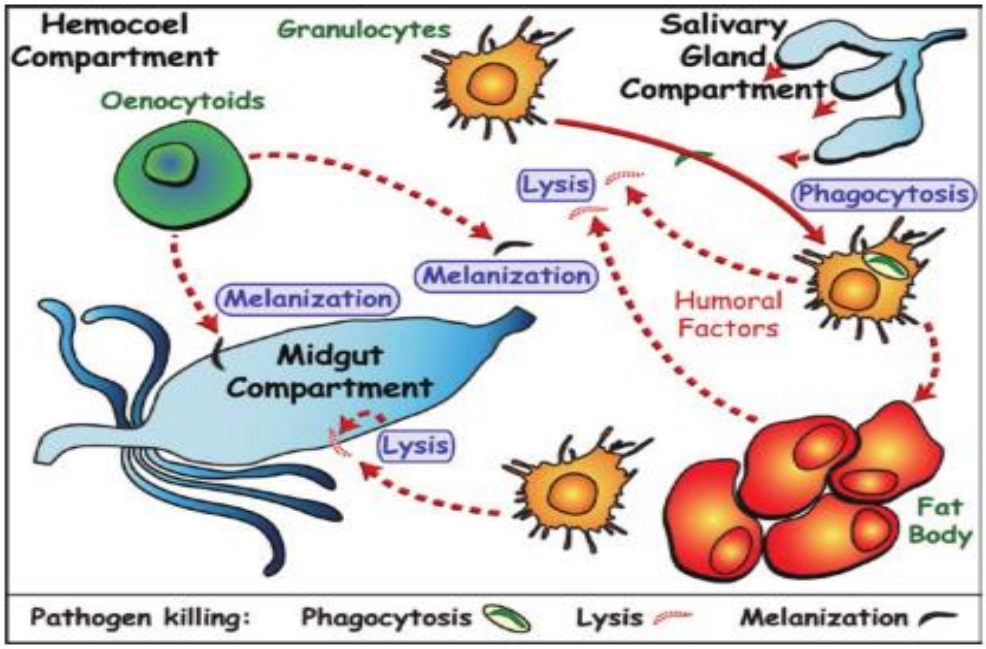


Figure 1.3 Mosquito immune responses in the three major immune compartments. In the hemocoel, granulocyte, oenocytoid hemocytes and fat body, kill pathogens via phagocytosis, lysis and melanization. In the midgut immune factors are produced by epithelial cells, hemocytes and possibly fat body, kill pathogens via lytic and melanization pathways (Figure adopted from Hillyer, 2010).

The humoral response include pattern recognition receptors, inducible antimicrobial peptides, the phenoloxidase cascade system of melanization, wound healing, reactive oxygen and nitrogen intermediates. The production and secretion of antimicrobial peptides (AMPs) in response to microbial challenge is the classical property of humoral immune response. These AMPs secreted locally as well as into the hemocoel (insect body cavity). The latter is referred to as the systemic immune response and is primarily regulated by the fat body, an organ comparable to the vertebrate liver and adipose tissues and the major production site of antimicrobials. Systemic responses are classically thought to turn on when

pathogens have managed to break the anatomical barriers of midgut which are also part of innate immune system. In the midgut, peritrophic matrix lines the intestinal lumen which creates physical barrier and comparable to mucous secretion in the vertebrate digestive tract (Lehane, 1997; Merzendorfer and Zimoch, 2003). This matrix is composed of secreted chitin, proteins, proteoglycans and serves as a semi-permeable barrier for gut pathogens.

1.3.1 Immune Signaling Pathways

Three major immune signaling pathways have been identified in insects: Toll, Immune Deficiency (IMD) and Janus kinase-signal transducer and activator of transcription (JAK-STAT). These pathways play key roles in defense against various pathogens (**Figure 1.4**).

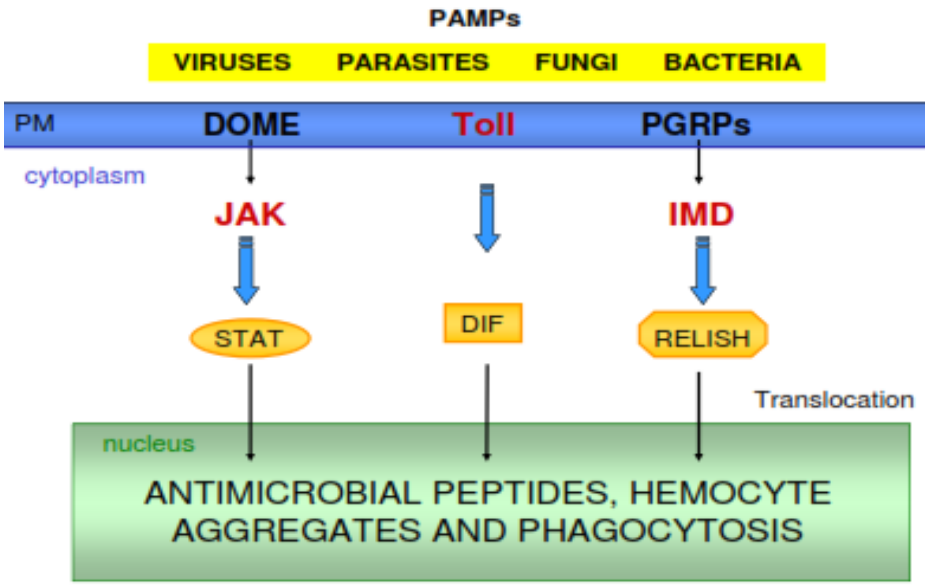


Figure 1.4 Toll, IMD and JAK-STAT pathways. Insect tissues recognize pathogen-associated molecular patterns (PAMPs) by transmembrane receptors (DOME, Toll and PGRPs) and activate the three pathways. The transcription factors (STAT, DIF and Relish) translocate to the nucleus through the nuclear membrane activating the expression of its transcriptional targets resulting in the production of antimicrobial peptides and other immune responses. (Figure adopted from Garcia et al., 2009, Parasit Vectors 2:33)

The Toll pathway is more specific for Gram-positive bacteria, fungi and viruses, whereas the Imd pathway is mainly active against Gram-negative bacteria and *Plasmodium* (Blandin et al., 2002 and Cirimotich et al., 2010). Toll pathway is also activated in response to dengue and chikungunya virus infection in *A. aegypti* and controls viral proliferation in the mosquito midgut (Shin et al., 2005 and 2006; McFarlane et al., 2014). Activation of the Imd pathway leads to the expression of anti-bacterial peptides, such as defensin (Xi et al., 2008).

Recently, the receptors involved in the specific recognition of pathogens have been identified. These molecules belong to the peptidoglycan-recognition protein (PGRP) family and are thought to mediate NF-kB signaling in the fat body cells (Gottar et al., 2002). The JAK-STAT pathway in insects has been proposed as functionally analogous to the interferon system in mammals (Kingsolver et al., 2013). STAT transcription factor has been implicated in biological events as diverse as embryonic development, programmed cell death, organogenesis, innate immunity, adaptive immunity and cell growth regulation in organisms ranging from slime molds to insects to man (Horvath, 2000). Although Toll, IMD and JAK-STAT pathways have been most investigated in few mosquitoes but other less characterized like, JNK pathway also contributed to the mosquito immunity (Ramphul et al., 2015).

1.3.2 The JAK-STAT pathway

The JAK-STAT signaling pathway is involved in various molecular events and mediated by cellular and humoral responses to viruses, some bacteria and other pathogens. STAT transcription factors are latently present in cytoplasm and activated by tyrosine phosphorylation in response to extracellular signals which is evolutionary conserved and canonical form of the pathway (Levy and Darnell, 2002). Recent findings in *Drosophila* have identified a non-canonical mode of JAK-STAT signaling, which directly controls heterochromatin stability (Li, 2008). In contrast to the canonical mode of signaling, in the non-canonical mode, a portion of the unphosphorylated-STAT pool is localized in the nucleus on heterochromatin in association with histone protein (HP1). The heterochromatin-associated unphosphorylated STAT is essential for maintaining HP1 localization and heterochromatin stability (Shi et al., 2008). This indicates that JAK-STAT pathway also controls cellular epigenetic status, which affects expression of genes beyond those under direct STAT transcriptional control.

Drosophila as a representative of the order diptera due to having completeness and simplicity, make it a good model to study the functions and regulation of the JAK-STAT signaling pathway (Zeidler et al., 2000). Despite numerous components discussed above, the signal transduction mechanism of JAK-STAT pathway is simpler in *Drosophila* compare to other vertebrates (Arbouzova and Zeidler, 2006). *Drosophila* genome contains single copy of JAK and STAT gene. The presence of single-copy gene simplifies interpretation of gene

function based on gain- or loss-of-function mutations. The complexity, interdependence and mutual redundancy often complicates this pathway in vertebrates but the situation in *Drosophila* is significantly more straightforward (Hoffmann and Reichhart, 2002). In mammals they are largely involved in a number of biological processes including immunity, hematopoiesis, inflammation and development (Ortmann et al., 2000). By contrast, *Drosophila* contains a simpler 'streamlined' pathway that is sufficient to mediate a multitude of different life processes. Likely to vertebrate system, in *Drosophila* the JAK-STAT pathway also contributes to the renewal of intestinal stem cells and to their differentiation into enterocytes after damage and stress-induced Upd production in the midgut (Jiang et al., 2009).

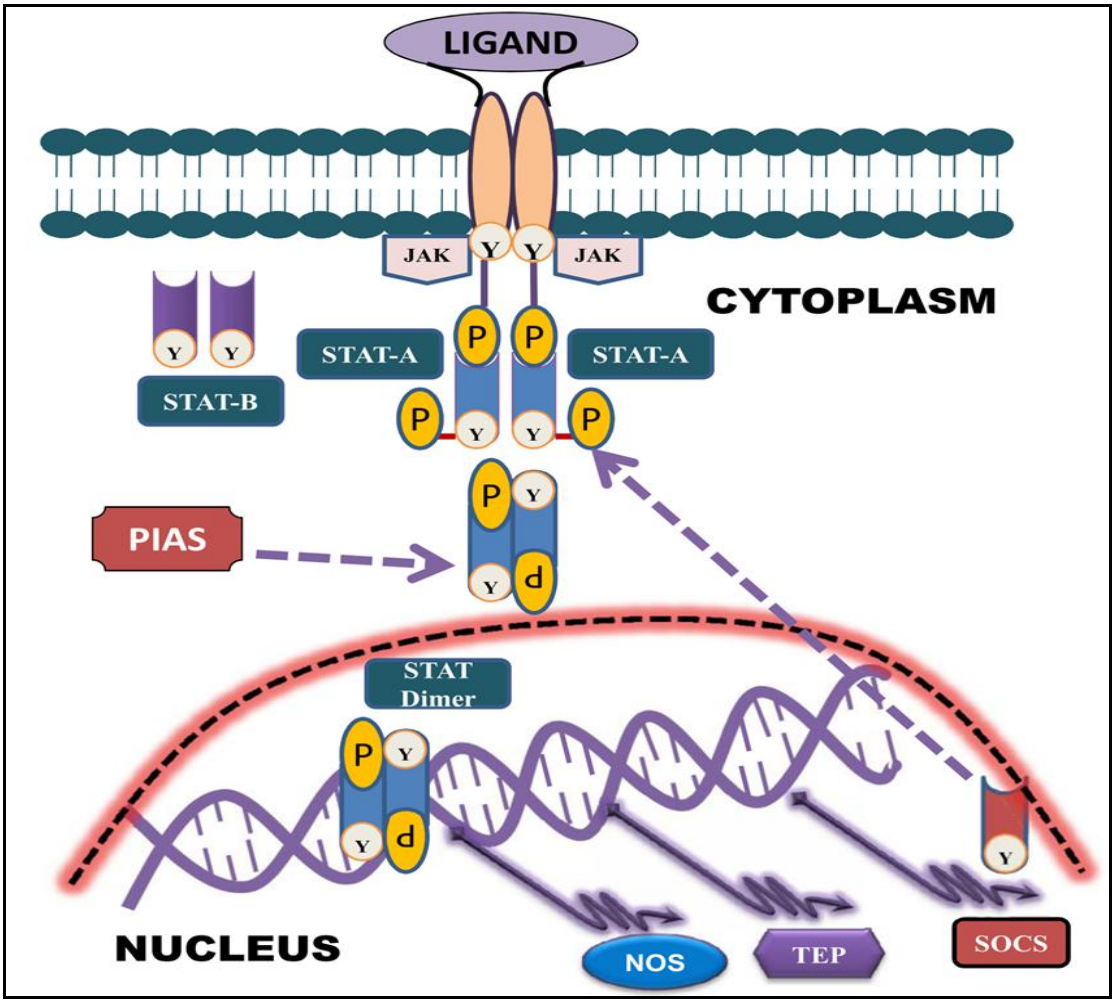


Figure 1.5 Mechanisms of JAK-STAT Signaling Pathway in mosquito: Binding of ligand to its receptor induces JAK-mediated STAT tyrosine phosphorylation and dimerization. STAT dimers translocate to the nucleus and bind to responsive DNA elements in the promoters of target genes and initiate gene expression.

Mechanistically, JAK-STAT pathway is activated with the binding of an extracellular ligand to a transmembrane receptor, which results in the activation of receptor-associated tyrosine kinase JAKs (**Figure 1.5**). The exact molecular mechanism, for how ligand binding activates receptor is still unclear. In the canonical model of the JAK-STAT pathway ligand binding induces rapid receptor dimerization followed by the activation of downstream targets (Agaisse et al., 2003). In *Drosophila*, ligand binding is predicted to cause conformational changes in the receptor homodimer, allowing interaction between associated Hop (JAK-Tyrosine Kinase) molecules (Brown et al., 2003). This tyrosine kinase then self-phosphorylated and their associated receptors generate docking sites for the Src homology2 (SH2) domains of STATs. According to the established models, STATs are normally present in the cytoplasm as inactive monomers before recruitment to the phosphorylated receptor/JAK complex. Once bound to the receptor/JAK complex, STAT molecules are themselves phosphorylated and dimerized (Chen et al., 1998). Phosphorylation in STATs is taken place on a single tyrosine (Y) residue, after which they form homo- or heterodimers with other phosphorylated STAT proteins. This dimerization is stabilized by the interaction between the SH2 domain of one molecule and the phospho-Tyr of the other molecule (Becker et al., 1998). These dimers detach from the receptor and translocate to the nucleus where it binds to a palindromic DNA sequence in the promoters of target genes to activate transcription (Agaisse et al., 2003). Three negative regulators, PIAS, SHP and SOCS have been shown to suppress the JAK-STAT pathway in *D. melanogaster*.

Comparative genomic analysis has confirmed the proximity and evolutionary conservation of orthologs for PIAS (SUMO) and SOCS in the JAK-STAT pathway between different mosquito species (Dimopoulos et al., 2000). Signal transduction components exhibit an unexpected mode of evolution even though they share a broadly similar body plan and a considerable number of other features, but they are also substantially different (Waterhouse et al., 2007). Although, in *An. gambiae*, this pathway is controlled through 2 STAT transcription factors, activation of the STAT pathway requires STAT-B mediated activation of STAT-A (Barillas-Mury et al., 1999; Gupta et al., 2009). Activation of this pathway leads to the induction of human complement-like protein known as thioester like proteins-1 (TEP-1) in *An. gambiae* and nitric oxide synthase (NOS) in *An. stephensi* and *An. gambiae*, upon *Plasmodium* infection (Levashina et al., 2001; Luckhart et al., 1998; Gupta et al., 2009).

Almost half of the genes of *Drosophila* and *Anopheles* genomes are interpreted as orthologs and show an average sequence identity of about 56%. However, there are low levels of redundancy in JAK-STAT pathway components; which gives the general idea to describe the conserved roles of JAK-STAT pathway (Zdobnov et al., 2002; Souza-Neto et al., 2009). Although two STAT genes were annotated and analyzed in the complete genome sequence of *An. gambiae*; only one STAT gene has been characterized in other mosquitoes. Similarly, there is only one STAT gene in the analogous insect *Drosophila*. The possibility of having another STAT gene in mosquito species requires further investigation. To gain insight into JAK-STAT pathway evolution, a comprehensive bioinformatics strategy was employed to identify and characterize STAT, PIAS and SOCS genes from diverged *Anopheles* species and functionally characterized in Indian malaria vector *An. stephensi*.

1.3.3 ABC Transporters

ATP-binding cassette (ABC) transporters gene family hydrolyzes the ATP and use energy to transport the compounds across the membrane (Dean, 2008). The functional protein contains two types of protein domains, the one which binds to ATP, known as ATP-binding domains or nucleotide-binding domain (NBD) and another transmembrane domain (TMD). The transmembrane domains form the membranous channel through which compounds are transported and important in determining the specificity of the transported substrate. Genes are classified as ABC transporters based on the sequence identity of the ATP-binding domain(s) which are highly conserved from bacteria to higher organism. Evolutionary analyses of ABC genes and the degree of identity of protein sequence of NBDs has allowed classification of the mammalian ABC genes into 7 subfamilies A-G. Subfamily ABCH has not reported in human (Dean and Allikmets, 1995).

These genes are found abundantly in the genomes of all organisms, consistent with the fact that regulation of transport is essential to all life forms (Dassa and Bouige, 2001). ABC genes are important in the biology of cells and cause many different human genetic diseases and also contribute to the resistance of tumor cells to chemotherapy drugs (Dean and Annilo, 2005). In general, most of its members are involved in transport of ions, amino acids, lipids, sugars, peptides, metals across the membrane (hence named as importers or exporters) and fewer members perform other cellular functions such as regulation of gene expression,

repairing of cell DNA and protein synthesis (Zhao et al., 2004). Importantly the ABC transporters which are exporters in nature are commonly found in eukaryotes however, the transporters with importing property are generally associated with prokaryotes (Wilkinson and Verschueren, 2003) The ABC genes encode a large family of transporters that are expressed on the outer membrane of cell or on intracellular organelles (Higgins, 1992).

This major activity of these transporters is not only helpful for the physiological functioning but also provides many other conditional advantages to the cells. For example, in bacteria ABC transporters concentrate intracellular nutrients, in eukaryotes export products out of cells as well as in immunity (Wu et al., 1991; Abele and Tampe, 2004). Insects, being the pathogens carriers, have also been aimed comprehensively by many researchers to understand the elite features of their ABC transporters so this knowledge can be helpful in designing tools to control the spread of diseases. A genome-wide analysis in some insects reveals the phylogenetic and structural specificity of ABC transporters and also their human transporters equivalent functioning such as transport of metabolic components, tolerance to xenobiotics and also, in addition, the insecticide resistance (Roth et al., 2003). ABC transporters also play a great role in manipulating the host metabolism for adaptations. For example, recent studies in hornworm (*Manduca sexta*) indicate that proficient excretion by P-glycoprotein like transporters reduces the neurotoxic effect of nicotine and provides an adaptation for survival on the tobacco leaves (Murray et al., 1994). Surprising facts from Bt toxins resistance in insects also impart the role of ABC transporter and casts a question to think ahead for the future of Bt crops (Gahan et al., 2010). Insect ABC transporters also play an important role in innate immunity and provide protection against various ranges of pathogens. In a way similar to mammalian ABCs, the insect ABC transporters also somehow regulate the anti-viral immunity. In *Drosophila* the sulfonylurea receptor (SUR), which is the part of $K^{+,}$ channel as well the target for benzoylurea-derived insecticides diflubenzuron, when silenced through RNA interference (RNAi) made the fly sensitive to Flock House Virus (FHV) (Nasonkin et al., 1999). These reports might be helpful in detailed analysis of ABC transporters in *Aedes aegypti* and establishing a comparative database to explore their exclusive aspects in vector immunity against pathogens in general.

The severity of dengue and other mosquito-borne viral infections has been increased for past few years. Some major concerns accelerating the severity and occurrence of these diseases

are associated with the abundance of higher vector population, close vector-human association, vector-friendly environment and evolvement of resistance in pathogens and mosquitoes against drugs and insecticides, respectively. High proximity of densely populated areas provides ample opportunities for the transmission of diseases. Asia's contribution is 70% (approximately 67 million) towards the apparent infections in the comprehensive global disease burden. India contributes 34% to the global infection which amounts to about 33 million infections (Chakravarti et al., 2012; Wichmann et al., 2011; Kakkar, 2012). These facts call for an urgency to explore the host-pathogens interactions either to develop potential drugs/vaccines or strategies to block disease transmission. This warrants the discovery of newer targets which can be potentially exploited and may be possible only after we understand the genomic structure of vectors in natural population.

The genome sequencing project of *Aedes aegypti* provides the scope to evaluate mosquito genes encoding different protein families and verifying them after comparing with the genome of other organisms, especially the insects (Waterhouse et al., 2008). The limitations of advances in the field of mosquito-pathogen interactions demand exploring the available information and analyzing it to find newer targets which can potentially manipulate host-pathogen interactions. The collective efforts of modern molecular techniques and mosquito genome sequencing data might be helpful in this direction. The comparative understanding of insect vectors genome such as *Aedes aegypti* (dengue vector) and *Anopheles gambiae* (malaria vector) reveals that the genome of former is almost 5 times larger than the later one (Nene et al., 2007). In addition, approximately half of the *Aedes* genome consists of transposable elements which are major driving force of genome evolution (Boulesteix and Biémont, 2005). In order to understand the evolution and organization of *Aedes* genomes, a comparative approach is required while keeping natural population polymorphism in mind.

1.4 Current efforts in developing novel control strategies for mosquito borne diseases

Conventional mosquito control has dependent upon killing the parasites using chemotherapy and/or the disease vector mosquitoes using insecticide (e.g. DDT etc.). However, mosquito borne disease has reappeared due to the following factors: the rapid spread of drug resistant parasites; the development of mosquito resistant to insecticides; huge genetic diversity exist between different mosquitoes species of same genus (Hemingway and Ranson, 2000;

Ranson et al., 2011; Manguin and Boëte, 2011). Thus new control strategies are urgently demanded. Current efforts toward new strategies for controlling these diseases are as follows: (1) Development of vaccines (Good et al., 2004; Aponte et al., 2007; The RTS,S Clinical Trials Partnership, (Krishna (Ed.), 2014) including transmission blocking vaccines directed at parasites in the mosquito or the mosquito itself (Dinglasan and Jacobs-Lorena, 2008). (2) Development of new drugs (Noedl, 2013) and (3) development of genetically modified mosquitoes (Harris et al., 2012). More recently, it has been argued that genetic manipulation of mosquito vectorial capacity is a promising by using the new bio weapon (Scholte et al., 2005; 2006; Fang et al., 2011). The first two attempts seem to be effective for a while but not have the longetivity due to drug resistance and even not have the broad spectrum due to huge genetic diversity in mosquito and parasite population worldwide (Takala and Plowe, 2009; Jallow et al., 2009).

The convinced approach for the mosquito control through next two above mentioned attempts was recently been made. The first successful germ line transformation of mosquitoes (transgenesis) was reported for *Aedes aegypti* (Jasinskiene et al., 1998) and later for *An. stephensi* and *An. gambiae* (Catteruccia et al., 2000; Grossman et al., 2001). Replacement of natural vector populations using transgenic parasite-refractory mosquitoes has been proposed as a novel way to control malaria although this strategy is still controversial in many aspects (Boëte, 2011).

Before such a strategy could be excluded, several research objectives must be solved. First, a stable transformation system for a model mosquito *An. gambiae* must be available and documented efficiently, heritable and stable integration of exogenous DNA into mosquito genome (Moreira et al., 2004; Fuchs et al., 2013). This is an essential requirement to investigate the efficacy of parasite-killing candidate genes on parasite development in mosquitoes. The lack of precise genetic tools, however, has been a serious limitation to the in-depth analysis of mosquito immune system (Curtis et al., 2006). Reverse genetic analysis, on the basis of RNAi and transgenic techniques; fill these deficiencies in the research of mosquito innate immunity (Blandin et al., 2002 and Shin et al., 2003).

Second, inducible stage/tissue-specific promoters are essential to express anti-parasite genes at appropriate times and in the correct tissues (James et al., 1999; Moreira et al., 2004). While these and other advances in mosquito transgenesis show the feasibility of generating

parasite-resistant vector (Marrelli et al., 2007; Fang et al., 2011). The characterization of homologous promoters that provide precise control of transgenic expression continues to be a requirement. Third, antiparasitic genes are a prerequisite for disease control through manipulation of vector competence is embedded under the understanding of mosquito-parasite interaction. Which lead to identification of target genes or gene product at the vulnerable points of the parasite`s life cycle in the mosquito vector (Reeves et al., 2012). The discovery of numerous bacterial nucleic acid modification systems has led to the recent development of two modular, precise genome editing tools. The TALEN (transcription activator-like effector nuclease) and CRISPR/Cas (clustered regularly interspaced short palindromic repeats) systems have recently been optimized for research use to site-specifically introduce mutations and manipulate transcriptional activation and repression in a variety of organisms (Mali et al., 2013; Gaj et al., 2013).

1.5 Genetic and genomic diversity as a major research gap

1. More popularly the laboratories data have been generated using model organisms have several limitations viz., the natural host–parasite relationship is different from one geographical area to another.
2. In order to unravel the mechanisms behind the evolution of drug-resistant strains of pathogens and identify better drug targets, a deeper and thorough understanding of the host and parasite's biology is vital.
3. Large evolutionary distance exists between different mosquito species (200 to150 mya) hence functional annotation based on comparative genomics analysis is demanded. Although the maintenance of genome database Flybase or Vectorbase are not completely supporting the end user requirement.
4. Completely understanding an organism at molecular level is quite challenging, because any cellular machinery is highly dynamic and involved in complex interactions with each other. Thus, simply listing out the genes at a particular time point would be insufficient unless the interactions between them are also traced to the level of individual interactions with pathogen.
5. Several genome sequences have rapidly add in NCBI database including 18 *Anopheles* species, 1 *Aedes* and 2 *Culex*. These whole genome shotgun (WGS) sequences are still

unannotated form which can provide immense information related to population dynamics and solve the mystery of genetic variation (**Figure 1.6**).

6. Unavailability of information in genetic diversity and variation among same species complexes are also a matter of mystery. The lack of comprehension of interactions between the two and potentially even three organisms (i.e. parasite and both the vertebrate and invertebrate hosts) is missing, that will become easier if WGS of all "partners" and species complexes become available.

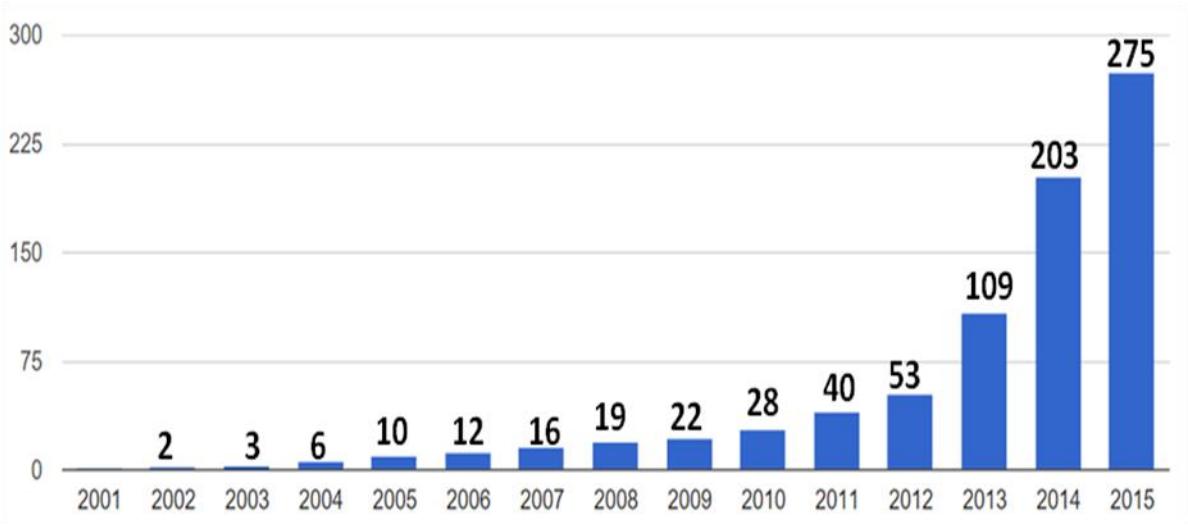


Figure 1.6 Year wise distribution of newly submitted genomes and their annotation, through NCBI. (Source: http://www.ncbi.nlm.nih.gov/genome/annotation_euk/#graphs)

1.6 Research Objectives

The specific objectives of thesis work are as follows-

1. *In silico* analysis of STAT signaling pathway genes in newly submitted genome assemblies of eighteen *Anopheles* species.
2. Characterization of STAT signaling pathway genes from Indian urban mosquito *An. stephensi* and their role in development and antiplasmodial immunity.
3. Identification and molecular characterization of *Aedes aegypti* mosquitoes from the natural environment of Pilani region Rajasthan.
4. Genome wide identification and characterization of ABC transporters gene family in *Aedes aegypti*.

1.7 References

Abele R, Tampe R. 2004. The ABCs of immunology: structure and function of TAP, the transporter associated with antigen processing. Physiology 19:216-224.

Agaisse H, Petersen UM, Boutros M, Mathey-Prevot B, Perrimon N. 2003. Signaling role of hemocytes in *Drosophila* JAK/STAT-dependent response to septic injury. Dev Cell 5(3):441-450.

Aponte JJ, Aide P, Renom M, Mandomando I, Bassat Q, Sacarlal J, et. al., 2007. Safety of the RTS,S/AS02D candidate malaria vaccine in infants living in a highly endemic area of Mozambique: a double blind randomized controlled phase I/IIb trial. Lancet 370:1543-1551.

Arbouzova NI, Zeidler MP. 2006. JAK/STAT signaling in *Drosophila*: Insights into conserved regulatory and cellular functions. Development 133:2605-2616.

Barillas-Mury C, Han YS, Seeley D, Kafatos FC. 1999. *Anopheles gambiae* AgSTAT, a new insect member of the STAT family, is activated in response to bacterial infection. EMBO J. 18:959-967.

Becker S, Groner B, Muller CW. 1998. Three-dimensional structure of the Stat3beta homodimer bound to DNA. Nature 394:145-151.

Blandin S, Moita LF, Kocher T, Wilm M, Kafatos FC, Levashina EA. 2002. Reverse genetics in the mosquito *Anopheles gambiae*: targeted disruption of the defensin gene. EMBO Rep. 3:852-856.

Boëte C. 2011. Scientists and public involvement: a consultation on the relation between malaria, vector control and transgenic mosquitoes. Trans R Soc Trop Med Hyg. 105(12):704-710.

Boulesteix M, Biémont C. 2005. Transposable elements in mosquitoes. Cytogenet Genome Res. 110:500-509.

Brown S, Hu N, Hombría JCG. 2003. Novel level of signaling control in the JAK/STAT pathway revealed by in situ visualization of protein-protein interaction during *Drosophila* development. Development 130:3077-3084.

Castillo JC, Robertson AE, Strand MR. 2006. Characterization of hemocytes from the mosquitoes *Anopheles gambiae* and *Aedes aegypti*. Insect Biochem Mol Biol. 36:891-903.

Catteruccia F, Nolan T, Loukeris TG, Blass C, Savakis C, Kafatos FC, et al., 2000. Stable germline transformation of the malaria mosquito *Anopheles stephensi* Nature 405(6789):959-962.

Chakravarti A, Arora R, Luxemburger C. 2012. Fifty years of dengue in India. Trans R Soc Trop Med Hyg. 106:273-282.

Chen X, Vinkemeier U, Zhao Y, Jeruzalmi D, Darnell JE Jr, Kuriyan J. 1998. Crystal structure of a tyrosine phosphorylated STAT1 dimer bound to DNA. Cell 93:827-839.

Christiansen-Jucht C, Parham PE, Saddler A, Koella JC, Basáñez MG. 2014. Temperature during larval development and adult maintenance influences the survival of *Anopheles gambiae* ss. Parasit Vectors. 7:489.

Christophers SR. 1915. The male genitalia of *Anopheles*. Ind J Med Res. 3:371-394.

Cirimotich CM, Dong Y, Garver LS, Sim S, Dimopoulos G. 2010. Mosquito immune defenses against *Plasmodium* infection. Dev Comp Immunol. 34:387-395.

Clements AN. 1992. The Biology of Mosquitoes. London: Chapman and Hall.

Cohuet A, Harris C, Robert V, Fontenille D. 2010. Evolutionary forces on *Anopheles*: What makes a malaria vector? Trends Parasitol. 26:130-136.

Curtis CF, Coleman P, Kelly DW, Campbell-Lendrum DH. 2006. Advantages and limitations of transgenic vector control: sterile males vs. gene drivers. In: Boëte C. (ed.) Genetically modified mosquitoes for malaria control. Georgetown, USA: Landes Biosciences-Eurekah. 60-78.

Dassa E, Bouige P. 2001. The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. Res Microbiol. 152:211-229.

Dean M, Allikmets R. 1995. Evolution of ATP-binding cassette transporter genes. Curr Opin Genetics Dev. 5:779-785.

Dean M, Annilo T. 2005. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet 6:123-142.

Dean M. 2008, ATP-binding Cassette (ABC) Transporter Supergene Family: Genetics and Evolution. In: Encyclopedia of Life Sciences (ELS). John Wiley and Sons, Ltd: Chichester.

Dinglasan RR, Jacobs-Lorena M. 2008. Flipping the paradigm on malaria transmission-blocking vaccines. Trends Parasitol. 24(8):364-370.

Dimopoulos G, Casavant L, Chang S, Scheetz T, Roberts C, Donohue M et al., 2000. *Anopheles gambiae* pilot gene discovery project: Identification of novel mosquito innate immunity genes from ESTs generated from immune competent cell lines. Proc Natl Acad Sci USA 97:6619-6624.

Dong X, Zhou H, Gong Z-D, Lu BL. 2004. Investigation of mosquito species in Yunnan province with some new species. Chin J Vec Biol Contr. 15:186-191.

Fang W, Vega-Rodriguez J, Ghosh AK, Jacobs-Lorena M, Kang A, St Leger RJ. 2011. Development of transgenic fungi that kill human malaria parasites in mosquitoes. Science 331:1074-1077.

Farnesi LC, Menna-Barreto RF, Martins AJ, Valle D, Rezende GL. 2015. Physical features and chitin content of eggs from the mosquito vectors *Aedes aegypti*, *Anopheles aquasalis* and *Culex quinquefasciatus*: Connection with distinct levels of resistance to desiccation. J insect physiol. 83:43-52.

Fuchs S, Nolan T, Crisanti A. 2013. Mosquito transgenic technologies to reduce *Plasmodium* transmission. Methods Mol Biol. 923:601-622.

Gahan LJ, Pauchet Y, Vogel H, Heckel D. 2010. An ABC transporter mutation is correlated with insect resistance to *Bacillus thuringiensis* Cry1Ac toxin. PLoS Genet. 6:e1001248.

Gaj T, Gersbach CA, Barbas CF. 2013. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31:397-405.

Garcia ES, Castro DP, Figueiredo MB, Genta FA, Azambuja P. 2009. *Trypanosoma rangeli*: a new perspective for studying the modulation of immune reactions of *Rhodnius prolixus*. Parasit Vectors. 2:33

Ghosh A, Mukherjee S, Dutta T, Roy S, Mondal FB. 2013. Community perceptions regarding mosquito borne diseases in some selected localities of Bankura, a peri-urban area of West Bengal, India. J Mosq Res. 3(7):54-57.

Good MF, Stanisic D, Xu H, Elliott S, Wykes M. 2004. The immunological challenge to developing a vaccine to the blood stages of malaria parasites. Immunol Rev. 201:254-267.

Gottar M, Gobert V, Michel T, Belvin M, Duyk G, Hoffmann JA, et al., 2002. The *Drosophila* immune response against gram-negative bacteria is mediated by a peptidoglycan recognition protein. Nature 416:640-644.

Grossman GL, Rafferty CS, Clayton JR, Stevens TK, Mukabayire O, Benedict MQ. 2001. Germline transformation of the malaria vector, *Anopheles gambiae*, with the piggyBac transposable element Insect Mol Biol. 10(6):597-604.

Gupta L, Molina-Cruz A, Kumar S, Rodrigues J, Dixit R, et al., 2009. The STAT pathway mediates late-phase immunity against *Plasmodium* in the mosquito *Anopheles gambiae*. Cell Host Microbe 5:498-507.

Harbach RE. 2004. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. Bull Entomol Res. 94:537-553.

Harris AF, McKemey AR, Nimmo D, Curtis Z, Black I, Morgan SA, et al., 2012. Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. Nature Biotechnol. 30:828-830.

Hay SI, Sinka ME, Okara RM, Kabaria CW, Mbithi PM, Tago CC, et al., 2010. Developing global maps of the dominant *Anopheles* vectors of human malaria. PLoS Med. 7(2):e1000209.

Hemingway J, Ranson H. 2000. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol. 45:371-391.

Higgins CF. 1992. ABC transporters: from micro-organisms to man. Annu Rev Cell Biol. 8:67-113.

Hillyer JF, Schmidt SL, Christensen BM. 2003a. Rapid phagocytosis and melanization of bacteria and *Plasmodium* sporozoites by hemocytes of the mosquito *Aedes aegypti*. J Parasitol. 89(1):62-69.

Hillyer JF, Schmidt SL, Christensen BM. 2003b. Hemocyte-mediated phagocytosis and melanization in the mosquito *Armigeres subalbatus* following immune challenge by bacteria. Cell Tissue Res. 313(1):117-127.

Hillyer JF. 2010. Mosquito immunity. Adv Exp Med Biol. 708: 218-238.

Hoffmann JA, Reichhart JM. 2002. *Drosophila* innate immunity: an evolutionary perspective. Nature Immunol. 3(2):121-126.

Horvath CM. 2000. STAT proteins and transcriptional responses to extracellular signals. Trend Biochem Sci. 25:496-502.

Impoinvil DE, Cardenas GA, Gihture JI, Mbogo CM, Beier JC. 2007. Constant temperature and time period effects on *Anopheles gambiae* egg hatching. J Am Mosq Control Assoc. 23(2):124.

Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, et al., 2009. Genome-wide and fine resolution association analysis of malaria in West Africa. Nature Genet. 41:657-665.

James AA, Beerntsen BT, Capurro Mde L, Coates CJ, Coleman J, Jasinskiene N, et al., 1999. Controlling malaria transmission with genetically-engineered, *Plasmodium*-resistant mosquitoes: milestones in a model system. Parassitologia 41:461-471.

Jarośík V, Kratochvíl L, Honék A, Dixon AF. 2004. A general rule for the dependence of developmental rate on temperature in ectothermic animals. Proc R Soc Lond B Biol Sci. 271(Suppl 4):S219-21.

Jasinskiene N, Coates CJ, Benedict MQ, Cornel AJ, Rafferty CS, James AA, et al., 1998, Stable transformation of the yellow fever mosquito, *Aedes aegypti*, with the Hermes element from the housefly. Proc Natl Acad Sci USA 95:3743-3747.

Jiang H, Patel PH, Kohlmaier A, Grenley MO, McEwen DG, Edgar BA. 2009. Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the *Drosophila* midgut. Cell 137:1343-1355.

Kakkar M. 2012. Dengue fever is massively under-reported in India, hampering our response. British Med J. 345: e8574.

Kingsolver MB, Huang Z, Hardy RW. 2013. Insect antiviral innate immunity: pathways, effectors and connections. J Mol Biol. 425:4921-4936.

Krishna S (Ed.) 2014. The RTS,S clinical trials partnership. Efficacy and safety of the RTS,S/AS01 malaria vaccine during 18 months after vaccination: A phase 3 randomized, controlled trial in children and young infants at 11 African sites. PLoS Med 11(7): e1001685

Krzywinski J, Besansky NJ. 2003. Molecular systematic of *Anopheles*: from subgenera to subpopulations. Ann Rev Entomol. 48:111-139.

Lehane MJ. 1997. Peritrophic matrix structure and function. Ann Rev Entomol. 42:525-550.

Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC. 2001. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae.* Cell 104:709-718.

Levy DE, Darnell JE Jr. 2002. Stats: transcriptional control and biological impact. Nat Rev Mol Cell Biol. 3:651-662.

Li WX. 2008. Canonical and non-canonical JAK-STAT signaling. Trends Cell Biol. 18(11): 545-551.

Luckhart S, Vodovotz Y, Cui L, Rosenberg R. 1998. The mosquito *Anopheles stephensi* limits malaria parasite development with inducible synthesis of nitric oxide. Proc Natl Acad Sci USA. 95:5700-5705.

Malhotra PR, Jatav PC, Chauhan RS. 2000. Surface morphology of the egg of *Anopheles stephensi stephensi* sensu stricto (Diptera, Culicidae). Ital J Zool. 67(2):147-151.

Mali P, Esvelt KM, Church GM. 2013. Cas9 as a versatile tool for engineering biology, Nat Methods, 10:957-963.

Manguin S, Carnivale P, Mouchet J. 2008. Biodiversity of Malaria in the World (John Libbey Eurotext, Paris, 2008)

Manguin S, Boëte C. 2011. Global impact of mosquito biodiversity, human vector-borne diseases and environmental change. INTECH Open Access Publisher; pp 27-50.

Marrelli MT, Li CY, Rasgon JL, Jacobs-Lorena M. 2007. Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on *Plasmodium*-infected blood. Proc Natl Acad Sci USA 104:5580-5583.

McFarlane M, Arias-Goeta C, Martin E, O'Hara Z, Lulla A, et al., 2014. Characterization of *Aedes aegypti* innate immune pathways that limit chikungunya virus replication. PLoS Negl Trop Dis. 8(7): e2994.

Mereta ST, Yewhalaw D, Boets P, Ahmed A, Duchateau L, Speybroeck N, et al., 2013. Physico-chemical and biological characterization of anopheline mosquito larval habitats (Diptera: Culicidae): implications for malaria control. Parasit Vectors. 6(1):320.

Merritt RW, Dadd RH, Walker ED. 1992. Feeding behavior, natural food and nutritional relationships of larval mosquitoes. Ann Rev Entomol. 37(1):349-374.

Merzendorfer H, Zimoch L. 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. J Exp Biol. 206:4393-4412.

Molina-Cruz A, Barillas-Mury C. 2014. The remarkable journey of adaptation of the *Plasmodium falciparum* malaria parasite to New World anopheline mosquitoes. Mem Inst Oswaldo Cruz 109(5):662-667.

Molina-Cruz A, Canepa GE, Kamath N, Pavlovic NV, Mu J, Ramphul UN, et al., 2015. *Plasmodium* evasion of mosquito immunity and global malaria transmission: The lock-and-key theory. Proc Natl Acad Sci USA 112(49): 15178-15183.

Moreira LA, Wang J, Collins FH, Jacobs-Lorena M. 2004. Fitness of anopheline mosquitoes expressing transgenes that inhibit *Plasmodium* development. Genetics 166:1337-1341.

Murray CL, Quaglia M, Arnason JT, Morris CE. 1994. A putative nicotine pump at the metabolic blood–brain barrier of the tobacco hornworm. J Neurobiol. 25:23-34.

Nasonkin I, Alikasifoglu A, Ambrose C, Cahill P, Cheng M, Sarniak A. et al., 1999. A novel sulfonylurea receptor family member expressed in the embryonic *Drosophila* dorsal vessel and tracheal system. J Biol Chem. 274:29420-29425.

Neafsey DE, Christophides GK, Collins FH, Emrich SJ, Fontaine MC, Gelbart W, et al., 2013. The evolution of the *Anopheles* 16 genomes project. G3 (Bethesda) 3:1191-1194.

Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al., 2015. Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. Science 347(6217):1258522.

Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al., 2007. Genome sequence of *Aedes aegypti* a major arbovirus vector. Science 316:1718-1723.

Noedl H. 2013. The need for new antimalarial drugs less prone to resistance Curr Pharm Des.19(2):266-269.

Ortmann RA, Cheng T, Visconti R, Frucht DM, O'Shea JJ. 2000. Janus kinases and signal transducers and activators of transcription: their roles in cytokine signaling, development and immunoregulation. Arthritis Res. 2:16-32.

Raikhel AS, Kokoza VA, Zhu J, Martin D, Wang SF, et al., 2002. Molecular biology of mosquito vitellogenesis: from basic studies to genetic engineering of antipathogen immunity. Insect Biochem Mol Biol. 32:1275-1286.

Ramphul UN, Garver LS, Molina-Cruz A, Canepa GE, Barillas-Mury C. 2015. *Plasmodium falciparum* evades mosquito immunity by disrupting JNK-mediated apoptosis of invaded midgut cells. Proc Natl Acad Sci USA. 112(5):1273-1280.

Ranson H, N'Guessan R, Lines J, Moiroux N, Nkuni Z, Corbel V. 2011. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? Trends Parasitol. 27(2):91-98.

Reeves RG, Denton JA, Santucci F, Bryk J, Reed FA. 2012. Scientific standards and the regulation of genetically modified insects. PLoS Negl Trop. 6:e1502.

Roth CW, Holm I, Graille M, Dehoux P, Rzhetsky A, et al., 2003. Identification of the *Anopheles gambiae* ATP-binding cassette transporter superfamily. Mol Cells. 15:150-158.

Robinson GG. 1939. The mouth parts and their function in the female mosquito, *Anopheles maculipennis*. Parasitol. 31: 212-242.

Scholte EJ, Ng'habi K, Kihonda J, Takken W, Paaijmans K, Abdulla S, et al., 2005. An entomopathogenic fungus for control of adult African malaria mosquitoes. Science 308:1641-1642.

Scholte EJ, Knols BG, Takken W. 2006. Infection of the malaria mosquito *Anopheles gambiae* with the entomopathogenic fungus *Metarhizium anisopliae* reduces blood feeding and fecundity. J Invertebr Pathol. 91:43-49.

Shi S, Larson K, Guo D, Lim SJ, Dutta P, Yan SJ, et al., 2008. *Drosophila* STAT is required for directly maintaining HP1 localization and heterochromatin stability. Nat Cell Biol. 10:489-496.

Shin SW, Kokoza VA, Raikhel AS 2003. Transgenesis and reverse genetics of mosquito innate immunity. J Exp Biol. 206:3835-3843.

Shin SW, Kokoza V, Bian G, Cheon HM, Kim YJ, Raikhel AS. 2005. REL1, a homologue of *Drosophila* dorsal, regulates toll antifungal immune pathway in the female mosquito *Aedes aegypti*. J Biol Chem. 280:16499-16507.

Shin SW, Bian G, Raikhel AS. 2006. A toll receptor and a cytokine, Toll5A and Spz1C, are involved in toll antifungal immune signaling in the mosquito *Aedes aegypti*. J Biol Chem. 281:39388-39395.

Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, et al., 2010a. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. Parasit Vectors 3(72):1-26.

Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, et al., 2010b. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. Parasit Vectors 3(117):1-34.

Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, et al., 2011. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. Parasit Vectors 4(89):1-46.

Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al., 2012. A global map of dominant malaria vectors. Parasit Vectors 5(1): 69.

Souza-Neto JA, Sim S, Dimopoulos G. 2009. An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. Proc Natl Acad Sci USA 106: 17841-17846.

Takala SL, Plowe CV. 2009. Genetic diversity and malaria vaccine design, testing and efficacy: preventing, overcoming vaccine resistant malaria. Parasite Immunol. 31:560-573.

Truman JW, Riddiford LM. 1999. The origins of insect metamorphosis. Nature 401:447-52.

Vilmos P, Kurucz E. 1998. Insect immunity: evolutionary roots of the mammalian innate immune system. Immunol Lett. 62:59-66.

Wahid I, Sunahara T, Mogi M. 2003. Maxillae and mandibles of male mosquitoes and female autogenous mosquitoes (Diptera: Culicidae). J Med Entomol 40: 150-158.

Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, et al., 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science 316(5832):1738-1743.

Waterhouse RM, Wyder S, Zdobnov EM. 2008. The *Aedes aegypti* genome: a comparative perspective. Insect Mol Biol. 17(1):1-8.

Wichmann O, Yoon IK, Vong S, et al., 2011. Dengue in Thailand and Cambodia: an assessment of the degree of under recognized disease burden based on reported cases. PLoS Negl Trop Dis. 5(3): e996.

Wilkinson AJ, Verschueren KHG. 2003. Crystal structures of periplasmic solute binding proteins in abc transport complexes illuminate their function. In ABC Proteins: From Bacteria to Man. Edited by Holland IB, Cole SPC, Kuchler K, Higgins CF. London: Academic Press; 187-208.

Wu CT, Budding M, Griffin MS, Croop JM. 1991. Isolation and characterization of *Drosophila* multidrug resistance gene homologs. Mol Cell Biol. 11:3940-3948.

Xi Z, Ramirez JL, Dimopoulos G. 2008. The *Aedes aegypti* toll pathway controls dengue virus infection. PLoS Pathog 4: e1000098.

Zdobnov EM, Mering CV, Letunic I, Torrents D, Suyama M, Copley RR, et al., 2002. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. Science 298:149-159.

Zeidler MP, Bach EA, Perrimon N. 2000. The roles of the *Drosophila* JAK/STAT pathway. Oncogene 19:2598-2606.

Zhao Z, Fang LL, Johnsen R, Baillie DL. 2004. ATP-binding cassette protein E is involved in gene transcription and translation in *Caenorhabditis elegans*. Biochem Biophys Res Commun. 323:104-111.

Objective: In silico analysis of STAT signaling pathway genes of different Anopheles species

Part of this chapter has been published in: Dhawan R[#], Gupta K[#], Kajla M, Kumar S, Gakhar SK, Kakani P, Choudhury TP, Gupta L, Molecular characterization of SOCS gene and its expression analysis on *Plasmodium berghei* infection in *Anopheles culicifacies*. **Acta Tropica** 2015, 152:170–175. (# Equal Contribution)

2.1 Introduction

2.1.1 STAT pathway: an evolutionary perspective

The Signal Transducer and Activator of Transcription (STAT) is a tyrosine kinase signaling pathway first identified in mammals through the study of interferon-α (IFN-α) and interferon γ (IFN-γ) induced transcriptional activation (Darnell et al., 1994). The STAT pathway has been found conserved during the evolutionary time and studied in many model organisms including mouse *Mus domesticus*, zebrafish *Danio rerio*, fruitfly *Drosophila melanogaster*, nematode *Caenorhabditis elegans* and the slime mould *Dictyostelium discoideum* (Dearolf, 1999). Thorough investigations of all model organisms have revealed that this pathway is regulated by the homologues of at least one STAT which is highly conserved (Hurst et al., 2004). The STAT signaling pathway was found in early metazoans to higher animals except plants (Aaronson and Horvath, 2002).

The availability of high-quality whole genome sequences from variety of organisms, including *Dictyostelium*, insects, mollusks, nematodes and various vertebrate animals, allowed us to systematically investigate the evolutionary process of the genes. The STAT family of transcription factors is involved in both differentiation and signaling processes. It is absent from the yeast genome but present in a facultative metazoan *Dictyostelium* (slime mold) in which it mediates the selective expression of an extracellular matrix protein in pre-stalk cells, in response to a differentiation-inducing factor (Kawata et al., 1997). Surprisingly, *Dictyostelium*, the only known non-metazoan organism having 4 STAT genes STATa, b, c and STATd which possesses protein kinases and Src2 homology (SH2) domains with remarkable divergent function which suggests a non-canonical mode of activation of STAT pathway (Kimmel et al., 2004). Therefore, molecular phylogenetic analysis indicated that the slime mold STATs form a distinct clade, which raises the possibility of mixed intensive and birth-and-death evolution of STAT family of transcription factors (Zhukovskaya et al., 2004). Consequently, the discovery of STAT family members in the slime mold placed this phosphotyrosine signaling pathway at the beginning of multicellular evolution.

