The Role of *Anopheles* Heme Peroxidases in Mosquito Immunity against Blood-Borne Antigens and Divergent Expression of anopheline Lineage-Specific Duplicated Genes

THESIS

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

PARIK KAKANI

Under the Supervision of
Prof. Sanjeev Kumar
and
Co-supervision of
Dr. Pankaj Kumar Sharma

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
2018

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis entitled **"The Role of *Anopheles* Heme Peroxidases in Mosquito Immunity against Blood-Borne Antigens and Divergent Expression of anopheline Lineage-Specific Duplicated Genes"** submitted by **Parik Kakani** ID No **2012PHXF0431P** for award of Ph.D. degree of the institute embodies original work done by her under my supervision.

Signature (Supervisor)

Name	**SANJEEV KUMAR, Ph. D.**
Designation	**Associate Professor, Department of Biological Sciences, BITS, Pilani**
Present address	**Professor and Head, Department of Biotechnology**
	Ch. Bansi Lal University (CBLU), Bhiwani, Haryana
Date:	

Signature (Co-supervisor)

Name	**PANKAJ KUMAR SHARMA, Ph. D.**
Designation	**Assistant Professor, Department of Biological Sciences, BITS, Pilani**
Date:	

ABSTRACT

Malaria, a vector-borne disease, affects approximately 3.2 billion people worldwide each year. The transmission of this disease is initiated by an infectious bite of female *Anopheles* mosquito which carries the parasite of genus *Plasmodium*. Malaria parasite requires human host to complete its asexual life cycle and female *Anopheles* to complete its sexual life cycle. A number of *Anopheles* species transmit human malaria parasite and in India, *Anopheles stephensi* is one of the major urban malaria vector.

The mosquito feeds on plant nectar. However, female mosquito requires blood meal for the development of eggs. During the blood meal, female *Anopheles* accidentally encounters the malaria parasite, *Plasmodium*. Several body compartments of mosquito are involved in the *Plasmodium* development and midgut is the first compartment that the parasite encounters in the mosquito. *Plasmodium* interacts with several molecules of mosquito midgut during its development. Molecules that positively regulate the development of *Plasmodium* are called as agonists. Also, *Plasmodium* undergoes massive losses during its development in the mosquito midgut and faces a sizeable bottleneck, suggested that midgut epithelial cells mount a potent immune response against it. Mosquito molecules that negatively regulate the *Plasmodium* development are known as antagonists. Currently, *Plasmodium* developmental stages of midgut are targeted to control malaria transmission. One such method is the development of vaccines against malaria in mosquito and known as Transmission Blocking Vaccines. Midgut-specific molecules can be used as vaccine candidates to control malaria transmission. Hence, targeting midgut molecules which are crucial for successful development of *Plasmodium* inside the mosquito host will prevent malaria transmission. Recently, in *Anopheles gambiae*, heme peroxidases such as HPX15, HPX2 and Duox are reported in modulating the midgut immunity. Functional studies of these heme peroxidases in *A. stephensi* may provide an opportunity to target these molecules as vaccine candidates in developing transmission blocking strategies.

In this thesis work, we carried out the molecular characterization of heme peroxidase gene HPX2 from Indian malaria vector *A. stephensi* and showed that it plays a crucial role in maintaining bacterial homeostasis and in limiting *Plasmodium* development inside the mosquito midgut. The AsHPX2 is a secreted protein of 692 amino acids. The orthologs of AsHPX2 are only present in mosquitoes. The expression

of this gene is reduced in blood fed midguts. RNAi based silencing of AsHPX2 gene in sugar fed midguts showed increased bacterial load. So, this gene is involved in maintaining the midgut bacterial homeostasis in the sugar fed mosquitoes. Silencing of AsHPX2 gene increased *Plasmodium