On the next hierarchical order, a demosponge *Amphimedon queenslandica*, a nematode *C. elegans* and other genome sequenced insects including *Drosophila, Apis* a single STAT gene have been identified (Wang and Levy, 2006). In round worm, *C. elegans* only STAT homologs (STA-1) are found which also shows very different and primitive functions (Wang and Levy, 2006; Zhukovskaya et al., 2004). Along with *Drosophila* (Order Diptera), honeybee *Apis mellifera*, (Order Hymenoptera) silk moth, *Bombyx mori*

(Order Lepidoptera) and other major insect orders only one homolog of STAT have been reported (Wang and Levy, 2012) (**Table 2.1**). In *Drosophila*, STAT gene was invented as a key player which involved in many developmental processes and maintained its conserved function as in other living organisms (Yan et al., 1996).

Table 2.1 STAT gene(s) in early metazoan, lower and higher invertebrate animals.

Organism	Common Name	Taxonomical category	No. of STAT gene in genome	Accession No.	Reference
Amphimedon queenslandica	Demo sponge	Porifera	1	XP_011402870	Srivastava et al., 2010
Caenorhabditis elegans	Round worm	Nematode	1	AAY18583	Wang and Levy, 2006
Penaeus monodon	Black tiger shrimp	Crustacea	1	AAQ94739	Chen et al., 2008
Artemia franciscana	Brine Shrimp	Crustacea	1	ACJ63721	Cheng et al., 2010
Bombyx mori	Silk Worm	Lepidoptera	1	ACR61178	Tanaka et al. 2008
Bombus terrestris	Buff-tailed Bumble Bee	Hymenoptera	1	XP_003401031	Sadd et al., 2015
Apis mellifera	Honey bee	Hymenoptera	1	XP_397181	NCBI's Annotation
Apis florae	dwarf honey bee	Hymenoptera	1	XP_012339763	NCBI's Annotation
Apis dorsata	Giant honeybee	Hymenoptera	1	XP_006608218	NCBI's Annotation
Aedes albopictus	Dengue mosquito	Diptera	1	AY299686.1	Lin et al., 2004
Culex tritaerhynchus	Filarial mosquito	Diptera	1	AY299687.1	Lin et al., 2004
Biomphalaria glabrata	Fresh water Snail	Mollusca	2	FJ804763 FJ804764	Zhang and Coultas 2011
Crassostrea gigas	Pacific oyster	Mollusca	2	EKC37808 XP_011437452	Zhang et al., 2012
Ciona intestinalis	Sea Squirt	Urochordata	2	BAE06716-A BAE06717-B	Hino et al., 2003
Saccoglossus kowalevskii	Acorn Worm	Hemichordata	2	XP_006814942 XP_006814944	NCBI's Annotation

Analysis of the first sequenced genome of a medically important *An. gambiae* (order diptera) mosquito revealed the early occurrence of two STAT genes in its genome. Later at above hierarchy, a mollusk *Biomphalaria glabrata* (Freshwater snail) and a tunicate *Ciona intestinalis* (Sea squirt) also show two STAT transcription factors (Zhang and Coultas, 2011; Hino et al., 2003). Another STAT gene contributes an additional support for the gene duplication theory, comes from the discovery of STAT5a

and STAT5b genes in zebrafish (Lewis and Ward, 2004). A canonical STAT signaling pathway in the mosquito *An. gambiae* revealed two stat genes originated through retro-duplication (Gupta et al., 2009). All insect STATs are predicted to be identical to mammalian STATs in domain structure, in contrast to the partial identities in the *Dictyostelium* and nematode STATs, suggesting that STAT evolution by domain accretion stopped before the rise of Deuterostomes over a billion years ago. The insect STATs also form a single clad in phylogenetic analysis, and constitute an ancient class of STATs with the clad consisting of STAT5s and 6 (Bach and Perrimon, 2003).

2.1.2 Regulation of STAT pathway

The fruit fly, *D. melanogaster* and the malaria mosquito *An. gambiae*, both are highly adapted, successful dipteran species that diverged about 250 million years ago (Gaunt and Miles, 2002). They share a broadly similar body plan and a considerable number of other features, but they are also substantially different in terms of ecology, morphology, life style and genome size (the *Anopheles* genome is twice the size of that of *Drosophila*) (Zdobnov et al., 2002). The detailed comparative study at genome and proteome level with malaria mosquito *An. gambiae* and *Drosophila* has revealed major differences in immunity-related and signaling pathway genes (Christophides et al., 2002).

It has long been anticipated that duplication of gene is a major driving force for genomic and organismal complexity during evolution (Liongue et al., 2012). However, the mechanism and evolutionary details of gene duplication related evidence remain largely unknown. The only way to get the direct insight into this dynamic process will be gain from individualized comparative genomic analyses, particularly those focusing on families of paralogous genes (Skrabanek and Wolfe, 1998). Gene duplication is thought to be generated by three types of mechanisms, chromosomal unequal crossing over, retrotransposition, and chromosomal (or genome) duplication, the outcomes of which are quite different (Zhang, 2003). With over one million representatives, arthropods have most diverse group of animals and likely shared the last predecessor with vertebrates at least one billion years ago. Evidence showed that during the evolution of vertebrates from early deuterostome ancestors entire genomes were duplicated through two rounds of duplications (the ‘one-to-two-to-four’ rule) (Meyer and Schartl, 1999). The first genome duplication in chordate evolution might predate the Cambrian explosion. The second genome duplication possibly dates back to the early Devonian (Sidow, 1996; Dehal and Boore, 2005).

The JAK-STAT pathway of *Drosophila* is quite simple and well studied; their molecular and functional data clearly indicate that a high level of conservation exists between the insect and mammalian pathways (Bach and Perrimon, 2003). In *Drosophila* the JAK-STAT pathway encodes, three unpaired ligands domain (upd), Upd1, Upd2 and Upd3 which secrete the glycosylated and extracellular matrix binding protein for the pathway (Wright et al., 2011; Harrison et al., 1998). The domeless gene encodes a transmembrane receptor protein called Domeless (dome) which shows most similarity to mammalian IL-6 receptor family (Brown et al., 2001; Chen et al., 2002). The hopscotch gene encodes a *Drosophila* JAK homologue, hop which is maternally essential for the establishment of normal embryonic segmentation. The hopscotch (hop) protein shares all the characteristics of mammalian JAK family non-receptor tyrosine kinases and similar to human JAK2 (27% identity) with high homology in kinase and kinase-like domains (Binari and Perrimon, 1994). Likely to human, in *Drosophila* STAT proteins (Stat92E) have been discovered with a highly conserved SH2 domain and a single tyrosine (Y) residue at the C-terminus which will be phosphorylated by JAKs upon activation (Yan et al., 1996). Stat92E is the only signal transducer and transcription activator of *Drosophila* which is highly homologous to human STAT5a (37% identity). Stat92E includes a SH2 domain, DNA binding domain and a single tyrosine residue around position 700 found in STAT-like genes (Hou et al., 1996).

Drosophila genome encodes various negative regulators of JAK-STAT pathway to suppress the signaling, including three suppressor of cytokine signaling (SOCS) proteins, termed SOCS16D, SOCS36E and SOCS44A. Of these, SOCS36E is the best-characterized family member, with closest homology to mammalian SOCS5. Transcription of SOCS36E mRNA is JAK-STAT pathway regulated (Karsten et al., 2002 and Stec et al., 2013) and the resulting protein has been shown to negatively regulate both JAK-STAT and epidermal growth factor receptor (EGFR) signaling *in vivo* (Stec et al., 2013). At the molecular level, all SOCS proteins are recognized by a C-terminally located SOCS-box domain, through which Elongin B and Elongin C interact and subsequently recruit Cullin-5 and Rbx-1 protein. This complex, termed the Elongin-Cullin-SOCS (ECS) complex which eventually help in the ubiquitinization of dome receptor (Babon et al., 2009). Centrally located SH2 domain mediates the interaction with phosphorylated tyrosine (pTyr) residue; SOCS molecule fulfills the role of substrate recognition in the ECS complex. By contrast, N-terminal of SOCS5 of *Drosophila* protein does not contain any recognizable domains and share low level of conservation

between family members (Stec et al., 2013). However, in mouse SOCS5 gene shows some conservation in N-terminal region known as N-terminal conserved region (NTCR) which is 70 amino acids long and have JAK interacting region (JIR) (Feng et al., 2012; Chandrashekharan et al., 2015).

Another possible means of directly suppress the pathway through protein inhibitor of activated STAT (dPIAS) which was first discovered as an inhibitors of activated STATs in vertebrate system (Darnell, 1997a). However, the physiological relevance of PIAS *in vivo* was first established with its *Drosophila* PIAS homologues, dPIAS (originally known as zimp) (Mohr and Boswell, 1999). Betz et al. demonstrated in *Drosophila*, the *in vivo* functional interaction of the dPIAS proteins with stat92E which bind and block the DNA binding domain of activated STATs and strongly inhibit STAT-driven transcription (Betz et al., 2001). Reduction or over expression of dPIAS in *Drosophila* leads to increases or decreases of JAK-STAT activity respectively, suggesting that dPIAS acts as a negative regulator of the pathway identified by RNAi experiments (Muller et al., 2005).

Interestingly, a study in mammalian systems suggests that PIAS proteins have miscellaneous functions including their roles as SUMO (small ubiquitin-like modifier)-ligases. These proteins have been shown to function as E3 ligase that promotes the SUMO modification of a number of transcription regulators (Jackson, 2001). While genetic studies in *Drosophila* have shown that dPIAS is in fact allelic to the Su(var)2-10 locus and regulates the chromosome structure and function (Hari et al., 2001). For most of the non-model organisms it is not feasible to dissect individual components of this complicated pathway. Because transcription factors are the key molecules in any regulatory pathway, and to investigate the other components will augment the complications. Mosquitoes are also belonging to the order diptera (class insecta) but their genomes are quite larger than *Drosophila*. Most of the *Anopheles* genomes have very recently been submitted and presently they are in unannotated form. These genomic resources provide the way for detailed comparative analyses within and across the species level. Here we tried to summarize the extension and evolution of STAT transcription factor along with two main negative regulators (SOCS5 and PIAS) in diverse *Anopheles* mosquitoes of different geographical location and belong to different non-ranked taxonomic groups.

2.2 Materials and Methods

2.2.1 Retrieval of genome information of sequenced *Anopheles* species

An. gambiae is the most studied malaria vector species throughout the research community but other *Anopheles* species are equally important to understand the host parasite interaction. Afrotropical major malaria vector, *An. gambiae* Pest strain genome assembly (AgamP3) is available for the malaria research community since 2002 and was used as a single reference till now (Holt et al., 2002; Sharakhova et al., 2007). Currently, whole genome sequences of different *Anopheles* species (major and minor) from various geographical regions are available at NCBI. Their whole genome shotgun (WGS) sequences are still in unannotated form which can be used to solve the mystery of genetic variation within the same genus (Neafsey et al., 2013). 16 *Anopheles* species were selected for genome sequencing with varying degrees of vectorial capacity and a wide range of geographic locations and ecological conditions (Besansky, 2008; Fontaine et al., 2015). The sequenced Anophelines fall under 3 main subgenera (*Cellia*, *Nyssorhynchus*, and *Anopheles*) (**Flowchart 1 of chapter 1**). The details related to their NCBI taxid number, assembly name, release date, GenBank assembly number and genome size have been mentioned in the **Table 2.2**.

Table: 2.2 Description of retrieved information related to sequencing of *Anopheles* mosquito genomes considered for present study

Mosquito Name	**Taxid No.**	**Genome Assembly Name**	**GenBank Assembly No.**	**No. of scaffold**	**Genome size (Mb)**
An. gambiae	62324	AgamP3	AgamP5	-	265.1
An. stephensi	30069	ASM30077v2	GCA_000300775.2	23371	221.3
An. minimus	112268	MINIMUS1_V1	GCA_000349025.1	678	201.8
An. quadriannulatus	34691	QUAD4_A_V1	GCA_000349065.1	2823	283.8
An. funestus	62324	FUMOZ_V1	GCA_000349085.1	1392	225.2
An. epiroticus	199890	epiroticus2_V1	GCA_000349105.1	2673	223.5
An. albimanus	7167	ALBI9_A_V1	GCA_000349125.1	204	170.5
An. dirus	7168	WRAIR2_V1	GCA_000349145.1	1266	216.3
An. christyi	43041	ACHKN1017_V1	GCA_000349165.1	30369	172.7
An. arabiensis	7173	DONG5_A_V1	GCA_000349185.1	1214	246.6
An. sinensis	74873	AS2	GCA_000441895.2	9592	220.8
An. maculatus	74868	maculatus3_V1	GCA_000473185.1	47797	141.9
An. culicifacies	139723	A-37_1_V1	GCA_000473375.1	16162	202.9
An. farauti	69004	FAR1_V2	GCA_000473445.2	310	183.1
An. atroparvus	41427	EBRO_V1	GCA_000473505.1	1371	224.3
An. melas	34690	CM101059_A_V2	GCA_000473525.2	20229	224.2
An. merus	30066	MAF_V1	GCA_000473845.2	2027	288.1
An. darlingi	43151	A_darlingi_V1	GCA_000211455.3	2220	136.9

2.2.2 *In silico* prediction and annotation of STAT pathway genes

Here, we have analysed four genes of STAT signaling pathway which includes two transcription factors STAT-A, STAT-B and their regulatory proteins PIAS and SOCS5. These four genes have been identified and annotated from all genome sequenced *Anopheles* species. For full gene prediction we used previously reported gene sequences (intentionally laboratory verified sequences only) of *Anopheles* species to get the maximum and reliable homology through genome-wide BLASTN program (**Table 2.3**). Here, lab verified sequence of *An. gambiae* and/or *An. stephensi*; *An. culicifacies* were used as a model *Anopheles* species to identify the STAT pathway genes into their neighboring *Anopheles* species. We selected AgSTAT-A, 3568bp (Accession No FJ792607) and AsSTAT-A, 2639 bp (KR779999) for retrieval of STAT-A gene from other Anophelines. AgSTAT-B, 2918bp (Accession No AJ010299) and AsSTAT-B, 2540bp (KR780000) sequences were used as a query to search the orthologs of STAT-B gene. For PIAS, lab verified gene sequence of *An. gambiae* was not available hence *An. aquasalis* (AqPIAS) sequence (Accession No HM851177) sequence was used as a query. For SOCS *An. stephensi*, 1194 bp (Accession No KU306401) or *An. culicifacies* partial sequence, 975 bp (Accession No KJ914628) were utilized to retrieve the full gene.

Table 2.3 List of lab-verified cDNA sequences used as a query for genome-wide homology searches of STAT pathway genes

Genes	GenBank ID	Gene size (bp)	5'UTR (bp)	CDS (bp)	3'UTR (bp)	Protein length (aa)	Reference
AgSTAT-A/ AsSTAT-A	FJ792607.1 KR779999	3568 2639	1064 227	2259 2259	245 153	752 752	Gupta et al. 2008 Present study
AgSTAT-B/ AsSTAT-B	AJ010299/ KR780000	2918 2540	450 271	2169 2231	299 38	722 743	Barillas-Mury et al., 1999; Present study
AqPIAS	HM851177	2407	211	1953	243	651	Bahia et al., 2011
AcSOCS5/ AsSOCS5/ AgSOCS5	KJ914628/ KU306401/ EF631979	- 1194 1617	- - 316	975 1194 1197	- - 110	325 397 398	Dhawan et al., 2015 Present study

Using these above cDNA sequences, different *Anopheles* species genome sequences were searched through BLAST for their respective contigs/supercontigs comprises the respective genes, individually. BLASTN searches were limited by *Anopheles* species taxid numbers in respective WGS dataset (**Table 2.2**). Matched region of the contigs with good query coverage and identity were retrieved and considered as predicted gene (gDNA sequence). These contigs were subjected to identify putative exons and introns

boundaries from Genscan programme (Burge and Karlin, 1997). Predicted CDS were used for translated BLAST search (tBLASTn) to verify the relevant genes. On confirmation of the CDS; sequences were then translated through ExPASY translate tool to get the deduced amino acid sequence (Gasteiger et al., 2003).

2.2.3 Sequence alignment and domain search

All deduced amino acid sequences were used for multiple sequence alignment using Clustal omega to fetch out their conserved domains and consensus motifs. Conserved domains and motifs in these predicted proteins sequences were analyzed using the Conserved Domain (CD) search tool at NCBI which integrates search with Pfam databases (Marchler-Bauer and Bryant, 2004; Marchler-Bauer and Anderson, 2007; Finn et al., 2008). Determination of their protein domain range through their related CDD link was done. Comparisons were made for the size of proteins domain along with genus specific conserved motif sequences for gene of interest. Amino acid and nucleotide sequences were aligned with the default parameters and assembled to generate amino acid and nucleotide percentage identity matrices for each gene of STAT pathway in Clustal omega web tool. (http://www.ebi.ac.uk/Tools/msa/clustalo) (Sievers et al., 2011). The sequence logos of multiple sequence alignments of the JAK-STAT pathway genes were generated by Weblogos 3.2 (Crooks et al. 2004) without any compositional adjustment. Sequence logos are a graphical representation of the multiple sequence alignment of a protein and/or DNA sequences.

2.2.4 Phylogenetic analysis

Phylogenetic analysis was performed using MEGA 5.2 software (Tamura et al., 2011). Phylogenetic studies carried out for all 4 STAT pathway genes (STAT-A, STAT-B, SOCS and PIAS) and compared to find a relationship between different subgenus and series. The evolutionary history of all these were inferred by using the Neighbor-Joining (NJ) method (Saitou and Nei, 1997). Reliability of the trees through branch support was assessed by using a bootstrap test with 1000 replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches (Tamura et al., 2011). In case of gaps or missing data, partial deletion and *p*-distance (amino acid differences) of the sequences parameters was chosen. Other parameters were kept as default.

2.3 Results

The JAK-STAT pathway is crucial for development and immunity of mosquito. The important key regulators of pathway include: transcription factors STAT-A, STAT-B; and negative regulators PIAS and SOCS5 genes. These genes were analysed for intron–exon region, full CDS and respective protein domains. Homology search were performed by WGS assembly option of BLAST using complete cDNA sequence of respective gene. The matching region on respective contigs were retrieved and used for putative exon-intron prediction through Genscan web tool (Burge and Karlin, 1998). It is worth to mention here that, the computer generated prediction algorithms were erroneous or partial hence we corrected them manually. It was found in many cases; a likely erroneous intron was projected at the end of the conserved sequence. It is fascinating that the prediction algorithms failed so frequently in this group of genes, and specifically in one site which was also the case in knirps gene family proteins of arthropods (Perl et al., 2014). To overcome this problem, the BLAST homology search matches were taken and compared it with the number of ranges in alignment section. Depending upon the predicted exonic/coding sequences, the ORF were formed and deduced amino acids were subjected to submit at CDD search. This elementary approach of gene annotation was utilized for all *Anopheles* species and their assembled sequences were submitted to NCBI under third party annotation (TPA) section. Following subsections is the finding of *in silico* prediction of JAK-STAT pathway genes and illustrate the genomic diversity exists if any among these mosquito species.

2.3.1 Identification and annotation of STAT-A and STAT-B genes in different *Anopheles* species

STAT-A and STAT-B genes were identified together as they frequently come concomitantly in each BLAST search because of the same domain architecture. The percentage of their query coverage, percentage identity and the priority of their occurrence get changed but they used to appear concurrently (**Figure 2.1**). AgSTAT-A (3568bp) and AgSTAT-B (2169bp) were separately used as a query for BLAST searches for all *Anopheles* WGS sequences one after another. Homology search in *An. stephensi* genome, through BLAST showed two distinct and diverge genomic scaffold_00038; KE388927.1 and scaffold_00093; KE388982.1 with initial match of 11573bp and 1397bp in the form of two contig. Result shows 65% and 39% query coverage and 95% and 66% identity respectively based on AgSTAT-A sequence. One striking outcome was

that, second contig had only 39% coverage but does not showed intronic sign, similar to *An. gambiae* AgSTAT-B. These two intact STATs were identified separately from their respective contigs and designated as AsSTAT-A (KE388927.1) and AsSTAT-B (KE388982.1) in case of *An. stephensi* (**Figure 2.1A and B)**.

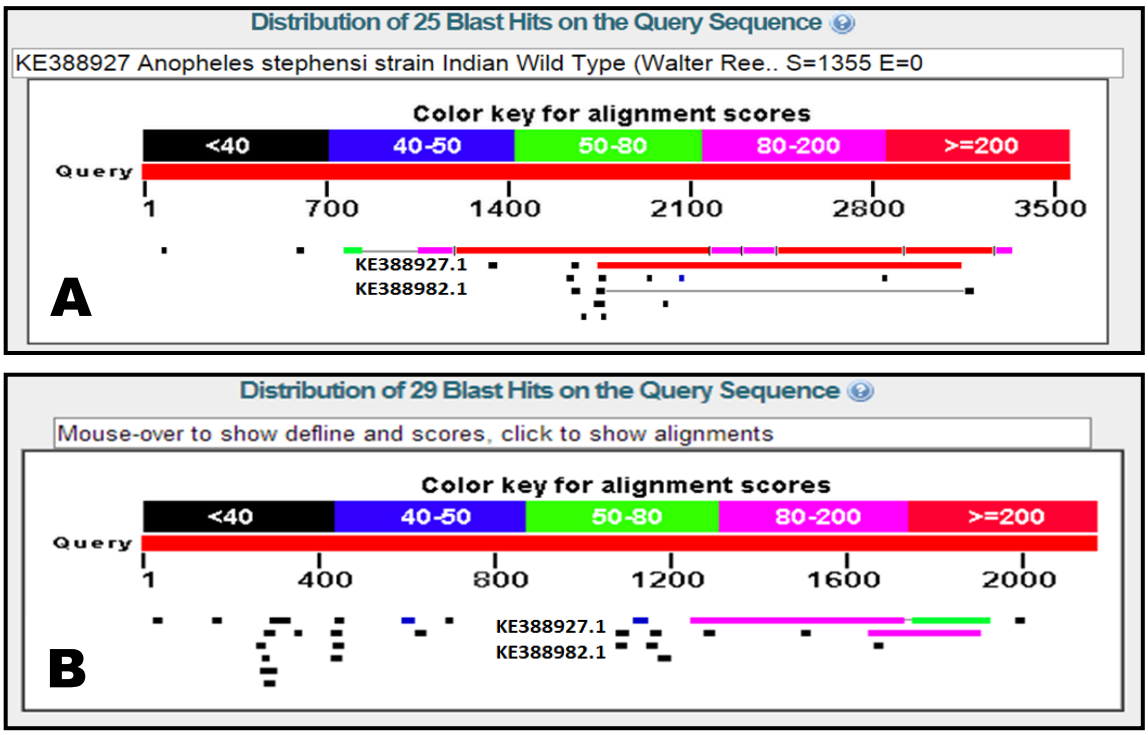


Figure 2.1 Somewhat similar BLASTn homology search result appear with AgSTAT-A as query showing good similarity and sequence coverage (**A**). AgSTAT-B sequence when used as query showing neither good similarity nor much sequence coverage, confer that STAT-B gene is highly divergent in nature (**B**).

Same exercise was also done for rest other *Anopheles* species. In some cases we found that the computer generated predictions were erroneous or partial. A presumably erroneous intron was predicted just after the end of the conserved core sequence of STAT-A. Manual curation was done to overcome the erroneous predictions and to deliver maximal reliability. Manual annotation through BLAST homology search revealed that STAT-A gene from all *Anopheles* species showed same genomic architectures irrespective of their intron lengths. All *Anopheles* species are conserved in number and length of exons while their introns are more or less diverged in size. STAT-A gene fully covered through homology searches showing 7 exons and 6 introns in the whole gene. Their constant exon size is 128bp, 974bp, 124bp, 124bp, 490bp, 340bp and 79bp in 5' to 3' direction which are similar in all species. The 6th intron (I6) of STAT-A is the largest one ranging from 4.5-9.0 kb in different species **(Table 2.4)**. Schematic representation of AsSTAT-A is given in **figure 2.2A**.

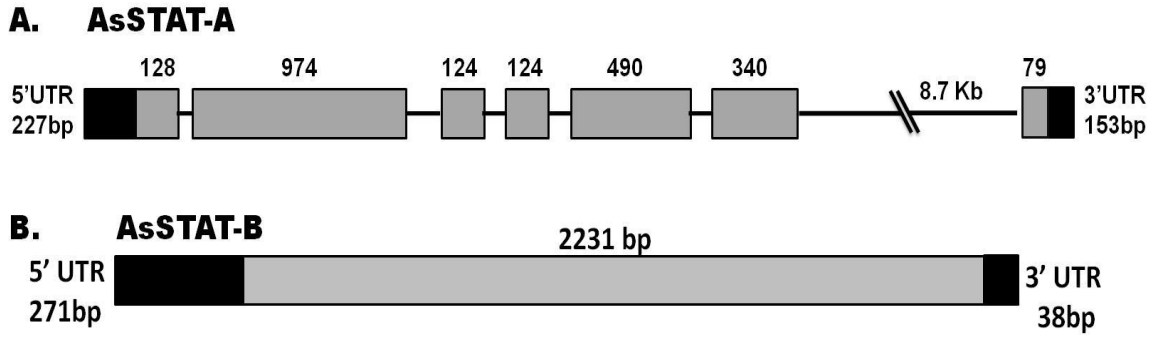


Figure 2.2 Schematic presentations of genomic architecture of AsSTAT-A (**A**) and AsSTAT-B (**B**) gene. Grey color box and black lines are showing exonic and intronic regions; Black color boxes at the ends are 5' and 3' UTRs.

Table 2.4 STAT-A gene architecture in 18 *Anopheles* species. The size of Exons (E1-E7) and Introns (I1-I6) in base pairs (bp) are mentioned for each species.

	E1	**I1**	**E2**	**I2**	**E3**	**I3**	**E4**	**I4**	**E5**	**I5**	**E6**	**I6**	**E7**
An. gambiae	**128**	80	**974**	68	**124**	477	**124**	81	**490**	93	**340**	8342	**79**
An. arabiensis	"	71	"	66	"	1319	"	80	"	78	"	8755	"
An. melas	"	82	"	91	"	88	"	78	"	77	"	8056	"
An. merus	"	78	"	90	"	295	"	77	"	93	"	8523	"
An. quadriannulatus	"	80	"	75	"	1284	"	83	"	78	"	8652	"
An. christyi	"	172	"	98	"	94	"	129	"	89	"	5742	"
An. funestus	"	110	"	88	"	88	"	129	"	83	"	8000	"
An. minimus	"	104	"	83	"	96	"	176	"	78	"	6669	"
An. culicifacies	"	88	"	71	"	82	"	177	"	71	"	6880	"
An. epiroticus	"	74	"	82	"	100	"	94	"	97	"	8516	"
An. maculatus	"	80	"	70	"	80	"	78	466	74	"	7147	"
An. stephensi	"	100	"	76	"	77	"	80	"	67	"	8683	"
An. sinensis	"	84	"	75	248	-	-	74	"	71	"	6416	"
An. atroparvus	"	77	977	148	"	84	"	118	"	132	"	5695	"
An. dirus	"	76	"	75	"	82	"	77	"	86	"	6681	"
An. farauti	"	92	"	87	"	72	"	81	"	84	"	6147	"
An. darlingi	"	84	"	84	"	77	"	94	"	131	"	4514	"
An. albimanus	"	81	"	81	"	86	"	89	"	130	"	4572	"

Few species from the cohort showed the occurrence of gene on more than one gDNA contigs. *An. maculatus*, *An. melas*, *An. merus*, *An. quadriannulatus* and *An. christyi* STAT-A gene is lying on two or three different contigs (**Table 2.5, third column**). Their annotation was done with extra care to join both the contig either in (+/+) or (-/-) direction for attainment of the full coding sequence. *An. maculatus* STAT-A relevant contig were not able to retrieve full gene sequences. Nucleotide sequence data of STAT-A of 16 *Anopheles* species reported here are available in the Third Party Annotation (TPA) section of the DDBJ/EMBL/GenBank databases under the accession numbers TPA: BK009250-BK009262; BK009380-BK009382 **(Table 2.5; last third column).**

Table 2.5: Table represents the genome attributes of **STAT-A gene** retrieved from their respective contigs: primary sequence IDs, %identity and coverage; start and end position for each analysed *Anopheles* mosquitoes through BLAST. After annotation the final submitted STAT-A gene reference IDs their size of gene and deduced proteins have also been shown in last three column.

Mosquito Name	Scaffold/ Supercontigs	Primary Sequence ID	% Query Coverage	% Identity	No. of Matches	Start	End	AFTER ANNOTATION	NCBI Reference ID	Gene Size (bp)	Amino acids (aa)
An. arabiensis	supercont1.22	APCN01000022.1	98	99	7	2077128	2089755		BK009250	2259	752
An. melas	cont2.19858 cont2.5179 cont2.5178	AXCO02019858.1 AXCO02005179.1 AXCO02005178.1	42 26 9	99 99 90	3 3 1	1235 5324 6633	2713 6435 6964		BK009382	2259	752
An. merus	cont2.3659 cont2.3661	AXCQ02003659.1 AXCQ02003661.1	77 20	99 85	8 2	38454 4635	50418 7117		BK009251	2259	752
An. quadriannulatus	cont1.4019 cont1.4015	APCH01004019.1 APCH01004015.1	77 18	99 91	8 3	5707 38009	18765 40478		BK009253	2259	752
An. christyi	cont1.3570 cont1.21783	APCM01003570.1 APCM01021783.1	63 2	93 96	7 1	4575 914	7350 1001		BK009252	2259	752
An. funestus	cont1.4905	APCI01004905.1	64	92	8	16786	27748		BK009254	2259	752
An. minimus	cont1.1601	APHL01001601.1	64	94	8	64897	74540		BK009256	2259	752
An. culicifacies	cont1.3811	AXCM01003811.1	61	88	7	6364	16170		BK009255	2259	752
An. epiroticus	cont1.469	APCJ01000469.1	99	91	7	8361	19580		BK009258	2259	752
*An. maculatus**	cont1.34279 cont1.20232 cont1.5479	AXCL01034287.1 AXCL01020239.1 AXCL01005484.1	54 34 3	93 93 96	4 2 1	1 1857 371	1588 2703 453		BK009380	2235	744
An. dirus	cont1.3648	KB672824.1	99	94	7	3890385	3899721		BK009257	2259	752
An. farauti	cont2.1351	AXCN02001351.1	63	92	7	22853	31684		BK009260	2259	752
An. sinensis	contig013419	KE524855.1	61	91	6	879020	888000		BK009381	2259	752
An. atroparvus	supercont1.14	KI421895.1	99	92	7	3042336	3050881		BK009259	2259	752
An. darlingi	cont8097	ADMH02002153.1	62	78	7	10384	17616		BK009261	2259	752
An. albimanus	cont1.1265	APCK01001266.1	55	79	7	15636	22705		BK009262	2259	752

* Full gene was not retrieved due to shortage of contig and/or not showing match with another contig.

In another episode, less query coverage was observed for STAT-B gene which pinpointed that only highly conserved domain was matched and confined only to DNA binding and SH2 domain in maximum incidents (**Figure 2.1B**). To elucidate the full STAT-B gene, searches were intensified in both directions of contig in continuous ORF to get all four relevant domains of STATs. For full identification of STAT-B CDS; AgSTAT-B (AJ010299.1) was selected as a query but it showed high divergence in sequence especially from Neotropical, Australasian and Oriental region *Anopheles* species (**Table 1.1; Chapter 1**). STAT-B gene was found to be intronless in every *Anopheles* species but their less query coverage was confined mainly to SH2 domain region except *gambiae* complex species and *An. atroparvus*. Remaining portion of the STAT-B gene was obtained by extending forward on the same contig or by using AsSTAT-A (KR790000) as a query. The whole procedure was done through manual annotation using BLAST homology searches. STAT-B gene was found highly diverged in CDS length irrespective to STAT-A (2259bp) of *Anopheles* species (**Table 2.6**).

Out of these two STATs, STAT-A has considered as ancestral gene and STAT-B as a duplicated gene, arisen through retrotransposition phenomenon. It has been hypothesized in case of *An. gambiae* that a spliced mRNA of AgSTAT-A gene have reverse-transcribed and integrated back into its genome, giving rise to a second 'intronless' STAT-B gene through retro-transposition (Zdobnov et al., 2002). Counting on, this hypothesis, we retrieved the STAT-B gene in the genome assemblies of 17 *Anopheles* WGS sequences. Schematic representation of AsSTAT-B gene (intronless) is given in **figure 2.2B**.

Out of 18 only 13 species showed the existence of two STAT genes in their genome. *An. dirus* (from Oriental region), *An. farauti* (from Australasian region), *An. darlingi and An. albimanus* from new world (Neotropical region) were deficient of STAT-B gene (due to lack of duplication) in their genome. Nevertheless, the size of protein among same genus of mosquito also varies from 726-750 amino acids (**Table 2.6; last third column**). Nucleotide data of all *Anopheles* STAT-B genes are reported in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession numbers TPA: BK009265-BK009275. Retrieval of gene sequence from the unannotated genome assembly is very cumbersome process and efforts were made to achieve the full gene sequences in most of the cases. In very few mosquito's species where contig size was smaller, we were only able to confer the partial gene sequences (e.g. *An. maculatus* STAT-A and *An. christyi* STAT-B).

Table 2.6: Data mining result of **STAT-B gene** of all *Anopheles* species. Locations of BLAST hits of *An. gambiae* AgSTAT-B gene (AJ010299.1) used as a query. Table represents their primary sequence IDs, %identity and coverage; start and end position for each analysed *Anopheles* mosquitoes. The final submitted STAT-B gene reference IDs with their size of gene and deduced protein have also been shown.

Mosquito Name	Scaffold/ Contig	Primary Sequence ID	% Query Coverage	% Identity	Start	End	AFTER ANNOTATION	NCBI Reference ID	Gene Size (bp)	Amino acids (aa)
An. arabiensis	cont1.5152	APCN01005153.1	100	98	2411	4597		BK009265	2187	728
*An. christyi**	cont1.18593	APCM01018593.1	58	74	1	1301		BK009268	1299	432
An. funestus	cont1.2772	APCI01002772.1	25	67	1712	3591		BK009270	2211	736
An. melas	cont2.19567	AXCO02019567.1	100	96	1718	3898		BK009266	2181	726
An. merus	cont2.4426	AXCQ02004426.1	100	96	562	2742		BK009267	2181	726
An. quadriannulatus	cont1.10110	APCH01010111.1	100	98	5744	7927		BK009269	2184	727
An. minimus	cont1.4283	APHL01004283.1	16	69	3557	4884		BK009272	2208	729
An. culicifacies	cont1.170	AXCM01000170.1	12	72	7829	9165		BK009271	2250	749
An. maculatus	cont1.13601	AXCL01013607.1	17	67	1388	2730		BK009273	2208	736
An. epiroticus	cont1.3133	APCJ01003133.1	57	65	56092	57311		BK009274	2229	739
An. atroparvus	cont1.2665	AXCP01002665.1	23	68	37586	39742		BK009275	2253	750

* Full gene was not retrieved due to shortage of contig and/or not showing match with another contig.

Table 2.7 Percentage nucleotide and amino acid identity matrix for STAT-A gene among 18 *Anopheles* species. Upper triangle comprises the % nucleotides identity (green color) and lower triangle is showing the percentage of protein identity (grey color). STAT-A gene has revealed the good percentage identity with each other at nucleotide as well as protein level.

	Mosquito Name	**% Nucleotide Identity**																	
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	***An. gambiae***		99	99	99	99	92	90	91	88	92	92	90	93	92	79	79	89	89
2	***An. arabiensis***	100		100	99	100	92	90	91	88	91	92	90	93	92	79	79	89	89
3	***An. merus***	100	100		99	99	92	90	91	88	91	92	90	93	92	79	79	89	89
4	***An. melas***	100	100	100		99	92	90	91	88	91	92	90	93	92	79	79	89	89
5	***An. quadriannulatus***	100	100	100	100		92	90	91	88	91	92	90	93	92	79	79	89	89
6	***An. christyi***	100	100	100	100	100		89	90	87	90	92	89	89	88	78	79	87	86
7	***An. funestus***	100	100	100	100	100	99		90	90	93	88	90	87	86	79	79	85	84
8	***An. stephensi***	100	100	100	100	100	99	100		89	91	90	93	89	88	79	79	86	86
9	***An. culicifacies***	96	96	96	96	96	96	96	96		93	87	89	87	86	78	78	84	84
10	***An. minimus***	99	99	99	99	99	99	99	99	96		91	91	90	88	79	79	87	86
11	***An. epiroticus***	99	99	99	99	99	99	99	99	96	99		89	91	89	80	80	88	87
12	***An. maculatus***	97	97	97	97	97	97	97	98	93	97	97		88	87	79	79	86	85
13	***An. dirus***	97	97	97	97	97	97	97	97	95	97	97	94		94	78	79	88	89
14	***An. farauti***	94	94	94	94	94	94	94	94	92	94	94	92	95		78	78	88	88
15	***An. darlingi***	93	93	93	93	93	93	93	93	90	93	93	91	93	90		93	80	79
16	***An. albimanus***	93	93	93	93	93	93	93	93	90	93	93	90	92	90	100		80	79
17	***An. atroparvus***	95	95	95	95	95	95	95	95	93	95	95	92	94	92	93	93		91
18	***An. sinensis***	93	93	93	93	93	93	93	93	91	93	93	90	93	90	92	92	96	
		% Amino Acid Identity																	

STAT-A and STAT-B partial or full nucleotide and protein sequences of all *Anopheles* species were utilized to build the percentage identity matrix. STAT-A was revealed the high percentage of sequence similarity at both nucleotide and protein level. STAT-A gene of all *Anopheles* species showed the conservation in sequences greater than 95% (**Table 2.7**). STAT-A nucleotides sequence diversity are in agreement with the rule that dissimilarity in the nucleotide at third position often code the same amino acid. Consequently, proteins of different species are maintaining the same structure and function; even there is a huge change in coding gene.

Contrary to the above fact, STAT-B gene was found to be very similar at nucleotide level but their protein sequences varied highly from species to species (**Table 2.8**). Although among all the 13 *Anopheles* species, few of them showed some similarity with neighboring species which renders to their taxonomical hierarchy or endemicity of particular location. But all of these were very peculiar and dissimilar with each other at nucleotide and even more at protein level. This percentage difference at protein level in different species of same genus provide the clue that these STATs possibly have some different function (independent or in same pathway) which compels us to believe their divergent evolution or neofunctionalization of the gene (Rottschaefer et al., 2015). This finding supports that STAT-B gene is still under selection pressure and have arisen through retro-duplication phenomenon.

Table 2.8 Percentage nucleotide and amino acid identity for STAT-B gene. Upper triangle showing % nucleotides identity (green color) and lower triangle is % identity of deduced amino acids (grey color).

	Mosquito Name	**% Nucleotide Identity**												
		1	2	3	4	5	6	7	8	9	10	11	12	13
1	***An. gambiae***		99	97	96	99	73	55	63	54	54	54	53	55
2	***An. arabiensis***	97		96	96	98	73	55	63	54	55	54	53	55
3	***An. melas***	94	94		96	97	73	55	63	54	54	54	53	55
4	***An.merus***	93	93	92		96	73	55	64	54	54	54	53	55
5	***An. quadriannulatus***	97	96	94	93		73	55	63	54	54	54	53	55
6	***An. christyi***	61	61	63	62	61		59	66	57	56	56	54	58
7	***An. funestus***	39	39	39	39	39	45		53	65	72	72	63	57
8	***An. epiroticus***	46	46	46	46	46	50	39		53	52	53	52	53
9	***An. maculatus***	39	38	38	38	39	42	50	36		63	63	78	58
10	***An. minimus***	35	35	35	36	35	44	56	37	48		74	62	58
11	***An. culicifacies***	36	36	36	37	36	42	57	37	48	59		62	58
12	***An. stephensi***	38	37	37	37	37	41	51	34	66	47	46		60
13	***An. atroparvus***	42	42	42	42	43	48	43	39	42	40	42	41	
		% Amino Acid Identity												

2.3.1.1 Comparative protein domain analysis of STAT-A and STAT-B

Conserved domains and motifs in these predicted proteins were determined using the programs, Pfam or Conserved Domain (CD) search. Protein sequences were aligned through Clustal Omega and their conserved motifs was highlighted through web logo tool. Alignments of the full translated amino acids illustrate the additional smaller conserved motifs consisting of few amino acids in all the sequences. These are located in each sequence at a particular site and called as signature motifs. Domains and motifs are two main entities that remain conserved during evolution and therefore, important factors for evaluating the evolution process. Protein domains of STAT-A and STAT-B were compared and their overall organization and domain size with their sequence length as well as amino acids present in each domain tabulated in **Table 2.9** and **2.10**, respectively. All *Anopheles* species have the same functional domain organization for STAT protein *viz.* STAT interaction domain, alpha domain, STAT binding domain and SH2 domain presented schematically in **Figure 2.3A**.

Table 2.9 Domain size of STAT-A protein from different *Anopheles* species retrieved through CD search. Size of each domain has been given in parentheses.

Mosquito species	**Protein Size (aa)**	**Signature Domains**			
		STAT interaction	**STAT alpha**	**DNA binding**	**SH2 domain**
An. gambiae	752	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. arabiensis	,,	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. melas	,,	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. merus	,,	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. quadriannulatus	,,	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. christyi	,,	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. funestus	,,	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. culicifacies	,,	2-127 (126)	135-322 (188)	324-571 (248)	562-676 (115)
An. minimus	,,	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. epiroticus	,,	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. stephensi	,,	2-127 (126)	141-322 (182)	324-571 (248)	578-677 (115)
An. maculatus	744	2-125 (124)	139-320 (182)	322-571 (246)	562-676 (115)
An. dirus	752	2-127 (126)	141-322 (182)	324-571 (248)	562-676 (115)
An. farauti	,,	2-126 (125)	141-322 (182)	324-571 (248)	562-676 (115)
An. sinensis	,,	2-128 (127)	141-322 (182)	324-571 (248)	562-676 (115)
An. atroparvus	,,	2-128 (127)	142-323 (182)	325-572 (248)	563-677 (115)
An. darlingi	,,	2-128 (127)	135-322 (188)	324-571 (248)	562-676 (115)
An. albimanus	,,	2-128 (127)	135-322 (188)	324-571 (248)	562-676 (115)

All tentative domains were also found in STAT-B protein of all anophelines (**Table 2.10**). The known STATs have highly conserved sequences among the DNA binding and SH2 domains present at C-terminal. However, the N-terminal STAT interaction and

alpha domain have lower homology, compare to C-terminal STAT domains. The conserved amino acid residues were also analysed. For example: A tyrosine residue at 685 position in STAT-A to be phosphorylated by JAKs during activation (Darnell 1997a, 1997b) was found constantly in each species while the position of tyrosine residue in STAT-B were highly variable (**Figure 2.3 B and C**).

Table 2.10 Size of STAT-B domains of *Anopheles* mosquitoes retrieved through Pfam search. Size of each domain has been given in parentheses.

Mosquito Name	Protein size (aa)	STAT interaction	STAT alpha	DNA binding	SH2 domain
An. gambiae	722	3-116 (114)	121-290 (170)	292-538 (247)	550-628 (79)
An. arabiensis	728	4-116 (113)	133-295 (163)	297-543 (247)	555-633 (79)
An. melas	726	4-116 (113)	121-295 (175)	297-543 (247)	555-633 (79)
An. merus	726	3-116 (114)	133-295 (163)	297-543 (247)	555-633 (79)
An. quadriannulatus	727	3-116 (114)	94-295 (202)	297-543 (247)	555-633 (79)
An. christyi	432	-	-	10-257 (248)	269-347 (79)
An. funestus	736	5-119 (115)	133-318 (186)	320-566 (247)	573-656 (84)
An. epiroticus	738	1-115 (115)	134-310 (177)	320-562 (243)	574-652 (79)
An. maculatus	735	6-120 (115)	131-316 (186)	318-563 (246)	575-653 (79)
An. stephensi	743	6-122 (117)	131-316 (186)	318-567 (248)	579-650 (72)
An. minimus	728	4-120 (117)	132-317 (186)	319-567 (247)	579-650 (72)
An. culicifacies	749	5-116 (112)	133-319 (187)	321-570 (250)	577-660 (84)
An. atroparvus	750	2-127 (126)	137-324 (188)	326-574 (249)	586-664 (79)

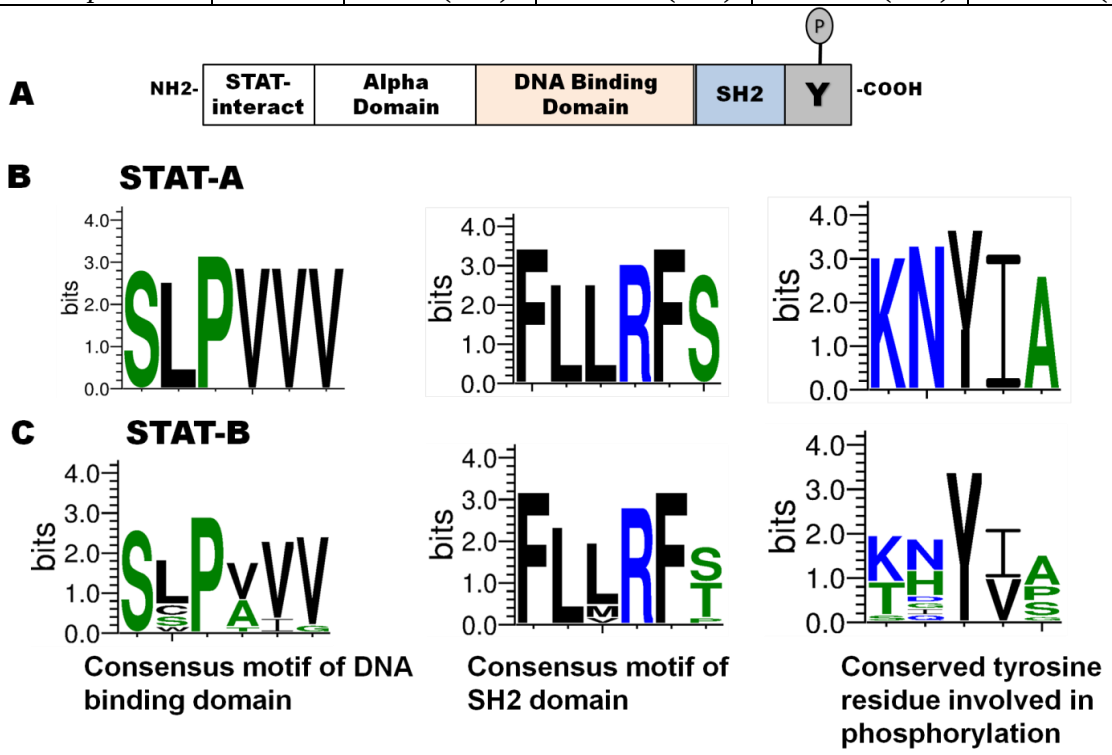


Figure 2.3 Conserved domains of STAT-A and STAT-B (**A**) showing consensus motif of DNA binding and SH2 domain respectively and phosphotyrosine residue (**B and C**) necessary for JAK activation.

Protein sequence analysis showed that consensus motif sequences of STAT-A were found constant in all *Anopheles* mosquitoes while there was a huge variation in STAT-B consensus motifs (**Figure 2.3**). Among highly conserved domains of STATs; STAT-binding and SH2 domains, consensus motif sequences are "SLPVVV" and "GTFLLRF" respectively which was found conserved for each species STAT-A. Whereas it varies a lot in case of STAT-B; *Anopheles* species belonging to Neocellia and Myzorhynchus series showed the difference among themselves but remain conserved in the same category (**Figure 2.3**). This variation showed that STAT-B gene is still under high selection pressure in above mentioned subgenera and may have some different role either in same pathway or different. The C-terminal region containing transactivation potential is poorly conserved among STAT-B proteins. Not only do they differ in length, ranging from 722-750 amino acids, they are sometimes missing entirely, such as splice variants of the mammalian STATs and the *Dictyostelium* sp. STATs (Zhukovskaya et al., 2004).

2.3.1.2 Phylogenetic correlation of STAT-A and STAT-B gene

Deduced amino acid sequences of all *Anopheles* mosquitoes were aligned through ClustalW and used for phylogenetic analysis through MEGA 5.2 software. Trees were obtained using NJ methods for 18 mosquitoes STAT-A and STAT-B genes, respectively. Their biogeographical landmarks support their genetic diversity and convergence relationship. Phylogenetic analysis of STAT-A gene also propounded the same genetic diversity in different *Anopheles* species as anticipated from their taxonomical categorization. The Asian mosquitoes, *An. dirus* and *An. farauti* belong to the series Neomyzomyia, consistently suggesting that this series is a phylogenetically separate sister clade from the remaining series: Pyretophorus (*An. gambiae*, *An. arabiensis*), Myzomyia (*An. funestus*) and Neocellia (*An. stephensi*) (**Figure 2.4**).

STAT-A genes are conserved at protein level therefore they clustered together according to their taxonomic groups. *An. albimanus* and *An. darlingi* belong to series Argyritarsis of subgenus Nyssorhynchus diverged first from the most common ancestor of subfamily Anophelinae. More advanced *An. sinensis* (series Myzorhynchus) and *An. atroparvus* (Series Anopheles) are diverged almost parallel with subgenus Nyssorhynchus. This type of diversification was common in both STAT-A and STAT-B genes. *An. farauti* and *An. dirus* of series Neomyzomyia diverged first from the last common ancestor of subgenus Cellia. Rest of the species of subgenus Cellia, *An. maculatus* and *An. stephensi* (series Neocellia), *An. minimus* and *An. culicifacies* (series Myzomyia) and gambiae complex

species along with *An. christyi* and *An. epiroticus* (series Pyretophorus) were clustered together in STAT-A and STAT-B (**Figure 2.4**).

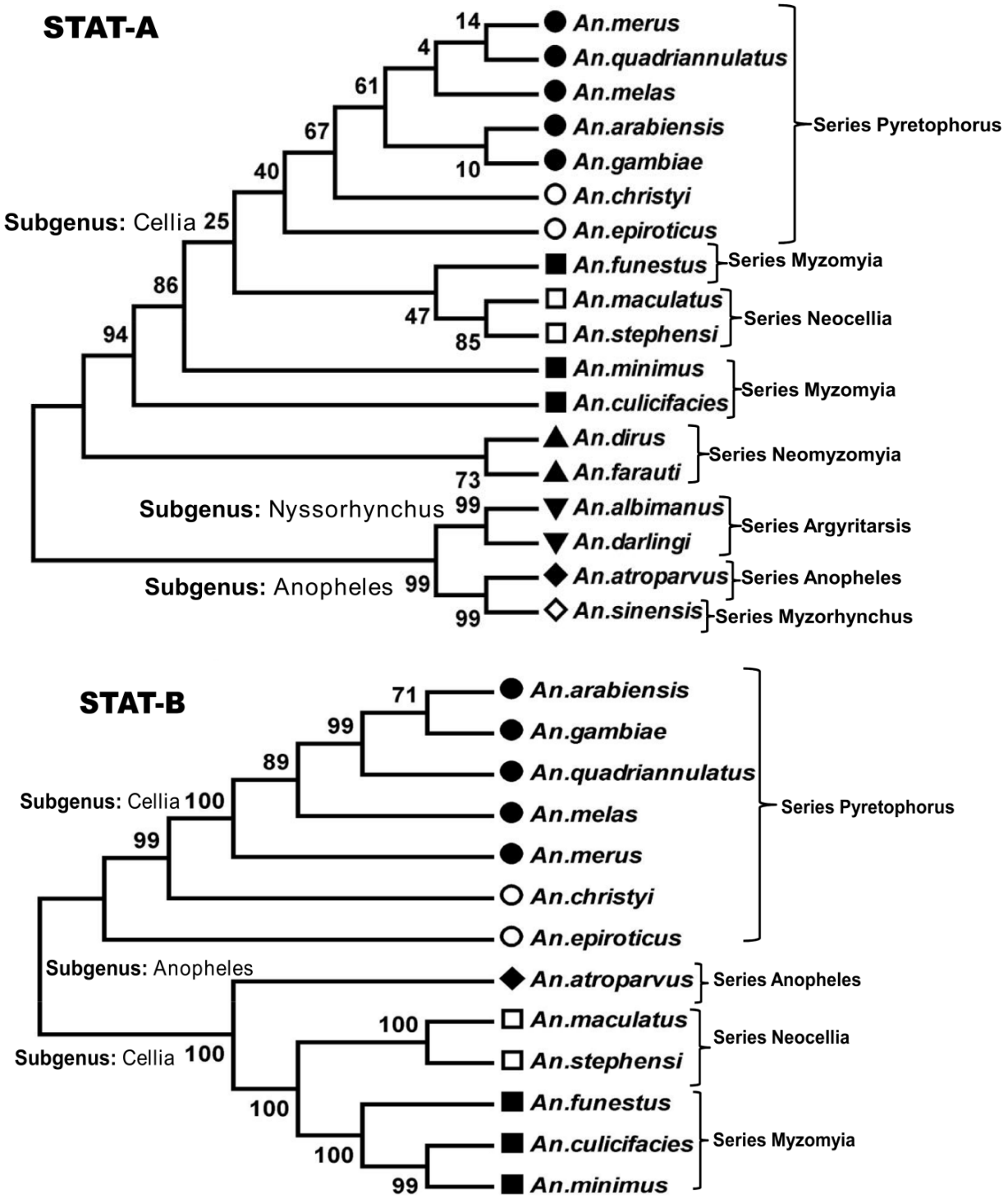


Figure 2.4 Phylogenetic tree of all *Anopheles* species of STAT-A and STAT-B gene obtained through NJ method. The clustering patterns were labeled with their specific taxonomical position; on left side subgenus and curly braces on right side are showing the series where they belong. The numbers on the branches represent the % of 1000 bootstrap. Black circle (●): Pyretophorus – *gambiae* complex and white circle (○): non-*gambiae* complex. Black square (■): Myzomyia and White square (□): Neocellia. Black triangle (▲): Neomyzomyia, Inverted black triangle (▼): Argyritarsis, Black diamond (♦): Anopheles, White diamond (◊): Myzorhynchus.

2.3.2 Identification and annotation of SOCS5 genes

Suppressors of cytokine signaling (SOCS) protein function as negative regulator of cytokine signaling and are involved in fine tuning of the immune responses. In contrast to other STAT pathway components, which exist as single members in invertebrates, three SOCS family members have been identified in fruit fly (SOCS36E, SOCS44A, SOCS16D) (Karsten et al., 2002). Among these three; SOCS36E is the orthologous to human SOCS5 and functions as negative regulator in JAK-STAT signaling pathway. SOCS5 nucleotide sequences from different *Anopheles* species were extracted from BLAST driven genome databases as described previously. Among all *Anopheles* species, *An. maculatus* is not showing complete match with any of the contigs, probably due to incomplete genomic assembly during the time when data was retrieved. In this case, the exon-I was only able to hunt through homology search. Similarly, *An. albimanus*, *An. sinensis*, *An. atroparvus* and *An. darlingi* SOCS5 gene showing high variability at 5'region; their exon II size was also found larger compared to other species **(Table 2.11**).The accession IDs of contigs and size of CDS for all SOCS5 of *Anopheles* are listed in **Table 2.12**.

Table 2.11 SOCS5 gene architecture in 18 *Anopheles* species. The size of exons and introns are mentioned for each species.

Mosquito Name	**Exon I (bp)**	**Intron I (bp)**	**Exon II (bp)**	**Intron II (bp)**	**Exon III (bp)**
An. gambiae	374	548	525	84	298
An. arabiensis	374	544	525	82	”
An. melas	365	744	525	114	”
An. merus	371	545	525	83	”
An. quadriannulatus	380	548	525	99	”
An. christyi	380	637	510	82	”
An. epiroticus	374	676	513	79	”
An. funestus	350	660	513	85	”
An. minimus	362	608	513	64	”
An. culicifacies	374	516	513	65	”
*An. maculatus**	389	187	-	-	-
An. stephensi	383	606	513	75	”
An. dirus	350	873	537	487	”
An. farauti	347	813	537	128	”
An. sinensis	398	1686	597	70	”
An. atroparvus	362	829	582	81	”
An. darlingi	344	291	559	89	”
An. albimanus	341	191	555	86	”

***** Full gene was not retrieved due to shortage of contig and/or not showing appropriate match with any other contigs.

Table 2.12 Data mining result of SOCS5 gene of all *Anopheles* species. Table represents the genome attributes of SOCS5 gene retrieved from their respective contigs: primary sequence IDs, %identity and coverage; start and end position for each analysed *Anopheles* mosquitoes through BLAST. After annotation the final SOCS5 gene size and deduced proteins have also been shown in last two columns.

Mosquito Name	Contigs/ Chromosome	Sequence ID	% Query Coverage	% Identity	Gene size (bp)	Start (bp)	End (bp)	AFTER ANNOTATION	CDS (bp)	amino acids (aa)
An. gambiae	chromosome 3L	NT_078267.5	100	99	1829	15673360	15671532		1197	398
An. arabiensis	supercont1.6	KB704895	99	99	2015	2671828	2674075		1197	398
An. melas	cont2.6542	AXCO02006542	98	94	2238	863	3314		1188	395
An. merus	supercont2.1	KI915156.1	100	96	2136	6566114	6568356		1194	397
An. quadriannulatus	supercont1.7	KB667888.1	100	96	2166	1043486	1045758		1203	400
An. christyi	supercont1.906	KB703082	93	87	2108	13996	16208		1188	395
An. epiroticus	supercont1.194	KB670658.1	90	91	2141	298990	301231		1185	394
An. funestus	supercont1.111	KB668345	80	84	2120	278126	280245		1161	386
An. minimus	supercont1.4	KB663943.1	100	94	1845	5639113	5640957		1173	390
An. culicifacies	cont1.17340 cont1.5070	AXCM01017340 AXCM01005070	51 35	84 70	2748	3228 17704	4124 18273		1185	394
An. maculatus*	cont1.37974	AXCL01037983	32	81	389	942	1330		389	**129***
An. stephensi	contig_3551	ALPR02003551	87	87	2029	52284	53972		1194	397
An. dirus	cont1.2079	APCL01002080	91	88	2743	72739	75582		1185	394
An. farauti	supercont2.6	KI915045.1	91	85	2321	11051573	11053963		1182	393
An. sinensis	scf7180000695811	KE524349.1	62	83	1376	504766	506042		1293	430
An. atroparvus	supercont1.12	KI421893.1	96	83	2152	1249127	1251278		1242	413
An. darlingi	cont7210	ADMH02001680	100	91	1581	126663	128243		1125	374
An. albimanus	supercont1.4	KB672424.1	55	79	2495	13624554	13625509		1194	397

* Full gene was not retrieved due to shortage of contig and/or not showing appropriate match with any another contig.

The nucleotides and protein sequences are highly variable for above mentioned mosquito species. Sequence analysis based on AgSOCS5 with all *Anopheles* genome database confirmed the presence of three exons separated by two introns (**Figure 2.5)**. The length of exon I and exon II speckled from one species to another species which leads to the variability in N-terminal domain of SOCS5 proteins (**Table 2.11 and 2.13**). Consequently, the size of SOCS5 gene, whose full genes are completely retrieved, varies from 1100bp to 1200bp. There are three exons and two introns were found common in all *Anopheles* species. Exon III, among all coding sequences was found to be analogous by having the 298bp sequence length which was found equivalent in all *Anopheles* species.

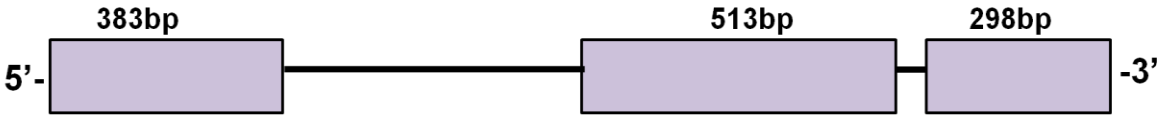


Figure 2.5 Schematic organization of SOCS5 gene in *An. stephensi* (5'→ 3'). The size of 1st and 2nd exons varies from species to species while and 3rd exon is 298bp equivalent in all *Anopheles* species.

2.3.2.1 Conserved domains of SOCS5

SOCS5 share a similar domain organization; central SH2 domain, conserved C-terminal SOCS box and the N-terminal domains that vary in length and amino acid sequence (**Figure 2.6A**). The SOCS-box motif is an essential protein domain, through which Elongin B and Elongin C interact and subsequently recruit Cullin-5 and Rbx-1 protein. Centrally located Src homology 2 (SH2) domains mediates the interaction with phosphorylated tyrosine (pTyr) residues. Protein sequence analysis of all *Anopheles* reveals the presence of highly conserved SH2 and SOCS box domains and least conserved N-terminal region (Dhawan et al., 2015) (**Figure 2.6 B, C and D**). Comparatively, the N-terminal of SOCS5 protein is larger in size and contain N-terminal conserved region (NTCR) which found to be conserved among family members (**Figure 2.6B**).

The SH2 and SOCS box domains showed 99-100% similarity with each other and 80-85% similarity while comparing the whole SOCS5 protein of *Anopheles* **(Table 2.14)**. It indicated that SH2 and SOCS box domains were highly conserved during evolution due to their important role in receptor signaling. These observations indicate that SOCS N-terminal amino acids identity is solely similar, rather limited to, anopheline SOCSs. This variability may indicate that all the domains of SOCS experienced differential selection

pressures and it provides the evidence that N-terminal domain is under least selection pressure (Wang et al., 2010).

Table 2.13 Size of SOCS5 protein domains from different *Anopheles* species through CD search.

Mosquitoes	Size of Protein	N-terminal region	Range (size-aa) of SH2 domain	Range (size) of SOCS box
An. gambiae	398	231	232-331 (101)	340-389 (50)
An. arabiensis	398	231	232-332 (101)	342-391 (50)
An. melas	395	228	229-329 (101)	339-388 (50)
An. merus	397	230	231-331 (101)	341-390 (50)
An. quadriannulatus	400	233	234-334 (101)	344-393 (50)
An. christyi	395	228	229-329 (101)	339-388 (50)
An. epiroticus	394	227	228-328 (101)	338-387 (50)
An. funestus	386	219	220-320 (101)	330-379 (50)
An. minimus	390	223	224-324 (101)	334-383 (50)
An. culicifacies	395	228	229-329 (101)	339-388 (50)
An. maculatus	129	129	-	-
An. stephensi	397	230	231-331 (101)	341-390 (50)
An. dirus	394	227	228-328 (101)	338-387 (50)
An. farauti	393	226	227-327 (101)	337-386 (50)
An. atroparvus	413	246	247-347 (101)	357-406 (50)
An. sinensis	430	310	311-411 (101)	421-470 (50)
An. albimanus	397	230	231-331 (101)	341-390 (50)
An. darlingi	374	207	208-308 (101)	318-367 (50)

Table 2.14 Percentage amino acid identity matrix of SOCS5 protein from different *Anopheles* species

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	***An.gambiae***	100																
2	***An.arabiensis***	99	100															
3	***An.melas***	98	98	100														
4	***An.merus***	99	99	98	100													
5	***An.quadriannul***	99	99	98	99	100												
6	***An.christyi***	88	89	88	89	88	100											
7	***An.epiroticus***	88	89	88	88	88	89	100										
8	***An.funestus***	85	85	85	85	84	89	88	100									
9	***An.minimus***	85	85	85	85	84	89	90	96	100								
10	***An.culicifacies***	85	85	85	85	84	88	88	96	96	100							
11	***An.stephensi***	86	86	86	86	85	89	87	93	95	94	100						
12	***An.farauti***	85	85	84	85	85	87	85	86	87	86	86	100					
13	***An.dirus***	86	87	86	87	86	88	86	88	89	88	88	94	100				
14	***An.atroparvus***	77	78	78	78	77	78	78	78	80	80	78	81	81	100			
15	***An.sinensis***	76	76	77	77	76	76	76	77	77	77	77	77	78	85	100		
16	***An.albimanus***	70	71	71	71	71	70	72	72	71	71	71	71	72	74	75	100	
17	***An.darlingi***	69	69	69	70	69	71	72	72	70	71	71	69	69	69	69	89	100

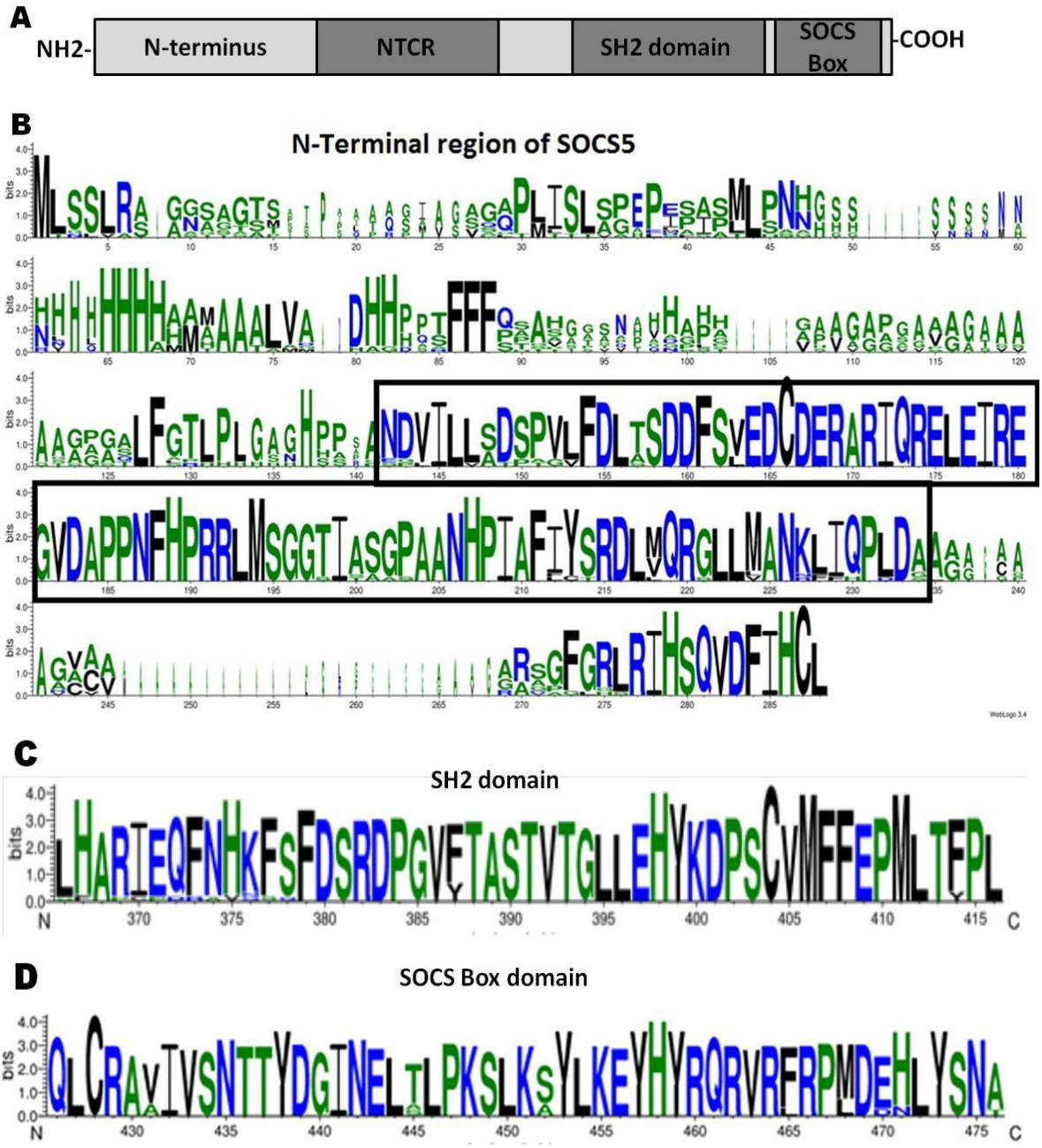


Figure 2.6 A. Schematic presentation of SOCS5 protein. B. Conserved region covered in black rectangle of N-terminal domains considered as a NTCR (N-terminal conserved region) among all *Anopheles* species which may act like JAK interacting region (JIR) orthologs of mammalian SOCS5 and SOCS4. **C & D.** SOCS protein has highly conserved SH2 and SOCS box domains in all *Anopheles* mosquitoes.

2.3.2.2 Phylogenetic correlation of SOCS5

The phylogenetic tree was constructed using the full length protein sequences of SOCS5 of all *Anopheles* species except *An. maculatus*. The evolutionary relationships among different proteins were inferred through Neighbor-Joining (NJ) method (Saitou et al., 1997). The branch support of the NJ phylogenetic tree was estimated using a bootstrap test with 1000 replicates. The percentage of replicate trees in which the associated taxa

clustered together in the bootstrap test is shown next to the branches. In case of gaps or missing data in sequences, partial deletion and *p*-distance of the sequences parameters was chosen. Other parameters were kept as default. Deduced amino acid sequences of SOCS5 gene of all *Anopheles* species were aligned by ClustalW in MEGA 5.2 software and further used for phylogenetic analysis (Tamura et al., 2011). The resulting phylogenetic tree was analyzed based on clusters and nodes formed.

Evolutionary trend of 17 *Anopheles* mosquito SOCS5 genes support their genetic diversity and convergence relationship. SOCS5 gene of all *Anopheles* species is clustered according to their taxonomical hierarchy (**Figure 2.7**). Series Pyretophorus mosquitoes are clustered in one group whereas another cluster has the mosquito species belong to series Myzomyia and Neocellia of subgenus *Cellia*. Evolutionary tree of SOCS5 gene of all 17 mosquito species reveal same taxonomical hierarchy which is according to *Anopheles* evolution.

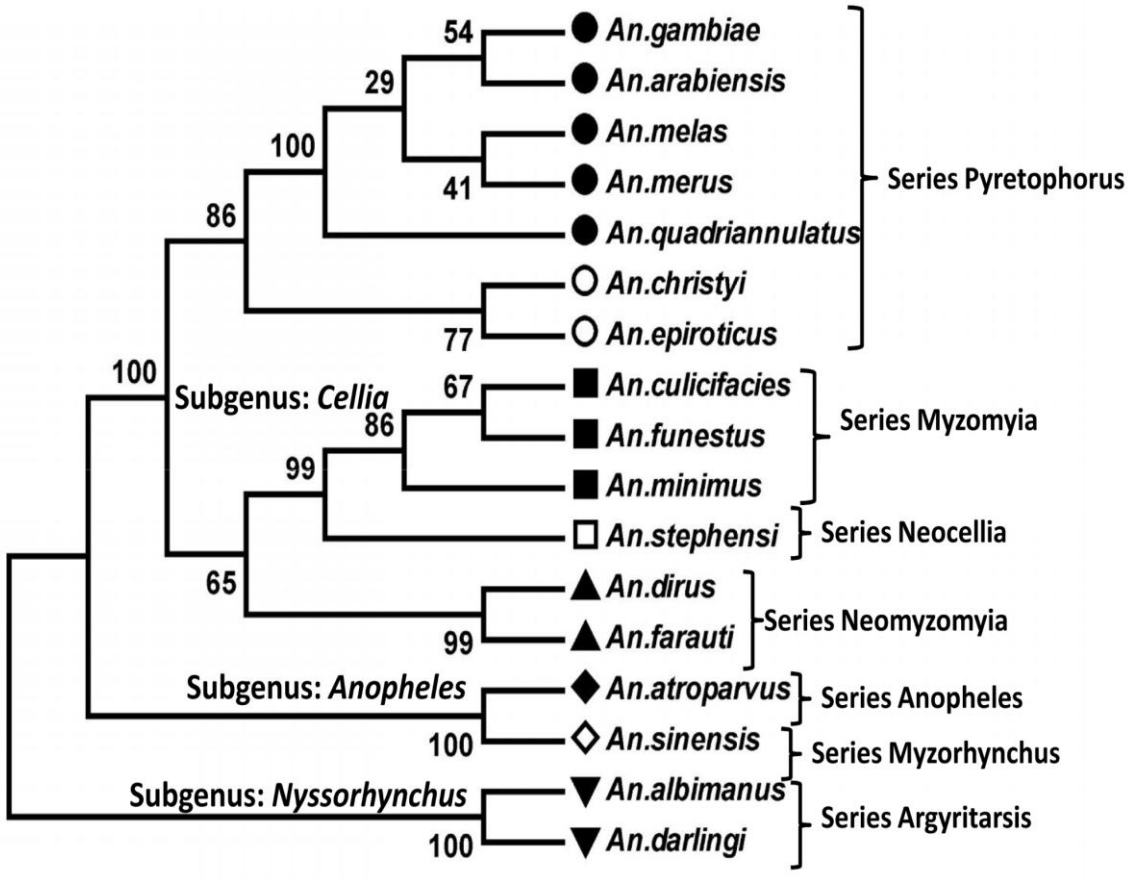


Figure 2.7 Phylogenetic tree of SOCS5 gene obtained through NJ method of all *Anopheles* species. The clustering patterns were further labeled with their taxonomic groups. The numbers on the branches represent the % of 1000 bootstrap. Black circle (●): Pyretophorus – *gambiae* complex and white circle (○): non-*gambiae* complex. Black square (■): Myzomyia and White square (□): Neocellia. Black triangle (▲): Neomyzomyia, Inverted black triangle (▼): Argyritarsis, Black diamond (♦): Anopheles, White diamond (◊): Myzorhynchus.

2.3.3 Identification and annotation of PIAS genes

Transcriptional suppression plays a central role in regulation of eukaryotic gene expression. Protein factors that obstruct in DNA binding activity of a transcription factor by corepressors control the basal transcriptional machinery or remodeling chromatin (Horwitz et al., 1996). The PIAS proteins have been shown to function as E3 ligases that facilitate the SUMO modification of a number of transcription factors (Rogers et al, 2003; Shuai, 2006). There is a single PIAS gene present in both *Drosophila* and *Aedes*, whilst four PIAS members in mammals (PIAS1, PIAS3, PIASx, PIASy) (Betz et al., 2001). In other insects including *Anopheles*, the probability of having one PIAS gene is definite. After the finding of two STAT genes, are present in majority of *Anopheles* species, then PIAS gene was taken into account to check the numbers of orthologous gene present in these species.

There is a lack of experimentally verified PIAS full gene for mosquitoes, even for *An. gambiae* hence *An. aquasalis* (Neotropical, new world) was first used as query. The full-length AqPIAS cDNA consists of 2407bp including a 1953bp CDS, which encodes a protein of 651 amino acid, as well as a 211bp 5'-UTR and 243bp 3'-UTR respectively (Bahia et al., 2011). We observed a huge genomic diversity among new world (Nearctic and Neotropical) and old world (Afrotropical and Oriental) mosquito species, that's why *An. stephensi* AsPIAS gene was used for query. Initially, *An. aquasalis* AqPIAS full gene sequence (HM851177) was used for predicting the *An. stephensi* full gene.

We followed the same strategy to retrieve the PIAS gene from the genome assembly of individual *Anopheles* mosquitoes. The full PIAS genes of all *Anopheles* species showed 10 exons and 9 introns, found common in all full annotated *Anopheles* species except *An. sinensis* (**Table 2.15**). The size of exons in *An. stephensi* in 5' to 3' direction are 42bp, 164bp, 661bp, 241bp, 106bp, 129bp, 164bp, 95bp, 196bp and 77bp respectively (**Figure 2.8)**. The annotated CDS and deduced amino acids sequence have been tabulated in **Table 2.16**. The sequence length of all *Anopheles* PIAS gene, whose full genes were annotated vary from 1750bp to 1953bp. *An. maculatus* and *An. darlingi* PIAS genes were not fully retrieved due to less query coverage or some of the sequence at 3' end was not matching with any of the contigs.

Table 2.15 Retrieved PIAS gene architecture of all *Anopheles* species. The size of exons and introns are mentioned for each species.

Mosquito Name	**E1**	**I1**	**E2**	**I2**	**E3**	**I3**	**E4**	**I4**	**E5**	**I5**	**E6**	**I6**	**E7**	**I7**	**E8**	**I8**	**E9**	**I9**	**E10**	**CDS**
An. gambiae	42	1149	164	94	631	94	241	75	106	78	129	89	164	136	137	40	181	1916	77	1872
An. arabiensis	”	1151	”	94	631	87	”	81	”	78	”	89	”	159	137	40	181	2898	92	1887
An. melas	”	1142	”	94	631	126	”	70	”	79	”	95	”	177	137	45	136	1636	77	1827
An. merus	”	1186	”	98	631	85	”	72	”	79	”	94	”	142	95	84	181	1627	77	1827
An. quadriannulatus	”	1135	”	91	631	87	”	74	”	78	”	90	”	177	54	28	240	1636	70	1841
An. christyi	”	1047	”	79	631	63	”	76	”	79	”	79	”	78	98	78	195	2109	75	1815
An. epiroticus	”	1246	”	122	574	79	”	60	”	96	”	104	”	473	98	81	196	1761	74	1792
An. stephensi	”	1136	”	88	661	89	”	75	”	77	”	89	”	133	95	80	196	1487	77	1875
An. maculatus*	”		”	155	499	91	”	70	”	76	”	80	”	100	95	-	-	-	-	-
An. culicifacies	”	1036	”	141	610	77	”	79	”	78	”	99	”	143	98	87	181	972	74	1809
An. minimus	”	1030	”	88	661	77	”	25	”	73	143	62	”	726	98	80	181	1241	74	1764
An. funestus	”	1025	”	82	634	98	”	90	”	89	133	88	”	268	98	2530	82	1102	137	1800
An. dirus	”	1160	”	94	664	86	”	86	”	224	”	731	161	515	95	72	184	527	77	1863
An. farauti	”	1130	”	98	661	78	”	64	”	377	”	313	154	620	95	87	184	560	70	1859
An. atroparvus	”	902	”	71	619	114	”	88	”	80	”	95	”	74	92	88	163	894	128	1848
An. sinensis#	”	846	”	117	616	72	”	86	”	82	”	71	”	113	92	77	166	-	-	1724
An. darlingi*	”	128	”	63	643	71	”	85	”	71	137	47	147	76	102	62	166	2388		1758
An. albimanus	”	108	”	54	637	72	”	68	”	69	”	58	”	71	98	67	166	2013	84	1953

***** Full gene was not retrieved due to shortage of contig and not showing match with any other contig. ***#*** *An. sinensis* contain only 9 exons and 8 introns.

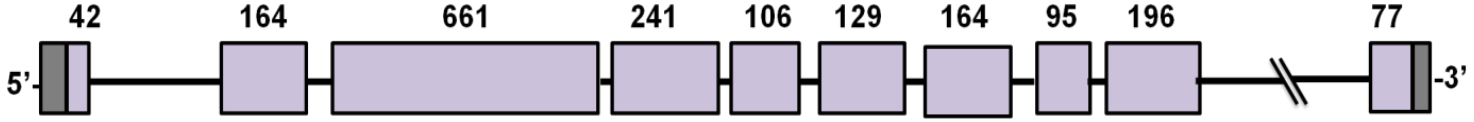


Figure 2.8 Generalized schematic organization of AsPIAS gene in 5’→ 3’direction. The genomic architecture is showing ten exons, separated by nine introns; first and ninth introns are larger in size. Grey colour boxes at the end are showing the 5’ and 3’ UTRs.

Table 2.16 Data mining result of PIAS gene of all *Anopheles* species. Locations of BLAST hits of *Anopheles* protein inhibitor of activated STAT genes. Table represents their primary sequence IDs, %identity and coverage; start and end position for each analysed *Anopheles* mosquitoes. The final annotated PIAS gene size and deduced protein have also been shown.

Mosquito Name	**Contigs/ Chromosome**	**Sequence ID**	**% Query coverage**	**% Identity**	**Gene size (bp)**	**Start (bp)**	**End (bp)**	**AFTER ANNOTATION**	**CDS (bp)**	**Amino acids (aa)**
An. gambiae	chromosome 2L	NT_078265.2	93	76	5536	8615382	8620917		1872	623
An. arabiensis	supercont1.15	KB704396.1	99	98	6564	5120699	5126594		1887	628
An. melas	cont2.8888	AXCO02008888	88	77	5284	7330	12613		1869	623
An. merus	supercont2.112	KI915267.1	88	76	5297	387000	386930		1827	609
An. quadriannulatus	supercont1.18	KB666287.1	88	76	5237	759912	765188		1841	613
An. christyi	supercont1.1024 supercont1.4376	KB674025.1 KB697878.1	96 4	82 93	6807	16384 2110	19722 2185		1803	600
An. epiroticus	supercont1.380	KB671599.1	89	75	5814	160413	166222		1794	597
An. funestus	supercont1.31	KB668848.1	66	74	7173	56306	62411		1800	599
An. minimus	supercont1.2	KB663721.1	95	81	5329	19059412	19064736		1911	589
An. culicifacies	cont1.5731	AXCM01005731	95	82	4521	7652	12168		1809	602
An. maculatus*	supercont1.13673	KI436596.1	73	91	1970	1022	2999		1398	**466***
An. stephensi	scaffold_00068	KE388957.1	73	76	5129	825853	831577		1875	624
An. dirus	supercont1.11	KB672602.1	87	78	5358	964955	970305		1860	620
An. farauti	supercont2.5	KI915044.1	99	84	5179	7449830	7455002		1859	618
An. sinensis	scf718000696048	KE525340.1	55	70	4552	18249	22800		1724	574
An. atroparvus	supercont1.5	KI421886.1	48	75	4254	2682241	2686049		1848	615
An. darlingi*	cont3176	ADMH02000560	63	69	2380	709	3275		1758	585
An. albimanus	supercont1.31	KB672415.1	89	91	4604	140496	145280		1953	650

***** Full gene was not retrieved due to shortage of contig and not showing match with any other contig.

2.3.3.2 Conserved domains of PIAS proteins

PIAS proteins were initially described as regulators of STAT signaling and were therefore named as protein inhibitor of activated STAT (PIAS). PIAS proteins regulate transcription through several mechanisms, including blocking the DNA-binding activity of transcription factors, recruiting transcriptional co-repressors and promoting protein sumoylation (Shuai and Liu, 2005). The size of PIAS protein in *Anopheles* varies from the 574 amino acids in *An. farauti* to 645 amino acids in *An. sinensis* but they share same conserved domain architecture (**Table 2.17**). The PIAS protein family contains several highly conserved regions (**Fig 2.9 A**).

Table 2.17 PIAS protein domains size from different *Anopheles* species

Mosquitoes	Protein Size (aa)	SAP domain (aa)	PINIT domain (aa)	MIZ/SP-RING zinc finger (aa)
An. gambiae	640	2-34 (33)	162-319 (158)	364-413 (50)
An. arabiensis	616	17-49 (33)	183-334 (152)	379-427 (48)
An. melas	623	17-49 (33)	183-334 (152)	379-428 (50)
An. merus	609	17-49 (33)	183-334 (152)	379-428 (50)
An. quadriannulatus	613	17-49 (33)	183-334 (152)	379-428 (50)
An. christyi	562	17-51 (35)	179-330 (152)	375-424 (50)
An. epiroticus	597	20-48 (29)	166-317 (152)	362-411 (50)
An. funestus	608	20-48 (29)	193-344 (152)	389-438 (50)
An. minimus	589	20-46 (27)	193-344 (152)	388-403 (16)
An. culicifacies	602	20-46 (27)	176-327 (152)	372-421 (50)
An. maculatus	466	6-34 (29)	125-276 (152)	321-370 (50)
An. stephensi	624	20-48 (29)	193-344 (152)	389-438 (50)
An. dirus	620	17-47 (31)	194-345 (152)	390-439 (50)
An. farauti	574	20-48 (29)	193-344 (152)	389-405 (17)
An. sinensis	645	17-45 (29)	175-326 (152)	371-420 (50)
An. atroparvus	615	17-49 (33)	179-330 (152)	375-424 (50)
An. albimanus	524	20-48 (29)	183-336 (154)	381-430 (50)
An. darlingi	585	20-48 (29)	185-338 (154)	383-432 (50)
An. aquasalis	650	20-48 (29)	181-338 (158)	383-432 (50)

The most striking conserved domain of the PIAS family is a MIZ/SP-RING-finger-like zinc-binding domain (RLD) and the SAP (scaffold attachment protein-A/B, acinus and PIAS) domain. The N-terminal SAP domain is evolutionarily conserved in proteins ranging from yeast to human in origin. They share the origin with other chromatin-binding proteins, such as scaffold attachment factor A and B (SAFA and B) (Aravind and Koonin 2000). The SAP domain can recognize and bind to AT-rich DNA sequences present in scaffold-attachment regions/matrix-attachment regions (S/MAR) and provide a unique nuclear microenvironment for transcriptional regulation (Kipp et al. 2000). Incorporated

within the SAP domain, PIAS proteins have an LxxLL signature motif that has been suggested to mediate interactions with the nuclear receptors and their co-regulators. (**Figure 2.9 B**) It has also been seen that the LxxLL motif in PIAS is essential for the PIAS mediated down regulation of STAT activity, but it is not required for the STAT1-PIASy interaction (Heery et al., 1997; Liu et al., 2001). In all *Anopheles* species the LQQLL motif is conserved except in subgenus *Nyssorhynchus* where LQHLL is present.

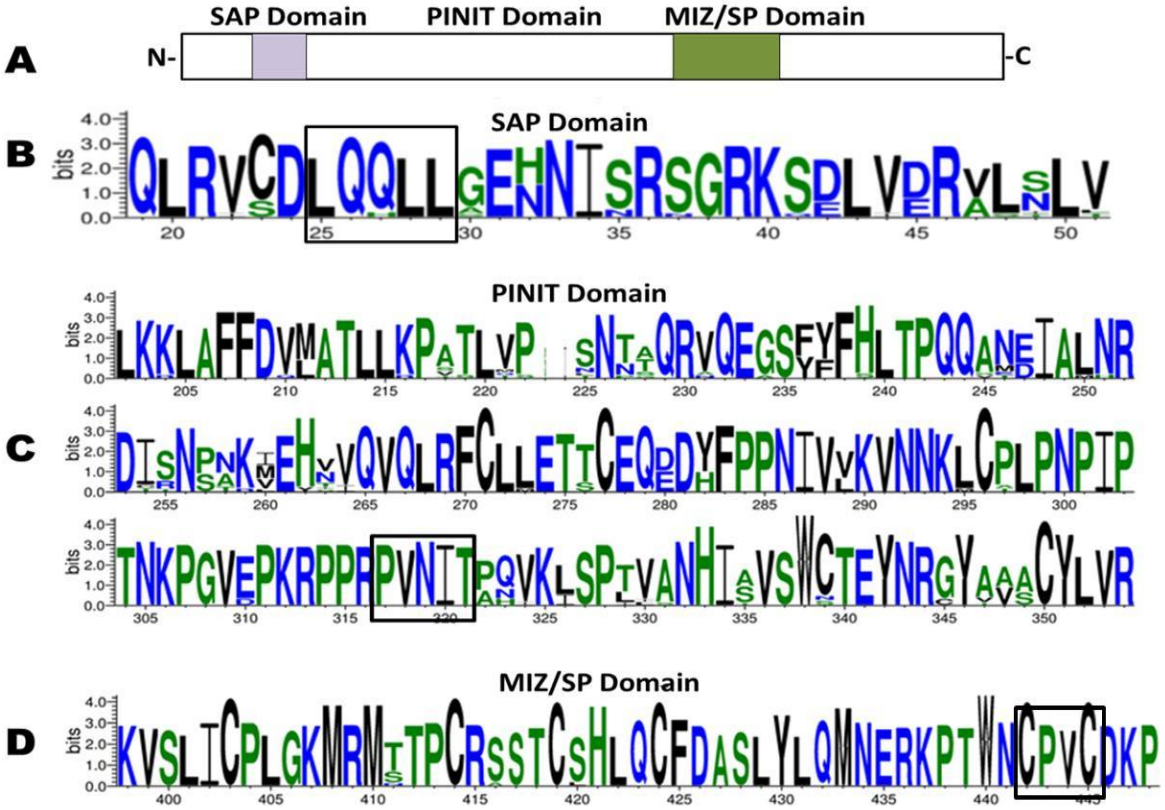


Figure 2.9 Conserved domains of PIAS protein, consensus motif and sequence variability. **A.** Schematic representation of three conserved domains of PIAS protein. **B.** Highly conserved SAP domain, showing consensus motifs (LxxLL) in all *Anopheles* mosquitoes captured in rectangular selection. **C.** Least conserved PINIT domain showing high variability in protein sequence have PVNIT motif. **D.** Third conserved MIZ/SP RING Zn Finger domain possibly mediates noncovalent interactions with SUMO (small ubiquitin-like modifier).

Next to the SAP domain, the Pro-Ile-Asn-Ile-Thr (PINIT) domain is attached, which represents a highly conserved region of PIAS proteins and involved in the nuclear retention of PIAS (Duval et al., 2003). This signature motif PINIT thought to regulate the subcellular localization found in each *Anopheles* species with some modification P**`V'**NIT (**Figure 2.9 C**). The centrally situated PINIT domain is least conserved and larger in size compared to other domain of the PIAS protein (**Table 2.17**). The C-terminal RING (Really Interesting New Gene)-finger-like zinc-binding domain (RLD) in the end of the protein is the most conserved region and is needed for the SUMO (small ubiquitin-like

modifier) E3 ligase function of PIAS proteins (**Fig. 2.9 D**) (Hochstrasser, 2001; Rytinki et al., 2009). Adjacent to the RING domain, a SUMO interacting motif is there, which mediates noncovalent interactions with SUMO proteins (Johnson, 2004; Sharrocks, 2006).

2.3.3.2 Phylogenetic correlation of PIAS

Similar to STAT-A gene, PIAS gene was also found conserved at protein level therefore they are clustered together according to the taxonomic groups. *An. albimanus*, *An. darlingi* along with *An. aquasalis* belong to series Argyritarsis of subgenus Nyssorhynchus that diverged first from the last common ancestor of subfamily Anophelinae. More advanced *An. sinensis* (series Myzorhynchus) and *An. atroparvus* (Series Anopheles) are diverged almost parallel with subgenus Nyssorhynchus (**Figure 2.10**). This diversification was found common in STAT-A, PIAS and SOCS5 gene. *An. farauti* and *An. dirus* of series Neomyzomyia have shown diverged first from the last common ancestor from subgenus Cellia. Rest of the species of subgenus Cellia, *An. maculatus* and *An. stephensi* (series Neocellia), *An. minimus* and *An. culicifacies* (series Myzomyia) and series Pyretophorus species were clustered together (**Figure 2.10**).

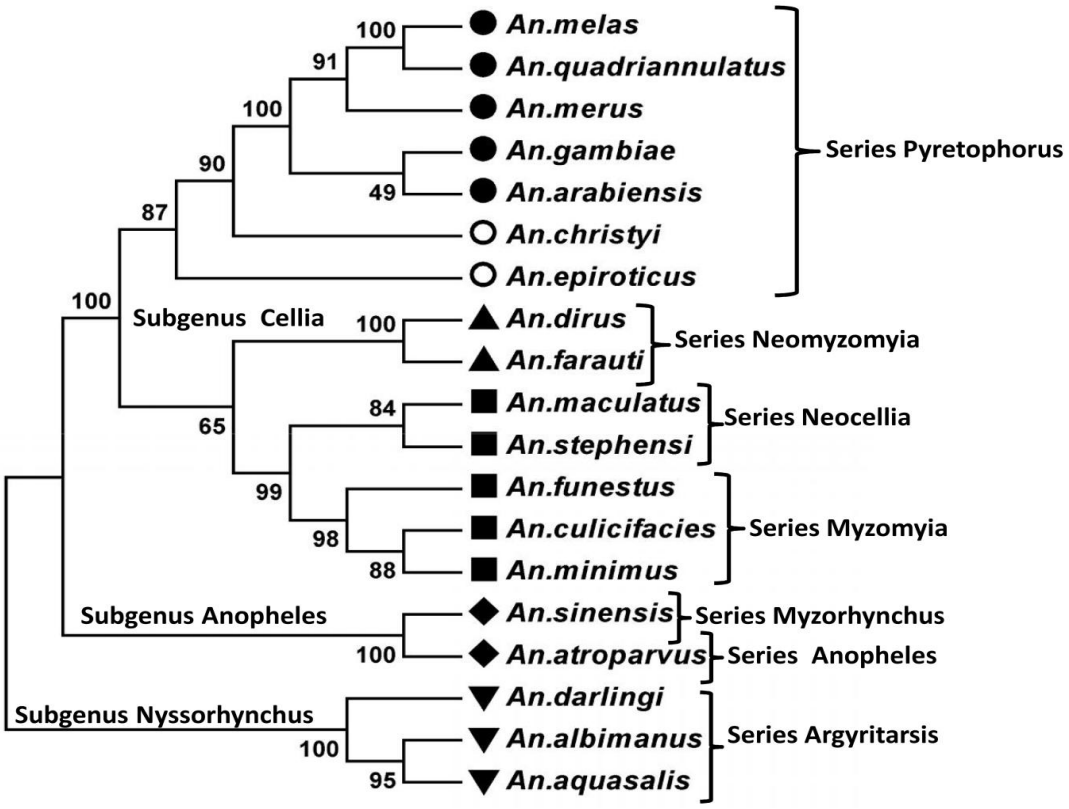


Figure 2.10 Phylogenetic tree of PIAS gene obtained through NJ method for all *Anopheles* species. The clustering patterns were further labeled with their taxonomic groups. The numbers on the branches represent the % of 1000 bootstrap. Color and shape codes are as follows: black circle (●): Pyretophorus – gambiae complex and white circle (○): non-gambiae complex. Black square (■): Myzomyia and Neocellia. Black triangle (▲): Neomyzomyia, Inverted black triangle (▼): Argyritarsis, Black diamond (♦): Anopheles, White diamond (◊): Myzorhynchus.

2.4 Discussion

Mosquitoes are the primitive metazoans in the living world where two STATs genes appeared for the first time in the taxonomical hierarchy. Among insects, *D. melanogaster* STAT pathway is well studied with their human orthologs for each component. *Anopheles* mosquitoes are separated from *Drosophila* through a common ancestor 260 million year ago and due to have divergent habit and habitat lead to different genome (Zdobnov et al., 2002). Report based comprehensive investigation exhibit that only *Anopheles* mosquitoes are the initial organism where two STAT genes appeared first. In *An. gambiae* two STAT genes showed its versatile effect on bacteria and *Plasmodium* infection (Barillas-Mury et al., 1999; Gupta et al., 2009).

Both STAT genes participate as the transcriptional regulator beginning to end (STAT-B regulates the STAT-A activation) and initiate the transcription of several effector genes to overcome the infection but their mechanism of action is still not clear (Gupta et al., 2009). *An. gambiae* has shown the late phase immune response against *P. falciparum* infection and contrary to that, another species from Neotropical region; *An. aquasalis* has shown invert effect (Early phase immunity) of STAT gene upon *P. vivax* infection (Bahia et al., 2011). Hence, it might be possible, two different type of mechanism exist in different subgenus of *Anopheles*. Analysis of 18 *Anopheles* genome, showed ancestral STAT gene (STAT-A) and retroduplicated STAT-B are present in the genome of 13 *Anopheles* species. While only one STAT gene (ancestral STAT-A) is present in 5 *Anopheles* species. Retro-duplicated STAT-B gene is not found in the genome of *An. darlingi*, *An. albimanus*, *An. sinensis*, *An. dirus* and *An. farauti*.

This suggested that retro-duplication event occurred more recently even before the divergence of the Anopheline and Culicine lineages (145-200MYA; Krzywinski et al., 2006), yet the two STAT copies are remarkably diverged. STAT-A deduced amino acid sequences are very much identical to each other and showed high similarity in the conserved motifs. Whereas STAT-B was found similar in domain architecture but sequentially it was very much different from one another especially in species belongs to subgenus *Nyssorhynchus* and Neomyzomyia of subgenus *Cellia*. Gene duplicates are often maintained in the genome because they acquire a function that is distinct from that of the ancestral gene. Even though the rapid divergence of STAT-B, there is no evidence of pseudogenization and both STAT-A and STAT-B appear to play a dependent role in immunity in the same pathway (Gupta et al. 2009). In this case, we may predict evidence of adaptive evolution in STAT-B, particularly when compared with STAT-A, due to high

amino acid divergence (Rottschaefer et al., 2015). According to the previous reports related to evolution of genome, it is an implicit fact that when genes are retrotransposed there will be introns in the parental copy, but no introns in the daughter copy. As far as the JAK-STAT signaling pathway is concerned, its involvement in defense mechanism of insects have been reported very piercingly which maintain the fact that immune pathway genes are more rapidly evolved than non-immune genes (McTaggart et al., 2012).

SOCS5 is an important molecule of the STAT pathway which regulates the developmental processes and immune responses (Dhawan et al., 2015). Particularly, STAT pathway regulates the expression of SOCS5 gene in a downstream process which bind to JAKs and negatively suppress the pathway of its own origin (Wang et al. 2011). SOCS5 proteins display a three-part architecture, the N-terminal, central SH2 and C-terminal SOCS box domain. The SH2 domain is involved in substrate binding through recognition of cognate phosphotyrosine motif. Anophelines SOCS5 domains (SH2 and SOCS box) exhibit considerable higher amino acids identity with each others. However, the N-terminal region is highly divergent and may indicate that different Anopheline SOCS5 domains experienced differential selection pressures. This diversified N-terminal of SOCS protein have a conseved 92 amino acid long NTCR which may function like the JIR (JAK inhibitory region) of mammalian SOCS5 (Feng et al., 2011; Chandrashekaran et al., 2015). Sequence analysis reveals the existence of Elongin B/C box in SOCS box domain (~50 amino acids) depicts how SOCS acts as adaptor protein to regulate signal transduction by linking their substrates to the ubiquitination machinery which eventually leads to proteosomal degradation of the target proteins in a way similar to other insects (Stec and Zeidler, 2011). The PIAS protein has been proposed to interact with many transcription factors involved in the immune system. PIAS protein revealed much conserved domains viz. SAP domain, PINIT domain and the MIZ/SP-RING zinc finger domain. Phylogenetic study of SOCS and PIAS genes reveal the same evolutionary pattern as it was followed in taxonomical hierarchy.

2.5 Conclusion

Two STAT transcription factors were only been recognized in *An. gambiae* genome. We here elucidated the occurrence of two STAT genes in other *Anopheles* species except *An. darlingi*, *An. albimanus*, *An. sinensis*, *An. dirus* and *An. farauti* in which only ancestral STAT was found. The significance of having two STAT genes in mosquito genome like vertebrates is very interesting because surprisingly it is not present in other insects

including *Drosophila* (Zdobnov et al., 2002). We showed that it is possible to reconstruct the taxonomic hierarchy of this gene family, taking advantage of its individual protein sequence. Both STATs are sharing the same STAT functional domain. This study was investigated to understand the differential evolution of STAT gene since the formation of the canonical JAK-STAT signaling pathway. This study will provide new insights into the evolutionary process, with implications for immune system evolution in different *Anopheles* mosquitoes. The STAT genes family of mosquito makes a remarkable study for molecular evolution, as it is fairly small, usually no more than two paralogs per species.

Interestingly, we found that *Anopheles* SOCS5 is present in all Anopheline mosquitoes with variable length due its N-terminal diversity. The SH2 and SOCS box domain exhibit considerable higher amino acids identity with each other. We have observed the N-terminal region of all *Anopheles* species have some conserved region called as NTCR and might be the orthologs of human JIR of SOCS5.

PIAS gene architecture confirms the presence of 10 exons, separated by 9 introns. Phylogenetic analysis indicates PIAS of same series of *Anopheles* are closer to each other. The study gives an insight into changes that may be taking place in the JAK-STAT pathway genes from *Anopheles* genome. The canonical cytokine receptor-JAK-STAT system, including its key negative regulators, evolved prior to the appearance of chordates, being observed in extant invertebrates such as fruit fly (Liongue et al., 2012). This study has sought to understand the subsequent evolution and diversification of this system during invertebrate evolution through the examination of the STAT, PIAS and SOCS gene in relevant species.

2.6 References

Aaronson DS, Horvath CM. 2002. A road map for those who don't know JAK-STAT. Science 296:1653-1655.

Aravind L, Koonin EV. 2000. SAP- a putative DNA-binding motif involved in chromosomal organization. Trends Biochem Sci. 25:112-114.

Babon JJ, Sabo JK, Zhang JG, Nicola NA, Norton RS. 2009. The SOCS box encodes a hierarchy of affinities for Cullin5: implications for ubiquitin ligase formation and cytokine signaling suppression. J Mol Biol. 387:162-174.

Bach EA, Perrimon N. 2003. Prime time for the *Drosophila* JAK/STAT pathway. In Signal Transducers and Activators of Transcription (STATs). *Activation and Biology* (Sehgal PB, Levy DE and Hirano T. eds) pp. 87-104, Kluwer Academic Publishers, Dordrecht, 87-104.

Bahia AC, Kubota MS, Tempone AJ, Araujo HRC, Guedes BAM, Orfanó AS. et al., 2011. The JAK-STAT pathway controls *Plasmodium vivax* load in early stages of *Anopheles aquasalis* infection. PLoS Negl Trop Dis. 5(11): e1317.

Barillas-Mury C, Han YS, Seeley D, Kafatos FC. 1999. *Anopheles gambiae* AgSTAT, a new insect member of the STAT family, is activated in response to bacterial infection. EMBO J. 18:959-967.

Besansky NJ. 2008. Genome analysis of vectorial capacity in major *Anopheles* vectors of malaria parasites. White paper submitted in August 2008 on behalf of the *Anopheles* Genomes Cluster Committee (AGCC).

Betz A, Lampen N, Martinek S, Young MW, Darnell JE, Jr. 2001. A *Drosophila* PIAS homologue negatively regulates stat92E. Proc Natl Acad Sci USA 98:9563-9568.

Binari R, Perrimon N. 1994. Stripe-specific regulation of pair-rule genes by hopscotch, a putative Jak family tyrosine kinase in *Drosophila*. Genes Dev. 8:300-312.

Brown S, Hu N, Hombria JC. 2001. Identification of the first invertebrate interleukin JAK/STAT receptor, the *Drosophila* gene domeless. Curr Biol. 11:1700-1705.

Burge C, Karlin S, 1997. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 268:78-94.

Chandrashekaran IR, Mohanty B, Linossi EM, Dagley LF, et al., 2015. Structure and functional characterization of the conserved JAK interaction region in the intrinsically disordered N-terminus of SOCS5. Biochemistry 54(30):4672-4682.

Chen HW, Chen X, Oh SW, Marinissen MJ, Gutkind JS, Hou SX. 2002. Mom identifies a receptor for the *Drosophila* JAK/STAT signal transduction pathway and encodes a protein distantly related to the mammalian cytokine receptor family. Genes Dev. 16:388-398.

Chen WY, Ho KC, Leu JH, Liu KF, Wang HC, Kou GH, et al., 2008. WSSV infection activates STAT in shrimp. Dev Comp Immunol. 32(10):1142-1150.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, et al., 2003. Multiple sequence alignment with the clustal series of programs. Nucl Acids Res. 31:3497-3500.

Cheng CH, Chen GD, Yeh MS, Chu CY, Hsu YL, Hwang PP, et al., 2010. Expression and characterization of the JAK kinase and STAT protein from brine shrimp, *Artemia franciscana*. Fish Shellfish Immunol. 28 (5-6):774-782.

Christophides GK, Zdobnov EM, Barillas-Mury C, Birney E, Blandin S, Blass C, et al., 2002. Immunity-related genes and gene families in *Anopheles gambiae.* Science 298:159-165.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188-1190.

Darnell JE Jr, Kerr IM, Stark GR. 1994. JAK-STAT pathway and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science 264:1415-1421.

Darnell JE Jr. 1997a. STATs and gene regulation. Science 277(5332):1630-1635.

Darnell JE Jr. 1997b. Phosphotyrosine signaling and the single cell: metazoan boundary. Proc Natl Acad Sci USA 94:11767-11769.

Dearolf CR. 1999. JAKs and STATs in invertebrate model organisms. Cell Mol Life Sci. 55:1578-1584.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3: e314.

Dhawan R, Gupta K**,** Kajla M, Kumar S, Gakhar SK, Kakani P, et al., 2015. Molecular characterization of SOCS gene and its expression analysis on *Plasmodium berghei* infection in *Anopheles culicifacies*. Acta Tropica 152:170-175.

Duval D, Duval G, Kedinger C, Poch O, Boeuf H. 2003. The 'PINIT' motif of a newly identified conserved domain of the PIAS protein family, is essential for nuclear retention of PIAS3L. FEBS Lett. 554:111-118.

Feng ZP, Chandrashekaran IR, Low A, Speed TP, Nicholson SE, Norton RS. 2012. The N-terminal domains of SOCS proteins: a conserved region in the disordered N-termini of SOCS4 and 5. Proteins 80:946-957.

Finn RD, Tate J, Mistry J, Coggill PC, Sammut JS, Hotz HR, et al., 2008. The Pfam protein families' database. Nucl Acids Res. 36 (Database Issue): D281-D288.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, et al., 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 346:1258524.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucl Acids Res. 31(13):3784-3788.

Gaunt MW, Miles MA. 2002. An insect molecular clock dates the origin of the insects and accords with paleontological and biogeographic landmarks. Mol Biol Evol. 19(5):748-761.

Gupta L, Molina-Cruz A, Kumar S, Rodrigues J, Dixit R, Zamora RE, Barillas-Mury C. 2009. The STAT pathway mediates late-phase immunity against *Plasmodium* in the mosquito *Anopheles gambiae*. Cell Host Microbe 5:498-507.

Hari KL, Cook KR, Karpen GH. 2001. The *Drosophila* Su(var)2-10 locus regulates chromosome structure and function and encodes a member of the PIAS protein family. Genes Dev. 15(11):1334-1348.

Harrison DA, McCoon PE, Binari R, Gilman M, Perrimon N. 1998. *Drosophila* unpaired encodes a secreted protein that activates the JAK signaling. Genes Dev. 12:3252-3263.

Heery DM, Kalkhoven E, Hoare S, and Parker MG. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature 387:733-736.

Hino K, Satou Y, Yagi K, Satoh N. 2003. A genome wide survey of developmentally relevant genes in *Ciona intestinalis*. VI. Genes for Wnt, TGFbeta, Hedgehog and JAK/STAT signaling pathways. Dev Genes Evol. 213:264-272.

Hochstrasser M. 2001. SP-RING for SUMO: New functions bloom for an ubiquitin-like protein. Cell 107(1):5-8.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science. 298:129-149.

Horwitz KB, Jackson TA, Bain DL, Richer JK, Takimoto GS, Tung L. 1996. Nuclear receptor coactivators and corepressors. Mol Endocrinol. 10:1167-1177.

Hou XS, Melnick MB, Perrimon N. 1996. Marelle acts downstream of the *Drosophila* HOP/JAK kinase and encodes a protein similar to the mammalian STATs. Cell 84:411-419.

Hurst LD, Pal C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet. 5:299-310.

Jackson PK. 2001. A new RING for SUMO: wrestling transcriptional responses into nuclear bodies with PIAS family E3 SUMO ligases. Genes Dev. 15:3053-3058.

Johnson ES. 2004. Protein modification by SUMO. Annu Rev Biochem. 73:355-382.

Karsten P, Häder S, Zeidler MP. 2002. Cloning and expression of *Drosophila* SOCS36E and its potential regulation by the JAK/STAT pathway. Mech Dev.117:343-346.

KawataT, Shevchenko A, Fukuzawa M, Jermyn KA, Totty NF, Zhukovskaya NV, et al., 1997. SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in *Dictyostelium*. Cell 89:909-916.

Kimmel AR, Firtel RA. 2004. Breaking symmetries: regulation of *Dictyostelium* development through chemoattractant and morphogen signal-response. Curr Opin Genet Dev. 14:540-549.

Kipp M, Gohring F, Ostendorp T, van Drunen CM, van Driel R, Przybylski M. et al., 2000. SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. Mol Cell Biol. 20:7480-7489.

Krzywinski J, Grushko OG, Besansky NJ. 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol Phylogenet Evol. 39:417-423.

Lewis RS, Ward AC. 2004. Conservation, duplication and divergence of the zebrafish stat5 genes. Gene 338:65-74.

Liongue C, O'Sullivan LA, Trengove MC, Ward AC. 2012. Evolution of JAK-STAT pathway components: mechanisms and role in immune system development. PLoS ONE 7(3): e32777.

Lin CC, Chou CM, Hsu YL, Lien JC, Wang YM, Chen ST, et. al., 2004. Characterization of Two Mosquito STATs, AaSTAT and CtSTAT differential regulation of tyrosine

phosphorylation and dna binding activity by lipopolysaccharide treatment and by japanese encephalitis virus infection. J Biol Chem. 279(5):3308-3317.

Liu B, Gross M, Ten Hoeve J, Shuai K. 2001. A transcriptional corepressor of Stat1 with an essential LXXLL signature motif. Proc Natl Acad Sci USA 98:3203-3207.

Marchler-Bauer A, Bryant SH. 2004. CD Search: protein domain annotations on the fly. Nucl Acids Res. 32(Web Server issue): W327-W331.

Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, et al., 2007. CDD: a conserved domain database for interactive domain family analysis. Nucl Acids Res. 35(Database Issue): D237-D240.

McTaggart SJ, Obbard DJ, Conlon C, Little TJ. 2012. Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex*. BMC Evol Biol. 12:63.

Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol. 11:699-704.

Mohr SE, Boswell RE. 1999. Zimp encodes a homologue of mouse Miz1 and PIAS3 and is an essential gene in *Drosophila melanogaster*. Gene 229(1-2):109-116.

Müller P, Kuttenkeuler D, Gesellchen V, Zeidler MP, Boutros M. 2005. Identification of JAK/STAT signaling components by genome-wide RNA interference. Nature 436(7052):871-875.

Neafsey DE, Christophides GK, Collins FH, Emrich SJ, Fontaine MC, Gelbart W, et al., 2013. The evolution of the *Anopheles* 16 genomes project. G3 (Bethesda) 3:1191-1194.

Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al., 2015. Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. Science 347(6217):1258522.

Perl TN, Schmid BGM, Schwirz J, Chipman AD. 2014. The evolution of the knirps family of transcription factors in arthropods. Mol Biol Evol. 30(6):1348-1357.

Rogers RS, Horvath CM, Matunis MJ. 2003. SUMO modification of STAT1 and its role in PIAS-mediated inhibition of gene activation. J Biol Chem. 278(32):30091-30097.

Rottschaefer SM, Crawford JE, Riehle MM, Guelbeogo WM, Gneme A, Sagnon N. et al., 2015. Population genetics of *Anopheles coluzzii* immune pathways and genes. G3 (Bethesda) 5:329-339.

Rytinki MM, Kaikkonen S, Pehkonen P, Jääskeläinen T, Palvimo JJ. 2009. PIAS proteins: Pleiotropic interactors associated with SUMO. Cell Mol Life Sci. 66:3029-3041.

Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, et al., 2015. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 16(1):76.

Saitou N, Nei M. 1997. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406-425.

Sharakhova MV, Hammond MP, Lobo NF, Krzywinski J, Unger MF, et al., 2007. Update of the *Anopheles gambiae* PEST genome assembly. Genome Biol. 8(1):R5.2

Sharrocks AD. 2006. PIAS proteins and transcriptional regulation-more than just SUMO E3 ligases? Genes Dev. 20:754-758.

Shuai K. 2006. Regulation of cytokine signaling pathways by PIAS proteins. Cell Res. 16:196-202.

Shuai K, Liu B. 2005. Regulation of gene-activation pathways by PIAS proteins in the immune system. Nat Rev Immuno. 5:593-605.

Sidow A. 1996. Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev. 6(6):715-722.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7(1):539.

Skrabanek L, Wolfe KH. 1998. Eukaryote genome duplication-where's the evidence? Curr Opin Genet Dev. 8:694-700.

Srivastava M, Simakov O, Chapman J, Fahey B, et al., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466:720-726.

Stec W, Vidal O, Zeidler MP. 2013. *Drosophila* SOCS36E negatively regulates JAK/STAT pathway signaling via two separable mechanisms. Mol Biol Cell. 24(18): 3000-3009.

Stec WJ, Zeidler MP. 2011. *Drosophila* SOCS proteins. J Signal Transduct. 2011:894510.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol Biol Evol. 28:2731-2739.

Tanaka H, Ishibashi J, Fujita K, Nakajima Y, Sagisaka A, Tomimoto K, et al., 2008. A genome-wide analysis of genes and gene families involved in innate immunity of *Bombyx mori*. Insect Biochem Mol Biol. 38(12):1087-1110.

Wang T, Gao Q, Nie P, Secombes CJ. 2010. Identification of suppressor of cytokine signaling (SOCS) 6, 7, 9 & CISH in rainbow trout *Oncorhynchus mykiss* and analysis of their expression in relation to other known trout SOCS. Fish Shellfish Immunol. 29(4):656-667.

Wang T, Gorgoglione B, Maehr T, Holland JW, Vecino JL,Wadsworth S, et al., 2011. Fish suppressors of cytokine signaling (SOCS): gene discovery, modulation of expression and function. J Signal Transduct. 2011:905813.

Wang Y, Levy DE. 2006. *C. elegans* STAT: evolution of a regulatory switch. FASEB J. 20(10):1641-1652.

Wang Y, Levy DE. 2012. Comparative evolutionary genomics of the STAT family of transcription factors. JAK-STAT. 1:23-36.

Wright VM, Vogt, Smythe E, Zeidler MP. 2011. Differential activities of the *Drosophila* JAK/STAT pathway ligands Upd, Upd2 and Upd3. Cell Signal. 23:920-927.

Yan R, Small S, Desplan C, Dearolf CR, Darnell JE, Jr. 1996. Identification of a Stat gene that functions in *Drosophila* development. Cell 84:421-430.

Zdobnov EM, Mering CV, Letunic I, Torrents D, Suyama M, Copley RR, et al., 2002. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. Science 298:149-159.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F. et al., 2012. The oyster genome reveals stress adaptation and complexity of shell formation Nature 490 (7418):49-54.

Zhang J. 2003. Evolution by gene duplication: an update. Trends Ecol Evol. 18:292-298.

Zhang S, Coultas KA, 2011. Identification and characterization of five transcription factors are associated with evolutionarily conserved immune signaling pathways in the schistosome-transmitting snail *Biomphalaria glabrata*. Mol Immunol. 48:1868-1881.

Zhukovskaya NV, Fukuzawa M, Tsujioka M, Jermyn KA, Kawata T, Abe T. et al., 2004. Dd-STATb, a *Dictyostelium* STAT protein with a highly aberrant SH2 domain, functions as a regulator of gene expression during growth and early development. Development 131(2):447-458.

Objective: Characterization of STAT signaling pathway genes in Anopheles stephensi mosquito and their role in development and antiplasmodial immunity

Part of this chapter has been published in: Gupta K**,** Dhawan R, Kajla M, Misra T, Kumar S and Gupta L. The evolutionary divergence of STAT transcription factor in different *Anopheles* species. **Gene.** 2016 (*In press*)

3.1 Introduction

3.1.1 Role of STAT signaling pathway in development of insects

Drosophila melanogaster has been extensively used as a representative organism in developmental research. These studies have revealed the essential role of the JAK-STAT pathway in many developmental processes (Harrison et al., 1995). The first evidence that the JAK-STAT pathway involved in development came from the characterization of dominant gain-of-function alleles of hop (Perrimon and Mahowald, 1986; Williams, 2000). The massive amount of roles played by this signal transduction pathway regulated through multiple regulatory mechanisms. Its role starts with sex determination and segmentation and subsequently the formation of the tracheal pits, elongation of intestinal tracks and formation of the posterior spiracles in *Drosophila* (Arbouzova and Zeidler, 2006). Other, less well defined roles include: development of axons within the central nervous system and embryonic dorsal/ventral patterning (Ihle et al., 1996). This pathway is also needed for eye development, cell proliferation, photoreceptor differentiation and establishment of equator and ommatidia polarity (Luo et al., 1999). Classical embryological studies in *Drosophila* revealed broad developmental similarities exist among diverse dipteran species (Goltsev et al., 2007). Here we consider the information about conserved roles of the STAT pathway during mosquito development in view of their similar and unwavering genomic array of developmental gene.

3.1.2 Role of STAT signaling pathway in insect immunity

Insects have an evolutionarily conserved first-line host defence that comprises the senses of pathogenic microorganisms through pattern-recognition receptor (PRR) molecules (Hoffmann and Reichhart, 2002; Shuai and Liu, 2003). Innate immune system not only combats foreign invaders but it also employed in wound healing, stress responses and the management of microbial symbiotic populations (Loker et al., 2004). However, memory-like responses have been reported in some insects that are termed as “immune priming” (Pham et al., 2007). There are examples of nonspecific and pathogen-specific priming in *An. gambiae* that can be long-lasting. Very recently a molecular mechanism has been elucidated, which allows the innate immune system to ‘remember’ a previous encounter with a pathogen (Ramirez et al., 2014). Challenged mosquitoes constitutively release a soluble haemocyte

differentiation factor (Lipoxin/Lipocalin complex) into their hemolymph, when transferred into naive mosquitoes, which ultimately induces priming (Rodrigues et al., 2010; Ramirez et al., 2015). The signaling pathway(s) that mediate the establishment of immune priming are not known.

Although Toll and Imd signaling pathways have been well studied against bacterial, fungal and *Plasmodium* infections ((Dimopoulos et al., 2002; Meister et al., 2005; Blumberg et al., 2013). Now many evidences have recently been cited to describe the role of JAK-STAT pathway in anti-bacterial and anti-viral infections (Gupta et al., 2009; Souza-Neto et al., 2009). The first evidence that the JAK-STAT pathway also plays a role in insect immunity was shown in the mosquito *Anopheles gambiae* where AgSTAT translocate into the nucleus of fat body cells in response to bacterial infection (Barillas-Mury et al., 1999). In *Drosophila*, STAT92E translocate into the nucleus of fat body cells upon bacterial challenge, which resulted into expression of several anti-microbial peptides including Tep and Tot protein families. Tep1 is one of the four members of the Tep family that contain thioester motifs and has high similarity to C3/α2-macroglobulin super family of human. Tep1 expression was found highly induced in fat body cells upon immune challenge in *Drosophila* and *An. gambiae* (Lagueux et al., 2000; Levashina et al., 2001). Tep1 is also induced upon *Plasmodium* infection and responsible for lysis of ookinetes during evasion of mosquito midgut (Blandin et al., 2004 and 2008).

Mosquitoes primarily rely on Toll, Imd, JAK-STAT and recently known JNK signaling pathways for limiting pathogen infection (Ramphul et al., 2014). STAT signaling in Indian *Anopheles* species against *Plasmodium* infection has not been characterized yet. Literature supports that there is much genetic diversity exist in *Anopheles* mosquitoes due to different geography and host-parasite interaction (Neafsey et al. 2015). Our study also revealed that STAT gene varies in different *Anopheles* species suggesting that it might act differently from one species to another species. Moreover, *An. gambiae* STAT pathway display the late phase immunity against *P. falciparum* and *P. berghei* while Brazilian species, *An. aquasalis* show the early immune response against *P. vivax* infection (Gupta et al., 2009, Bahia et al., 2011). Immune genes in general evolve at faster rate than whole genome, which is explained by the persistent selective pressures posed by new type of pathogens (Obbard et al., 2009). Therefore, we analysed STAT pathway genes in *An. stephensi* against *P. berghei* infection.

3.2 Materials and Methods

3.2.1 *Anopheles* mosquito characterization and colony maintenance

The *Anopheles stephensi* mosquitoes used in the present research work were collected from nearby locality of Pilani, Rajasthan region (28°22′N 75°36′E). Mosquito larvae were collected from the field and reared inside the insectary. These larvae kept in deionized water in plastic pans or tray (white or transparent) in sterile condition according to manual for mosquito rearing and experimental techniques published by American Mosquito Control Association California (AMCA) (Gerberg, 1979; Feldmann et al., 1989). All the stages of mosquito were reared inside the insectary with 28±0.5^{0}C temperature, 80±5.0% relative humidity and 12-hour light and 12-hour dark cycle. For better growth and adaptation, we kept approximately 200 larvae per pan with a fine powder of tetramin (TetraWerke, Melle, Germany), fish food (Gold Tokyo, India) and dog feed (Pet Lover`s crunch milk biscuit, India) in 1:1:1 ratio until pupation. Pupae were kept in standard cages for hatching in to the adult form.

Adults were maintained on 10% sucrose-soaked cotton balls which were changed every day. Supply of sucrose was eliminated 12 hours prior to blood-feeding. Blood feeding of female mosquitoes was performed on anaesthetized swiss albino mice (*Mus domesticus*) with the prior permission of the Institutional animal ethical committee (Protocol No IAEC/RES/15/01). After feeding, egg cups covered with cellulose paper was kept inside the mosquito cage for egg laying. Life-cycle of mosquito has been maintained throughout the study.

The laboratory reared *Anopheles stephensi* mosquitoes were characterized morphologically and at molecular level before the commencement of experiments. Mosquitoes were initially identified using morphological identification keys customary used for larva and adults' identification (Glick, 1992). Morphological characterization was done through microscopic examination. The key feature include palpi with speckling, apical and sub apical band are equal separated by a dark band, speckled appendages, 4 dark bands on costa and sub costa, 6th vein has two dark spots as shown in **Figure 3.1A** with black arrow heads.

The molecular marker of second internal transcribed spacer (ITS-2) and third domain (D3) of the rDNA was used for molecular characterization. For that, lab designed ITS-2 and D3

primers from the reported sequence for *An. stephensi* conserved region were used (Alam et al., 2008). These primer sequences were as follows:

AsITS2-Fwd 5'-ATCGGACGACTCAACCCAAC-3'

AsITS2-Rev 5'-CTGTCGTGCTAACCTCACTCAC-3' and

AsD3B-Fwd 5'-TCGGAAGGAACCAGCTACTA-3'

AsD3A-Rev 5'-GACCCGTCTTGAAACACGGA-3'

PCR was done from gDNA taken directly from the mosquito leg and the same mosquito was use to maintain the pure-line. Desired sizes of band from agarose gel had been eluted and sent for sequencing (**Figure 3.1 B**). BLAST result confirmed the 99% identity and 100% query coverage with the already reported sequences for *An. stephensi*.

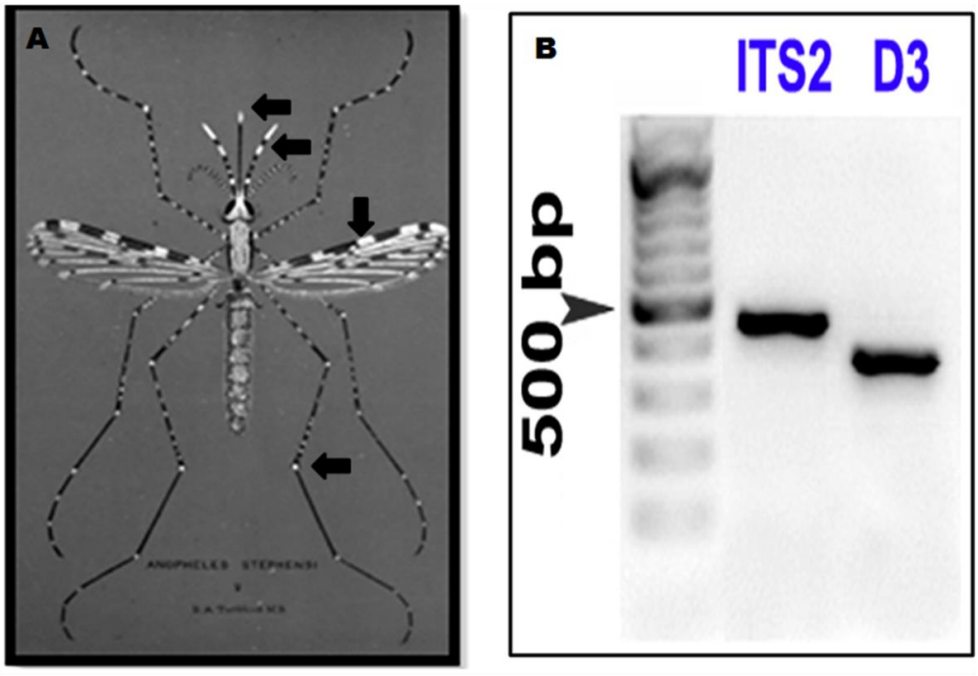


Figure 3.1 Morphological and molecular identification of field collected *An. stephensi* mosquito **A.** *Anopheles stephensi* mosquito were identified through peculiar morphological features comprises banding pattern on labium, palpi, wings and hind leg marked with black arrows (Das et al., 1990). **B.** Molecular characterization through ITS-2 and D3 markers showing desired band size and sequencing revealed 100% similarity with reported sequences of *Anopheles stephensi*.

3.2.2 Maintenance of *Plasmodium* in mouse and mosquitoes infection

Mice were maintained in central animal facility (CAF) of institute at standard condition. Only 25-30gm and ≤ 3 month old female mice were used in the present study with proper handling and care. To infect the mice, *Plasmodium berghei* (ANKA strain) infected frozen P_o blood

was intraperitoneally (ip) injected immediately after thawing it from -80°C deep freezer. After 1 week, when infection appears, blood from these mice subsequently passed in another mouse. For mosquito infection study, we used mice having 5-6% parasitemia and 1-2 exflagellations/field (Billker et al., 1997).

For infection study 3-4 days old female mosquitoes were fed on anesthetized *P. berghei* infected mice. After infected blood feeding, mosquitoes were maintained at 19-21°C temperature inside the incubator (as this is the optimum temperature for *P. berghei* sporogonic development) (Vandenbergh and Yoeli, 1966) and provide 10% sucrose solution *ad libitum*. Control (only blood-fed) and *Plasmodium* infected mosquito after 24 hours were dissected on ice chilled 1x Ash burner's PBS. Their midgut and carcass (remaining body parts) were collected in RNA later (Qiagen) and stored at -80°C. For developmental stages, samples were collected from different life forms *viz.* eggs, 4th-instar larvae, late pupae (tanned stage) and non blood fed adult mosquitoes (male and females) and snap freezed at -80°C till further use.

3.2.3 Degenerate and gene specific PCR and cloning

Initially due to unavailability of genome sequence of *An. stephensi* (before 2013) degenerate primers were designed on the basis of conserved regions of STAT, SOCS and PIAS with the help of known amino acid sequences of *An. gambiae*, *An. aquasalis*, *Aedes aegypti*, *Apis mellifera* and *D. melanogaster*. Only insect species have been taken into the consideration to design more accurate and less degeneracy in primers (**Table 3.1**).

Table 3.1 List of degenerate primers of STAT pathway

S.No.	Deg Primers	Primer Sequence (5'→3')
1.	DegSTAT-Fwd	GARAADCARCCRCCRCARGTSATG
2.	DegSTAT-Rev	GTRAAYGGCTGRATGTGYAGDAYCTG
3.	DegPIAS-Fwd	GCCGCTCGAGCARGTICARYTIAGRTTYTGYYT
4.	DegPIAS-Rev	GCCGCTCGAGCCAIGTIGGYTTYYTYTCRTTCATYTG
5.	DegSOCS-Fwd	GCCGCTCGAGCCIGHIGGIACITTYYTIGTIMGNGA
6.	DegSOCS-Rev	GCCGCTCGAGTTYTGYTTRTARTGRTAYTCYTT

For cloning of the STAT, SOCS and PIAS genes, gDNA or midgut cDNA were used as template. The PCR cycles were used as follows: two cycles (1 min steps at 95^{0}C, 55^{0}C and 72^{0}C, and 95^{0}C, 42^{0}C and 72^{0}C) followed by 40 cycles at moderate stringency (1 min steps

at 95^{0}C, 52^{0}C and 72^{0}C) and a final 10 min extension at 72^{0}C. We followed the same PCR protocol for all degenerate amplification as per the earlier reports (Barillas-Mury et al., 1999). All amplicons generated were purified through Qiaquick PCR purification kit (Qiagen Cat No. 28104). If purified products were not used immediately for cloning, 3' poly-A overhang was inserted and cloned into PCR-II TOPO TA-Vector® (Invitrogen Cat No K46001-01) following the manufacture instruction. Recombinant plasmid was utilized to transform high efficiency DH5-α TOP10 *Escherichia coli* chemi-competant cell. Few white colonies (anticipate to contains insert) were screened through colony PCR with M13 universal primers present in the vector and checked on a gel. The standard DNA molecular marker (Fermentas, catalogue no. SM0331) was used as reference to identify the size of amplified DNA fragments in each gel image. Recombinant colony containing inserts were used for plasmid isolation and sent for sequencing (either to Delhi University or Excleris Lab Pvt. Ltd. Ahmadabad). Sequencing results were analysed using Chromas software (http://www.technelysium.com.au/) and confirmed using BLAST search.

To clone full genes of STAT-A, STAT-B, SOCS and PIAS, annotated sequences of *An. stephensi* was used as given in chapter 2. The majority of primers were: 19–22 base pairs long, selected to have a GC content of >50% with fewer than 4 contiguous identical bases and melting temperatures between 55°C and 60°C. Total cDNA and gDNA were used as template to confirm the size of exon and introns. Following primer sequences (5'→3') (primer numbers 1-8 used in result section) were utilize to amplify the full gene (**Table 3.2**).

3.2.4 Total RNA isolation and cDNA synthesis

Total RNA was isolated from eggs, IVth instar larvae, pupae, female and male adult mosquito. Additionally, RNA was isolated from Pb-infected midgut and remaining whole body except midgut (Carcass) using the RNeasy Mini Kit (Qiagen catalogue No. 74104) according to the manufacturer instructions. RNA was eluted in nuclease free water already having the RNAse inhibitor (Genei) (1µl in 100µl of water) and stored in -80°C deep freezer until further use. First-strand cDNAs were synthesized from 1-2µg of total RNAs using QuantiTect Reverse Transcription kit, (Qiagen Catalogue No. 205311) according to kit manual instruction which contains gDNA wipe-out buffer to remove gDNA contamination. The cDNA was diluted with equal volume 1x TE buffer and used 1µl as template for amplification of desired gene using gene-specific primers.

Table 3.2 List of gene specific primers used in the study

Primer No.	Primer Name	Primer Sequence (5'→3')
	STAT-A	
1	AsSTAT-A-PW Forward:	TCGCTTCATTAGAGCGGGAC
2	AsSTAT-A-Seq Forward	CAATCCGGCCCAGCTGTAC
3	AsSTAT-A-GS Reverse:	GTGCGCACCATTATCTGCAG
4	AsSTAT-A Realtime Forward:	CCTGTGTGAAAAAGCGTTCA
5	AsSTAT-A Realtime Reverse:	ATAATCCTCCGCCTTCGACT
6	AsSTAT-A-PW Reverse:	CGGGCACCTGACCGTAATTG
7	AsSTAT-A-3'UTR Reverse:	CGTAAACCTTAAGCTACCGTGC
	STAT-B	
1	AsSTAT-B-5'UTR Forward:	TAGTGTTTTCCCCGCTGCAT
2	AsSTAT-B-WG Forward:	GCTTTGCAGTCATCGACCTG
3	AsSTAT-B-GS Reverse:	TGCCGTATTGATCGTTGCCT
4	AsSTAT-B-PW Forward:	TATCGGTGCGTATGCCAAGC
5	AsSTAT-B-WG Reverse:	CGAAGAATTTGGGGCGGTTG
6	AsSTAT-B-RT Forward:	GCCAGTTGTAAGCCGGGCACA
7	AsSTAT-B-PW Reverse:	CAGCTCCGTCGCGATATAGT
8	AsSTAT-B-3'UTR Reverse:	CTTTTCATCATCGTAAGCTCCG
	SOCS5	
1	AsSOCS-WG1 Forward:	CCATATGCTAAGCTCGCTCCG
2	AsSOCS-WG Forward:	GACCATCATCCGTCGTTCTT
3	AsSOCS-GS Reverse:	TTGTTCGCCATCAGCAGCCC
4	AsSOCS-GSP Forward:	CGTCGTACGTCGTATTGCTC
5	AsSOCS-GSP Reverse:	CGGAAGTACAATCGGTCGTT
6	AsSOCS-WG Reverse:	CCGCTGCCGGTAGTGATACTC
7	AsSOCS-WG1 Reverse:	CATCGAGCAGCTACGCGTTGG
	PIAS	
1	AsPIAS-PW Forward:	TAGCCGACGAGTTTCCTGTG
2	AsPIAS Realtime Forward:	ACAACGACGCATCAAAGCAC
3	AsPIAS Realtime Reverse:	GTGTCCGTTGCCGAATCCTA
4	AsPIAS-PW Reverse:	TACACCAACTGCGAGTCTGC
	NOS	
	AsNOS Forward:	ACATCAAGACGGAAATGGTTG
	AsNOS Reverse:	ACAGACGTAGATGTGGGCCTT
	Other Primers	
	S7 Forward:	GGCGATCATCATCTACGT
	S7 Reverse:	GTAGCTGCTGCAAACTTCGG
	Pb (28srRNA) Forward:	CGTGGCCTATCGATCCTTTA
	Pb (28srRNA) Reverse:	GCGTCCCAATGATAGGAAGA
	M13 Forward:	TGTAAAACGACGGCCAGT
	M13 Reverse:	CAGGAAACAGCTATGAC
	Xho-T7-M13 Reverse: CTCGAGTAATACGACTCACTATAGGGCAGGAAACAGCTATGAC	

3.2.5 Expression analysis of STAT pathway genes through real time PCR

The resulting cDNA was analyzed for expression of different genes by semi-quantitative real time PCR using the BioRad iCycler Optical Module (Biorad). Reactions were prepared in volumes of 20µl using 2 X IQ SYBR Green supermixes (Biorad) and 1pmoles/µl of final primer concentration with respective cDNA templates. PCR cycle parameters involved an initial denaturation at 95°C for 5 min, 40 cycles of 10 s at 95°C, 20 s at 58°C and 30 s at 72°C. Fluorescence readings were taken at 75°C after each cycle. A final extension at 72°C for 10 min was completed before deriving a melting curve, to confirm the identity of the PCR product. Each reaction was performed in triplicates. Ribosomal protein subunit S7 (rpS7; housekeeping gene) was used as an internal loading control for normalization of the gene expression (Salazar et al., 1993). The fold expression was analysed using $^{\Delta\Delta}C_t$ method (Livak and Schmttgen, 2001). All the data were expressed as mean ± standard deviation. Statistical significance was analyzed by Student's t-test or one-way ANOVA using GraphPad Prism 5.0 software (Motulsky, 1999).

3.2.6 dsRNA synthesis and gene silencing

For silencing of STAT-A or STAT-B genes, dsRNA was synthesized using MEGAscript RNAi Kit (Applied Biosystems, AM1626). To achieve gene specific silencing, highly mismatched region of ASSTAT-A and AsSTAT-B was selected from their whole gene sequence. Fragment of 681bp of AsSTAT-A and 534bp of AsSTAT-B were cloned in PCRII TOPO TA vector. This vector contains T7 promoter sequence at one end. To generate a template having T7 promoters at both ends, PCR was carried out using above STAT-A or STAT-B plasmid DNA with M13 forward and Xho-T7-M13 reverse primers (sequence is given in **table 3.2**). The 1-2µg of this PCR purified template was used for *in vitro* dsRNA synthesis as described in the MEGAscript RNAi kit manual. Similarly, dsRNA of LacZ gene (500bp from bacterial origin) was also prepared and used as a control. The synthesized LacZ, STAT-A and STAT-B dsRNA were further purified with Millipore column (Microcon YM1000) and concentrated up to 3µg/µl with nuclease free water. The concentration of synthesized dsRNAs was quantified by measuring the absorbance at 260nm in Nano-spectrophotometer and their integrity was assessed by electrophoresis in 1% agarose gel.

20-20 female mosquitoes of one day old were used in each case to inject LacZ, STAT-A or STAT-B dsRNA. 69nl dsRNA of 3µg/µl stock (0.21µg/mosquito) was injected into the hemolymph of each mosquito using the nanoinjector (Drummond) attached to the micromanipulator under dissecting microscope. After injection, mosquitoes were placed slowly into their respective cages for 4 days provided with sucrose and kept at 28°C temperature with 80% humidity. To check the gene silencing after 4 days of injection, 10 mosquitoes were taken from each cage to make RNA and cDNA. Real time PCR was performed to check the effect of silencing.

3.3 Results and Discussion

3.3.1. Identification and cloning of *An. stephensi* STAT-A gene

To identify STAT pathway genes in *An. stephensi*, initially degenerate primers were designed based on conserved sequence of different mosquitoes. From the multiple sequence alignment of STAT protein sequences of various mosquitoes, we selected QPPQVMK and LHIQPFTARD conserved motifs that belong to STAT-DNA binding and Src Homology 2 (SH2) domains respectively (**Figure 3.2A**).

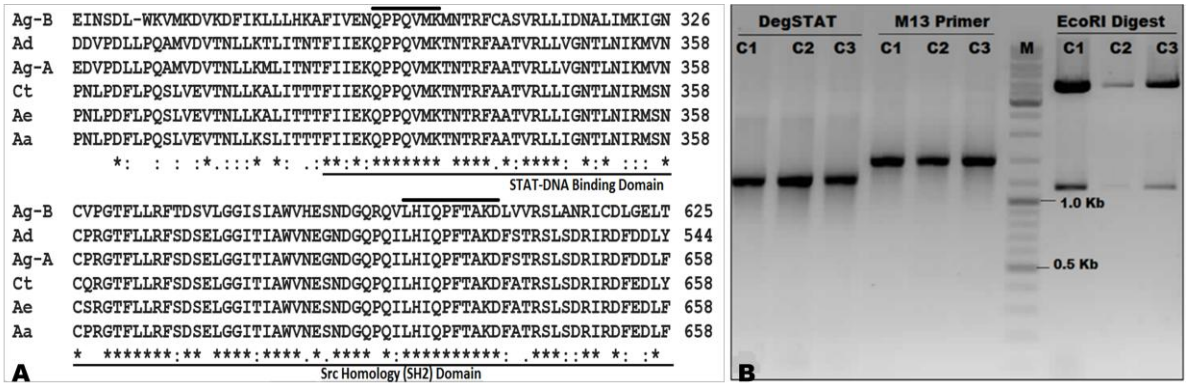


Figure 3.2 A. Degenerate primers sequences of STAT pathway genes designed from conserved region of the protein. The numbers on right side correspond to amino acid residue in particular protein. Multiple sequence alignment of STAT protein of different mosquito species include Ag-B, *Anopheles gambiae* STAT-B (CAA09070.1); Ad, *Anopheles darlingi* (EFR26562.1); Ag-A, *Anopheles gambiae* STAT-A (ACO05014.1); Ct, *Culex tritaeniorhynchus*; Ae, *Aedes aegypti* (ABO72629.1) and Aa, *Aedes albopictus* (AAQ64662.1). Thin under-line in bottom represent the boundaries of STAT-DNA binding and Src Homology 2 (SH2) domains. Thick line on top reveals conserved amino acids sequences where forward and reverse degenerate primers (**DegSTAT**) were designed. **B.** Confirmation of *An. stephensi* STAT-A (AsSTAT-A) clones. All clones (C1, C2, C3) were confirmed by colony PCR with the help of DegSTAT primer (lanes 1, 2 and 3; band size 1234 bp) and universal M13 (lanes 4, 5 and 6; band size ~1450 bp) primers. Last three lanes are showing the EcoRI digestion of plasmids isolated by the same colony.

With the help of newly designed DegSTAT primers (**Table 3.1**), only ~1200bp AsSTAT-A gene fragment was amplified from gDNA (**Figure 3.2B**). The eluted PCR product was cloned in PCR-II TOPO TA vector (Invitrogen) and sequenced. The sequence of AsSTAT-A showed the maximum identity of 98% with *An. gambiae* AgSTAT-A gene (Gene ID: FJ792607.1). AsSTAT-B gene sequence was found extremely diverged from AsSTAT-A (46% at amino acid level and 63.5% at nucleotides level) hence the amplification through degenerate primers could not be achieved simultaneously (**Table 2.8, chapter 2**).

Analysis of full length STAT-A gene

To clone and confirm the predicted cDNA of *An. stephensi* STAT-A gene given in chapter 2, **Figure 2.2A**, different primers were designed and PCR was performed the using gDNA or midgut cDNA as template. *An. stephensi* predicted cDNA sequence of AsSTAT-A has 7 exons and primers are lying only in the exonic region. Different forward and reverse primer sets were used to clone and confirm the size of exons and introns lying in that region (**Figure 3.3A**). Numbering of each primer set (1-7) correlates with the primer names as given in materials and methods section (**Table 3.2)**. Each of the PCR products was checked on agarose gel to verify the existence and size of introns.

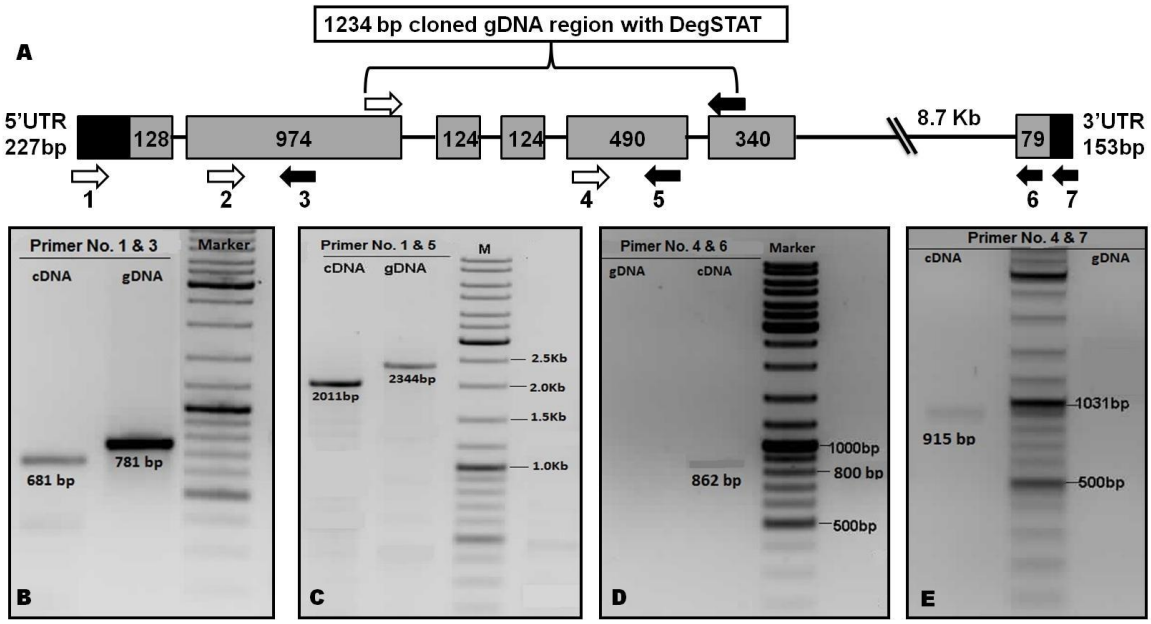


Figure 3.3 A. Schematic presentations of exonic (box) and intronic regions (lines) of AsSTAT-A gene. Grey color boxes are exonic region and their size given inside it; black color boxes are 5' and 3' UTRs. Black color lines are showing the length of introns and arrows represent the forward and reverse primers locations in the gene used for cloning/sequencing. **B, C, D and E** are amplified PCR products with cDNA and gDNA and their sizes have been analyzed through respective gel images.

The full AsSTAT-A gene was not amplified through its distantly situated 5' and 3' terminal primer sets hence full gene was not cloned in a single effort. Overall STAT-A gene was cloned in two partial and overlapping fragments; 2011bp and 862bp from cDNA (**Figure 3.3C and D**). 8.7kb long intron was present at the 3' end therefore; fragments were not amplifiable with gDNA using primer number 4 and 6; 4 and7 (**Figure 3.3 D and E**). For dsRNA synthesis one small region with cDNA (681bp) was also cloned which has maximum dissimilarity in that region compare to STAT-B (**Figure 3.3.B and 3.12**) Full-length AsSTAT-A cDNA consists a 2259bp coding region, which encodes a protein of 752 amino acid residues, plus 227 bp 5' upstream and 153bp 3' downstream UTR. With the confirmation of result, a gene map was prepared and cDNA sequence of AsSTAT-A was submitted to GenBank with accession numbers KR779999.

3.3.2 Identification and full gene organization of *An. stephensi* STAT-B gene

The putative AsSTAT-B gene sequence was extracted by using *An. gambiae* AgSTAT-B CDS 2918bp (Accession no. AJ010299.1) sequences as query through BLASTN based homology searches (**Figure 2.1B**). To elucidate the full AsSTAT-B gene; searches were expanded in both directions in scaffold_00093 with continuous ORF finding and CDD searches to get all four relevant domains of STATs. With the pronouncement of results, this contigs were termed as AsSTAT-B and further validated. Similar to AsSTAT-A, AsSTAT-B was also not feasible to amplify from its terminal 5' and 3' primers set. Since AsSTAT-B is intronless and have some similarity with AsSTAT-A, all AsSTAT-B primers (listed in material and method) were designed in such a manner so that they do not share any stretch of identical nucleotide sequence longer than 7bp as well as highly mismatched at 3'end. AsSTAT-B was found intron-less gene hence every time it gave same amplicons size with gDNA or cDNA in each PCR reaction with all designed primers (**Figure 3.4A**). The full coding region of AsSTAT-B gene was cloned with blood fed carcass cDNA in two overlapping fragments. These two cloned sequences of AsSTAT-B cDNA were 983bp and 1312bp (**Figure 3.4D and E**). For STAT-B silencing, dsRNA was synthesized from one small cloned region with cDNA has dissimilarity with AsSTAT-A towards 5' end (**Figure 3.4B and 3.12**). To report the full gene sequence of AsSTAT-B; primers were designed from the predicted UTRs and amplified PCR products were sent for sequencing without cloning

(**Figure 3.4B and F**). The final 2231bp full cDNA sequence along with UTRs of AsSTAT-B was deposited to GenBank with accession numbers KR780000.

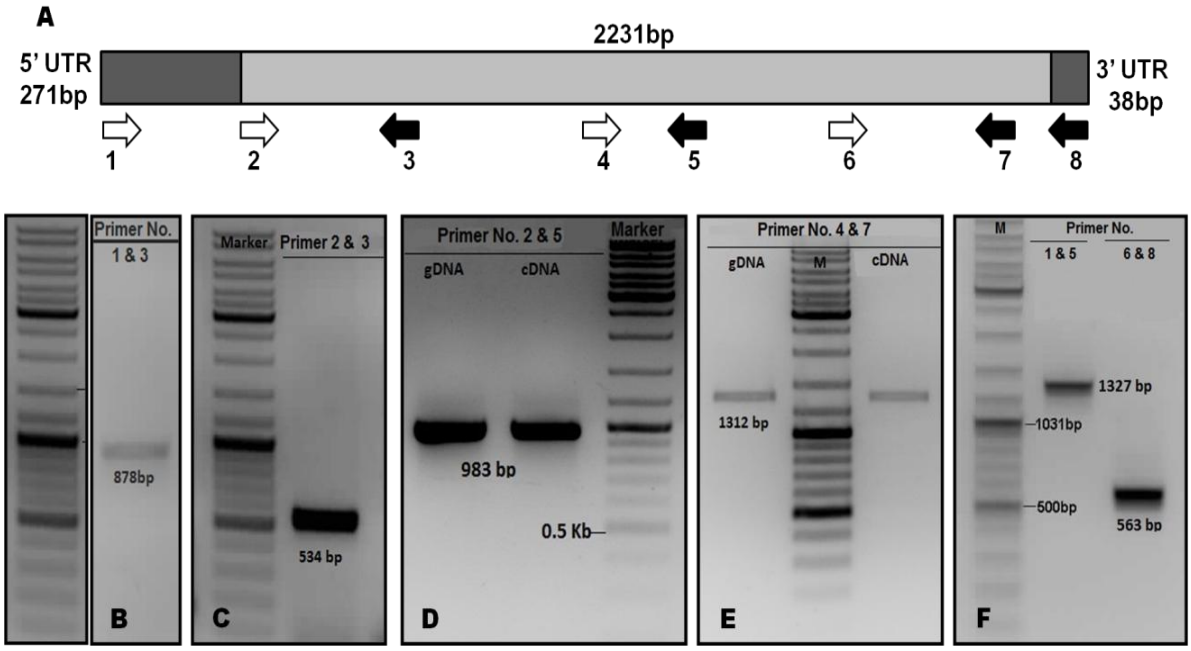


Figure 3.4 A. Schematic presentations of intronless AsSTAT-B gene. Grey color box is the exonic region which is 2231 bp long and black color boxes are 5' and 3' UTRs. White and black arrows represent the forward and reverse primers locations in the gene used for cloning/sequencing. Numbering of each primer set correlates with the primer names given in materials and methods section. **B and F** are PCR products amplified with cDNA and gDNA and sequenced. **C** 534bp region of cDNA was used for cloning and utilized in dsRNA preparation. **D and E** PCR products were used for cloning in two overlapping fragments and sequenced. Their sizes have verified through respective gel images.

3.3.3 Identification and cloning of *An. stephensi* SOCS5 gene

Cloning of SOCS5 gene was done through degenerate primers (degSOCS) designed from most conserved domains containing SH2 and SOCS box from *An. stephensi* gDNA. These primers were designed to amplify sequences corresponding to two terminal stretches of 7-9 amino acids residues, PEGTFLLRD and KEYHY(K/R)Q, which are highly conserved among many insect SOCS genes and amplified 436 bp fragment using gDNA (**Figure 3.5 A and B**). The cloned fragment was sequenced and its nucleotide identity was confirmed through BLAST with available gene sequences in NCBI. The sequence thus obtained was used to design gene specific primers (AsSOCS) that revealed the presence of intron on comparing the amplicons obtained from gDNA (315bp) and cDNA (240 bp) (**Figure 3.6E)**. The cloned SOCS fragment (436 bp) contains partial SH2 domain and full length SOCS box domain.

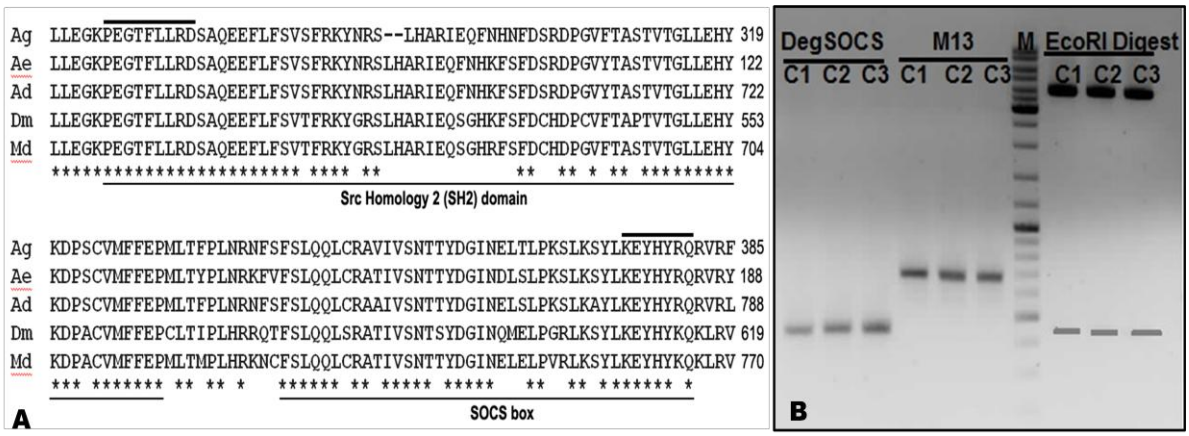


Figure 3.5 A. Protein sequence alignment of SOCS protein of different insect species includes Ag, *Anopheles gambiae* (ABV01933.1); Ae, *Aedes aegypti* (XP_001656067.1); Ad, *Anopheles darlingi* (AND_003361); Dm, *Drosophila melanogaster* (AAO39436.1); Md, *Musca domestica* (XP_005179585.1). Thin underline in bottom represent the boundaries of SH2 and SOCS box domains. Thick line on the top reveals conserved amino acids sequences where forward and reverse degenerate (**DegSOCS**) primers were designed. **B.** Verification of AsSOCS5 clones through degenerate SOCS (DegSOCS) primers (lanes 1, 2 and 3; expected size 436bp) and with M13 primers (expected size 656bp) using three bacterial colonies (C1, C2 and C3) as templates. Last three lanes are showing the EcoRI digestion of plasmids isolated by the same colony. 1 Kb ladder was used as a reference.

Analysis of full length SOCS5 gene

In order to annotate and identify the full SOCS5 gene, the genomic sequence that includes the conserved area of the SOCS gene were extracted. We selected AgSOCS5, 1617bp (GenBank accession No EF631979.1) sequence as a query for BLAST homology search. AgSOCS5 comprises 1191bp of CDS which encodes 396 amino acids and 316bp of 5' and 110bp 3' plausible UTR region. The obtained contig for AsSOCS was ALPR02003551; contig_3551. Finally, the obtained putative cDNA sequence of AsSOCS was also confirmed by tBLASTn program for their respective protein domains (**Table 2.12**). These predicted cDNA sequences were used for primer designing to acquire full gene sequences and further gene cloning. 972 bp and 1653 bp fragment were amplified from cDNA (blood-fed midgut) and gDNA, respectively suggesting the presence of intron(s) in this region (**Figure 3.6C and D**). AsSOCS5 gene of *An. stephensi* has the 3 exons and two introns; 3 exons indicate the size of 383, 513 and 298bp respectively separated by 2 intron of size 606 and 75bp (**Figure 3.6A**). Only one 972bp region was cloned and remaining 5' and 3' terminal region were apmlified by primer pairs 1 and 2; 2 and 7, purified PCR product were sequenced to get the full length

gene (**Figure 3.6B and F**). The sequence of AsSOCS5 gene has been submitted to the Genbank under the accession number KU306401.

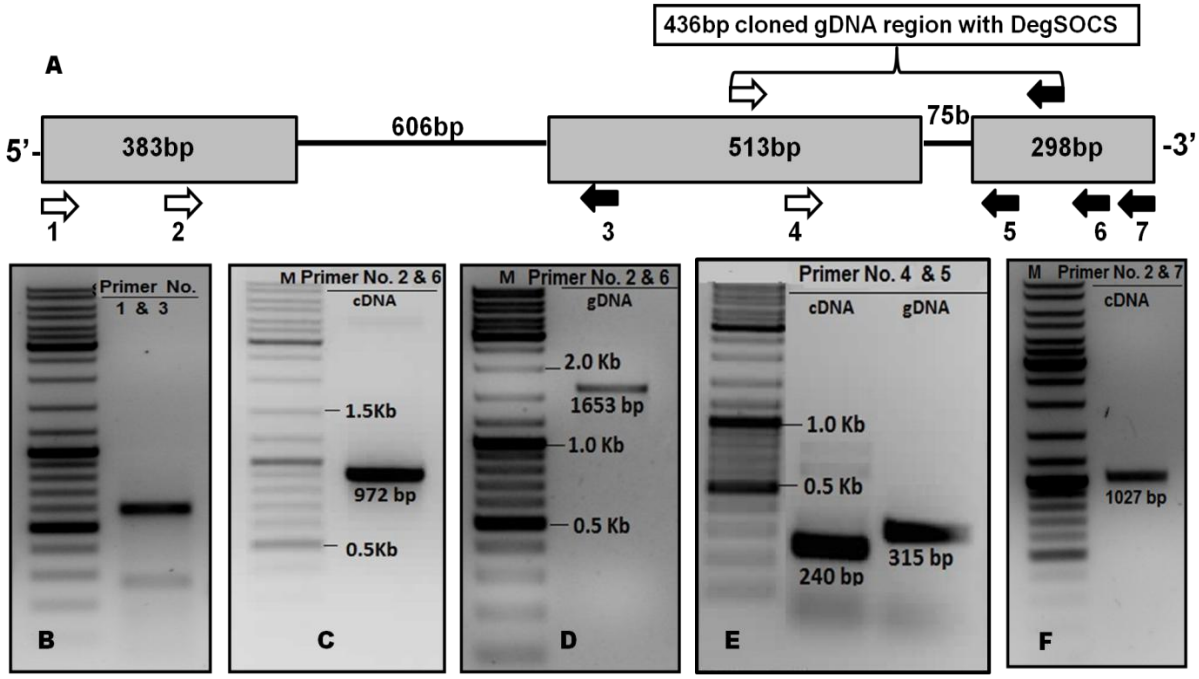


Figure 3.6 A. Schematic presentations of exonic (box) and intronic regions (lines) of AsSOCS5 gene. Grey color boxes are exonic region and their size given on above the boxes and black color lines are showing the length of introns. White and black colour arrows represent the forward and reverse primers locations in the gene used for cloning/sequencing. Numbering of each primer set correlates with the primer names given in materials and methods section. **B, C, D, E and F** are the PCR products amplified with cDNA or gDNA and their sizes have been verified through respective gel images.

3.3.4 Identification and cloning of *An. stephensi* PIAS gene

Based on sequence alignment of the highly conserved domain of PIAS proteins, degenerate oligonucleotides were designed and used to amplify *An. stephensi* genes using cDNA as template (blood fed midgut). For PIAS, PINIT domain conserved sequence VEPKRPPRPVN and Zn finger motif NLVIDGYFQ conserved region were targeted (**Figure 3.7A**). We successfully amplify the 528bp AsPIAS fragment from cDNA template which was further purify and cloned into TOPO TA Vector (**Figure 3.7B**). The sequencing result of these clones was shown the maximum identity with *A. gambiae* and *Ae. aegypti* and *A. aquasalis* of respective genes. The sequencing details were utilized for designing of gene specific primers for real time analysis (**Figure 3.8C**).

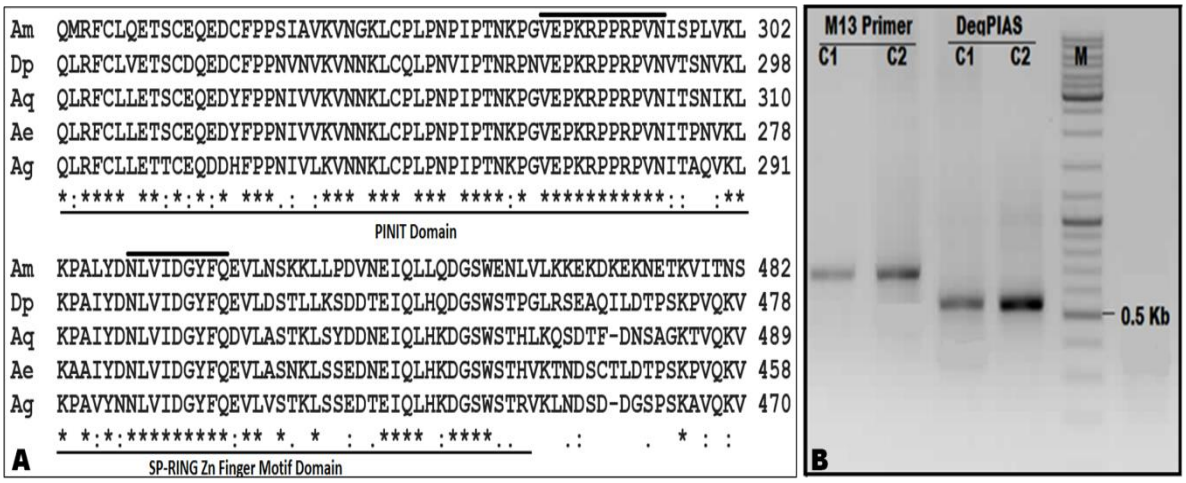


Figure 3.7 A. Degenerate primers sequences of PIAS genes designed from conserved region of the protein. Protein sequences for different insects were obtained from NCBI database. The numbers on left side correspond to amino acid residue in particular protein. Sequence alignment of PIAS protein of different insect species include Am, *Apis mellifera* (XP_623571.3); Dp, *Drosophila pseudoobscura* (XP_002138569.1); Aq**,** *Anopheles aquasalis* (AEK26394.1); Ae, *Aedes aegypti* (XP_001647815.1); Ag, *Anopheles gambiae* (XP_001688469.1). Thin underline in bottom represents the boundaries of PINIT and SP-RING Zn finger motif domains. Thick line on top reveals conserved amino acids sequences where forward and reverse degenerate (**DegPIAS**) primers were designed. **B.** Confirmation of *An. stephensi* PIAS (AsPIAS) clones through M13 and DegPIAS primers. 1 Kb ladder (SM0331) was used as a reference for each gel image.

Analysis of full length PIAS gene

As done earlier, the cDNA sequences of AgPIAS (XM_001688417.2) 2050bp partial CDS of *An. gambiae* were used to retrieve the complete sequence of AsPIAS gene (**Table 2.16 and Figure 2.8**). The obtained contig for AsPIAS was KE388957.1; scaffold_00068. Finally, the obtained putative cDNA sequence for PIAS was also confirmed by tBLASTn program for their respective protein domains. These predicted cDNA sequences were used for primer designing to get full gene sequences and further gene cloning. AsPIAS gene of *An. stephensi* has the maximum number of exon and intron among other genes of this pathway (**Figure 3.8A**). Primer no. 2 is lying on two consecutive exons therefore, not amplified the gDNA and give 908bp fragment with cDNA using reverse primer no. 4 (**Figure 3.8D**). Primer pair 1 and 3 were used to get 5' end of PIAS and PCR product 1209bp obtained from cDNA (**Figure 3.8B**). Amplified PCR products were outsourced for sequencing.

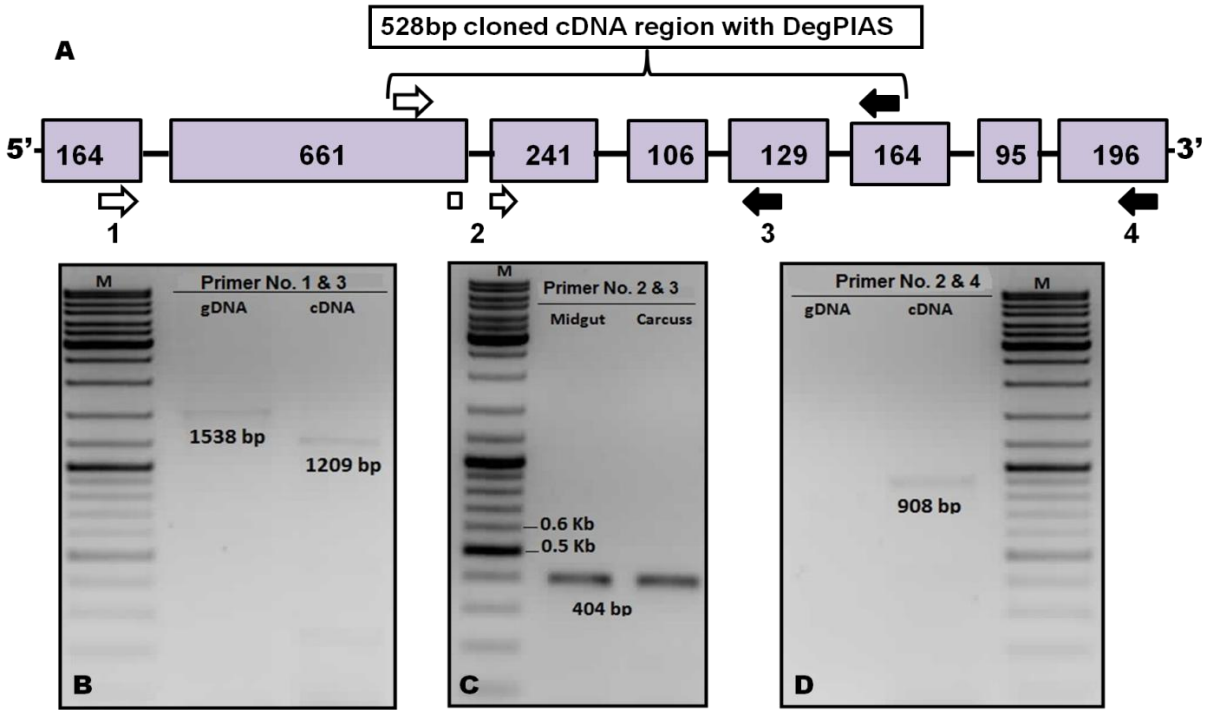


Figure 3.8 A. Schematic presentations of exonic (box) and intronic regions (lines) of AsPIAS gene. Grey color boxes are exonic region and their size given on above the boxes and black color lines are showing the length of introns. White and black arrows represent the forward and reverse primers locations in the gene used for cloning/sequencing. Numbering of each primer set correlates with the primer names given in materials and methods section. **B, C and D** are the PCR products amplified with cDNA and gDNA and their sizes have been rectified through respective gel images.

3.3.5 Expression profiles of STAT pathway genes in different developmental stages

It was well established fact that, STAT pathway participates in the developmental process of many insects including mosquitoes (Baumer et al., 2011; Bahia et al., 2011). Therefore, the involvement of STAT pathway in different developmental stages of *An. stephensi* was investigated. Different developmental stages of mosquito life cycle *viz.* egg, larva, pupae and adult male and female were collected. Total RNA was isolated and their cDNA were prepared and these samples used for gene expression analysis. For expression analysis; primer pair 4 and 5 for STAT-A; primer pair 6 and 7 for STAT-B; primer pair 4 and 5 for SOCS5 and primer pair 2 and 3 were used for PIAS gene. Ribosomal protein subunit S7 (rpS7; housekeeping gene) was used as an internal loading control for normalization of the gene expression (Salazar et al., 1993).

Semi quantitative real time PCR (qPCR) analysis indicated that STAT pathway gene expression was detected in all life stages of *An. stephensi* (**Figure 3.9**). The comparison of

relative mRNA expression of two STATs (STAT-A and STAT-B) in different developmental stages of *An. stephensi* indicated that the expression levels of STAT-A are higher than STAT-B in all life stages. The expression of STAT-A is significantly high in egg, larva, pupa, adult male and female stages (P<0.001) compare to STAT-B using one way ANOVA (**Figure 3.9A**). It is in agreement that the least expression of STAT-B in *An. gambiae* is sufficient to regulate the STAT pathway (Gupta et al., 2009). SOCS5 gene showed the constitutive gene expression in all stage from egg to adult female. AsNOS is continuously expressing higher (P<0.01) in all developmental stages except larvae and female mosquitoes. As far as the PIAS and NOS genes are concerned, pupae and adult male mosquito have the highest expression level with respect to eggs and larvae (**Figure 3.9**).

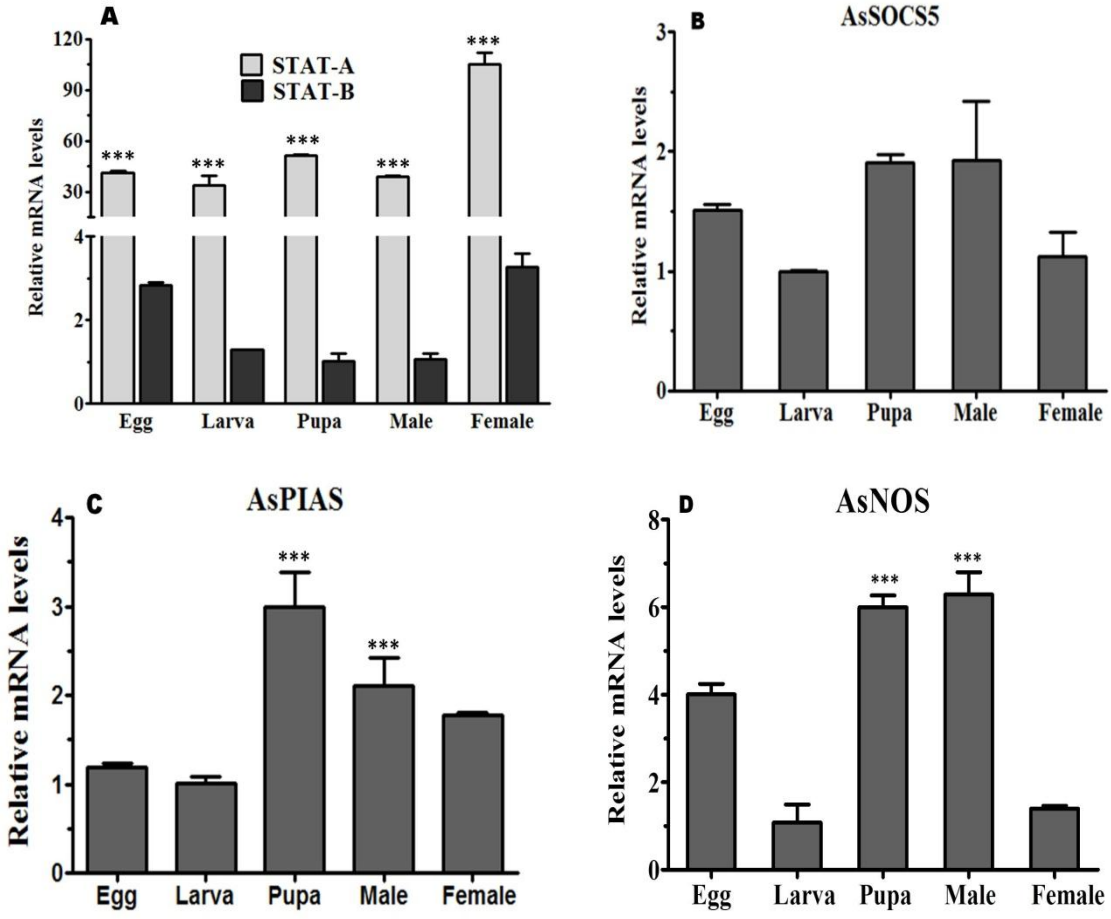


Figure 3.9 Expression profiles of STAT pathway genes in different developmental stages of *An. stephensi*. Relative expression level of AsSTAT-A, AsSTAT-B, AsSOCS, AsPIAS and AsNOS mRNA at different developmental stages (Eggs, Larvae-IV, Pupae, Males and Females) were determined based on qPCR amplification of this gene as described in Material and Method. Relative gene expression values were normalized using *An. stephensi* ribosomal protein S7 mRNA levels as internal loading controls. Results are mean±SD of duplicates for each developmental stage.

We also observed that during pupal stages of development, expression levels of STAT, PIAS and NOS are high compare to larvae. These findings correspond with other reports where STAT pathway is upregulated during insect metamorphosis and also induce the expression of its own regulators. For example, in *Drosophila* SOCS36E mRNA is developmentally expressed, with relatively high expression in 2 and 12 hour embryonic stages. In addition, SOCS36E mRNA levels rapidly decline at the end of embryogenesis and return to basal level during larval stages of development (Mukherjee et al., 2006; Callus et al. 2002). In Brazilian mosquito *An. aquasalis,* expression levels of both STAT and PIAS genes are higher in pupa compare to larval stages (Bahia et al., 2011).

Mosquito pupa is in fact a transient stage of development where almost whole body reorganization takes place under the influence of hormones (Truman and Riddiford, 1999). This is a stage where cell death and regeneration occurs extensively. These findings also supported by others studies where the expression of STAT pathway and negative regulators are induced in mosquito pupae (Bahia et al., 2011; Callus et al., 2002). The induced expression of genes during *An. stephensi* pupal development indicated that STAT pathway is also responsible for reorganization of the internal body parts during metamorphosis. The expression of SOCS genes is comparatively higher in males than females that correlate this finding with other reports where SOCS is found to be a male biased gene in mosquito *An. gambiae* and *An. culicifacies* (Magnusson et al., 2011; Dhawan et al., 2015). These findings may support the concept that sex-regulated genes control gender dimorphism and determine the ability of male or female mosquitoes to perform specific behaviors such as, female precision to transmit malaria parasite. The study of such genes may certainly help in understanding the sexual development and also provides valuable targets for controlling fertility, altering the sex ratio and reducing mosquito population through genetic control measures.

3.3.6 Expression profile of STAT pathway genes in *Plasmodium berghei* infected mosquitoes

The *Plasmodium* cycle commences when female *Anopheles* mosquito ingests gametocytes-infected blood. Ingested gametocytes differentiated into macro- and micro-gametes in the mosquito midgut immediately after blood feeding. These motile micro (male) gametes

emerge from erythrocytes through a process called exflagellation (Billker et al., 1997). This is followed by the fertilization of macrogamete which leads to the formation of non-motile zygote. During 15-20 h post fertilization, the zygote differentiates into motile ookinete and penetrates the nonmembranous peritrophic matrix. This matrix surrounds the blood bolus and acts like a barrier between the gut epithelium and the bolus antigens (Kumar et al., 2010; Kajla et al., 2015). *Plasmodium* invades mosquito gut cells and encounters the immunity, which limits their growth and development (Baton and Ranford-Cartwright 2005). It may be a general phenomenon of interaction of *Plasmodium* with vector *Anopheles* species and defined as a strategy for parasite survival in host (Vinetz et al., 2005; Vlachou et al., 2006). Among the variety of immune responses, STAT pathway plays a fundamental role by inducing the expression of NOS, which catalyzes the formation of nitric oxide (NO), a highly reactive immune molecule (Han et al., 2000). The NO gets converted to NO_2 which mediates protein nitration in a peroxidase-catalyzed reaction (Kumar et al., 2004; Oliveira et al., 2012). The alleviated production of NO not only bears out toxicity to the parasite, it can also be potentially deleterious to the host itself (Luckart et al., 1998; Han et al., 2000). Therefore, to ease the overall rigorousness of these toxic molecules, STAT pathway should be negatively regulated by PIAS and SOCS through a negative feedback loop.

These antiplasmodial roles of STAT pathway have been extensively studied in other insects and mosquitoes however, it is unknown in *An. stephensi*. To understand the involvement of STAT pathway genes in anti-*Plasmodium* response, adult female mosquitoes were fed on a non-infected (control) or *Plasmodium berghei (Pb)*-infected mouse. After 24h of blood feeding (it is the time when *Plasmodium* ookinetes invade midgut epithelium) midgut and carcass were collected separately and the expression of STAT pathway genes were analyzed. The one-way ANOVA test was used as statistics method to calculate the significant difference. We investigated the effect of *P. berghei* (Pb) infection on expression of these five genes (STAT-A, STAT-B, PIAS, SOCS and NOS).

Quantitative PCR analysis revealed that AsSTAT-A and AsSTAT-B transcription factors mRNA levels are not affected after Pb infection. As mentioned earlier that their expression is quite low and their requirement is to switch-on the pathways. They are expressing constantly in both control and infected samples but difference in their mRNA level is nonsignificant (*P* value > 0.5). In case of AsSOCS5 near about 6 times higher expression was observed in the

infected midguts ($P < 0.001$), whereas in carcass samples it was three times ($P < 0.01$) highly expressed compare to controls (**Figure 3.10**).

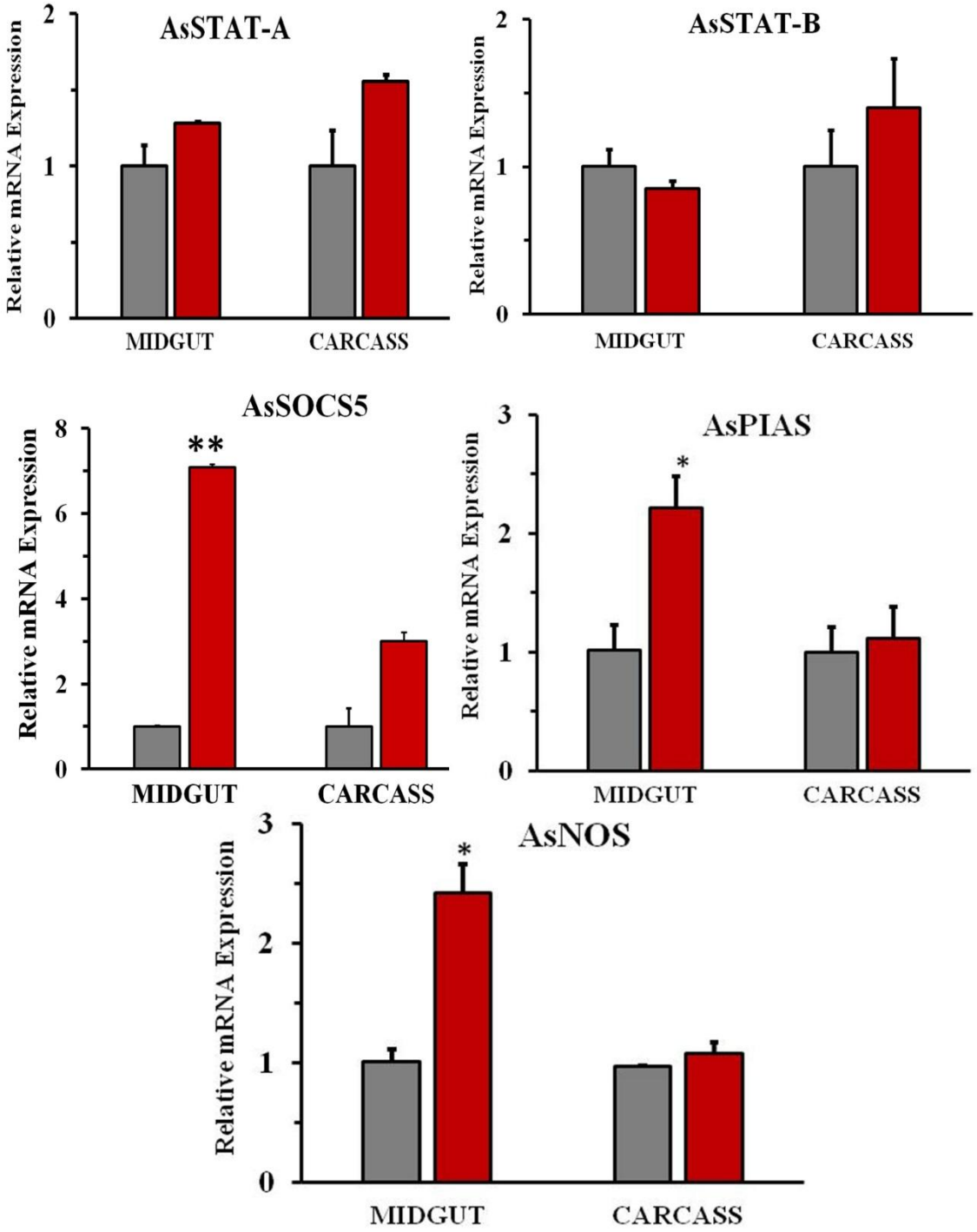


Figure 3.10 Expression profiles of STAT pathway genes in control and infected tissue samples of *An. stephensi.* Relative expression level of AsSTAT-A, AsSTAT-B, AsSOCS, AsPIAS and AsNOS mRNA were determined based on qPCR amplification of this gene as described in Material and Method. Grey and red color bars represented the expression levels of genes in control and infected mosquito samples respectively. Results are mean±SD of duplicates for each sample.

When ookinetes invade the mosquito midgut epithelium numerous immune responses are induced, which are not exclusively limited to this compartment of the mosquito body. However, another compartment of body, the body wall also exhibits the induction of immunity in parallel (Dixit et al., 2009; Hillyer, 2010). This reveals the cross talk between these two compartments. Therefore, 24 hours after infection when ookinetes are crossing the midgut epithelium, AsSOCS5 is induced approximately three folds ($P< 0.05$) in the carcass of infected mosquitoes in comparison to uninfected controls. This clearly implies a very decisive task being played by AsSOCS5 gene to curb the production of lethal effector molecules that formerly were produced in response to ookinete invasion of midgut cells.

The expression of AsPIAS was found significantly high in midgut ($P < 0.01$) compare to blood-fed control. Likely, to AsPIAS, AsNOS was not found significantly expressed in midgut compare to control. There is no effect of AsPIAS and AsNOS ($P > 0.05$) genes on wall tissues of infected mosquitoes (**Figure 3.10**). This indicates that STAT pathway might be induced during ookinete invasion and in parallel the suppressor of this pathway is also active to counter balance the immune reactions. These findings are in agreement with other reports where STAT pathway genes and its suppressors are induced simultaneously at 24h after *Plasmodium* infection (Bahia et al., 2011; Gupta et al., 2009).

Studies carried in *An. aquasalis* mosquitoes indicate the induction of NOS is also paralleled with PIAS in midgut during *P. vivax*'s ookinete invasion (Bahia et al., 2011). Studies carried in *An. gambiae* also revealed similar findings that both NOS and SOCS5 are induced during *P. falciparum* and *P. berghei* invasion of mosquito midgut (Gupta et al., 2009). Although those two studies used different mosquito species and *Plasmodium* combinations however, these observations clearly indicate the involvement of STAT pathway in antiplasmodial immunity and PIAS and SOCS5 being suppressants of this pathway is also induced during this process. Induction of mRNA expression of PIAS indicates that STAT pathway is on during ookinete invasion and in parallel the suppressor of this pathway is also active to counter balance the immune reactions. Moreover, in another *Anopheles* species the induction of NOS and SOCS5 is also reported during midgut invasion by diverse *Plasmodium* species (Noh et al., 2006). Interestingly, the silencing of SOCS causes hyper activation of NOS which mediates *Plasmodium* killing in mosquito midgut (Gupta et al., 2009; Sharma et al., 2010; Vijay et al., 2011). Dengue viruses (DENV-2) in *Aedes* mosquitoes exploit similar

mechanism to evade host immunity (Souza-Neto et al., 2009). The midgut compartment of mosquito represents one of the most exigent environments for the survival and development of *Plasmodium*.

3.3.7 Functional characterization of STAT-A and STAT-B through RNAi experiments

RNA interference (RNAi) is a gene regulatory mechanism that controls the coding transcript level (mRNA) by either suppressing transcription (transcriptional gene silencing or TGS) or by activating a homology based mRNA degradation process (post-transcriptional gene silencing or PTGS) (Zamore et al., 2000; Tomari and Zamore, 2005).

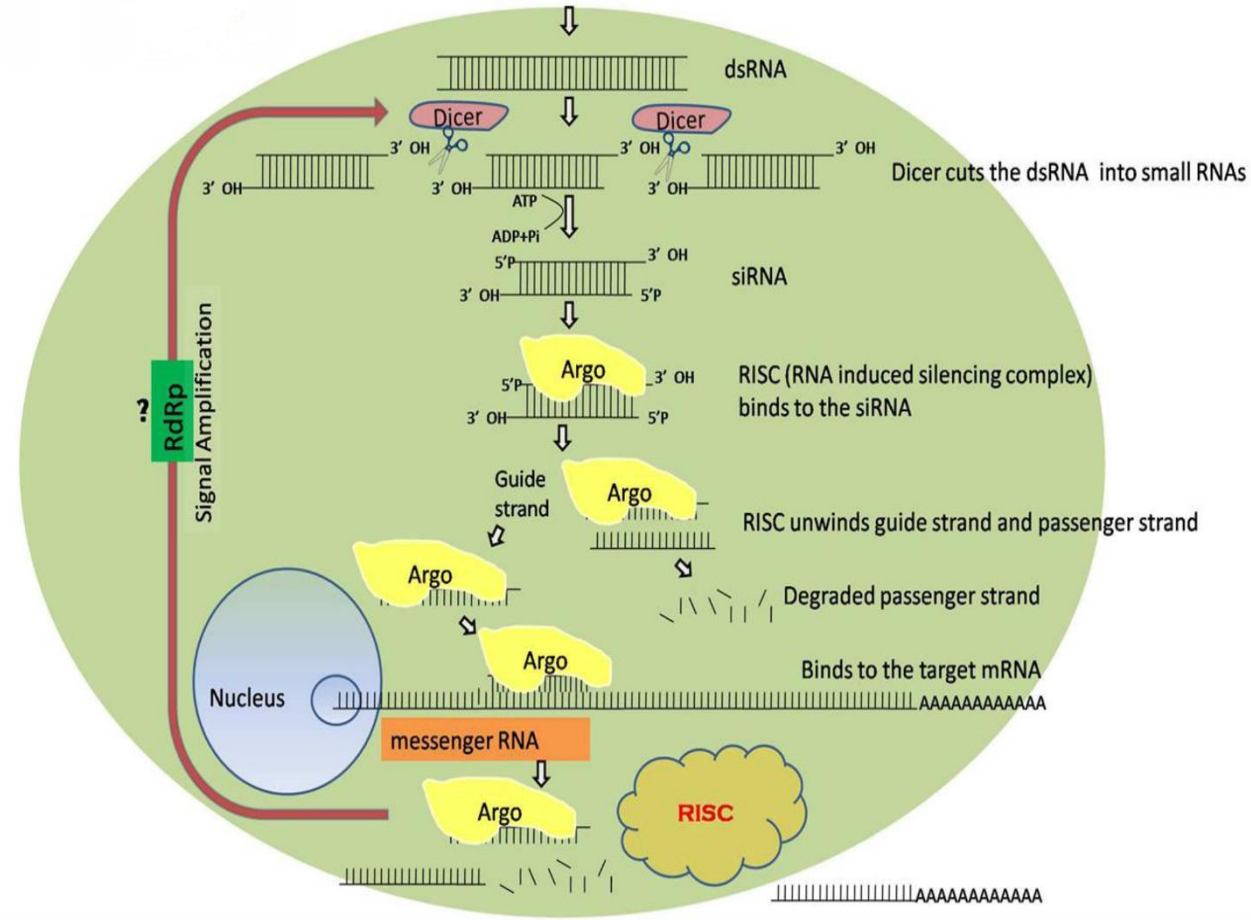


Figure 3.11 Systematic illustration of gene silencing by RNAi inside the mosquito cells. dsRNA of target genes trigger RNAi machinery and processed by RNAse III enzyme 'Dicer' to synthesize 21-24 bp small interfering RNAs (siRNAs). RNAi inducing silencing complex (RISC) binds these siRNAs. The guide strand of siRNAs helps RISC to target the corresponding mRNA. Argonaute protein (Argo) present in RISC complex cleaves the target mRNA. The cleaved target mRNA is amplified by RNA dependent RNA polymerase (RdRp) to form dsRNAs, which enter in RNAi pathway and amplify the signal. However, the signal amplification step in insects is not yet very well understood. (Figure adapted from Kola et al., 2015)

The process involves the production of double stranded RNAs (dsRNA) of target gene which is processed into 21–24 nucleotides RNA duplexes by the RNase III enzyme dicer and its homologs. These siRNAs are then incorporated into a multi-subunit endonuclease silencing complex called RNA-induced silencing complex (RISC). Argonaute proteins, the core catalytic components of RISC, use small interfering RNAs (siRNA) as a guide to recognize and degrade the complementary gene or mRNA (Hammond et al., 2000) (**Figure 3.11**).

```
AsSTAT-A    TCGCTTCATTAGAGCGGGACCTTTGGCGTGTTAATTAAGCAATCAGACCGAAAGCCAGCC  60
AsSTAT-B    ------------------------------------------------------------  0
AsSTAT-A    GGACGGACCACCAGCAGCAGCAGTAGTAGAGAGCCTGTGGAAGACCAGCAGAACACTTCC  120
AsSTAT-B    ------------------------------------------------------------  0
AsSTAT-A    TCCCCGTACCAGACAAGTAGTTAGGGCGTGCGCGGCTGTGTGCAGTCCGTCGACGAGCTC  180
AsSTAT-B    ------------------------------------------------------------  0
AsSTAT-A    TTTTGAGTTTTTGGTGCAGTAAGGCAAATAAACACTAAACGGCAAGATGTCACTGTGGGC  240
AsSTAT-B    ----------------------------------------GCTTTGCAGTCATCGAC  17
                                                     **      * * *
AsSTAT-A    GCGTGTGAATCAACTACCCCAGCCGATACTGGAGCAGATACGCTTCATCTATGGTAGCAA  300
AsSTAT-B    CTGGCTGGGTCAGCTTCCCCAACCCACGCTGGAACCATTTCGTTTGCTGTATGGTAACAA  77
              *  **  *** ** ***** ** *  ***** *   * ** **  * ******* ***
AsSTAT-A    CTTCCCGATCGAGGTGCGCCACTATCTAGCAGATTGGATCGAAGAACGATTGCTCAATGC  360
AsSTAT-B    TATACCGGAGGACGTGCGCAATCATATAGCAAATTGGATCCAGGCTCAGTTTATCAATGC  137
              * ***   ** ****** *  ** ***** ******** * *  *  **  *******
AsSTAT-A    ACCAGTCTACACGAACGATCAGGAAGCGGTGTACGAGCAG------GATGCGGCCAACTT  414
AsSTAT-B    ACCGGCTTTCTACGACGATCAGAATGCTATTGATGATATGTATGAAGAAGCGGCCAAATT  197
            *** *  * *    ******** * **  *  * **   *      ** ******** **
AsSTAT-A    TCTTAACCAGCTGATCATGGAGCTGGAGCGGACCGCGATCAATCTGCCGGAGACGAACTT  474
AsSTAT-B    TCTCAACAAGCTCATAGACAAGCTGAGGTGCACGGGTGCG---------CTCTATGCAAA  248
            *** *** **** **     *****  * * ** *
AsSTAT-A    CACCATCAAGATCCGTCTGAACGAGTCGGCCCGGAACTTCCGCCAGCTGTTCTCGCACAA  534
AsSTAT-B    CGATATCAAGCACGAATTGGACGGGACGGCCTGCATCCTACTACAGCTGTTTTCGCGCGA  308
            *   ******  *    ** *** * ***** * * * * *  ******** **** * *
AsSTAT-A    TCCGGCCCAGCTGTACCAGCATCTGATGAACTGTCTGCACCGCGAACGGCAGTGTGTCGC  594
AsSTAT-B    CCCGATCGCACTCTACCAGGAGCTGATCAACTGTTTGCAGCAGCACCCGTACTATTGCAT  368
             ***  *   ** ****** * ***** ****** **** *   * * * * * *  *
AsSTAT-A    CTACCCGGATGAGTGCGTGAACGTG-----------------CAGGATCCGGAGGTGAC  636
AsSTAT-B    CAATCTGGACGCGATCGAGAGGACCGAGGTGTGGCGAGAATTCCAGGAGCTGAAGGTGAT  428
            * * * *** * *  ** **                     ***** * * ******
AsSTAT-A    G-------------------------------------------------------GA  639
AsSTAT-B    GGTGCATGGCAATGAAAACGATAAAGGTAGCTTGATAATGGTGCACGGCTACTTGGCGCA  488
            *                                                        *
AsSTAT-A    GGTGTTCAATGCGGTCCAGCAGCTGCAGATAATGGTGCGCAC----        681
AsSTAT-B    TGCGATAAGTGAGGTTCAGAATGTACAGGCAACGATCAATACGGCA        534
             * * * * ** *** *** *  * ***  ** * *    **
```

Figure 3.12 Sequence alignment of AsSTAT-A and AsSTAT-B from highly mismatched region (N-terminal region) used for dsRNA synthesis. Double underlined sequence at the beginning showing forward primer and bold underlined sequence are the reverse primers.

To check the functional role of a particular gene in the pathway or against to infection, RNAi technology of gene silencing were utilized. First, double-stranded RNAs (dsRNAs) were synthesized *in vitro* for a particular gene of interest and later injected in to the mosquito with the help of a Nano-injector. Here, we synthesize the dsRNA of AsSTAT-A (681bp) and AsSTAT-B (534bp) from the most divergent regions (N-terminal) of the gene and used to

silence the gene expression. These probes were designed so that they do not share any stretch of identical nucleotide sequence longer than 8bp (**Figure 3.12**). This strategy was used to specifically silence the expression of particular gene only. dsRNAs for AsSTAT-A, AsSTAT-B or LacZ (500bp from bacterial origin as control) were produced from PCR-amplified fragments previously cloned in PCR-II TOPO vector having the T7 promoter at one end. Template for dsRNA synthesis using the T7 MEGAscript kit has already been described in Material and Method section.

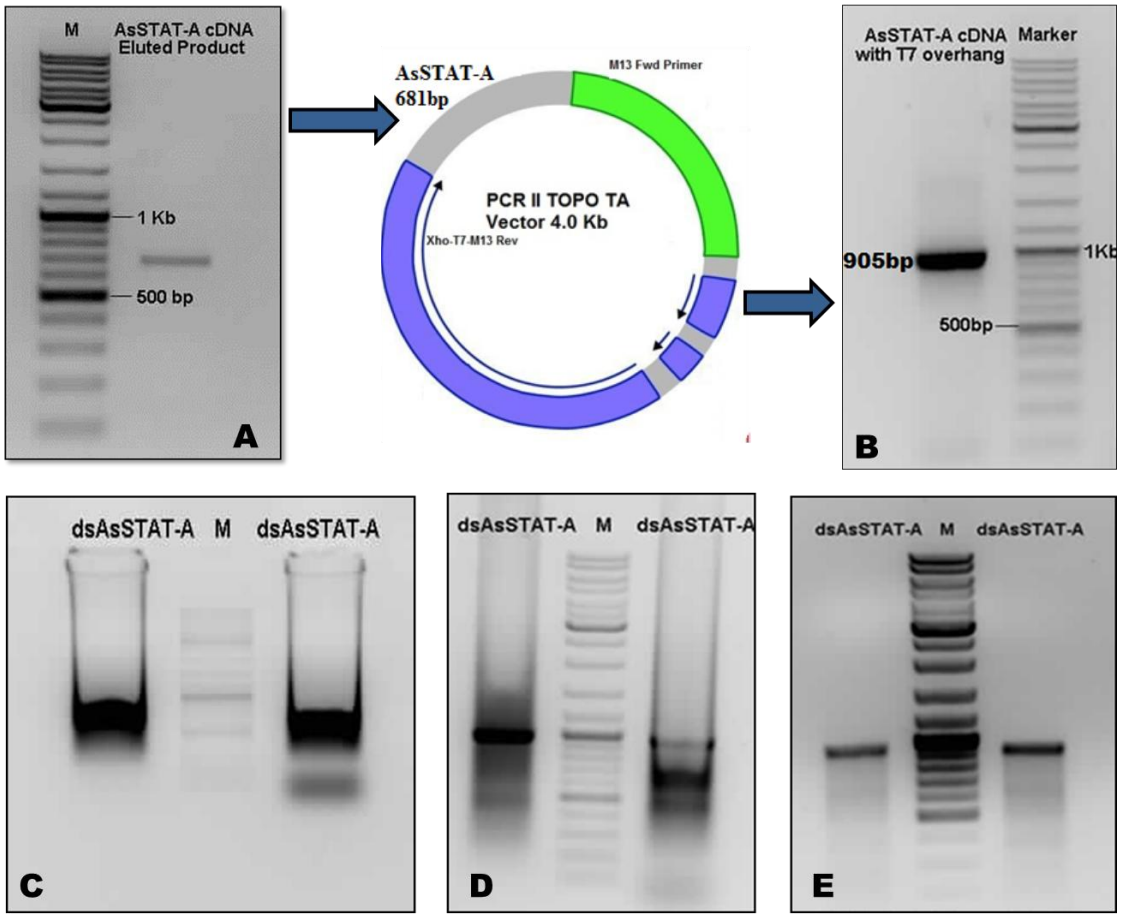


Figure 3.13 Preparation of dsRNA for AsSTAT-A gene. **A.** AsSTAT-A gene were amplified with primer no. 1 and 3 (STAT-A-5'UTR Fwd and STAT-A-GS-Rev) and eluted PCR product cloned in to PCR-II TOPO TA vector. **B.** Addition of T7 overhangs in cloned plasmid template through M13 Fwd. and Xho-T7 primers. **C.** PCR product after in vitro transcription reaction. **D.** After annealing **E.** Purified dsRNA of AsSTAT-A through Millipore column.

Amplicons for dsSTAT-A and dsSTAT-B were produced using plasmid as a templates already cloned reverse transcriptase PCR (RT-PCR) products (681bp and 534bp), from blood-fed female cDNA, respectively. Each of the plasmid was further used for addition of

T7 overhang in both the end of cDNA as shown in the **Figure 3.13 and 3.14**. PCR purified eluted products (905bp and 768bp) were used for overnight transcription reaction. Transcribed products were used for annealing reaction at 75°C and further for ssRNA digestion. Final products were eluted in to nuclease free water and checked for concentration on gel as well as Nano-spectrophotometer. The final concentration of dsSTAT-A and dsSTAT-B 3µg/µl were achieved and stored in -20°C freezer for further use.

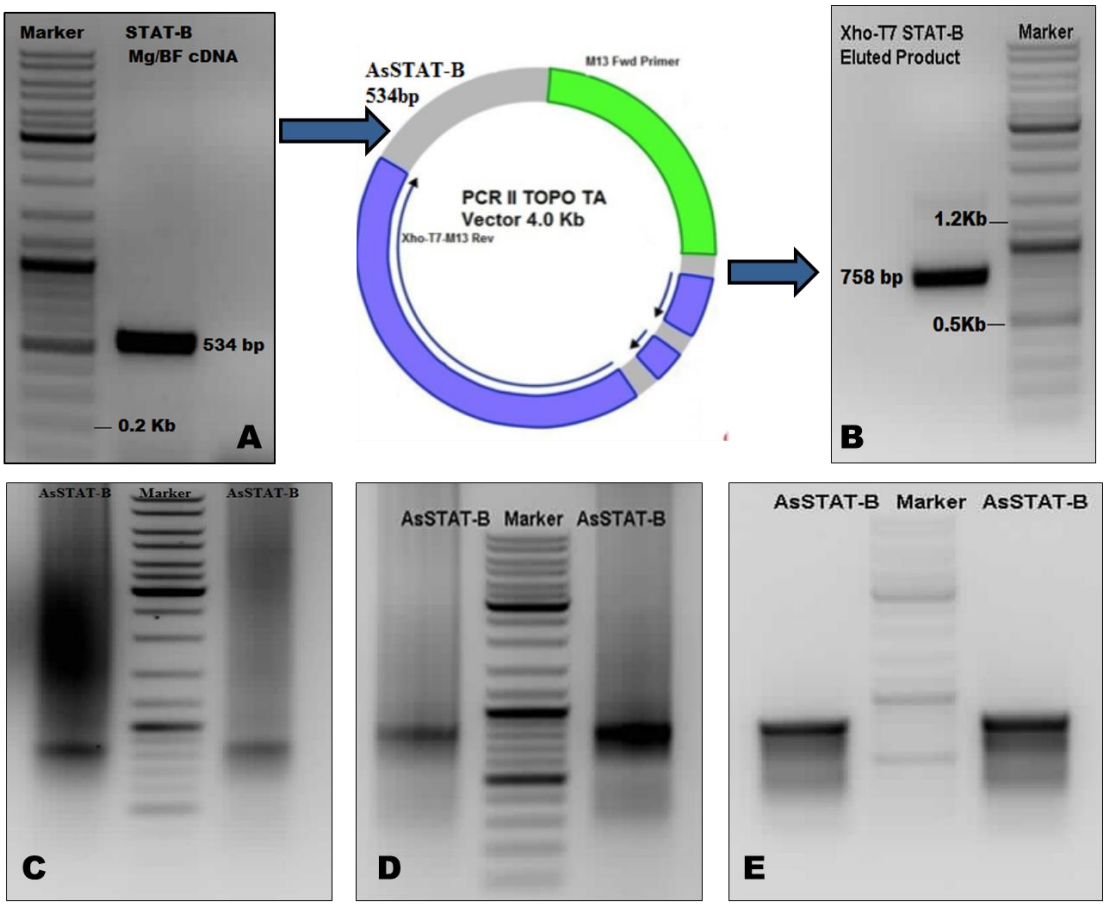


Figure 3.14 Preparation of dsRNA for AsSTAT-B gene. **A.** AsSTAT-B gene was amplified with primer no. 2 and 3 (STAT-B-WG-Fwd and STAT-B-GS-Rev) and eluted PCR product (534bp) cloned in to PCR-II TOPO TA vector. **B.** Addition of T7 overhang in STAT-B cloned plasmid template through M13 Fwd and Xho-T7 Rev primers. **C.** PCR product after in vitro transcription reaction. **D.** After annealing **E.** Purified dsRNA of AsSTAT-B through Millipore column.

Silencing of STAT-A and STAT-B genes

To test how these two STAT genes controls the downstream genes of JAK-STAT pathway in *An. stephensi* or else one another. The effect of silencing of transcription factors AsSTAT-A and AsSTAT-B was evaluated by systemic injection of dsRNA separately. As a control,

female mosquitoes were injected with dsLacZ, a gene not present in mosquito genome. At 4 days after injections, 10 mosquitoes were collected for RNA isolation and cDNA preparation. Silencing of STAT-A and STAT-B was checked, through real time PCR using S7 as a internal control. Expression levels of downstream genes were checked in the STAT-A and STAT-B silenced cDNA samples respectively.

3.3.7.1 Gene silencing of STAT-A and effect on downstream genes

The transcription levels of STAT-A were reduced in mosquitoes injected with dsSTAT-A, relative to those injected with dsLacZ. The expression level of AsSTAT-A mRNA were significantly reduced 84% (P value < 0.05) when mosquitoes were injected with their respective dsRNA (**Figure 3.15**). AsSTAT-A silencing was effective but did not reduce AsSTAT-B expression in *An. stephensi* females in which the gene was silenced by systemic injection of dsRNAs. To rule out the possibility of non-specific cross-silencing, an independent dsRNA was designed from the 5'UTR region of AsSTAT-A, which bears no sequence homology to AsSTAT-B (**Figure 3.12**). Expression levels of downstream genes were checked through real time PCR in the STAT-A silenced cDNA samples.

The effect of AsSTAT-A silencing on the expression of other downstream gene which supposed to regulate through STAT-A has no effect compared with dsLacZ injected female mosquitoes of same age. There was no effect on expression of AsPIAS, AsSOCS and AsNOS genes when experiment was repeated several times. This suggested that either it may not participate in canonical JAK-STAT pathway or it works together with STAT-B and make heterodimers to control the expression of SOCS, PIAS or NOS. Insects have comparatively simpler JAK-STAT pathway having one STAT gene in most of the insects (Lemaitre and Hoffmann, 2007). However, only *An. gambiae* possess two STATs and effect of these two STATs only analyzed in this species. STAT-A was found to be regulated through STAT-B and controls the expression of SOCS and NOS (Gupta et al., 2009). As present study reveals that STAT-A may regulate other effector molecule, which need to be require further investigation.

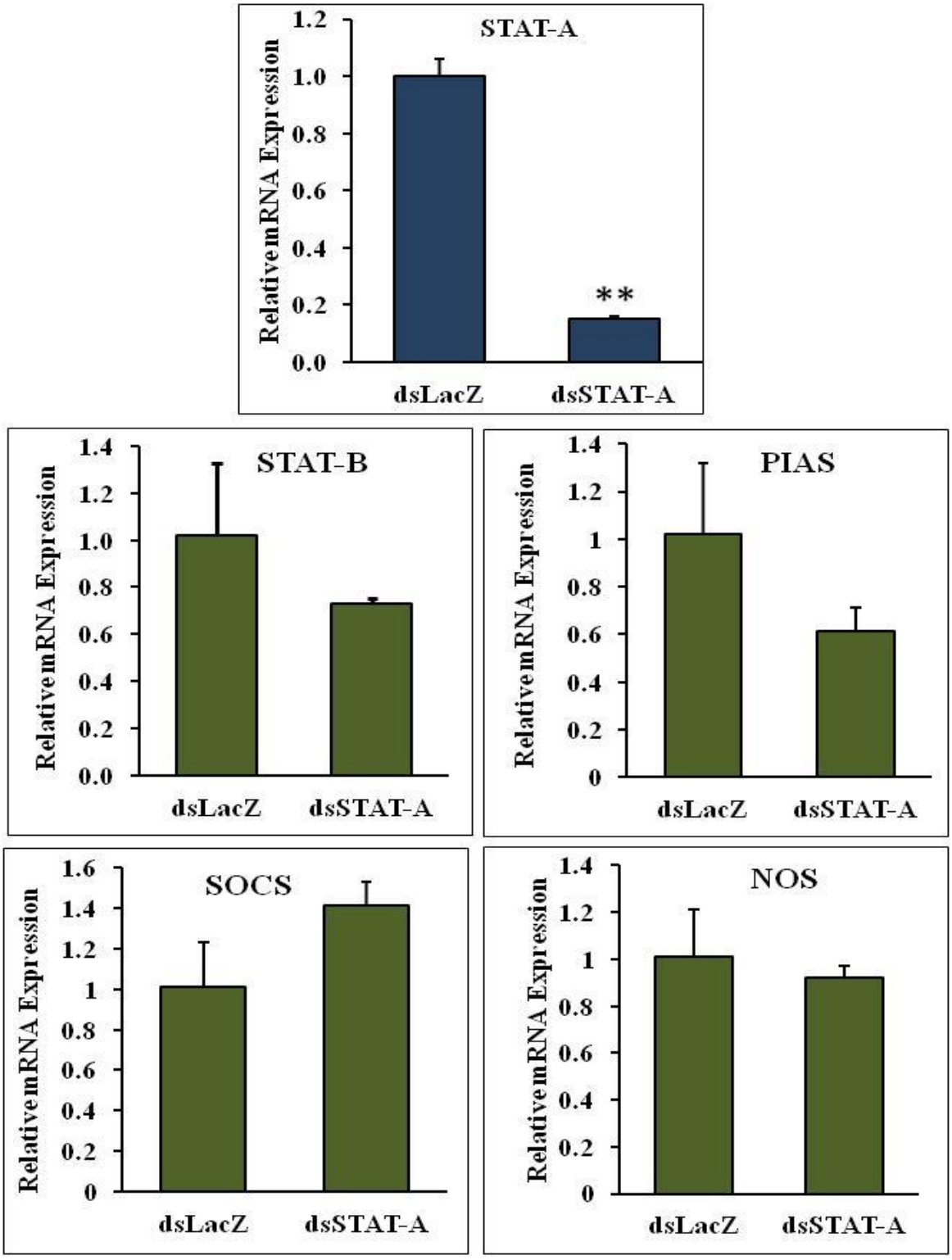


Figure 3.15 Silencing of *An. stephensi* female mosquito with AsSTAT-A dsRNA significantly reduces (*) the AsSTAT-A (P< 0.05) expression, but does not affect the AsSTAT-B, AsPIAS, AsNOS and AsSOCS mRNA levels (nonsignificant). Silencing with AsSTAT-B dsRNA significantly reduces (*) the AsSTAT-A (P< 0.05) expression. The AsSTAT-B reduces the expression of AsSTAT-A, AsPIAS, AsNOS and AsSOCS mRNA levels (P value < 0.05).

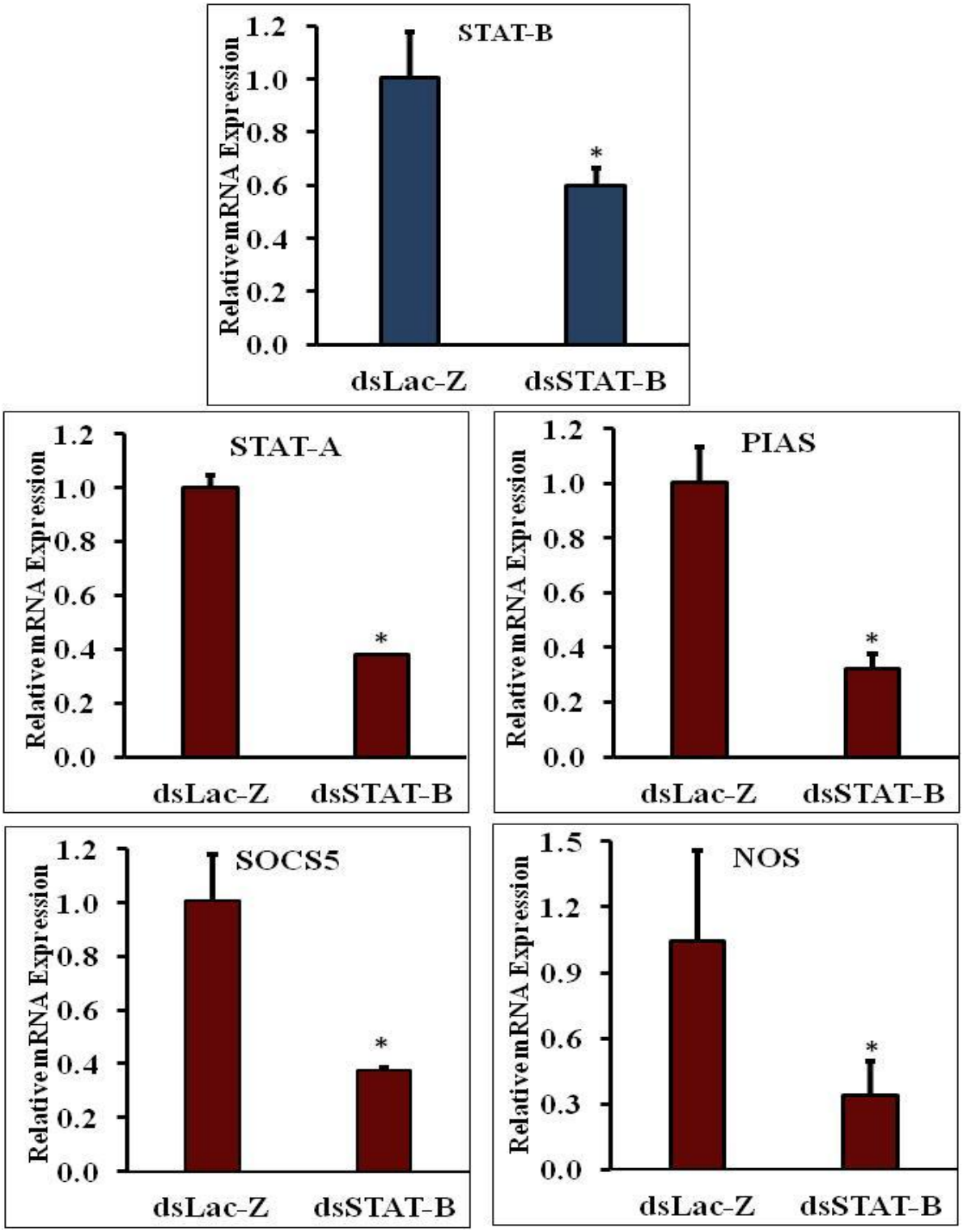


Figure 3.16 Silencing of *An. stephensi* female mosquito with AsSTAT-A dsRNA significantly reduces (*) the AsSTAT-A (P< 0.05) expression, but does not affect the AsSTAT-B, AsPIAS, AsNOS and AsSOCS mRNA levels (nonsignificant). Silencing with AsSTAT-B dsRNA significantly reduces (*) the AsSTAT-A (P< 0.05) expression. The AsSTAT-B reduces the expression of AsSTAT-A, AsPIAS, AsNOS and AsSOCS mRNA levels (P value < 0.05).

3.3.7.2 Gene silencing of STAT-B and effect on downstream genes

To silence the AsSTAT-B gene, dsRNA of STAT-B was injected in to the hemolymph of female mosquitoes. Silencing was checked after 4 days of nanoinjection. Expression of STAT-B was significantly lower (46%) then LacZ control indicated that silencing is working. Unexpectedly, AsSTAT-B silencing also reduced AsSTAT-A mRNA levels by about 70% (**Figure 3.16**). Expression levels of downstream genes were checked through real time PCR in the STAT-B silenced cDNA samples. AsSTAT-B gene also silenced AsSTAT-A expression in adult females, indicating that AsSTAT-B regulates the basal levels of AsSTAT-A mRNA. Further analysis of downstream gene indicated that silencing of STAT-B also lower down the mRNA expression of SOCS (63%), PIAS (68%) and NOS (70%). These findings support a model for the STAT pathway in *An. stephensi* in which both STAT genes are in the same signaling cascade, with AsSTAT-B acting upstream and regulate AsSTAT-A expression as it was observed in *An. gambiae* (Gupta et al., 2009). Recent studies of *An. clouzzi* (formerly known as *An. gambiae* M form) revealed adaptive evolution and neofunctionalization of STAT-B (Rottschaefer et al., 2015). It may be possible that in *An. stephensi*, STAT-B may have new functions and regulated the ancestral STAT-A. Further investigation required to understand this pathway more clearly. However due to time constraint, it is not possible to be reported in present study.

3.4 Conclusions

Among many known genes, STAT pathway genes are important as they have been involved in regulating developmental processes and immunity against a variety of pathogens (Agaisse and Perrimon, 2004). STAT pathway also regulated by the constitutive expression of SOCS and PIAS genes in a downstream process to negatively suppress the pathway. NOS is an important effector molecule of STAT pathway and subsequently exploited to regulate the transmission of vector borne pathogens. (Gupta et al., 2009) At this point, we characterize the components of evolutionary conserved STAT pathway genes from major Indian malaria vector *An. stephensi*. We focused to characterize the STAT pathway genes and analysed its expression during developmental and *Plasmodium* infection. Gene organization and primary sequence of *Anopheles* mosquito were found to be more or less identical to already reported *An. gambiae* pathway components. Comparative pathway studies with other mosquito species

provide further direction to work on immune system and verified the *in silico* prediction. We found two diverse STAT genes in *An. stephensi* which may be due to retro-duplication event because AsSTAT-B was found intronless. This is the principal mechanism for generating this additional diversity in this mosquito species. Expression analysis of each studied pathway components showing significant change after *Plasmodium* infection. The specific upregulation of effector gene in response to *Plasmodium* infection is suggestive of its role in contributing in immunity to the *An. stephensi*.

Expression of STAT pathway gene throughout the developmental stages reveals its important role in development. Expression of STAT-A was higher in all developmental stages compare to STAT-B. In *An. gambiae* STAT-A mRNA expression was not found at pupal stage (Gupta et al., 2009). However, in *An. stephensi* it is present in pupa stage and higher than the STAT-B expression. Recent report revealed that STAT-B gene was expressed at higher level in *An. clouzzi* larvae than in *An. gambiae* larvae (Cassone et al., 2014). Larval habitat preferred by *An. clouzzi* from biotically to abiotically environment condition leads to differential expression of genes. This may also be true that mosquito of different origin may have different expression pattern of both genes. Interestingly, the SOCS5 gene at larval stage was found as male biased gene in *An. culicifacies* and play an important role in controlling gender dimorphism and gender-based specific behaviors (Magnusson et al. 2011; Dhawan et al., 2015). These findings are common with *An. stephensi* and other mosquitoes and may be targeted to alter the sex ratio and reducing mosquito population through genetic control measures.

Our results clearly demonstrate the upregulation of SOCS, PIAS and NOS after *Plasmodium* infection in these mosquitoes. Silencing of STAT-B, down regulate the expression of these genes indicated that they are under the control of STAT pathway. Further, investigations of STAT-A and STAT-B gene may be better picture of overall regulation of this pathway. Moreover, refinement of the cross talk between signaling pathways and various immune elicitors will make the scenario of *Plasmodium* bottleneck more lucrative. Advanced analysis may assist in finding out the way for manipulation of mosquito immune system towards effective strength and in blocking the pathogenic cycle in mosquito.

3.5 References

Agaisse H, Perrimon N. 2004. The roles of JAK/STAT signaling in *Drosophila* immune responses. Immunol Rev. 198:72-82.

Alam MT, Bora H, Das MK, Sharma YD. 2008. The type and *Mysorensis* forms of the *Anopheles stephensi* (Diptera: Culicidae) in India exhibit identical ribosomal DNA ITS2 and domain-3 sequences. Parasitol Res.103:75-80.

Arbouzova NI, Zeidler MP. 2006. JAK/STAT signaling in *Drosophila*: Insights into conserved regulatory and cellular functions. Development 133:2605-2616.

Bahia AC, Kubota MS, Tempone AJ, Araújo HRC, Guedes BAM, et al., 2011. The JAK-STAT pathway controls *Plasmodium vivax* load in early stages of *Anopheles aquasalis* infection. PLoS Negl Trop Dis. 5: e1317.

Barillas Mury C, Han YS, Seeley D, Kafatos FC. 1999. *Anopheles gambiae* Ag STAT, a new insect member of the STAT family, is activated in response to bacterial infection EMBO J. 18(4):959-967.

Baumer D, Trauner J, Hollfelder D, Cerny A, Schoppmeier M. 2011. JAK-STAT signaling is required throughout telotrophic oogenesis and short-germ embryogenesis of the beetle *Tribolium*. Dev Biol. 350(1):169-182.

Baton LA, Ranford-Cartwright LC. 2005. How do malaria ookinetes cross the mosquito midgut wall? Trends Parasitol. 21:22-28.

Billker O, Shaw MK, Margos G, Sinden RE, 1997. The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of *Plasmodium berghei in vitro*. Parasitology 115(1):1-7.

Blandin S, Shiao SH, Moita LF, Janse CJ, Waters AP, Kafatos FC, et al., 2004. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. Cell 116(5):661-670.

Blandin SA, Marois E, Levashina EA. 2008. Antimalarial responses in *Anopheles gambiae*: From a complement-like protein to a complement-like pathway. Cell Host Microbe 3(6):364-374.

Blumberg BJ, Trop S, Das S, Dimopoulos G. 2013. Bacteria- and IMD pathway independent immune defenses against *Plasmodium falciparum* in *Anopheles gambiae*. PLoS ONE 8(9): e72130.

Callus BA, Mathey-Prevot B. 2002. SOCS36E, a novel *Drosophila* SOCS protein, suppresses JAK/STAT and EGF-R signaling in imaginal wing disc. Oncogene 21:4812-21.

Cassone BJ, Kamdem C, Cheng C, Tan JC, Hahn MW, Costantini C, Besansky NJ. 2014. Gene expression divergence between malaria vector sibling species *Anopheles gambiae* and *An. coluzzii* from rural and urban Yaounde Cameroon. Mol. ecol. 23(9):2242-59.

Das BP, Rajagopal R, Akiyama J. 1990. Pictorial key to the species of Indian Anopheline mosquitoes. Zoology 2:131-162.

Dhawan R, Gupta K**,** Kajla M, Kumar S, Gakhar SK, Kakani P, et al., 2015. Molecular characterization of SOCS gene and its expression analysis on *Plasmodium berghei* infection in *Anopheles culicifacies*. Acta Tropica 152:170-175.

Dimopoulos G, Christophides GK, Meister S, Schultz J, White KP, Barillas-Mury C, et al., 2002. Genome expression analysis of *Anopheles gambiae*: Responses to injury, bacterial challenge and malaria infection. Proc Natl Acad Sci. USA. 99(13):8814-8819.

Dixit R, Sharma A, Mourya DT, Kamaraju R, Patole MS. et al., 2009. Salivary gland transcriptome analysis during Plasmodium infection in malaria vector *Anopheles stephensi*. Int J Infect Dis. 13(5):636-646.

Feldmann AM, Brouwer W, Meeussen J. 1989. Rearing of larvae of *Anopheles stephensi*, using water replacement, purification and automated feeding. Entomol Exp Appl. 52:57-68.

Gerberg EJ. 1979. Manual for mosquito rearing and experimental techniques, American Mosquito Control Association (AMCA bulletin no. 5), California.

Glick JI. 1992. Illustrated key to the female *Anopheles* of south western Asia and Egypt (Diptera: Culicidae). Mosq Syst. 24:125-153.

Goltsev Y, Fuse N, Frasch M, Zinzen RP, Lanzaro G, Levine M. 2007. Evolution of the dorsal-ventral patterning network in the mosquito, *Anopheles gambiae*. Development 134(13):2415-2424.

Gupta L, Molina-Cruz A, Kumar S, Rodrigues J, Dixit R, Zamora RE, et al., 2009. The STAT pathway mediates late-phase immunity against *Plasmodium* in the mosquito *Anopheles gambiae*. Cell Host Microbe 5:498-507.

Hammond SM, Bernstein E, Beach D, Havvon GJ. 2000. An RNA directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. Nature 404:293-296.

Han YS, Thompson J, Kafatos FC, Barillas-Mury C. 2000. Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes. EMBO J 19:6030-6040.

Harrison DA, Binari R, Nahreini TS, Gilman M, Perrimon N. 1995. Activation of a *Drosophila* Janus kinase (JAK) causes hematopoietic neoplasia and developmental defects. EMBO J. 14:2857-2865.

Hillyer JF. 2010. Mosquito immunity. In: Söderhäll K, editor. Invertebrate Immunity: Landes BioScience.

Hoffmann JA, Reichhart JM. 2002. *Drosophila* innate immunity: an evolutionary perspective. Nature Immunology 3(2):121-126.

Ihle JN. 1996. STATs: signal transducers and activators of transcription. Cell 84(3):331-334.

Kajla M, Gupta K, Gupta L, Kumar S. 2015. A fine-tuned management between physiology and immunity maintains the gut microbiota in insects. Biochem Physiol. (4):182.

Kola VSR, Renuka P, Madhav MS, Mangrauthia SK. 2015. Key enzymes and proteins of crop insects as candidate for RNAi based gene silencing. Front Physiol. 6:119.

Kumar S, Gupta L, Han YS, Barillas-Mury C. 2004. Inducible peroxidases mediate nitration of *Anopheles* midgut cells undergoing apoptosis in response to *Plasmodium* invasion. J Biol Chem. 279(51):53475-53482.

Kumar S, Molina-Cruz A, Gupta L, Rodrigues J, Barillas-Mury C. 2010. A peroxidase/dual oxidase system modulates midgut epithelial immunity in *Anopheles gambiae*. Science 327:1644-1648.

Lagueux M, Perrodou E, Levashina EA Capovilla M, Hoffmann JA. 2000. Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*. Proc Natl Acad Sci. USA 97(21):11427-11432.

Lemaitre B, Hoffmann J. 2007. The host defense of *Drosophila melanogaster*. Annu Rev Immunol. 25:697-743.

Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC. 2001. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. Cell 104(5):709-718.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) method. Methods 25:402-408.

Loker ES, Adema CM, Zhang SM, Kepler TB. 2004. Invertebrate immune systems–not homogeneous, not simple, not well understood. Immunol Rev. 198:10-24.

Luckart S, Vodovotz Y, Cui L, Rosenberg R. 1998. The mosquito *Anopheles stephensi* limits malaria parasite development with inducible synthesis of nitric oxide. Proc Natl Acad Sci. USA 95:5700-5705.

Luo H, Asha H, Kockel L, Parke T, Mlodzik M, Dearolf CR. 1999. The *Drosophila* Jak kinase hopscotch is required for multiple developmental processes in the eye. Dev Biol. 213:432-441.

Magnusson K, Mendes AM, Windbichler N, Papathanos PA, Nolan T, Dottorini T, et al., 2011. Transcription regulation of sex-biased genes during ontogeny in the malaria vector *Anopheles gambiae*. PLoS ONE 6: e21572.

Meister S, Kanzok SM, Zheng XL, Luna C, Li TR, Hoa NT, et al., 2005. Immune signaling pathways regulating bacterial and malaria parasite infection of the mosquito *Anopheles gambiae*, Proc. Natl. Acad. Sci. USA 102(32):11420-11425.

Motulsky HJ. 1999. Analyzing data with GraphPad Prism. GraphPad Software Inc.,San Diego, CA www.graphpad.com.

Mukherjee T, Schafer U, Zeidler MP. 2006. Identification of *Drosophila* genes modulating janus kinase/signal transducer and activator of transcription signal transduction. Genetics 172:1683-1697.

Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al., 2015. Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. Science 347(6217):1258522.

Noh MY, Han MD, Kim IS, Han YS, JoYH et al., 2006. Induction patterns of suppressor of cytokine signaling by immune elicitors in *Anopheles sinensis*. Int J Indust Entomol. 12(2).

Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. PLoS Genetics 5(10):1-13.

Oliveira GdeA, Lieberman J, Barillas-Mury C. 2012. Epithelial nitration by a peroxidase/NOX5 system mediates mosquito antiplasmodial immunity. Science 335(6070):856-859.

Pham LN, Dionne MS, Shirasu-Hiza M, Schneider DS. 2007. A specific primed immune response in *Drosophila* is dependent on phagocytes. PLoS Pathog. 3: e26.

Perrimon N, Mahowald AP. 1986. l (1) hopscotch, a larval–pupal zygotic lethal with a specific maternal effect on segmentation in *Drosophila*. Dev Biol. 118:28-41.

Ramirez JL. Garver LS, Brayner FA, Alves LC, Rodrigues J, Molina-Cruz A, et al., 2014. The role of haemocytes in *Anopheles gambiae* antiplasmodial immunity. J Innate Immun. 6(2):119-128.

Ramirez JL, Oliveira GA, Calvo E, Dalli J, Colas RA, Serhan CN, et al., 2015. A mosquito lipoxin/lipocalin complex mediates innate immune priming in *Anopheles gambiae*. Nat Commun. 6:7403.

Ramphul UN, Garver LS, Molina-Cruz A, Canepa GE, Barillas-Mury C. 2015. *Plasmodium falciparum* evades mosquito immunity by disrupting JNK-mediated apoptosis of invaded midgut cells. Proc Natl Acad Sci. USA 112(5):1273-1280.

Rodrigues J, Brayner FA, Alves LC, Dixit R, Barillas-Mury C. 2010. Haemocyte differentiation mediates innate immune memory in *Anopheles gambiae* mosquitoes. Science 329:1353-1355.

Rottschaefer SM, Crawford JE, Riehle MM, Guelbeogo WM, Gneme A, Sagnon N. et al., 2015. Population genetics of *Anopheles coluzzii* immune pathways and genes. G3 (Bethesda) 5:329-339.

Salazar CE, Mills-Hamm D, Kumar V, Collins FH. 1993. Sequence of a cDNA from the mosquito *Anopheles gambiae* encoding a homologue of human ribosomal protein S7. Nucleic Acids Res. 21(17):4147.

Sharma A, Vijay S, Rawat M, Raghavendra K. 2010. Nitric oxide synthase immunity in malaria non-vector *An. culicifacies* species B: a putative transmission blocking *Plasmodium vivax* immune responsive mechanism for refractoriness. Int J Infect Dis. 14(1):111-112.

Shuai K, Liu B. 2003. Regulation of JAK-STAT signaling in the immune system. Nat Rev Immunol. 3:900-911.

Souza-Neto JA, Sim S, Dimopoulos G. 2009. An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. Proc Natl Acad Sci. USA 106:17841-17846.

Tomari Y, Zamore PO. 2005. Perspective: machines for RNAi. Genes Dev. 19:517-529.

Truman JW, Riddiford LM. 1999. The origins of insect metamorphosis. Nature 401:447-452.

Vandenbergh JP, Yoeli M. 1966. The effects of temperature on sporogonic development of *Plasmodium berghei*. J Parasitol. 52:559-564.

Vijay S, Rawat M, Adak T, Dixit R, et al., 2011. Parasite killing in malaria non-vector mosquito *Anopheles culicifacies* species B: implication of nitric oxide synthase upregulation. PLoS One 6(4): e18400.

Vinetz JM. 2005. *Plasmodium* ookinete invasion of the mosquito midgut. Curr Top Microbiol Immunol. 295:357-382.

Vlachou D, Schlegelmilch T, Runn E, Mendes A, Kafatos FC. 2006. The developmental migration of *Plasmodium* in mosquitoes. Curr Opin Genet Dev. 16:384-391.

Williams JG. 2000. STAT signaling in cell proliferation and in development. Curr Opin Genet Dev. 10:503-507.

Zamore PD, Tuschl T, Shorp PA, Bartel DP. 2000. RNAi: double stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotides intervals. Cell 101:25-33.

Objective: Identification and molecular characterization of Aedes aegypti mosquitoes from the natural environment of Pilani region of Rajasthan, India

This work has been published in a paper: Gupta K, Dhawan R, Kajla M, Kumar S, Jnanasiddhy B., Singh NK, Dixit R, Bihani A and Gupta L. Molecular Identification of *Aedes aegypti* mosquitoes from Pilani region of Rajasthan, India. **J Vector Borne Dis.** 2016 (*In press*).

4.1 Introduction

The reinfestation of dengue fever (DF) and dengue hemorrhagic fever (DHF) in South Asian countries including India have become a major public health concern (Bhatt et al., 2013; Vikram et al., 2016). *Aedes aegypti* and *Aedes albopictus* are invasive species as they transmit various serotypes of viral pathogens like dengue and chikungunya to human population (Higgs and Vanlandingham, 2015). Almost whole world, especially the tropical countries, lies in the danger zone of these infectious diseases (Thiboutot et al., 2010). Recent reports from WHO indicated that up to 50-100 million infections are now estimated to occur annually in over 100 endemic countries, putting almost half of the world's population at risk (WHO report 2014). India is also at the risk of these diseases where more than 97,000 confirmed cases and approximately 200 deaths have been reported due to dengue and chikungunya in year 2015 (NVBDCP report 2015).

Aedes being the important vector is a prime target for disease surveillance program in India (Gupta et al., 2012). However, information regarding its distribution, density, disease transmission and seasonal prevalence is still very fragmentary (Angel and Joshi, 2008). In addition, parallel occurrence of cryptic species of *Aedes* also make the situation worrisome as reported from some parts of the world (Cook et al., 2005). Out of the 334 species of *Aedes,* only some of them have been characterized at molecular levels. Nevertheless, very few of these characterized mosquitoes were further lab colonized to determine their vectorial capacity against viral pathogens and identifying as susceptible and resistant strains (Caicedo et al., 2013)

Aedes aegypti (Linnaeus) (Diptera: Culicidae), vector competence varies from one geographical region to another and affected by both genetic and environmental factors (Sim et al., 2013). A detailed study on *Ae. aegypti* population of Southeast Asia, Africa, America and Latin American countries illustrated that there is a local genetic variation and gene flow among same species which is responsible for different diseases transmission rate (Urdaneta-Marquez and Failloux, 2011). According to earlier reports, certain variants of *Ae. aegypti* species are known to carry only specific virus of a family but not others and *vice-versa* (Wallis et al., 1985). Therefore, understanding the molecular mechanisms that influence vector competence as well as molecular identification of *Aedes* species may help in developing novel strategies to control vector-borne diseases.

Aedes aegypti is the major disease vector which spreads many viral pathogens of different families (arboviruses and alphaviruses) and serotypes in human population. Genetic selection of *Ae. aegypti* strain to carry the virus load and successively its transmission to next host is defined by their degree of susceptibility.

Species identification is the first step to understand the vector competence. Conventionally, mosquito identification was carried based on their morphological characteristics but since it was time consuming and needs professional expertise, only few species had been identified throughout the world, morphologically (Reinert et al., 2004; Besansky et al., 2003; Pennisi, 2003). In recent years, *Aedes* with overlapping morphological characteristics was best differentiated by DNA-based methods (Cameron et al., 2010, Kumar et al., 2007). The nuclear ribosomal and mitochondrial genes have frequently been used as molecular markers in species identification and evolutionary studies. To study the genomic diversity within and among *Ae. aegypti* species from different world population, the second internal transcribed spacer (ITS-2) of rRNA gene had been used frequently (Mousson et al., 2005). The lack of recombination in mitochondrial genome and high copy number of gene in each cell, mitochondrial cytochrome c oxidase (mtCO) had also been used as a universal entity to differentiate the mosquitoes at species level or lower (Hlaing et al., 2009).

The intergenic spacer (IGS) and nuclear ribosomal internal transcribed spacers (ITS-1 and ITS-2) within the genome have become very popular targets for addressing taxonomic issues among Aedini (**Figure 4.1A**). Beside this, mitochondrial genes are often superior choice for phylogenetic studies as they evolve faster than nuclear DNA (Higa et al., 2010). Cytochrome c oxidase subunit I (COI) is the terminal catalyst in mitochondrial respiratory chain and is involved in electron transport and proton translocation across the membrane (Morlais et al., 2002). COI is the largest gene among 3 mitochondrial-encoded cytochrome oxidase subunits and the nucleotide sequence of this region is often much more polymorphic between two species than within species (Behura et al. 2011; Manonmani et al., 2013) This makes this region of genome (**Figure 4.1B**) useful for delineating molecular differences between cryptic species by length or sequence polymorphism (Musters et al., 1990).

The nuclear ribosomal ITS-1 and ITS-2 and mitochondrial COI genes are not only one of the most important markers in molecular systematic but also used in species barcoding and DNA array technologies (Kumar et al., 2007; Landis and Gargas, 2007). In the present

study we identified the *Aedes aegypti* at molecular level collected from Pilani region of Rajasthan and continuous rear them in insectary. Moreover, we also compare the molecular variation of ITS-2 and COI genes of *Aedes aegypti* strains present in different parts of world.

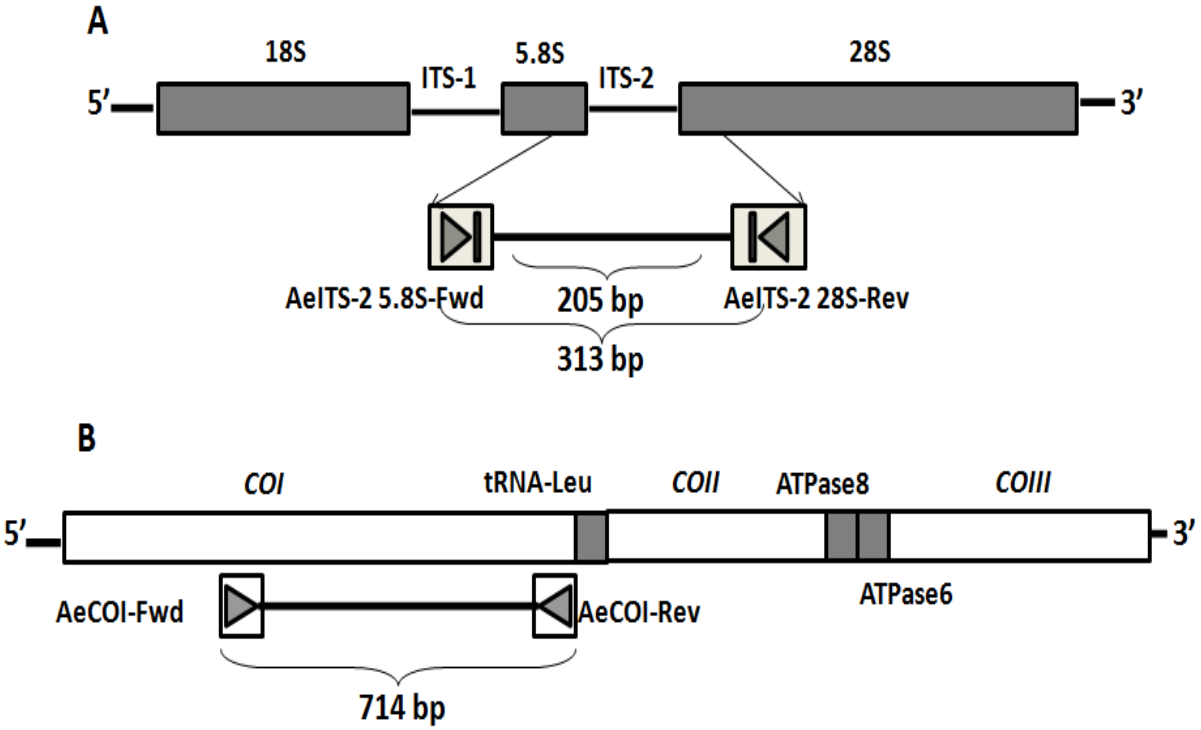


Fig.4.1. Schematic representation of rDNA and mtDNA of *Aedes aegypti* with positions of primers: **A**. Genomic organization of rDNA showing 18S, 5.8S, 28S ribosomal subunit and their intergenic spacers; ITS-1 & ITS-2 **B**. Partial gene organization of mtDNA for COI, COII and COIII genes. (Symbol ► and ◄: indicate primer pair used in the present study).

4.2 Materials and Methods

4.2.1 Collection of *Aedes* mosquito and pure line maintenance

Morphologically identified *Aedes aegypti* larvae were collected from temporary and semi-permanent groundwater pools, fields and water storage containers from Pilani region of Rajasthan (28°22′N 75°36′E). These field collected larvae were reared in the laboratory under specific conditions according to manual for mosquito rearing and experimental techniques published by American Mosquito Control Association California (AMCA) (Gerberg, 1979). Briefly, laboratory conditions: 28±2°C temperature, 80±10% relative humidity and 12:12 hour light: dark cycle was maintained. Larvae were reared in plastic containers with RO water and fed daily with *Spirulina* rich fish food. Pupae were collected

in small plastic cups and kept in big close plastic container having net on the top, to emerge into adult. Adults were fed with a 10% sucrose solution daily; deprived only on blood feeding day for few hours. Blood feeding is necessary to obtain the eggs and for establishment of colony. Use of mice for rearing and maintenance of mosquito culture were followed according to the approved guidelines (IAEC/RES/18/01).

The *Aedes aegypti* adult female and male were identified morphologically with the help of pictorial keys for the identification of mosquitoes (Diptera: Culicidae) associated with dengue virus transmission guide published by Magnolia Press (Rueda et al., 2004). Characterizations and colonization of mosquito colony were done as reported by Hoshino et al., 2010 and Benedict et al., 2009 after slight modifications as follows. For establishing pure line colony of *Ae. aegypti*, Individual female mosquitoes (total 20 females) were separated and kept individually in falcon tubes and provided the blood meal to lay eggs. All females laid eggs; however eggs of only 5 females, were able to complete their entire life cycle. DNA from 1st instar larvae of their F1 generation from all 5 viable pure lines were used for PCR analysis and sequencing. As all of them produce same size of ITS-2 fragment and having same sequence suggesting similar mosquito species, therefore only one line called AePL-2 mosquito colony was used for self-mating to produce F2 generation. Morphologically and molecularly confirmed single mosquito colony (AePL-2) was propagated in insectary.

4.2.2 Morphological identification and molecular markers

Genomic DNA was extracted from pure line (AePL-2) of IVth instar larvae using the method of Gupta and Preet (2012). Precisely, ten 4th instar *Aedes* larvae were grounded in 100µl lysis buffer (100mM Tris-HCl, pH 8.0; 0.5% SDS; 50mM NaCl; 100mM EDTA) and the mixture was treated with 5 µl of proteinase K (20mg/ml) for 1 hour at 55°C. To this cell lysate 5µl of RNAse was added (10mg/ml) and kept for 30 min for incubation at 37°C in water bath. The suspension was extracted twice with equal volume of phenol-chloroform, and DNA was extracted by the addition of 3M sodium acetate (pH 5.6) and isopropanol. Centrifuge at high speed for 5 min and wash the pellet in 70% ethanol, air dry, and resuspend in 50µl TE buffer. Purity of DNA was checked by absorbance ratios A_{260}/A_{280}.

ITS-2 sequences of *Ae. aegypti* from different world isolates, available at NCBI database (**Table 4.1**) were aligned to get a common conserved region. Selected common region of oligonucleotides were used for primer designing. Primers were checked on oligoanalyser and Primer 3 platform for their compatibility. The designed primers of ITS-2 were lying in 5.8S region AeITS-2-Fwd: 5'-ATCACTCGGCTCGTGGATCG-3' and AeITS-2-Rev: 5'-ATGCTTAAATTTAGGGGGTAGT-3' that ends at 28S region **(Figure 4.1A)**. For mtCOI, primers were taken from published reports on Indian *Aedes* mosquitoes (Kumar et al., 2007). The primer sequence was AeCOI-Fwd: 5'-GGATTTGGAAATTGATTAGTTCCTT-3' and AeCOI-Rev: 5'-AAAAATTTTAATTCCAGTTGGAACAGC-3' **(Figure 4.1B)**. Primers were used at a final concentration of 1.0pmol/µl.

4.2.3 PCR amplification and cloning

Nuclear ITS-2 and mitochondrial COI region were amplified through PCR reaction carried out in a volume of 50µl. Each reaction tube contained 50ng of genomic DNA, 1U/µl of Taq DNA polymerase (Genei), 1µl of 10mM dNTP (Genei), 5µl of 10x PCR buffer having 2.5mM $MgCl_2$ and 1µl of each 20pmol of forward and reverse primers. The PCR conditions were subjected to single cycle of Pre-denaturation at 95°C for 3 minutes then 35 cycles of 95°C for 30 sec, 55°C for 30 sec, and 72°C for 45sec. A final 72°C extension was carried out for 10 min and stopped the reaction at 4°C. Amplicons were resolved and visualized on 1.2% agarose gel according to standard procedures and photographed on gel doc. Single band of ~313bp and ~700bp long fragment was purified through Qiaquick PCR purification kit (Qiagen Cat No. 28104) according to the manufacturer instruction. COI purified gene product was sent directly for sequencing with their respective primers while ITS-2 PCR product was used for cloning.

The purified PCR product of ITS-2 was cloned using PCR-II TOPO TA-Vector® (Invitrogen, Cat No K46001-01) as per the protocol. Recombinant colonies were further screened through colony PCR with universal M13 vector specific primers as well as gene specific primers. Each colony was analysed for their accurate amplification in agarose gel electrophoresis and one selected colony was used for plasmid isolation using Qiaprep Spin Miniprep Kit (Qiagen Cat No. 27104). Purified plasmid was sent for sequencing to Delhi University (India) through vector specific M13 universal primers.

4.2.4 Genetic polymorphism and phylogenetic analysis

Sequencing data of Pilani, Rajasthan AeITS-2 and mtCOI were analyzed using Chromas software (http://www.technelysium.com.au/) and confirmed through blast search. Other available ITS-2 and COI sequences of *Ae. aegypti* mosquito's isolates were retrieved from the NCBI **(Table 4.1)**. The DNA sequences were subjected for alignment using Clustal omega platform as before (Sievers et al., 2011) and their sequence variability had been recorded. Sequence divergences among individual species were quantified using the Kimura two-parameter (K2P) distance model (Kimura, 1980) obtained by 100 bootstrap replicates. This is a quantitative approach to identify the genetic variation among different population of same species. Here we used 2 genes; one from nuclear origin ITS-2 and another mitochondrial gene COI to identify the genetic variability. Same lengths of nucleotides (~191 bp and ~621 bp respectively) from each variant were taken and aligned together. The alignment session was exported in MEGA format and estimated the average evolutionary divergence, conducted using the Neighbor-Joining (NJ) method (Tamura et al., 2011). The average evolutionary divergence was estimated as a number of base substitutions per site from averaging over all sequence pairs within and between each group. All results are based on the pair-wise analysis of known sequences of ITS-2 and COI of other reported isolates **(Table 4.1)**. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic tree was constructed by MEGA using NJ method with 1000 bootstrap value and complete deletion option.

Table 4.1 NCBI retrieved sequences of ITS-2 and COI of *Aedes aegypti* from different geographical locations of India and world

***Aedes aegypti* ITS-2 sequences from world**			
Variant	**Accession number**	**Sequence Length**	**Submitted by**
Ryukyu, Japan	AB548800	207bp	Higa et al., 2010
Cajamarca, Peru	AY512665	274bp	Leiva and Caceres, 2003
Hosta, Russia	HE820724	314bp	Ganushkina et al., 2012
Saudi Arabia	JX423807	378bp	Alhudaib et al., 2012
Mayotte, France	KF135506	213bp	Le Goff et al., 2013
Uganda	M95126	948bp	Wesson et al., 1992
Rajasthan, India	KJ862124	313bp	Gupta et al., 2014 Present study
***Aedes aegypti* COI sequences from India**			
Variant	**Accession number**	**Sequence Length**	**Submitted by**
Ennore, Tamilnadu1	DQ424949	656bp	Kumar et al., 2007
Agricultural farm Tamilnadu2,	AB907183	705bp	Veeramani et al., 2014

Pondicherry	AY729987	510bp	Kumar et al., 2007
Mamulapusi, Orissa	HM807269	710bp	”
Thiruvananthapuram, Kerala	HM807268	708bp	”
Thirumala, Andhra Pradesh P7	HM807267	668bp	”
Thirumala, Andhra Pradesh P6	HM807266	669bp	”
Thirumala, Andhra Pradesh P5	HM807265	660bp	”
Thirumala, Andhra Pradesh P4	HM807264	671bp	”
Thirumala, Andhra Pradesh P3	HM807263	676bp	”
Thirumala, Andhra Pradesh P2	HM807262	677bp	”
Puthur, Andhra Pradesh P1	HM807261	665bp	”
Rajasthan, India	KP121340	609bp	Gupta et al., 2014 present study
***Aedes aegypti* COI sequences from world**			
Variant	**Accession number**	**Sequence Length**	**Submitted by**
Cambodia	JQ926688	764bp	Paupy et al., 2012
Bolivia	JQ926679	”	”
Brazil	JQ926703	”	”
Cameroon	JQ926702	”	”
Ivory Cost	JQ926694	”	”
Guinea	JQ926700	”	”
Tanzania	JQ926704	”	”
Thailand	JQ926692	”	”
USA	JQ926684	”	”
Venezuela	JQ926701	”	”
Vietnam	JQ926687	”	”
Martinique	JQ926696	”	”
Mexico	JQ926698	”	”
France	HQ688296	948bp	Fort et al., 2012
Madagascar	HQ688298	948bp	Fort et al., 2012
Portugal	KF909122	1098	Seixas wt al., 2013

4.3. Results

4.3.1 Molecular characterization of *Aedes*

Field collected mosquito larvae were brought in bulk to insectary and maintained in standard conditions as described in methodology. After hatching into the adult stage, some females were visualized under microscope for their unique morphological feature as described in the key guide for *Aedes* mosquitoes (Rueda et al., 2004). Particularly, scales on vertex and wing structure specific to *Ae. aegypti* in male and female mosquitoes had been observed. Microscopic examination showed middle of vertex has silvery white flat scales with erect forked bristles which are restricted to occiput. Wings were oval in shape and posterior hind margin has erect fringe scales. Those females who followed the above morphological features were selected for molecular characterization and pure line generation.

Each identified female mosquito was housed in individual cages and fed with mice blood to lay eggs and colony propagation. Total gDNA from 1st instar larva from each individual female cage was used as a template for the amplification of ITS-2 gene from gene-specific primers. PCR amplified ~300 bp fragments from AePL-1 to AePL-5 colonies respectively are shown in **Figure 4.2A**. Larvae of lane 2 (AePL-2) were allowed to complete metamorphosis and established in insectary. Isolated gDNA from this colony was further used for amplification of ITS-2 and COI gene which are the reliable gene marker to confirm the species **(Figure 4.2B and 4.2C)**.

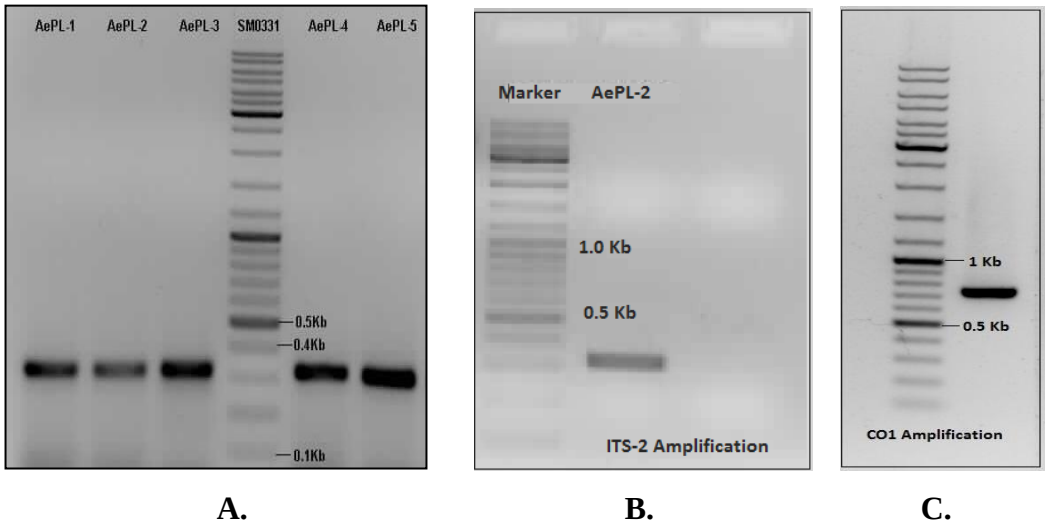


Fig.4.2 A. Molecular characterization of morphologically identified individual mosquito pure lines (AePL 1-5), **B.** Gel image showing the desired band size (313bp) of ITS-2 for one colony (AePL-2) that was finally selected for colonization in the insectary **C.** PCR amplification of COI gene (714 bp) of the same colonized mosquito pure line AePL-2.

4.3.2 Cloning and sequencing of ITS-2 and COI genes

Purified PCR product of ITS-2 (~313bp) of AePL-2 was cloned in PCR-II TOPO TA vector as described above. White colonies with ITS-2 as insert were screened through colony PCR using M13 universal primers. One of the positive colonies were used for plasmid isolation and sent for sequencing and 205bp region of ITS-2 was obtained. In case of COI ~714bp partial gene was amplified and sequenced. Through sequencing only 621bp sequence was obtained. ITS-2 and COI sequences were checked for its nucleotides identity and confirmed through BLAST. ITS-2 gene showed highest identity with *Ae. aegypti* of Saudi Arabia

(JX423807), as none of the ITS-2 gene sequence was reported from India. While COI gene showed 99% identity with *Ae. aegypti* Indian isolate from Thirumala, Andhra Pradesh region (HM807266). These two sequences were reported as complete ITS-2 and partial COI with accession numbers KJ862124 and KP121340 to NCBI respectively. The morphological characteristics of field collected mosquito were matched with molecular markers and the species was confirmed as *Aedes* (Stegomyia) *aegypti*.

4.3.3 Genetic polymorphism among *Aedes aegypti* species

To compare the genetic diversity among *Ae. aegypti* of Rajasthan (India ITS-2, mentioned in Figure 4.3A) with others species located in different geographical location, 204bp fragment of ITS-2 was examined. Since no ITS-2 sequence of Indian isolate was available, comparison was made with different *Ae. aegypti* isolates from various geographical locations of world **(Table 4.1)**. Sequence analysis with other world isolates indicates highest identity to *Ae. aegypti* of Saudi Arabia (JX423807) origin and farthest to Mayotte, France, isolate. Amongst 204 nucleotides of ITS-2, we found 9 polymorphic sites (4.4%) in all 6 world isolates at base pair position number 22, 26, 28, 29, 37, 110, 116, 150 and 154 (**Figure 4.3A**)

For the COI analysis, 609bp was selected for multiple sequence alignment. A total of 12 isolate of Indian origin and 16 isolates from different countries **(Table 4.1)** were compared with Rajasthan isolate. Unlike ITS-2, we observed less genetic difference within Indian isolate as well as other globally reported isolates. There are 19 polymorphic sites (3.1%) were observed from Indian isolates at different base pair position as given in the figure starting from 51 bp position to 465 bp position **(Figure 4.3B).** While compared at worldwide we found 36 sites (5.9%) started at base pair position number 6 and ended at 591 base pair position **(Figure 4.3C)**. The above observation is based on, the numbers of base differences per site from averaging over all sequence pairs were 0.011 and 0.014 respectively. Interestingly, lab colonized *Ae. aegypti* of Rajasthan, COI gene showed only one difference with *Ae. aegypti* Indian isolate from Thirumala, Andhra Pradesh region (HM807262.1), while other isolates of Andhra Pradesh showed high genetic variation. Likely, India Rajasthan COI showed similarity with Thailand, except one difference at 51bp position.

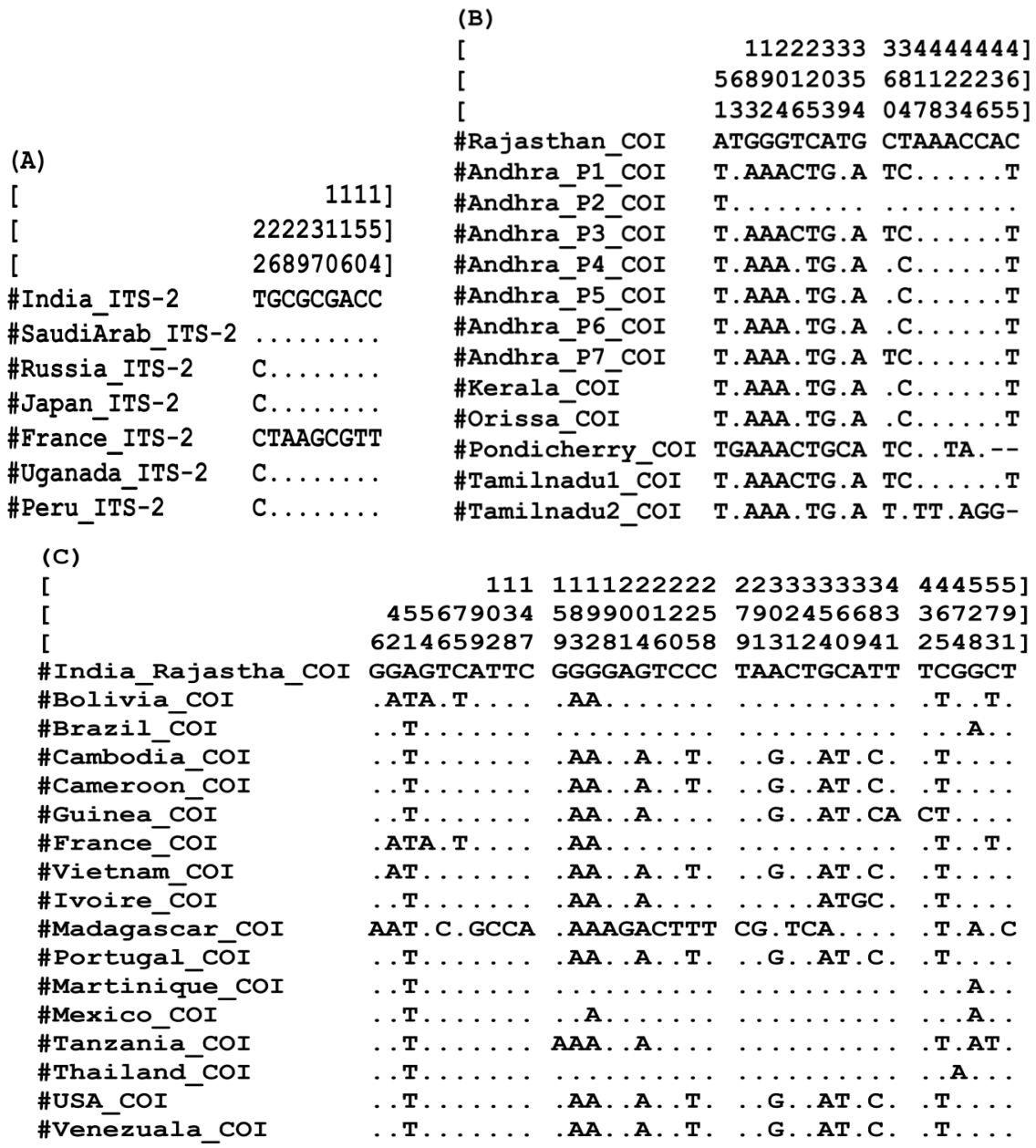


Figure 4.3 Genetic polymorphic sites of ITS-2 and COI for Indian isolate compared with NCBI retrieved sequence as given in **Table 4.1** respectively. At the top of the alignment, vertical number digits are showing the polymorphic sites at particular base pair position **(A)** 9 nucleic acid polymorphic sites for Indian *Aedes aegypti* ITS-2 compared with other world isolates **(B)** Variable nucleic acid sites in Indian *Aedes aegypti*, showing 19 polymorphic sites for COI gene (**C)** 36 nucleic acid polymorphic sites for Indian *Aedes aegypti* COI compared with world isolates.

4.3.4 Phylogenetic analysis

Phylogenetic analysis of Indian *Ae. aegypti* ITS-2 with 6 world isolates showed high similarity with Saudi Arabia and diverged from France isolate ITS-2 sequence **(Figure**

4.4A). Russia and Japan *Ae. aegypti* strains are more closer to Indian *Ae. aegypti* whereas Peru and Uganda are very closer to each other.

The phylogenetic analysis of COI sequence involved 13 *Ae. aegypti* sequences from India and 17 sequences globally. Rajasthan *Ae. aegypti* showed high similarity with Andhra Pradesh isolate (AndhraP2COI) of India and Thailand isolates (**Figure 4.4B and 4.4C**). The COI dendrogram of *Ae. aegypti* clustered in to two clearly shows that based on Indian variant of *Ae. aegypti* specimens; it is clustered into two groups. The first consists of Island species and other continental species of subtropical and tropical origin **(Figure 4.3C)**. Madagascar COI is highly diverged showing in separate clad suggesting much different from other *Ae. aegypti* mosquitoes.

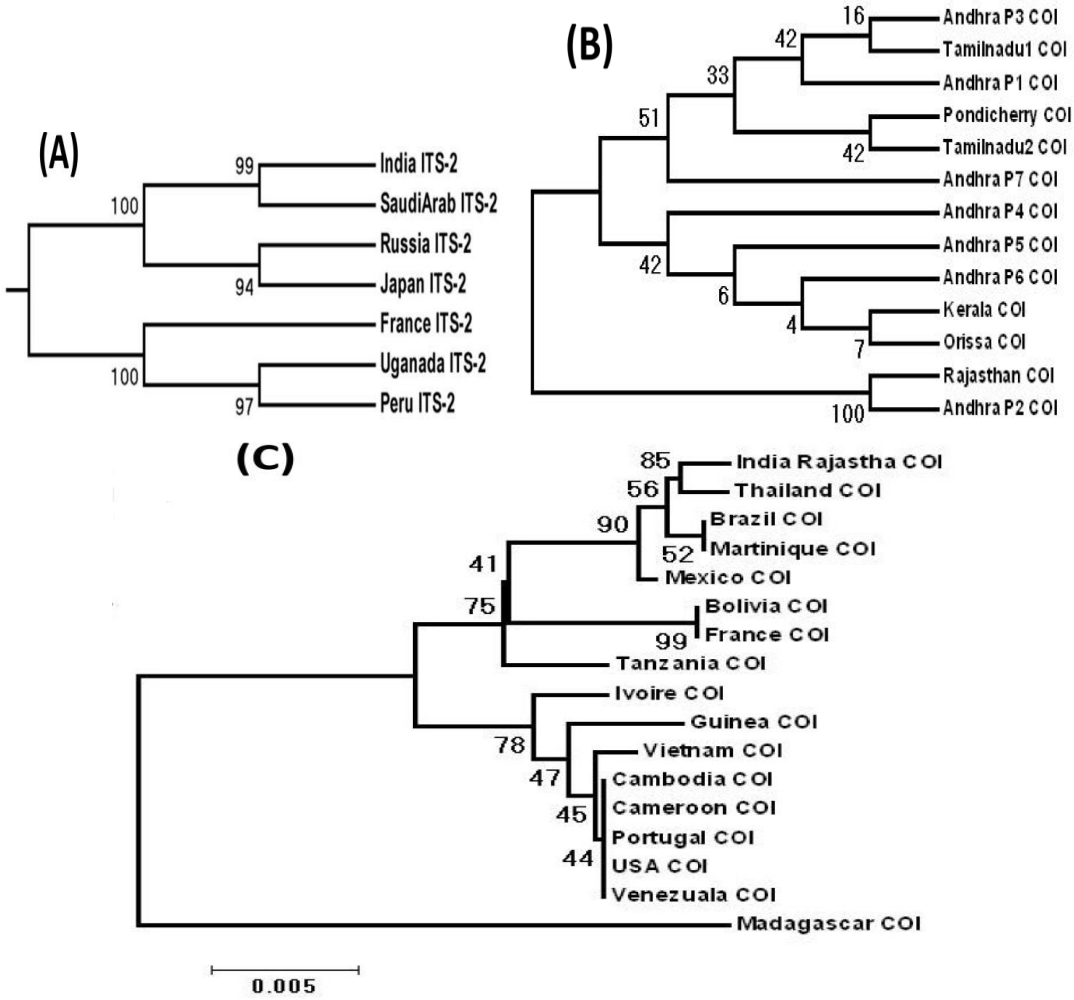


Figure 4.4 Evolutionary relationship of *Ae. aegypti* based on the aligned region of ITS-2 sequences **(A)**. Phylogenetic tree deriving the relationship of AePL-2 COI sequence from Indian isolate **(B)** and other world isolates **(C)**, respectively. Previously published sequence data are shown together with their GenBank accession numbers in **Table 4.1**.

4.4 Discussion

Ae. aegypti is an important disease vector causing millions of death every year and therefore justifies the focus that has been given to its genetic diversity (Kawada et al., 2014; Vikram et al., 2016). Wide combinations of genetic markers are used to examine population structure, genetic differentiation and gene flow of the species. These are, therefore, an essential component of vector-borne diseases management strategies as geographic origins of mosquito populations have epidemiological importance. This has been shown by several studies where relevance among the geographic origin of vectors with vector competence and insecticide resistance has been established (Lourenço-de-Oliveira et al., 2004). Similarly, sequences of COI and ITS-2 of *Ae. aegypti* from Pilani region of Rajasthan have been phylogenetically analysed and compared with their counterparts of India and World. The results showed that COI from Pilani isolate shows a striking polymorphism in 51st position that may be a result of the arid climate interactions. On the other hand, ITS-2 showed close relatedness with the mosquitoes from Saudi Arabia. This close relationship can be attributed to the similar climatic condition and arid environment that led to the similar evolutionary event within two mosquito strains for its successful adaptation. The results presented here are the first genetic analysis of *Ae. aegypti* from north western India (Pilani region of Rajasthan). Sequence analysis of ITS-2 and COI from morphologically identified *Aedes* confirmed the identity of the species that were subsequently colonized in the laboratory. Correct vector identification is crucial aspect for the designing of strategies that could effectively manage the vector-borne diseases.

4.5. Conclusion

Aedes has been classified as one of the "100 of the World's Worst invaders species" and its close contacts with humans mainly in urban areas could be a threat for the emergence of several diseases (Lowe et al., 2000). The results presented here are the first evaluation of the genetic analysis of *Ae. aegypti* from Pilani region of Rajasthan as well as from north western India. The present study is fruitful to colonize the mosquito colony and further work on mosquito immunity. In present study, we first amplified the ITS-2 region in order to potentially discriminate between *Aedes* sp. such molecular diagnostic may also facilitate larger scale studies. This molecularly identified *Ae. aegypti* are now continuously reared in

the insectary. Further we are performing different experiment to understand mosquito immunity to understand specifically ABC transporters and its role against bacterial and viral infections.

4.6 References

Angel B, Joshi V. 2008. Distribution and seasonality of vertically transmitted dengue viruses in *Aedes* mosquitoes in arid and semi-arid areas of Rajasthan, India. J Vector Borne Dis. 45:56-59.

Behura SK, Lobo NF, Haas B, deBruyn B, Lovin DD, Shumway MF, et al., 2011. Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. Insect Biochem Mol Biol. 41(10):770-777.

Benedict MQ, Knols BGJ, Bossin HC, Howell PI, Mialhe E, et al., 2009. Colonization and mass rearing: learning from others. Malaria J. 8(2): S4

Besansky NJ, Severson DW, Ferdig MT. 2003. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. Trends Parasitol. 19:545-546.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. 2013. The global distribution and burden of dengue. Nature 496(7446):504-507.

Caicedo PA, Baron OL, Perez M, Alexander N, Ocampo CB. 2013. Selection of *Aedes aegypti* (Diptera: Culicidae) strains that are susceptible or refractory to Dengue-2 virus. Can Entomol. 145:273-282.

Cameron EC, Wilkerson RC, Mogi M, Miyagi I, Toma T, et al., 2010. Molecular phylogenetics of *Aedes japonicus*, a disease vector that recently invaded Western Europe, North America and the Hawaiian Islands. J Med Entomol. 47(4):525-535.

Cook S, Diallo M, Sall AA, Cooper A, Holmes EC. 2005. Mitochondrial markers for molecular identification of *Aedes* mosquitoes (Diptera: Culicidae) involved in transmission of arboviral disease in West Africa. J Med Entomol 42(1):19-28.

Gerberg EJ. 1979. Manual for mosquito rearing and experimental techniques, American Mosquito Control Association (AMCA bulletin no. 5), California

Gupta N, Srivastava S, Jain A, Chaturvedi UC. 2012. Dengue in India. Indian J Med Res. 136(3):373-390.

Gupta S, Preet S. 2012. Protocol optimization for genomic DNA extraction and RAPD-PCR in mosquito larvae (Diptera: Culicidae). Ann Biol Res 3:1553-1561.

Higa Y, Toma T, Tsuda Y, Miyagi I. 2010. A multiplex PCR-based molecular identification of five morphologically related, medically important subgenus Stegomyia mosquitoes from the genus Aedes (Diptera: Culicidae) found in the Ryukyu Archipelago, Japan. Jpn J Infect Dis. 63(5):312-316.

Higgs S, Vanlandingham D. 2015. Chikungunya virus and its mosquito vectors. Vector Borne Zoonotic Dis. 15(4): 231-240.

Hlaing T, Tun-Lin W, Somboon P, Socheat D, Setha T, Min S, et al. 2009. Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. BMC Genetics 10:11.

Hoshino K, Isawa H, Tsuda Y, Kobayashi M. 2010. Laboratory colonization of *Aedes japonicus japonicus* (Diptera: Culicidae) collected in Narita, Japan and the biological properties of the established colony. Jpn J Infect Dis. 63(6):401-404.

Kawada H, Oo SZ, Thaung S, Kawashima E, Maung YN, Thu HM, et al. 2014. Co-occurrence of point mutations in the voltage-gated sodium channel of pyrethroid-resistant *Aedes aegypti* populations in Myanmar. PLoS Negl Trop Dis. 8(7): e3032.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16:111-120.

Kumar NP, Rajavel AR, Natarajan R, Jambulingam P. 2007. DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). J Med Entomol. 44(1):1-7.

Landis FC, Gargas A. 2007. Using ITS-2 secondary structure to create species-specific oligonucleotides probes for fungi. Mycologia 99(5):681-692.

Lourenço-de-Oliveira R, Vazeille M, de Fillippis AM, Failloux AB. 2004. *Aedes aegypti* in Brazil: genetically differentiated populations with high susceptibility to dengue and yellow fever viruses. Trans R Soc Trop Med Hyg. 98:43-54.

Lowe S, Browne M, Boudjelas S, De Poorter M. 2000. 100 of the world's worst invasive alien species. A selection from the global invasive species database. The invasive species specialist group (ISSG)/ Species Survival Commission (SSC)/World Conservation Union (IUCN), 12 pp http://www.issg.org/booklet.pdf.

Manonmani AM, Mathivanan AK, Sadanandane C, Jambulingam P. 2013. Evaluation of the mtDNA-COII region based species specific assay for identifying members of the *Anopheles culicifacies* species complex. J Arthropod-Borne Dis. 7(2):154-163.

Morlais I, Severson DW. 2002. Complete mitochondrial DNA sequence and amino acid analysis of the cytochrome C oxidase subunit I (COI) from *Aedes aegypti*. DNA Sequence 13(2):123-127.

Mousson L, Dauga C, Garrigues T, Schaffner F, Vazeille M, Failloux AB. 2005. Phylogeography of *Aedes (Stegomyia) aegypti* (L.) and *Aedes (Stegomyia) albopictus (Skuse)* (Diptera: Culicidae) based on mitochondrial DNA variations. Genet Res. 86(1):1-11.

Musters W, Boon K, Sande C, Heerikhuizen HV, Planta RJ. 1990. Functional analysis of transcribed spacers of yeast ribosomal DNA. EMBO J. 9(12):3989-3996.

National vector borne disease control programme (NVBDCP), Directorate general of health services, Ministry of Health and family welfare 2015.

Pennisi E. 2003. Modernizing the tree of life. Science 300:1692-1697.

Reinert JF, Harbach RE, Kitching IJ. 2004. Phylogeny and classification of Aedini (Diptera: Culicidae), based on morphological characters of all life stages. Zool J Linn Soc-Lond. 142(3):289-368.

Rueda LM. 2004. Zootaxa 589, Pictorial keys for the identification of mosquitoes (Diptera: Culicidae) associated with dengue virus transmission. Magnolia Press, Auckland, New Zealand 60 pp; 30 cm.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:539.

Sim S, Jupatanakul N, Ramirez JL, Kang S, Romero-Vivas CM, et al., 2013. Transcriptomic profiling of diverse *Aedes aegypti* strains reveals increased basal-level immune activation in dengue virus-refractory populations and identifies novel virus-vector molecular interactions. PLoS Negl Trop Dis. 7(7): e2295.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731-2739.

Thiboutot MM, Kannan S, Kawalekar OU, Shedlock DJ, Khan AS, Sarangan G, et al., 2010. Chikungunya: A potentially emerging epidemic? PLoS Negl Trop Dis. 4(4): e623.

Urdaneta-Marquez L, Failloux AB. 2011. Population genetic structure of *Aedes aegypti* the principal vector of dengue viruses. Infect Genet Evol. 11(2):253-261.

Vikram K, Nagpal BN, Pande V, Srivastava A, Saxena R, Anvikar A, et al., 2016. An epidemiological study of Dengue in Delhi, India. Acta Trop 153:21-27.

Wallis GP, Aitken TH, Beaty BJ, Lorenz L, Amato GD, Tabachnick WJ. 1985. Selection for susceptibility and refractoriness of *Aedes aegypti* to oral infection with yellow fever virus. Am J Trop Med Hyg. 34(6):1225-1231.

World Health Organization report, 2014. World health statistics. World Health Organization http://www.who.int/malaria/publications/world malaria report 2014/report/en/

Objective: Genome wide identification of ATP-Binding Cassette (ABC) transporters gene family in the Aedes aegypti mosquito

5.1 Introduction

The ATP-binding cassette (ABC) transporter family is one of the largest families of membrane proteins identified in all living life from bacteria to human (Higgins, 1992; Schuller et al., 2003). This family usually functions in an ATP-dependent manner to transport various substrates across plasma membranes, such as inorganic ions, sugars, amino acids, lipids, lipopolysaccharides, peptides, metals, and chemotherapeutic drugs (Schuller et al., 2003). ABC transporters have been classified and analyzed in several insect species, such as *Drosophila melanogaster* (Dean et al., 2001) *Anopheles gambiae* (Roth et al., 2003) *Bombyx mori* (Liu et al., 2011; Xie et al., 2012) *Tribolium castaneum* (Broehan et al., 2013). However, comparisons of ABC transporter family members among taxonomically related species are rare. The genome sequences of *An. gambiae*, *Aedes aegypti* and *Culex quinquefasciatus* (Holt et al., 2002; Nene et al., 2007; Arensburger et al., 2010) are available now for intensive genome wide analyses which are also important vectors for several human diseases. *An. gambiae* ABC transporters have been analyzed and classified at the genome level (Roth et al., 2003). But, the mechanism by which ABC transporter family evolves in other mosquito species remains unknown.

The ABC transporters are represented with four functional units: two nucleotide-binding domains (NBDs) and two transmembrane domains (TMD) encoded by a single peptide chain. ABC genes can be organized either as full transporters or half transporters. A typical ABC full transporter consists of two NBDs and two TMD, each of which contains 5–6 transmembrane helices and provides substrate specificity. While an ABC half transporter consists of one TMD and one NBD or either one (**Figure 5.1 B-D**). Either two identical (homo) or two different (hetero) half transporters come together as a dimer to assemble a functional transporter (Wilkinson and Verschueren, 2003). The highly conserved region of ABC protein is their NBDs which contains three characteristic motifs, namely, Walker A, Walker B, and the ABC signature (LSGGQ) motif that links the two Walker motifs also known as C motif (Walker et al., 1982; Smith and Rayment, 1996; Bianchet et al., 1997) (**Figure 5.1 A**). In general, Walker A and B have found to be approximately 100 amino acids apart from each other. In contrast, the transmembrane domains of these transporters may have 6-11 transmembrane helices composed of less conserved amino acids (Schneider and Hunke, 1998; Holland and Blight, 1999).

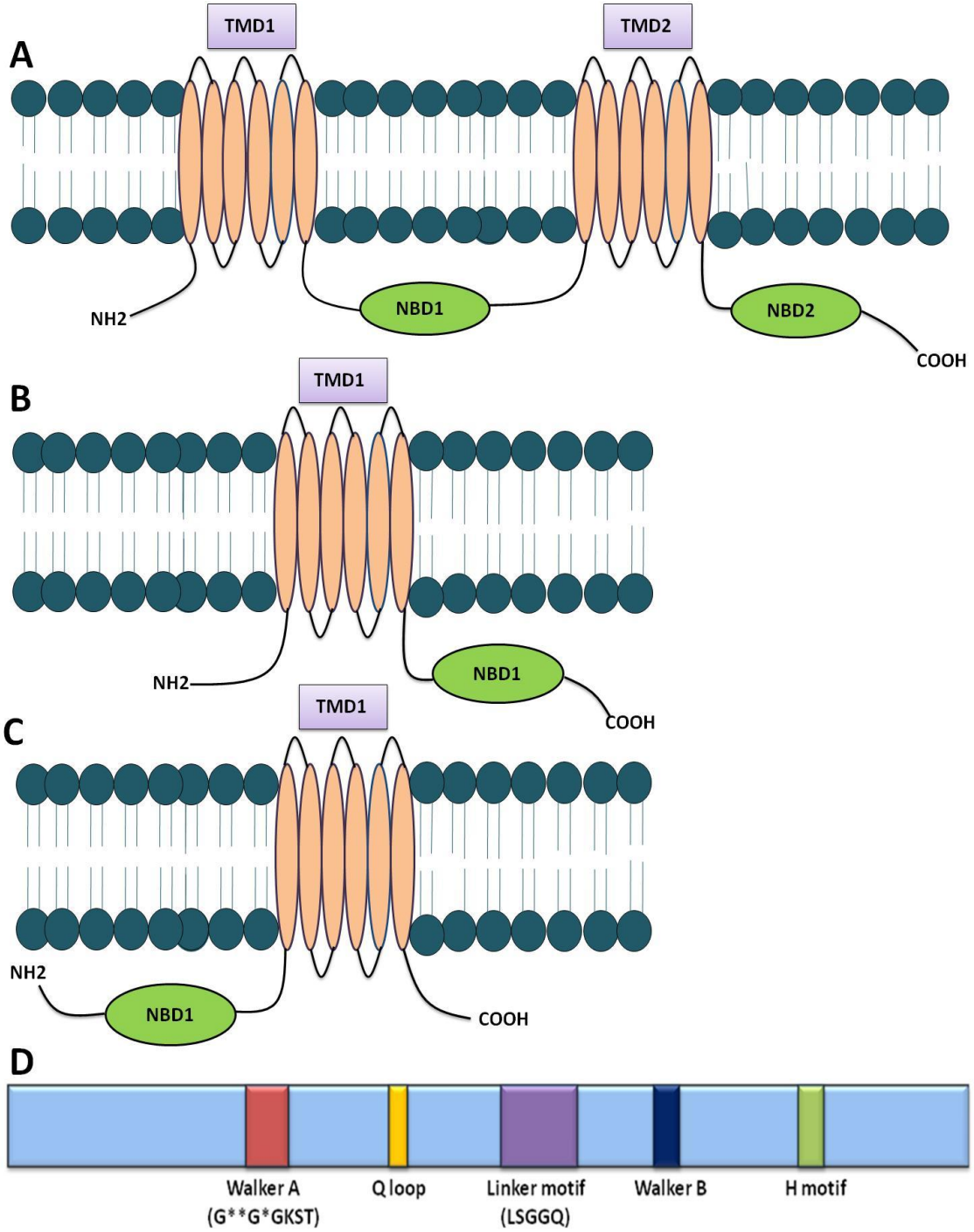


Figure 5.1: Schematic representation of different types of ABC transporters **A.** Consensus motifs in NBDs. Different colors distinguish the conserved segments: Walker A motif (*red)*, LSGGQ motif (*purple)*, Walker B motif *(blue)*, and the Q loop *(yellow)*, H motif *(green)*. **B.** A typical ABC full transporter showing two TMD and two NBD (ATPase domain) on a lipid bilayer membrane with 6 transmembrane helices. **C.** ABC half transporters showing one TMD at N-terminal and one NBD at C-terminal **D.** Another type of half transporter showing NBD at N-terminal and TMD at C-terminal.

Nucleotide-binding domain (NBD) is in fact the ATP-binding domain which provides the energy for transportation of a diverse array of molecules against concentration

gradients through ATP hydrolysis (Saurin et al., 1999; Hollenstein et al., 2007). In addition, substrate-binding proteins (SBP) also comprise the part of ABC transporters for substrate uptake in ion gradient-driven transporters found in both prokaryotes and eukaryotes (Berntsson et al., 2010). With the conservation of NBDs, ABC transporters have been divided into eight subfamilies (A–H) (Jones & George, 2004). Members of these subfamilies are commonly present throughout the known genomes starting from bacteria to human however; some of them are exclusively limited to the genome of specific organisms (Schneider and Hunke, 1998). The human ABC transporters were the first to be identified and 48 ABCs were classified into 7 subfamilies (ABCA to ABCG) (Dean et al., 2001). The eighth subfamily, ABCH is absent in the genome of plants, worms, yeast and mammals including human and was first identified in *D. melanogaster* and appears to be present in *Dictyostelium* (soil-dwelling amoeba), all insects and zebra fish (Dean and Annilo, 2005; Liu et al., 2011). Thus, the evaluation of ABCs, exclusively confined to insects, may provide an insight into their relationship to vector biology of pathogens development and transmission in human population.

The genes within a subfamily also show considerable identity in their TMDs and have similar intervening sequence location (gene organization) (Wilkinson and Verschueren, 2003). Six of the ABC gene subfamilies are found in the yeast genome, indicating that these groups were established early in the evolution of eukaryotes, are essential to all plants and animals and therefore have been retained. The ABCA, ABCB and ABCC subfamilies are the most abundant and each have 10-12 genes in the human genome. Genes within the same subfamily often have very different functions, and genes from different sub-families can have related functions (Schuller et al., 2003). In general, most of its members are involved in transport of ions, amino acids, lipids, sugars, peptides, metals across the membrane (hence named as importers or exporters) and fewer members perform other cellular functions such as regulation of gene expression, repairing of cell DNA and protein synthesis (Zhao et al., 2004). Importantly, ABC transporters which are exporters in nature are commonly found in eukaryotes however, the transporters with importing property are generally associated with prokaryotes (Wilkinson & Verschueren, 2003; Dawson & Locher, 2006). These transporters are not only helpful in physiological functions but also provide other conditional advantages to cells (Abele and Tampe, 2004).

The importance of ABC transporters increases due to their involvement in transport of xenobiotics, chemotherapeutic drugs and cancer therapy (Misra et al., 2011; Kennedy and Tierney, 2013). The first characterized eukaryotic ABC transporter (P-glycoprotein) present on cancer cell surface mediates multidrug resistance (MDR) through efflux and prevents the accumulation of effective chemotherapeutic drugs inside these cells (Kartner et al., 1983). In addition, the over-expression of ABC transporters in human cancer cells also enhances their drug resistance. Moreover, mutations in human ABC transporters are linked to many disorders such as, cystic fibrosis, hypercholesterolemia and diabetes (Klein et al., 1999; Gottesman and Ambudkar, 2001). Also near about 17 human ABC transporters, out of total 48 ABCs, have been found to be involved in hereditary diseases including cystic fibrosis, adrenoleukodystrophy, Stargardt disease and disorders of cholesterol metabolism (Dean et al., 2001; Dean and Annilo, 2005).

Along with P450 monooxygenases, glutathione-S-transferases and carboxylesterases, the ABC transporter family is also considered as a major detoxification gene family due to its functions in transporting drugs or insecticides (Srivastava et al., 2010). Several ABC members reportedly participate in important roles like insecticide resistance in lepidopterans; these members include the P-glycoproteins from the ABCB subfamily and the multidrug-resistance associated proteins from the ABCC subfamily (Labbe et al., 2011). Insect P-glycoproteins and MDR associated proteins are frequently linked to pesticide resistance. Some ABC transporter genes are upregulated in pyrethroid-resistant strains of *An. gambiae*, *Ae. aegypti* and *An. stephensi* mosquitoes (Bonizzoni et al., 2012; Epis et al., 2014; Bariami et al., 2012). In the present study, we have classified and compared the functions of ABC transporters of *Ae. aegypti* (Liverpool strain), to explore the gene expansion or deletion in the evolution of the ABC transporter family. The comparatively conserved nature of NBD and their phylogenetic relationship is helpful to categorize eukaryotic ABC transporters into subfamilies.

5.2 Materials and Methods

5.2.1 Identification of ABC transporters in *Aedes* genomes

To identify the total number of ABC transporters from *Aedes* genome, *Drosophila*, ABC gene accession numbers were used from previous studies (Roth et al., 2003; Sturm et al., 2009). The highly conserved NBDs were extracted for each *Drosophila*'s ABC protein by using the Scanprosite software at Expasy (**Table 5.1**). As a query for homology

searches, we used only N-terminal NBDs from the *Drosophila* genomes, if the ABC transporter had two NBDs. Individual NBD of *Drosophila* ABC transporters was used to perform BLASTP against *Aedes* genome. Hits from individual member were only considered when the E-values were less than 10^{-6}. Among these hits, identified putative *Aedes* ABC transporter proteins were also validated by VectorBase protein databases. With the help of Microsoft Excel programme, the redundancy in the current repository was minimized and unique protein ID was used.

Table 5.1 Inventory of *Drosophila melanogaster* N-terminal NBDs used as a query for genome wide homology search through BLASTP

Subfamily	Accession no.	Start – End	No. of AAs in NBD 1	Protein Size
ABCA (10)	CG1494	541 – 773	233	1695
	CG1718	538 – 767	230	1714
	CG1801	500 – 733	234	1655
	CG1819	841 – 1065	225	1981
	CG6052	515 – 745	231	1700
	CG8908	434 – 633	230	1625
	CG31213	491 – 720	230	1809
	CG31731	408 – 638	231	1544
	CG33173	438 – 668	231	1459
	CG32186	339 – 568	230	1459
ABCB (8)	AAF 47525	485 – 719	235	743
	AAF 48177	462 – 699	238	761
	AAF 50670	414 – 650	237	1320
	AAF 55241	576 – 810	235	866
	NP 476831	405 – 641	237	1302
	NP 523724	402 – 638	237	1302
	NP 523740	431 – 667	237	1313
	NP 569844	445 – 684	240	692
ABCC (14)	AAF 46706	428 – 655	228	1312
	AAF 52639	621 – 850	230	1487
	AAF 52648	445 – 668	224	1355
	AAF 53223	649 – 871	223	1549
	AAF 53950	265 – 487	223	1145
	AAF 54656	422 – 645	224	1316
	AAF 55707	443 – 666	224	1362
	AAF 56312	495 – 722	228	1402
	AAF 56869	448 – 673	226	1374
	AAF 56870	420 – 649	230	1320
	AAF 58947	499 – 722	224	1374
	NP 477472	785 –1014	230	2171
	NP 609930	434 – 656	223	1307
	NP 724148	434 – 657	224	1323
ABCD (2)	AAF 49018	446 – 665	220	665
	AAF 59367	505 – 727	223	730
ABCE (1)	AAF 50342	78 – 323	246	611
ABCF (3)	AAF 48069	352 – 596	245	911
	AAF 48493	74 – 315	242	611
	AAF 49142	178 – 424	247	708
ABCG (15)	AAF 45826	93 – 341	249	687
	AAF 47020	34 – 261	228	675
	AAF 49455	69 – 316	248	666
	AAF 51027	145 – 384	240	832

	AAF 51122	1 – 207	208	634
	AAF 51130	148 – 388	241	808
	AAF 51131	23 – 261	239	609
	AAF 51341	106 – 351	246	766
	AAF 51548	46 – 285	240	698
	AAF 51552	54 – 294	241	677
	AAF 56361	10 – 251	242	604
	NP 722827	15 – 248	234	615
	NP 729728	15 – 246	232	623
	NP 733058	303 – 536	234	911
ABCH (3)	AAF 52284	4 – 238	235	711
	AAF 56807	74 – 308	235	808
	NP 001034071	51 – 282	232	777

5.2.2 Analysis of conserved domain in ABC protein

The protein IDs of each unique ABC transporters of *Aedes* were used to retrieve the full protein sequence from NCBI. From the supporting information of Nene et al., 2007, their genome coordinates like position of these transporters on genomic contigs were retrieved. The sequences were re-evaluated with NCBI's Conserved Domain (CD) search tool from conserved domain database (Marchler-Bauer and Bryant, 2004; Marchler-Bauer et al., 2007) to search the corresponding domain of particular subfamily. The peptide sequences of each distinct subgroup of all annotated NBDs were analysed using multiple sequences alignment program Clustal omega (Sievers et al., 2011).

5.2.3 Subfamily classification of *Aedes* ABC transporters and structural topology

The respective NBDs of each protein was analysed to classify these particular ABC transporters in relevant subfamily. For naming the ABC proteins of *Aedes aegypti*, following nomenclature convention were followed. *Ae* indicating the species *Aedes aegypti*, ABC the protein superfamily abbreviation, A-H the subfamily name, followed by the Arabic numbers according to their increased occurrence of NCBI accession numbers designating its arbitrary place in the subfamily. All amino acid sequences were used to extract the topology by performing Pfam searches at the Sanger Institute website (Finn et al., 2008). Pattern of distribution of these conserve domain further decides their topology in form of full and half transporters. Number of helices in the ABC protein was determined using TMHMM tool (Krogh et al., 2001).

5.2.4 Phylogenetic relationship of ABC transporter subfamilies

The NBD sequences of ABC transporters for *An. gambiae* were also retrieved from NCBI following the similar strategies as described above. Only N-terminal NBD

domains of *Aedes* along with *An. gambiae* and *D. melanogaster* were used for phylogenetic analysis. Phylogenetic analysis was performed using MEGA 5.2 software (Tamura et al., 2011). Phylogenetic studies was carried out for all 8 subfamilies genes individually whereas ABCD, ABCE, ABCF and ABCH altogether due to having least number of members in these subfamilies. The evolutionary history was inferred using the Neighbor-Joining (NJ) method (Saitou and Nei, 1997). The branch support of the NJ phylogenetic tree was estimated using a bootstrap test with 1000 replicates. The percentage of replicate trees in which the associated taxa clustered together is shown next to the branches (Tamura et al., 2011). In case of gaps or missing data in sequences, partial deletion and *p*-distance of the sequences parameters was chosen.

5.3 Results

5.3.1 Identification of *Aedes aegypti* ABC transporter

The conserved NBDs of all 56 *D. melanogaster* genes were used as queries for homology searches performed by BLASTP in RefSeq database of updated genome assembly of *Ae. aegypti*. All putative ABC transporter with E-value lower than 10^{-6} were assembled together and duplicates were eliminated from these clusters. All the obtained accession numbers of ABC transporters were verified through vectorbase data search. Ultimately, a total of 71 putative ABC transporters genes were confirmed in the *Aedes* genome. ABC transporters were further classified in subfamilies ABCA-ABCH as mentioned in **Table 5.2**. These *Aedes* ABC transporters were compared with ABCs inventories of other arthropods available in public database.

Table 5.2 Classification of ABC transporters in 7 arthropod species, Dm: *Drosophila melanogaster*, Ag: *Anopheles gambiae*, Bm: *Bombyx mori*, Tc: *Tribolium castaneum*, Dp: *Daphnia pulex*, Tu: *Tetranychus urticae* and Ae: *Aedes aegypti* and the % of ABC genes in genome.

Subfamilies	***Dm***	***Ag***	***Bm***	***Tc***	***Dp***	***Tu***	***Ae***
A	10	6	9	9	4	9	13
B	8	5	9	6	7	4	7
C	14	14	15	31	7	39	21
D	2	1	2	2	3	2	3
E	1	1	1	1	1	1	1
F	3	3	3	3	4	3	3
G	15	12	12	13	23	23	20
H	3	2	2	3	5	22	3
Total ABCs	**56**	**52**	**53**	**68**	**64**	**103**	**71**
Total gene	17215	13184	14436	16540	30613	18414	17339
% of ABC	0.32%	0.39%	0.37%	0.41%	0.21%	0.56%	0.41%

Among arthropods, an arachnid *Tetranychus urticae* genome revealed the maximum member in ABC superfamily. This organism has 103 ABC transporters that constitute 0.56% of whole protein coding genes (Dermauw et al., 2013). After that, *Aedes aegypti* (order diptera) with 71 and *Tribolium castaneum* (red flour beetle, belongs to order coleoptera) with 68 comprise maximum number of members in this superfamily (Broehan et al., 2013). Interestingly, among the known genomes of insects, *Aedes* has the highest number of ABC gene compare to others. Comparative genome study clearly showed that, one coleopteran *T. castaneum* (0.41%) and a dipteran *Ae. aegypti* (0.41%) have the maximum number of ABC transporter genes in so far known insect genomes (**Table 5.2**).

5.3.2 Genome wide comparison and protein domain analysis

Subfamily assignment of ABC transporters were evaluated on the basis of the known architecture of conserved domains as before (Dassa and Bouige, 2001). All the *Aedes* ABC-binding proteins were grouped according to their functional domains as mentioned in **Table 5.3**. We analysed the conserved domains (CD search) in each ABC protein and grouped them together. Proteins were named accordance with the above mentioned guidelines. The ABCA, ABCC and ABCG transporter subfamilies of *Aedes* have the maximum members among the other category; these subfamilies have 13, 21 and 20 proteins, respectively (**Table 5.2**). Subfamily ABCG is the third in abundance after an arachnid and crustacean species but again higher among insects.

The ABCE and ABCF subfamily genes encode proteins with ATP-binding domains (only NBD) but no transmembrane segments (TMD). These two families are comparatively short having 1 and 3 proteins respectively which also maintains the group size as found in other arthropods as well as insects. The ATP-binding domains of these genes are clearly related to other genes in the superfamily, and these proteins either bind compounds related to ATP (ABCE1) or are thought to be regulatory subunits for other proteins (ABCF). Subfamily ABCD (3) and ABCH (3) are also small in size and maintain the number of members as found in other species except *T. urticae* (**Table 5.2**).

All ABC protein sequences of *Ae. aegypti* were used to extract the topology by performing Pfam searches. Various pattern of array of NBDs or TMDs were observed on a protein sequence which inferred their structure. Pattern of distribution of these conserve domains further decide their topology in form of full and half transporters. Those ABC peptide sequences having 2 NBDs and 2 TMDs or 1 TMD and 2 NBDs were

considered as full transporters. While protein sequences have 1 TMD and 1 NBD or 1 NBD and 2 TMD or either one were listed as half transporters.

Table 5.3 ABC protein domains identified through CD search depict their plausible functions

Subfamily	Domains
ABCA	▪ Energy Coupling Factor Transporter ATP binding Protein (ECFA2) ▪ Retinal Specific rim ABC Transporter (RIM) ▪ PQQ Dependent Alcohol Dehydrogenase System (PQQ) ▪ Multidrug Transport System, ATPase Component (CCM) ▪ Na+ Transport System (NAT)
ABCB	▪ Eye Pigment Precursor Transporter Protein (EPP) ▪ ABC Transporter G-25 ▪ Multidrug Transport System Permease Component (YadH) ▪ Cyclic Beta 1,2 Glucan Transporter ▪ Polar Amino Acid Transport System (GlnQ) ▪ Conjugate transporter 2 Family Protein ▪ Lipoprotein Export System (LolD) ▪ Cytochrome C biogenesis Protein
ABCC	▪ Multidrug Resistance Associated Protein (MRP) ▪ Thiamine ABC transporter (thiQ) ▪ ABC Transporter C family member (PLN03130) ▪ Sulfate ABC Transporter, ATP binding Protein (3a0106s01)
ABCD	▪ Putative Bacteriocin Export ABC Transporter (Locin 972) ▪ Peroxysomal Fatty Acyl CoA transporter (3a01203)
ABCE	▪ Translation Inititation Factor RLI 1
ABCF	▪ Arginine Transport System (ArtP)
ABCG	▪ Phosphonate ABC Transporter (phnC) ▪ CCM A Multidrug Transporter System
ABCH	▪ Cobalamine Siderophores Fe3+ Transport System (fepC) ▪ Nodulation Protein (Nod 1) ▪ Daunorubicin resistance ABC Transporter ATP Binding Unit (drrA)

There are 28 full transporters which belong to subfamily ABCA-D and 43 half transporters which are present in other subfamilies. Full repository of ABC transporter with their genome coordinates have been provided in to two exclusive tables as full transporters (**Table 5.4**) and half transporters (**Table 5.5**). The largest ABC transporter has 2056 amino acids and smallest one with 252 amino acids encoded by XP_001663398.1 (AeABCC18) and XP_001662304.1 (AeABCC15), respectively. However, they might have different chromosomal locations which are not known in this particular genome. We identified their genomic occurrences in respective supercontigs or contigs through the genome/proteome inventory submitted by Nene et al. 2007.

Table 5.4 Genome coordinates of full ABC transporters of *Aedes aegypti*

Subfamily	NCBI-ID	Vector base ID	Amino acids	Topology	No of TM Helices	Supercontigs	Start	End
A1	XP_001649361.1	AAEL014699-PA	1480	NBD-TMD-NBD	10	1.1224	80946	112357
A3	XP_001650080.1	AAEL004968-PA	1582	NBD-TMD-NBD	7	1.136	297771	303977
A4	XP_001650791.1	AAEL015146-PA	1549	2(TMD-NBD)	12	1.1520	51863	59635
A5	XP_001653233.1	AAEL008388-PA	1666	2(TMD-NBD)	13	1.321	644618	664804
A6	XP_001653234.1	AAEL008384-PA	1660	2(TMD-NBD)	13	1.321	675803	697600
A7	XP_001653235.1	AAEL008386-PA	1569	2(TMD-NBD)	12	1.321	698372	730927
A8	XP_001654315.1	AAEL001938-PA	1673	2(TMD-NBD)	14	1.46	792516	818527
A9	XP_001662813.1	AAEL012702-PA	1669	2(TMD-NBD)	14	1.726	372101	377726
A10	XP_001662814.1	AAEL012700-PA	1669	2(TMD-NBD)	14	1.726	372101	377726
A11	XP_001662815.1	AAEL012701-PA	1669	2(TMD-NBD)	12	1.726	372101	377726
A12	XP_001662816.1	AAEL012698-PA	1669	2(TMD-NBD)	12	1.726	372101	377726
B1	XP_001648931.1	AAEL004331-PA	1419	2(TMD-NBD)	12	1.115	240545	271476
B2	XP_001652165.1	AAEL006717-PB	744	2(TMD-NBD)	6			
B4	XP_001654492.1	AAEL010379-PA	1307	2(TMD-NBD)	11	1.474	313030	327570
C3	XP_001650217.1	AAEL005043-RA	1505	2(TMD-NBD)	16	1.139	1140679	1145559
C4	XP_001650218.1	AAEL005026-RA	1384	2(TMD-NBD)	15	1.139	1168407	1184363
C5	XP_001650219.1	AAEL005045-PA	1514	2(TMD-NBD)	17	1.139	1184563	1195380
C6	XP_001650220.1	AAEL005030-PA	1396	2(TMD-NBD)	17	1.139	1233513	1252972
C8	XP_001651693.1	AAEL005918-PA	1312	2(TMD-NBD)	11	1.180	664096	681744
C11	XP_001652123.1	AAEL006622-PA	1540	2(TMD-NBD)	13	1.213	838086	915438
C12	XP_001656872.1	AAEL013567-PA	1311	2(TMD-NBD)	12	1.871	281423	317150
C14	XP_001662303.1	AAEL012192-PA	1345	2(TMD-NBD)	9	1.664	660781	670973
C16	XP_001662529.1	AAEL012395-PA	1357	2(TMD-NBD)	7	1.688	67831	72390

C17	XP_001662530.1	AAEL012386-PA	1351	2(TMD-NBD)	9	1.688	87463	91714
C18	XP_001663398.1	AAEL013215-RA	2056	2(TMD-NBD)	12	1.806	207278	448223
C19	XP_001664022.1	AAEL013833-PA	807	NBD-TMD-NBD	5	1.936	247719	261239
C20	XP_001664023.1	AAEL013834-PA	1235	2(TMD-NBD)	8	1.936	291553	353031
D1	XP_001651026.1	AAEL005499-PA	1382	NBD-TMD-NBD	11	1.160	1362499	1398139

Table 5.5 Genome coordinates of *Aedes aegypti* half ABC transporters

Subfamily	NCBI-ID	Vector base ID	Amino acids	Topology	No of TM Helices	Supercontigs	Start	End
A13	EJY57605.1	AAEL017572-PA	347	NBD	0	1.176	1628836	1629879
A2	XP_001647575.1	AAEL015644-RA	551	TMD	5	1.413	1036	2842
B3	XP_001652166.1	AAEL006717-PA	734	TMD-NBD	6	1.219	178589	203717
B5	XP_001655364.1	AAEL002468-RA	703	TMD-NBD	3	1.58	1203051	1224141
B6	XP_001656464.1	AAEL000434-RA	693	TMD-NBD	5	1.8	3711414	3730662
B7	XP_001658950.1	AAEL008134-PA	848	TMD-NBD	10	1.302	73729	107503
C1	XP_001649738.1	AAEL004743-PB	1069	TMD-NBD	10	1.129	994901	1030978
C2	XP_001649739.1	AAEL004743-PA	1089	TMD-NBD-TMD	10	1.129	994901	1030978
C7	XP_001650467.1	AAEL015067-PA	571	TMD-NBD	2	1.1442	8853	20230
C9	XP_001651697.1	AAEL005937-PA	1300	TMD-NBD	11	1.180	724473	765746
C10	XP_001651698.1	AAEL005929-PA	1413	TMD-NBD	12	1.180	786121	801780
C13	XP_001657174.1	AAEL013854-PA	1013	TMD-NBD-TMD	7	1.941	98439	117196
C15	XP_001662304.1	AAEL012189-PA	252	TMD	2	1.664	689083	689967
C21	EJY57511.1	AAEL017209-PA	903	TMD-NBD	9	1.107	820177	825969
D2	XP_001654181.1	AAEL010047-PA	753	TMD-NBD	3	1.449	843528	895566
D3	XP_001662679.1	AAEL002913-PA	659	TMD-NBD	4	1.71	1617561	1676168

E1	XP_001660605.1	AAEL010059-PA	609	NBD-NBD	0	1.450	713084	727146
F1	XP_001652415 .1	AAEL001101-PA	894	NBD-NBD	0	1.23	2941514	2961984
F2	XP_001654470.1	AAEL010359-PA	712	NBD-NBD	0	1.473	244386	260577
F3	XP_001661220.1	AAEL010977-PA	602	NBD-NBD	0	1.529	122943	143748
G1	XP_001653354.1	AAEL008627-PA	264	NBD	0	1.337	9155	10003
G2	XP_001653355.1	AAEL008624-PB	593	NBD-TMD	6	1.337	23491	61022
G3	XP_001653356.1	AAEL008624-PA	593	NBD-TMD	6	1.337	23491	61022
G4	XP_001653357.1	AAEL008632-PA	607	NBD-TMD	7	1.337	68512	71099
G5	XP_001653358.1	AAEL008628-PA	571	NBD-TMD	5	1.337	85665	99799
G6	XP_001653359.1	AAEL008625-PA	606	NBD-TMD	5	1.337	119628	131014
G7	XP_001653360.1	AAEL008629-PA	723	NBD-TMD	7	1.337	131034	224797
G8	XP_001653363.1	AAEL008631-PA	759	NBD-TMD	7	1.337	276979	394542
G9	XP_001653364.1	AAEL008635-PA	676	NBD-TMD	6	1.337	470559	525334
G10	XP_001655903.1	AAEL012170-PA	275	NBD	0	1.662	13565	20660
G11	XP_001657117.1	AAEL003703-RA	616	NBD-TMD	5	1.94	984346	994398
G12	XP_001658974.1	AAEL008138-RA	773	NBD-TMD	7	1.303	412767	432830
G13	XP_001659406.1	AAEL008672-PA	689	NBD-TMD	6	1.340	378892	469513
G14	XP_001661538.1	AAEL011265-PA	787	NBD-TMD	6	1.561	434947	481728
G15	XP_001662000.1	AAEL011863-PA	547	NBD-TMD	7	1.624	277315	291745
G16	XP_001663554.1	AAEL013372-PA	599	NBD-TMD	6	1.830	256110	301680
G17	EJY57442.1	AAEL016999-PA	692	NBD-TMD	6	1.310	837158	838440
G18	EJY57661.1	AAEL017188-PA	692	NBD-TMD	6	1.225	1495602	1528709
G19	EJY58108.1	AAEL017106-PA	686	NBD-TMD	6	1.1174	142758	145023
G20	AAC04894.1	Not listed in Vectorbase	692	NBD-TMD	6	U88851.1	-	-
H1	XP_001648758.1	AAEL014428-PA	727	NBD-TMD	6	1.1111	126773	184623
H2	XP_001650571.1	AAEL005249-PA	872	NBD-TMD	7	1.147	1132338	1176150
H3	XP_001650952.1	AAEL005491-PA	783	NBD-TMD	7	1.159	901905	920193

5.3.3 Topology and pattern of occurrence of NBD Proteins

To analyze the structural topology of *Aedes* ABC transporters, positions of the NBDs and TMDs were identified for each protein. The characteristic pattern of NBDs and TMDs decide the function of ABC transporters as a full or half transporters. Through this analysis total eight types of pattern of TMDs and NBDs were found (**Figure 5.2**).

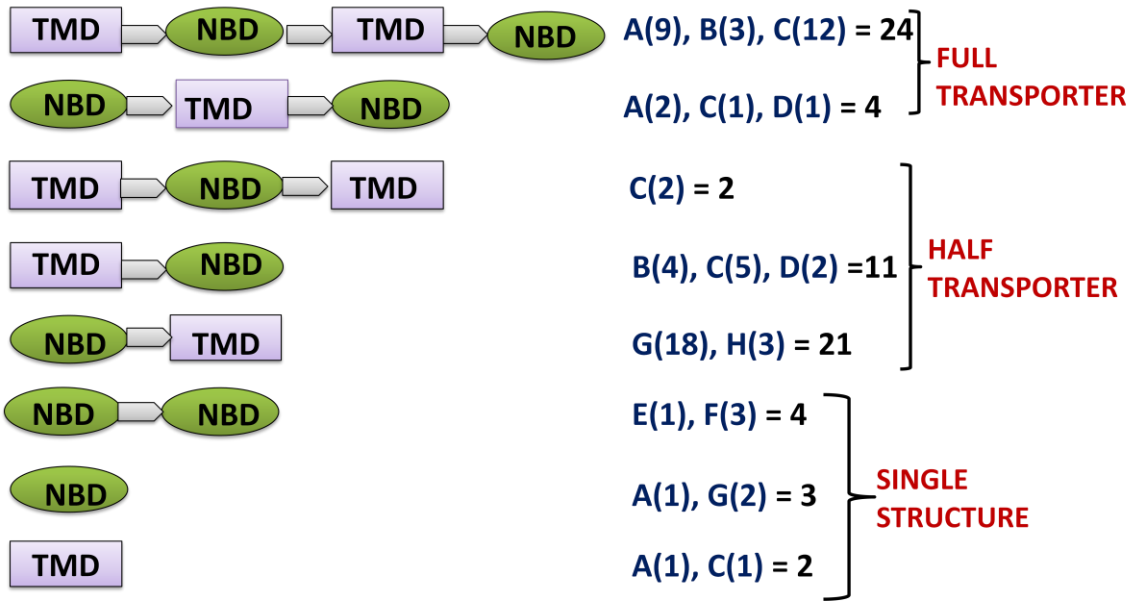


Figure 5.2: Topological distribution of domains on ABC proteins in ABC transporters inventory of *Aedes aegypti*. There are 8 types of structural patterns have found which is further revealed by their subfamily name and number of members is presented in parentheses.

Full transporter topological architecture comprises either 2 NBDs and 2 TMDs which includes 9 members from ABCA, 3 from ABCB and 12 ABCC subfamily proteins or 2 NBDs and 1 TMDs comprising 2 members of ABCA, 1 ABCC and 1 ABCD subfamily. Half transporters in *Aedes* have various configurations *viz.* **(a)** 2 TMDs and 1 NBD including 2 members of ABCC **(b)** 1 TMD and 1 NBD including 4 members from ABCB, 5 members from ABCC and 2 ABCD **(c)** 1 NBD and 1 TMD including 18 members from ABCG and 3 from ABCH **(d)** only 2 NBDs including 1of ABCE and 3 of ABCF **(e)** only 1 NBD including 1 member from ABCA and 2 ABCG **(f)** only 1 TMD having one members from ABCA which is not present in any other insect repository.

5.3.4 Phylogenetic analysis

To analyse the evolutionary position of the 71 putative *Aedes* ABC transporters, phylogenetic analysis was performed using NBDs of *Aedes* and with existing dipterans genomes fruit fly (*D. melanogaster*) and malaria mosquito (*An. gambiae*). For two reasons, separate analyses were carried out for each subfamily. First, the domain

architecture of ABC transporters is highly variable among subfamilies, which complicates bioinformatics analyses on full length sequences of the protein. Second, the NBDs of different transporters show a similar length and organization; their sequence is too conserved to provide a meaningful degree of resolution in phylogenetic analysis.

The ABCA transporters in mammals perform critical functions in the control of cellular lipid-transport processes (Kaminski et al., 2006) yet little is known about their physiological functions in insects. The ABCA subfamily proteins were well-characterized as full transporters. A total of 8, 13 and 10 ABCA genes were identified in *An. gambiae*, *Ae. aegypti*, and *Drosophila*, respectively. The *Aedes* ABCA subfamily contains thirteen members: eleven full transporters and two half-size members comprising one member with only an NBD domain (EJY57605.1) and one with only TMD (XP_001647575.1), suggesting that the ABCA subfamily has underwent rapid divergence. The speculated duplication of genes within the ABCA subfamily were observed between AeABCA1 and AeABCA13; AeABCA4 and AeABCA11; AeABCA6 and AeABCA7; AeABCA9 and AeABCA10 (**Figure 5.3**). AeABCA proteins are the largest among the 8 subfamilies, with sizes ranging from 1480 to 1673 amino acids.

The evolutionary relationships of ABCA transporters proteins with *An. gambiae* and *D. melanogaster* are shown in **Figure 5.3.** Phylogenetic analysis of the ABCA subfamily revealed a clear orthologous relationship between *D. melanogaster* CG31731 (DmABCA8) with AeABCA1 and AeABCA13; DmABCA4 with AeABCA3 proteins. Except for CG31731, reported to be down regulated in the salivary glands of an E93 mutant of *D. melanogaster* (Dutta, 2008), no information is available on the function of these insects ABCA transporters. Little is known about the function of human ABCA5, 6, 8 and 10 but several studies have suggested roles of these transporters in lipid transport (Wenzel et al., 2007). Human ABCA12 on the other hand has been thoroughly characterized and is known as a keratinocyte lipid transporter (Akiyama, 2013). The role of arthropod orthologues of these human ABCAs is unknown, but they might play a role in lipid transport. To date, only one study has investigated the function of ABC proteins in this insect-specific ABCA group. Injection of dsRNA targeting TcABCA-9A or TcABCA-9B of *T. castaneum* resulted in 30% mortality during the pupae and adult molting showed severe wing defects and elytra shortening (Broehan et al., 2013). Future studies might confirm that this group of insect-specific ABCAs is involved in development.

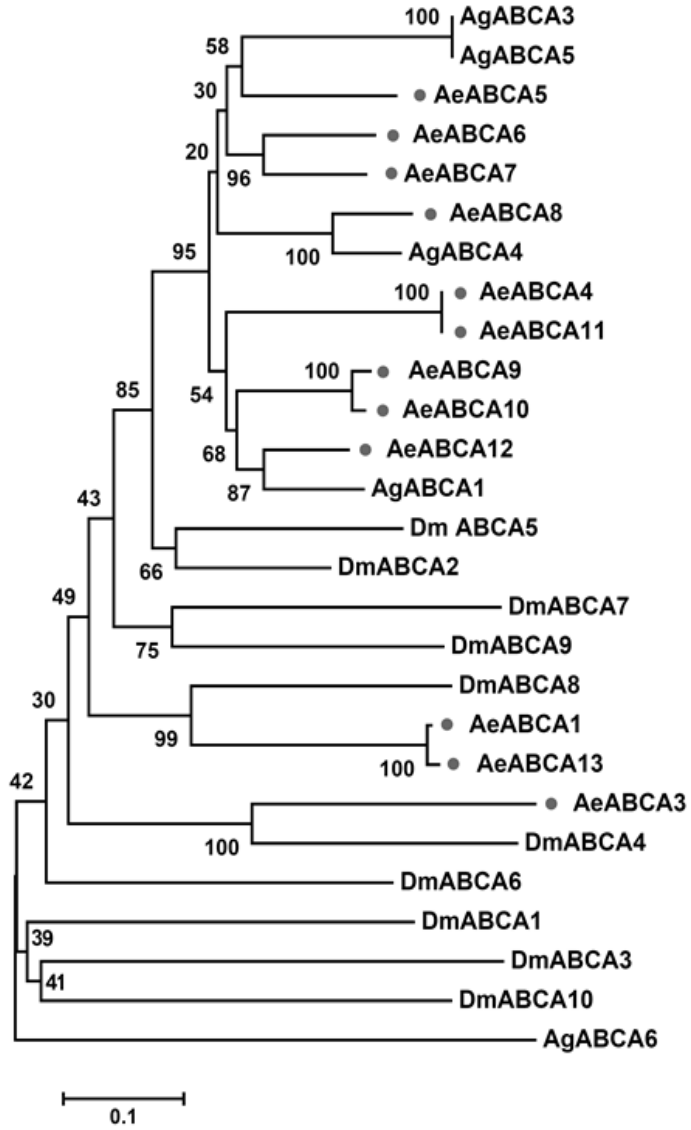


Figure 5.3 Phylogenetic tree of ABCA subfamily of proteins in three insect genomes. N-terminal NBDs sequences were used to generate a phylogenetic tree using the neighbor-joining (NJ) method. Bootstrapping (1,000 replicates) was used to determine the relative support of various branches. Dm: *Drosophila melanogaster*; Ae: *Aedes aegypti* (●); Ag: *Anopheles gambiae*.

The protein size of *Aedes* ABCB subfamily proteins ranges around 693-1419 amino acids, include both half and full transporters **(Table 5.4)**. The ABCB subfamily of *Aedes*, consists of 3 full and 4 half transporters. Unlikely to 5 ABCB members of *An. gambiae*, *Ae. aegypti* have seven members which might be due to the duplication of AeBCB5 and AeABCB6 genes. Different from *Ae aegypti* and *An. gambiae*, *D. melanogaster* has 8 ABCB proteins, three of which are called Mdr49, Mdr50, and Mdr65. Mdr49 and Mdr65 confer multidrug resistance phenotype (Wu et al., 1991; Begun and Whitley, 2000). Phylogenetic analysis of ABCB subfamily showed that the MDR transporters of *Drosophila* were clustered together with *Aedes* AeABCB5, AeABCB6 having the same

domain for drug resistance (**Figure 5.4**). Both two full ABCB transporters appeared to be most closely related to the *Drosophila* biochemical defense genes (Mdr49, Mdr50 and Mdr65). All these genes formed a closer cluster with human Pgp (MDR/PGP) subfamily, which have a wide range of functions including bile salt excretion from the liver and export of hydrophobic molecules and steroids (Dermauw and Leeuwen, 2014).

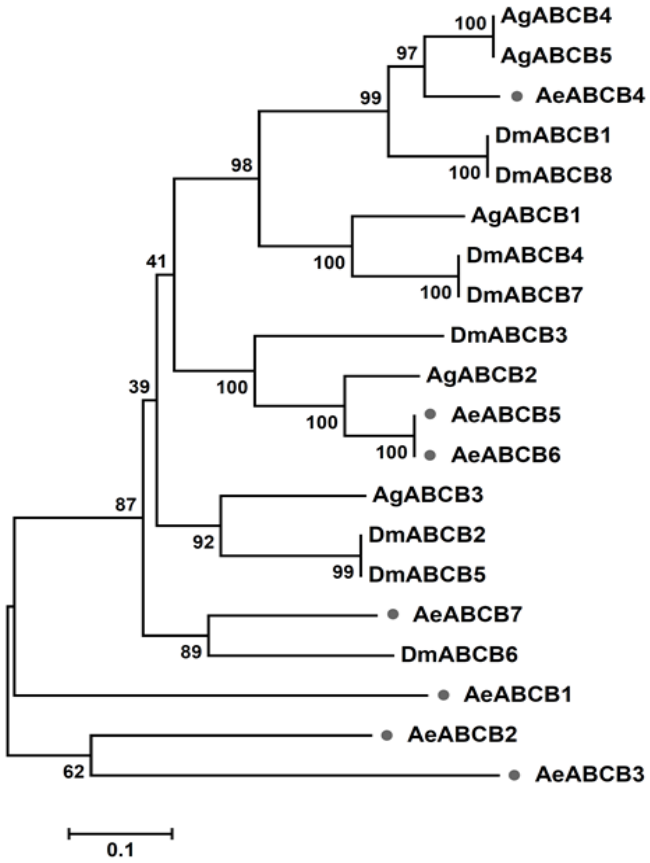


Figure 5.4 Phylogenetic tree of ABCB subfamily of proteins in three insect genomes. Dm: *Drosophila melanogaster*; Ae: *Aedes aegypti* (●); Ag: *Anopheles gambiae*.

The ABCB subfamily is involved in multidrug resistance (MDR) in human, ABCB1 known as P-glycoprotein was first identified and characterized to confer a MDR phenotype in mammalian cancer cell lines (Kartner et al., 1983). In mosquitoes, the P-glycoprotein inhibitor verapamil can increase the toxicity of cypermethrin, endosulfan and ivermectin to *Culex pipiens* 4th instar larvae but not that of chlorpyrifos. The expression of *Ae aegypti* P-glycoprotein (AeABCB4-AAEL010379) increases 8-fold in temephos-treated larvae and silencing of this gene expression significantly increases temephos toxicity (David et al., 2014; Figueira-Mansur et al., 2013). ABCB6, ABCB7, ABCB8 and ABCB10 are members of mitochondrial ABC systems, which play roles in the metabolism of iron and the transport of Fe/S protein precursors (Dean et al., 2001).

D. melanogaster ABCB6 (CG4225) known as heavy metal tolerance factor one, confers resistance against cadmium (Sooksa-Nguan et al., 2009). *T. castaneum* ABCB7 (TcABCB-5A) affects development and female reproduction (Broehan et al., 2013). *Ae. aegypti* ABCB3 (AAEL006717) is reportedly upregulated in a pyrethroid-resistant strain (Bariami et al., 2012).

In the *Aedes* genome, ABCC proteins form the largest subfamily with 21 members while 14 in *An. gambiae* and 14 in *D. melanogaster* have been reported. In human the ABCC subfamily contains 12 full transporters with a diverse functional spectrum that includes ion transport, cell-surface receptor and toxin secretion activities. *Ae. aegypti* ABCC subfamily contain 13 full transporters and 8 half transporters. The huge burst in the ABCC subfamily of *Aedes* is obvious due to high duplication rate in their genome. As a result of this, it could be speculated from the phylogenetic relationships of ABCC proteins that 6 duplication events have occurred within the subfamily. This duplication include following members: (1) AeABCC1 and AeABCC2; (2) AeABCC5 and AeABCC13; (3) AeABCC10 and AeABCC11; (4) AeABCC18 and AeABCC20; (5) AeABCC12 and AeABCC16; (6) AeABCC19 and AeABCC21. Phylogenetic analysis showed that AeABCC1, 2, 3, 4, 9, 21 grouped with *Anopheles* AgABCC1, AgABCC2, AgABCC3, and AgABCC4 have MRP1 domain **(Figure 5.5)**.

HsABCC4 proteins are organic anion transporters that mediate the cellular efflux of a wide range of exogenous and endogenous compounds, such as cyclic nucleotides and anti-cancer drugs (Schuetz et al., 1999; Borst et al., 2007). ABCC4 also functions in insecticide detoxification. For example, DDT and lindane induces the expression of zebrafish ABCC4; in addition, over expression of zebrafish ABCC4 significantly decreases the cytotoxicity and accumulation of DDT and lindane in LLC-PK1 cells and developing embryos (Lu et al., 2014). An *Ae aegypti* ABCC9 gene (AAEL005937) is upregulated in pyrethroid resistant strains (Bariami et al., 2012). In *Pediculus humanus* (bedbug), ABCC1 is highly expressed after exposure to the insecticide ivermectin exposure (Yoon et al., 2011).

The ABCD, ABCE ABCF and ABCH subfamilies in *Aedes* contained a small number of members with clear orthologous relationships. These subfamilies are comparatively conserved in insects and even higher organisms. *Aedes* ABCD transporters contain 3 members, two half transporters and one half transporters have orientation TMD-NBD and NBD-TMD-NBD.

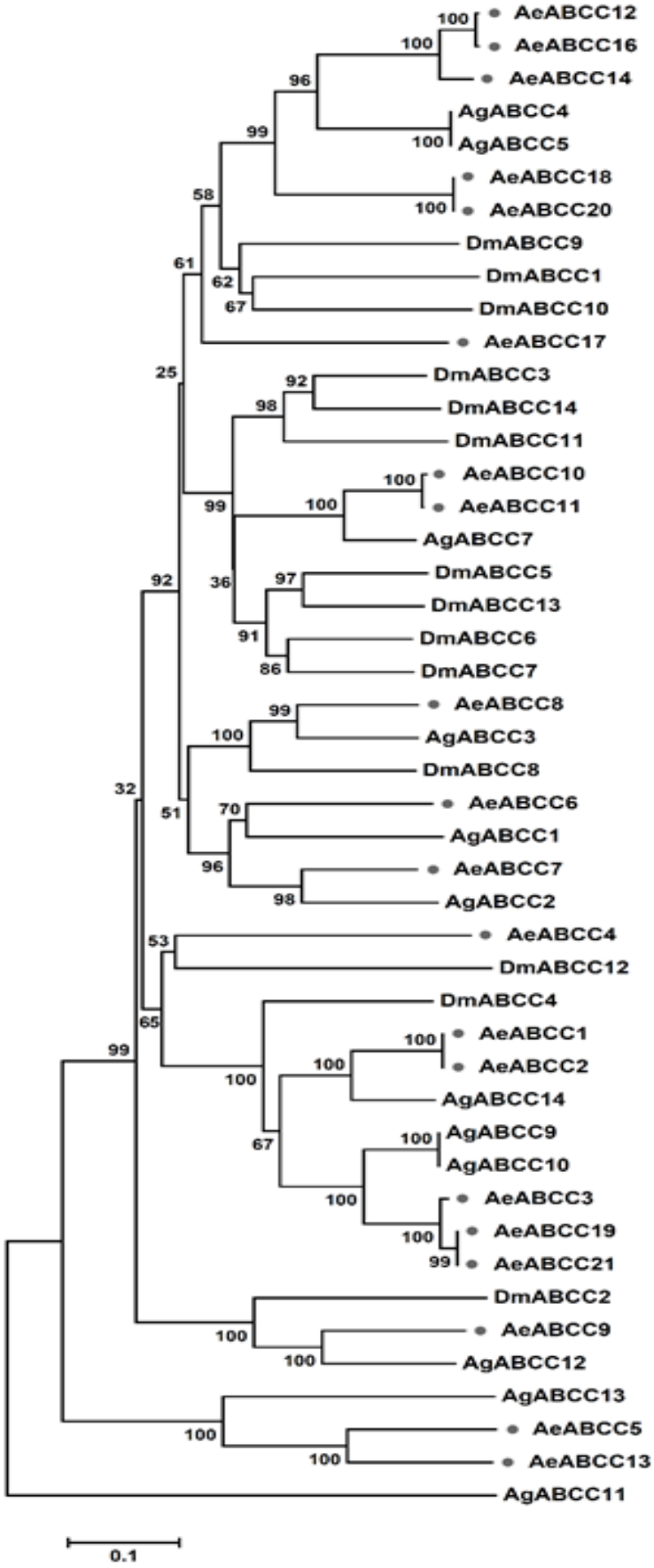


Figure 5.5 Phylogenetic tree of ABCC subfamily of proteins in three insect genomes. Dm: *Drosophila melanogaster*; Ae: *Aedes aegypti* (●); Ag: *Anopheles gambiae*

In *Tetranychus urticae* have two ABCD proteins, which are orthologous to human ABCD1/2 and ABCD3 (Dermauw and Leeuwen, 2014). In yeast and mammals, ABCD transporters are located in the peroxisome membrane and involved in transporting acyl coenzyme A esters across (fatty acid transport) their membrane (Wanders et al., 2007; Morita and Imanaka, 2012). However, the role of ABCDs has not been studied in insects.

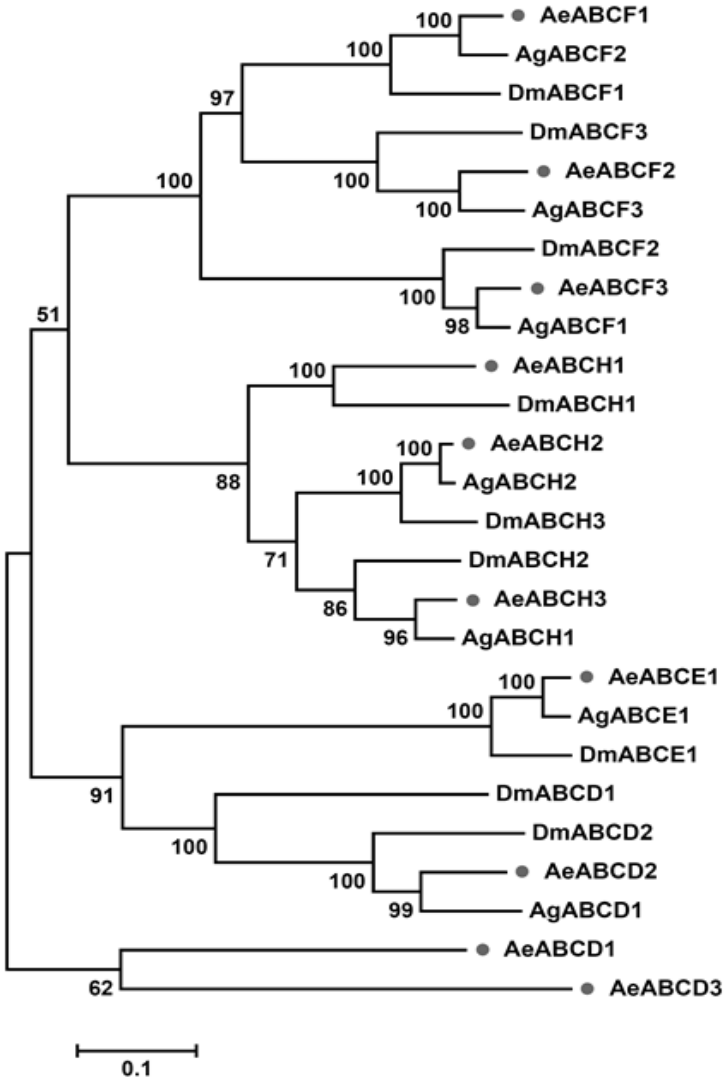


Figure 5.6 phylogenetic tree of ABCD, ABCE, ABCF and ABCH subfamilies of proteins in three insect genomes. Dm: *Drosophila melanogaster*; Ae: *Aedes aegypti* (●); Ag: *Anopheles gambiae*

The ABCE and ABCF subfamilies are involved in biological processes rather than transportation because of lacking transmembrane domains (Dean et al., 2001). The proteins of subfamilies E and F are characterized by two NBD domains and they are unlikely to function as transmembrane transporters. Most of the higher organisms only have one ABCE member in this smallest subfamily that is highly conserved during evolution (Dermauw and Leeuwen, 2014). **Figure 5.6** reveals the close relationship of

AeABCE1 with AgABCE1 and DmABCE1 with high bootstrap value. In human ABCE1 was identified as an RNAse L inhibitor which is an important component of interferon regulated 2-5A pathway (Bisbal et al., 1995). The *Drosophila* ABCE1 is called as "pixie" which is essential for translation initiation through binding of 40S ribosomal subunit in an ATP dependent manner (Andersen and Leevers, 2007).

Each of the insect species mentioned in **table 5.2** has three ABCF proteins except *Daphnia pulex*, which belong to ABCF1, ABCF2 and ABCF3. By phylogenetic analysis of ABCF transporters, AeABCF1 and AeABCF3 were found orthologous with human ABCF1 and HsABCF3 (61% and 63% amino acid identity). In addition, AeABCF2 is a homolog of *Drosophila* CG9281 (DmABCF3), but its function is not clear (**Figure 5.6**). Each ABCF protein has two NBD domains and no TM domain. In mammals, ABCF1 (ABC50) plays a key role in the control of translation initiation (Paytubi et al., 2009). In *T. castaneum*, interference of ABCE1 (TcABCE-3A) or ABCF1 (TcABCF-2A) expression produces the lethal phenotype in larvae (Broehan et al., 2013).

The ABCH subfamily was first identified in *D. melanogaster* and later found to exist in other insects and zebrafish (Popovic et al., 2010). *D. melanogaster* and *Ae. aegypti* have three ABCH genes, whereas *An. gambiae* has two ABCH genes, which formed three paralogous groups (**Figure 5.6**). The function of ABCH is poorly understood, although it is closely related to the ABCG subfamily. A recent study has shown that RNAi knockdown of the *T. castaneum* TcABCH9C gene (homologous to *An. gambiae* AGAP003680) and TcABCG-4C gene (homologous to *An. gambiae* AGAP009850) causes similar phenotypes: arrested development, rapid death during the quiescent stage, rough cuticles and reduced reproduction (Broehan et al., 2013). This evidence suggests ABCH genes function in transport of cuticular lipids and is deposited in epicuticular layer to prevent water loss.

The *Aedes* ABCG proteins are the second largest subfamily after ABCC, with 20 members, which were half ABC transporters with a reverse domain organization (NBD-TMD). Two ABCG proteins (AeABCG1: AAEL008627-PA and AeABCG9: AAEL012170-PA) has only a single NBD domain. The size of ABCG protein is ranging from 547-787 amino acids excluding 2 above mentioned protein have only NBD domain. This genomic architecture supports the duplication of gene and explodes in number within the subfamily. These duplicated genes include AeABCG1 and AeABCG2; AeABCG4 and AeABCG5; AeABCG15 and AeABCG16; AeABCG17 and AeABCG20.

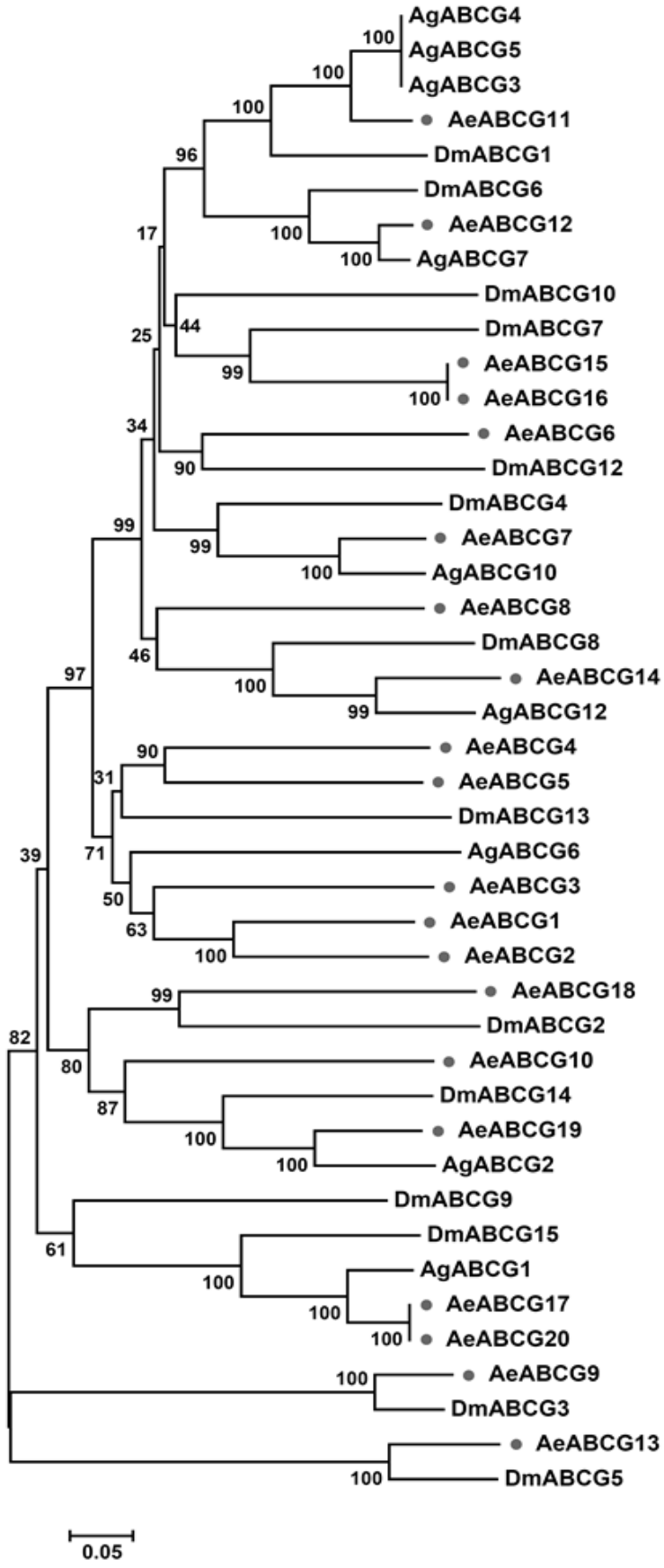


Figure 5.7 Phylogenetic tree of ABCG subfamily of proteins in three insect genomes. Dm: *Drosophila melanogaster*; Ae: *Aedes aegypti* (●); Ag: *Anopheles gambiae*

These mosquito ABCG genes were homologous to human ABCG1, ABCG4, ABCG5, and ABCG8, which are primarily involved in the transport of endogenous and dietary lipids in mammals (Dean et al., 2001). The functions of ABCG1, -G4, -G5, and -G8 genes in insects are unknown. In *Ae aegypti*, two ABCG genes (AAEL008138-AeABCG12, AAEL008624-AeABCG2) were upregulated in pyrethroid-resistant strains (Bariami et al., 2012). In humans, ABCG2 is a multidrug resistance protein that is over expressed in cell lines cultured under mitoxantrone selection pressure (Miyake et al., 1999). However, ABCG2 homologs were not found in mosquitoes. The three best-studied insect ABCG subfamilies, i.e., white, scarlet, and brown, were present in mosquito species. Each mosquito had one orthologous white or brown gene and two orthologous scarlet genes. Phylogenetic analysis revealed that AeABCG1-6 had clear orthologous relationships with *Drosophila* (**Figure 5.7**). In *D. melanogaster*, white, brown and scarlet genes are involved in pigment transport in eyes by transporting guanine or tryptophan, the precursors of the red and brown eye pigments (Mackenzie et al., 1999; Borycz et al., 2008).

5.4 Discussion

The ABC families of five insect species, along with one crustacean (*Daphnia pulex*) and a chelicerate (*T. urticae*) have been previously annotated in some details. We have identified 71 unique ABC transporters in *Ae aegypti* genome through homology searches which was found comparatively higher in order Diptera **(Table 5.2)**. Comparative analyses of sequenced insect genomes including: fruitfly (*D. melanogaster*), malaria mosquito (*An. gambiae*), a lepidopteron silkworm (*B. mori*), a coleopteran beetle (*T. castaneum*) and water-flea (*D. pulex*) show that these species also contain 50–80 ABC genes with almost the same basic subfamily organization. Among insects, *Ae aegypti*, 71 ABC genes form the largest ABC data set after *T. castaneum* according to the available reports. However, the percentage of ABC genes in total genes is almost similar in *Aedes* and *Tribolium* (**Table 5.1**). This huge number might be evident because *Aedes* genome is known to have 50% gene duplication event which was also observed in voltage-gated sodium channel genes. Nevertheless, a huge portion of mitochondrial genome of *Aedes aegypti* is integrated to the nuclear genome several million years ago (Severson and Behura, 2012; Martins et al., 2013). Similar results were obtained by Harland et al., (2005) when comparing bacteria with different respiratory requirements since aerobic

bacteria have larger genomes and display greater numbers of ABC systems than anaerobes. These finding support that number of ABC genes encoded in bacterial genomes correlates with genome size and also with physiological niche in which bacteria live (Harland et al., 2005).

In vertebrates, a remarkable decrease in the number of ABC genes was observed: for instance, only 48 ABC genes have been identified in humans (Dean and Annilo, 2005). Unlikely to human, insects have 8 subfamilies of ABC transporters (A-H). A novel subgroup ABCH was identified in *Drosophila* ABCs (Dean et al., 2001; Vasiliou et al., 2009). Although the number of ABC members in different subgroups is generally similar. *Drosophila* has three times more ABCG-type proteins than human (Dean et al., 2001). It may indicate the expansion and deletion of specific ABC transport during evolution and may indicate differential adaptation for individuals. The identification and domain analysis of ABC transporters in *Ae aegypti* genome has revealed interesting traits of these proteins in this medically important mosquito species. The high conservation in *Aedes* ABC proteins involved in fundamental cellular processes (such as the mitochondrial ABCB half transporters and the members of ABCD, ABCE and ABCF subfamilies) suggesting an evolutionary ancestral origin of these proteins (Dermauw and Leeuwen, 2014). The complement of ABC proteins shows parallel between two insects *Anopheles* and *Drosophila* in that both lacking homologues to the human proteins TAP (translocations involved in antigen processing) and CFTR (chloride channel regulated by ATP) (Abele and Tampe, 2004). Moreover, the importance of these transporters increases due to their involvement in transport of xenobiotics, chemotherapeutic drugs and also in MDR and immunity (Labbe et al., 2011). The first characterized eukaryotic ABC transporter (P-glycoprotein) which is present on the cancer cell surface mediates MDR through efflux and prevents the accumulation of effective chemotherapeutic drugs (Kartner et al., 1983; Van-Veen et al., 1998). These facts provide evidences that the bacterial and human transporters are functionally compatible and conserved from bacteria to human.

The members of E and F subfamilies are not transporters and diverge from other subfamilies proteins since they possess two NBDs but lack TMDs. In *C. elegans* ABCE proteins are inhibitors of RNAse L and involved in the assembly of preinitiation complex while ABCF proteins have roles in ribosome assembly and protein translation (Zhao et al., 2004; Dong et al., 2004; Bisbal et al., 1995; Marton et al., 1997; Tyzack et al., 2000).

The comparatively conserved nature of NBD domains and their phylogenetic relationship is employed to categorize ABC transporters into 8 subfamilies (subfamily A-H). The uniformity of the ABCs, NBD domains between diverse groups of organisms is helpful to understand their phylogenetic relationship. This statement can be supported with examples where white eye pigments protein in *Drosophila* and mosquito are 65% identical; both of them are near about 35% identical to human white protein (Waterhouse et al., 2008).

5.5 Conclusion

Mosquitoes, being the pathogens carriers, have also been aimed exhaustively to understand the selected features of their ABC transporters for designing tools to control the spread of diseases. ABC transporters in insects are known to regulate the transport of metabolic components, pigments, chemoattractant, uric acid metabolism and development. Moreover, ABCs-mediated insecticide resistance and tolerance to xenobiotics also increases their importance (Bates et al., 2000). Taken together, our analyses indicate that *Ae*. *aegypti* ABC transporters family members are conserved and may play multiple roles. Thorough phylogenetic analysis of three dipteran ABC protein subfamilies, infer homologous relationships which suggest its conserved function. The majority of ABC proteins function as primary transporters that bind and hydrolyze ATP while transporting a large diversity of substrates across lipid membranes. The diverse functions of ABC proteins has been well studied in insects for their role in drug resistance but, less is known about its role in the transport of endogenous and exogenous substances in arthropods.

5.6 References

Abele R, Tampe R. 2004. The ABCs of immunology: structure and function of TAP, the transporter associated with antigen processing. Physiology (Bethesda) 19:216-224.

Akiyama M. 2014. The roles of ABCA12 in epidermal lipid barrier formation and keratinocyte differentiation. Biochimica et Biophysica Acta-Molecular and Cell Biology of Lipids. 1841(3):435-440.

Andersen DS, Leevers SJ. 2007. The essential Drosophila ATP-binding cassette domain protein, pixie, binds the 40S ribosome in an ATP-dependent manner and is required for translation initiation. J Biol Chem. 282:14752-14760.

Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, et al., 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science 330:86-88.

Bariami V, Jones CM, Poupardin R, Vontas J, Ranson H. 2012. Gene amplification, ABC transporters and cytochrome P450s: unraveling the molecular basis of pyrethroid resistance in the dengue vector, *Aedes aegypti*. PLoS Negl Trop Dis. 6: e1692.

Bates SE, Robey R, Knutsen T, Honjo Y, Litman T, Dean M. 2000. New ABC transporters in multi-drug resistance. Expert Opin Ther Targets 4(5):561-580.

Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B. 2010. A structural classification of substrate-binding proteins. FEBS Lett. 584:2606-2617.

Begun DJ, Whitley P. 2000. Genetics of alpha-amanitin resistance in a natural population of *Drosophila melanogaster*. Heredity 85 (Pt 2):184-190.

Bianchet MA, Ko YH, Amzel M, Pedersen PL. 1997. Modeling of nucleotide binding domains of ABC transporter proteins based on a F1-ATPase/recA topology: structural model of the nucleotide binding domains of the cystic fibrosis transmembrane conductance regulator (CFTR). J Bioenerg Biomembr. 29(5):503-524.

Bisbal C, Martinand C, Silhol M, Lebleu B, Salehzada T. 1995. Cloning and characterization of a RNAse L inhibitor. A new component of the interferon-regulated 2-5A pathway J Biol Chem. 270:13308-13317.

Borst P, de Wolf C, Van de Wetering K. 2007. Multidrug resistance-associated proteins 3, 4 and 5. Pflügers Archiv-European Journal of Physiology. 453(5):661-673.

Borycz J, Borycz JA, Kubow A, Lloyd V, Meinertzhagen IA. 2008. *Drosophila* ABC transporter mutants white, brown and scarlet have altered contents and distribution of biogenic amines in the brain. J Exp Biol. 211: 3454-3466.

Bonizzoni M, Afrane Y, Dunn WA, Atieli FK, Zhou G, Zhong D, et al., 2012. Comparative transcriptome analyses of deltamethrin-resistant and -susceptible *Anopheles gambiae* mosquitoes from Kenya by RNA-Seq. PLoS One 7: e44607.

Broehan G, Kroeger T, Lorenzen M, Merzendorfer H. 2013. Functional analysis of the ATP-binding cassette (ABC) transporter gene family of *Tribolium castaneum*. BMC Genomics 14:6.

Dassa E, Bouige P. 2001. The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. Res Microbiol. 152:211-229.

David JP, Faucon F, Chandor-Proust A, Poupardin R, Riaz MA, Bonin A, et al., 2014. Comparative analysis of response to selection with three insecticides in the dengue mosquito *Aedes aegypti* using mRNA sequencing. BMC Genomics 15:174.

Dawson RJ, Locher KP. 2006. Structure of a bacterial multidrug ABC transporter. Nature 443:180-185.

Dean M, Hamon Y, Chimini G. 2001. The human ATP-binding cassette (ABC) transporter superfamily. J Lipid Res. 42:1007-1017.

Dean M, Rzhetsky A, Allikmets R. 2001. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res. 11:1156-1166.

Dean M, Annilo T. 2005. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet. 6:123-142.

Dermauw W, Osborne EJ, Clark RM, Grbic M, Tirry L, Leeuwen TV. 2013. A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*. BMC Genomics 14(317):1-21.

Dermauw W, Leeuwen TV. 2014. The ABC gene family in arthropods: Comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol. 45:89e110.

Dong J, Lai R, Nielsen K, Fekete CA, Qiu H Hinnebusch AG. 2004. The essential ATP-binding cassette protein RLI1 functions in translation by promoting preinitiation complex assembly. J Biol Chem. 279:42157-42168.

Dutta S. 2008. Genetic regulation of autophagic cell death in *Drosophila melanogaster*.

Epis S, Porretta D, Mastrantonio V, Comandatore F, Sassera D, Rossi P, et al., 2014. ABC transporters are involved in defense against permethrin insecticide in the malaria vector *Anopheles stephensi*. Parasit Vectors 7:349.

Figueira-Mansur J, Ferreira-Pereira A, Mansur JF, Franco TA, Alvarenga ES, Sorgine MH, et al., 2013. Silencing of P-glycoprotein increases mortality in temephos-treated *Aedes aegypti* larvae. Insect Mol Biol. 22:648-658.

Finn RD, Tate J, Mistry J, Coggill PC, Sammut JS, Hotz HR, et al., 2008. The Pfam protein families' database. Nucl Acids Res. 36(Database Issue): D281-D288.

Gottesman MM, Ambudkar SV. 2001. Overview: ABC transporters and human disease. J Bioenerg Biomembr. 33:453-458.

Harland DN, Garmory HS, Brown KA, Titball RW. 2005. An association between ATP binding cassette systems, genome sizes and lifestyles of bacteria. Res Microbiol. 156:434-442.

Higgins CF. 1992. ABC transporters: from microorganisms to man. Annu Rev Cell Biol. 8:67-113.

Holland B, Blight MA. 1999. ABC-ATPases, adaptable energy generators fueling transmembrane movement of a variety of molecule in organism from bacteria to human. J Mol Biol. 293:381-399.

Hollenstein K, Dawson RJP, Locher KP. 2007. Structure and mechanism of ABC transporter proteins. Curr Opin Struc Biol. 17:412-418.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science. 298:129-149.

Jones PM, George AM. 2004. The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci. 61:682-699.

Kaminski WE, Piehler A, Wenzel JJ. 2006. ABC A-subfamily transporters: structure, function and disease Biochim Biophys Acta. 1762:510-524.

Kartner N, Riordan JR, Ling V. 1983. Cell surface P-glycoprotein associated with multidrug resistance in mammalian cell lines. Science 221:1285-1288.

Kennedy C, Tierney K. 2013. Xenobiotic protection/resistance mechanisms in organisms. In: Laws, E.A. (Ed.), Environ. Toxicol. Springer, New York, pp. 689e721.

Klein I, Sarkadi B, Váradi A. 1999. An inventory of the human ABC proteins. Biochem Biophys Acta. 1461:237-262.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305(3):567-580.

Labbe, R., Caveney, S, Donly, C. 2011. Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera. Insect Mol Biol. 20:243-256.

Liu S, Zhou S, Tian L, Guo E, Luan Y, Zhang J, Li S. 2011. Genome-wide identification and characterization of ATP-binding cassette transporters in the silkworm, *Bombyx mori*. BMC Genomics 12:491.

Lu X, Long Y, Lin L, Sun RZ, Zhong S, Cui ZB. 2014. Characterization of zebrafish abcc4 as an efflux transporter of organochlorine pesticides. PLoS One 9.

Mackenzie SM, Brooker MR, Gill TR, Cox GB, Howells AJ, Ewart GD, 1999. Mutations in the white gene of *Drosophila melanogaster* affecting ABC transporters that determine eye coloration. Biochim Biophys Acta 1419:173-185.

Marchler-Bauer A, Bryant SH. 2004. CD Search: protein domain annotations on the fly. Nucl Acids Res. 32 (Web Server issue): W327–W331.

Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, et al., 2007, CDD: a conserved domain database for interactive domain family analysis. Nucl Acids Res. 35 (Database Issue): D237-D240.

Martins AJ, Brito LP, Linss JGB, Rivas GBS, Machado R, Bruno RV, et al., 2013. Evidence for gene duplication in the voltage-gated sodium channel gene of *Aedes aegypti*. Evol Med & Public Health. 148-160.

Marton MJ, Vazquez de Aldana CR, Qiu H, Chakraburtty K, Hinnebusch AG. 1997. Evidence that GCN1 and GCN20, translational regulators of GCN4, function on elongating ribosome in activation of eIF2alpha kinase GCN2. Mol Cell Biol. 17:4474-4489.

Misra JR, Horner MA, Lam G, Thummel CS. 2011. Transcriptional regulation of xenobiotics detoxification in *Drosophila*. Genes Dev. 25: 1796e1806.

Miyake K, Mickley L, Litman T, Zhan ZR, Robey R, Cristensen B, et al., 1999. Molecular cloning of cDNAs which are highly over expressed in mitoxantrone-resistant cells: Demonstration of homology to ABC transport genes. Cancer Res. 59:8–13.

Morita M, Imanaka T, 2012. Peroxisomal ABC transporters: structure, function and role in disease. Biochim Biophys Acta 1822:1387-1396.

Nene V., Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al., 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316:1718-1723.

Paytubi S, Wang XM, Lam YW, Izquierdo L, Hunter MJ, Jan E, et al., 2009. ABC50 promotes translation initiation in mammalian cells, J. Biol. Chem. 284:24061-24073.

Popovic M, Zaja R, Loncar J, Smital T. 2010. A novel ABC transporter: the first insight into zebrafish (Danio rerio) ABCH1, Mar Environ Res. 69 (Suppl.):S11-S13.

Roth CW, Holm I, Graille M, Dehoux P, Rzhetsky A, Wincker P, et al., 2003. Identification of the *Anopheles gambiae* ATP-binding cassette transporter superfamily genes. Mol Cells 15(2):150-158.

Saitou N, Nei M. 1997. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406-425.

Saurin W, Hofnung M. & Dassa E. 1999. Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. J Mol Evol. 48(1):22-41.

Schneider E, Hunke S. 1998. ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. FEMS Microbiol Rev. 22(1):1-20.

Schuetz JD, Connelly MC, Sun D, Paibir SG, Flynn PM, Srinivas RV, et al., 1999. MRP4: a previously unidentified factor in resistance to nucleoside-based antiviral drugs. Nat Med. 5:1048-1051.

Schuller C, Bauer BE, Kuchler K: Inventory and evolution of fungal ABC protein genes. In ABC Proteins: From Bacteria to Man. Edited by Holland IB, Cole SPC, Kuchler K, Higgins CF. Elsevier; 2003:279-293.

Severson DW, Behura SK. 2012. Mosquito genomics: progress and challenges. Annu Rev Entomol. 57:143-166.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:539

Smith CA, Rayment I. 1996. Active site comparisons highlight structural similarities between myosin and other P-loop proteins Biophys J. 70(4):1590-1602.

Sooksa-Nguan T, Yakubov B, Kozlovskyy VI, Barkume CM, Howe KJ, Thannhauser TW, et al., 2009. *Drosophila* ABC transporter, DmHMT-1, confers tolerance to cadmium. DmHMT-1 and its yeast homolog, SpHMT-1 are not essential for vacuolar phytochelatin sequestration. J Biol Chem. 284:354-362.

Srivastava H, Sharma M, Dixit J, Das A. 2010. Evolutionary insights into insecticide resistance gene families of *Anopheles gambiae*. Infect Genet Evol. 10:620-628.

Sturm AP, Cunningham P, Dean M. 2009. The ABC transporter gene family of *Daphnia pulex*. BMC Genomics 10:170.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731-2739.

Tyzack JK, Wang X, Belsham GJ, Proud CG. 2000. ABC50 interacts with eukaryotic initiation factor 2 and associates with the ribosome in an ATP-dependent manner. J Biol Chem. 275:34131-34139.

Vasiliou V, Vasiliou K, Nebert DW. 2009. Human ATP-binding cassette (ABC) transporter family. Hum Genet. 3(3):281-290.

Van-Veen HW, Callaghan R, Soceneantu L, Sardini A, Konings WN, Higgins CF. 1998, A bacterial antibiotic resistance gene that complements the human multidrug-resistance P-glycoprotein gene. Nature 391:291-295.

Walker JE, Saraste M, Runswick MJ, Gay NJ. 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1(8):945-951.

Wanders RJ, Visser WF, van Roermund CW, Kemp S, Waterham HR. 2007. The peroxisomal ABC transporter family. Pflugers Arch. 453:719-734.

Waterhouse RM, Wyder S, Zdobnov EM. 2008. The *Aedes aegypti* genome: a comparative perspective. Insect Mol Biol. 17(1):1-8.

Wenzel JJ, Piehler A, Kaminski WE. 2007. ABC A-subclass proteins: gatekeepers of cellular phospho- and sphingolipid transport. Front Biosci. 12:3177-3193.

Wu CT, Budding M, Griffin MS, Croop JM. 1991. Isolation and characterization of *Drosophila* multidrug resistance gene homologs. Mol Cell Biol. 11:3940-3948.

Wilkinson AJ, Verschueren KHG. 2003. Crystal structures of periplasmic solute binding proteins in abc transport complexes illuminate their function. In ABC Proteins: From

Bacteria to Man. Edited by Holland IB, Cole SPC, Kuchler K, Higgins CF. London: Academic Press 187-208.

Xie X, Duan J, Cheng T, Niu W, Wang G, Xia Q. 2012. Genome-wide analysis of the ATP-binding cassette (ABC) transporter gene family in the silkworm, *Bombyx mori*. Mol Biol Rep. 39:7281-7291.

Yoon KS, Strycharz JP, Baek JH, Sun W, Kim JH, Kang JS, et al., 2011. Brief exposures of human body lice to sublethal amounts of ivermectin over-transcribes detoxification genes involved in tolerance. Insect Mol Biol. 20:687-699.

Zhao Z, Fang LL, Johnsen R, Baillie DL. 2004. ATP-binding cassette protein E is involved in gene transcription and translation in *Caenorhabditis elegans*. Biochem Biophys Res Commun. 323:104-111.

Future Prospectives

The scope of this thesis was to identify the important immune genes which are playing a crucial role in mosquito immune system, against pathogens. In this thesis comprehensive bioinformatics analysis for STAT pathway and the interaction between *An. stephensi* mosquito and *Plasmodium* parasite was analyzed. Beside this a complete inventory of *Ae. aegypti* ABC transporters were also envisaged. A further research will include the functional elucidation of following future perspectives:

- If STAT-A and STAT-B are playing a role in same canonical pathway then mechanistically how these STATs are working in the same cascade is a matter of high interest.
- The occurrence of two STATs only in *Anopheles* species of mosquitoes (especially in few series of the subgenus) is due to high rate of host pathogen interaction which needs to be checked in other species whose genome is not available.
- Whether STAT-A and STAT-B are also playing a non-canonical mode of action which is not yet reported in any insect species. As well as other downstream and effector gene of the STAT pathway are further needs to be characterized.
- Characterization of important ABC subfamily members against blood-borne pathogens needs to be identify in the laboratory colonized *Ae. aegypti* mosquitoes. Important candidate genes from each subfamily further needs to be elucidated for their role in insecticide resistance or blood-borne pathogens.

Overall Publications

1. **Kuldeep Gupta**, Rini Dhawan, Mithilesh Kajla, Tripti Misra, Sanjeev Kumar and Lalita Gupta. 2016. The divergent evolution of signal transducer and activator of transcription factors in *Anopheles* mosquitoes. **Genes (In press)**

2. **Kuldeep Gupta[#],** Rini Dhawan[#], Mithilesh Kajla, Sanjeev Kumar, B. Jnanasiddhy, Naveen K Singh, Rajnikant Dixit, Ashish Bihani, Lalita Gupta. 2016. Molecular identification of *Aedes aegypti* mosquitoes from Pilani region of Rajasthan, India. **J Vector Borne Dis.** (**In press**) (**# Joint first authors**)

3. Rini Dhawan[#], **Kuldeep Gupta[#]**, Mithilesh Kajla, Sanjeev Kumar, Surendra K. Gakhar, Parik Kakani, Tania Pal Choudhury, Lalita Gupta. 2015. Molecular characterization of SOCS gene and its expression analysis on *Plasmodium berghei* infection in *Anopheles culicifacies*. **Acta Tropica** 152:170–175. (**# Joint first authors**)

4. Mithlesh Kajla, Parik Kakani, Tania P Choudhury, **Kuldeep Gupta,** Lalita Gupta, Sanjeev Kumar. 2016. Characterization and expression analysis of gene encoding heme peroxidase HPX15 in major Indian malaria vector *Anopheles stephensi* (Diptera: culicidae) **Acta Tropica** 158: 107–116.

5. Mithilesh Kajla, Kurchi Bhattacharya, **Kuldeep Gupta**, Ujjwal Banerjee, Parik Kakani, Lalita Gupta, Sanjeev Kumar, 2015. Identification of the temperature induced larvicidal efficacy of *Agave angustifolia* against *Aedes*, *Culex* and *Anopheles* larvae. **Frontier in Public Health** 3:286. doi: 10.3389/fpubh.2015.00286

6. Mithilesh Kajla, **Kuldeep Gupta**, Parik Kakani, Rini Dhawan, Tania Pal Choudhury, Lalita Gupta, Surendra K. Gakhar, Sanjeev Kumar. 2015, Identification of an *Anopheles* lineage-specific unique heme peroxidase HPX15: A plausible candidate for arresting malaria parasite development. **Phylogenetics & Evolutionary Biology**. 3(4) 160.doi:10.4172/23299002.1000160

7. Mithilesh Kajla, **Kuldeep Gupta**, Lalita Gupta, Sanjeev Kumar. 2015, A fine-tuned management between physiology and immunity maintains the gut microbiota in insects. **Biochemistry and Physiology**. 4(4):182. doi: 10.4172/2168-9652.1000182.

8. Mithilesh Kajla, Parik Kakani, Tania Pal Choudhury, **Kuldeep Gupta**, Rini Dhawan, Surendra K. Gakhar, Lalita Gupta, Sanjeev Kumar. 2014. Characterization of Anopheline unique peroxidase and its role in the regulation of *Plasmodium* development. **Malaria Journal**, 13(Suppl 1): P49.

Research Articles under Review

1. **Kuldeep Gupta**, Rini Dhawan, Mithilesh Kajla, Sanjeev Kumar and Lalita Gupta. Comparative analysis of Anopheline SOCS5 and expression analysis upon *Plasmodium* infection in *Anopheles stephensi.*

Brief Bio-data of Supervisor

Prof. Gupta received the Ph.D. degree in Molecular Biology from the Banaras Hindu University in 1999. Prof. Lalita Gupta has extensive research experience of diverse research area of contemporary biological science world. She started her post-doc career at Colorado State University, Fort Collins, Colorado, USA and later moved to NIAID, NIH, USA as a research scientist with Prof. Carolina Barillas-Mury. She was working on molecular interaction of mosquitoes with *Plasmodium* or Dengue virus infection and investigating the effect of mosquito immune system on malaria disease transmission. During this period she earned several research papers of international reputation including Science, PNAS and Cell host and Microbes. She also received several awards including best performance award from NIH.

She Joined BITS-Pilani in 2008 as an Assistant Professor of Biological Sciences group and promoted to Associate Professor of Department of Biological Sciences in 2013. In BITS-Pilani she is involved in interactive teaching of first degree (M.Sc. Hons.), higher degree (M. E. Biotech) and Master of Public Health courses. She acquires the designation of convenor of departmental research committee and senate member of institutional research board and proctorial committee. Till yet, she co-supervised one PhD student from Shankar Nethralaya in 2015 and several higher degree students for their research projects. Prof. Gupta established a lab in the department with a huge amount of funds from various national funding agencies. She has successfully completed two of her research projects from ABG and DST-SERB as a principal investigator. Currently she has a major grant from ICMR, DBT and DST.

Distinguished Publications:

1. **Gupta L**, Noh JY, Jo YH, Oh SH, Kumar S, Noh MY, Lee YS, Cha SJ, Seo SJ, Kim I, Han YS, Barillas-Mury C. Apolipophorin-III mediates antiplasmodial epithelial responses in Anopheles gambiae (G3) mosquitoes. **PLoS One. 2010,** 5(11):e15410
2. Kumar S, Molina-Cruz A, **Gupta L**, Rodrigues J, Barillas-Mury C A peroxidase/dual oxidase system modulates midgut epithelial immunity in Anopheles gambiae. **Science 2010,** 327(5973): 1644-1648.
3. **Gupta L**, Molina-Cruz A, Kumar S, Rodrigues J, Dixit R, Zamora R, Barillas-Mury C. The STAT Pathway Mediates Late-Phase Immunity against Plasmodium in the Mosquito *Anopheles gambiae*. **Cell Host & Microbe**, **2009,** 5(5):498-507.
4. **Gupta L**, Kumar S, Han YS, Pimenta PF, Barillas-Mury C. Midgut epithelial responses of different mosquito-Plasmodium combinations: the actin cone zipper repair mechanism in *Aedes aegypti*. **Proc Natl Acad Sci USA**. **2005,** 102(11):4010-4015.

Brief Bio-data of Student

Mr. Gupta completed his Bachelors and Master degree in Zoology from CSJM Kanpur University. He has achieved the distinctions in genetics and molecular cytogenetics course in his M.Sc. degree as a specialization. Later he did the Advanced Post diploma in Bioinformatics from University Institute of Technology and Science (UIET), Kanpur University. He has qualified the GATE-2007 examination and earned pre-doctoral degree, MPhil in Zoology from Annamalai University. He had worked with Prof. D.N. Rao as a research assistant in Biochemistry department at the All India Institute of Medical Science (AIIMS), New Delhi. Mr. Gupta has joined the BITS-Pilani, Biological Sciences group as a PhD student under the mentor-ship of Dr. Suman Kapur. He was awarded junior research fellowship from University Grant Commission (UGC) since 2010 and senior research fellow since 2012. Due to the delocalization of the mentor from BITS-Pilani campus to Hyderabad campus he has changed his supervisor and started working under the guidance of Dr. Lalita Gupta since 2011. Mr. Gupta has presented 3 research papers in national conferences and recipient of two times best poster awards. He has published 8 research papers in year 2015-2016 and two as first author during his PhD tenure.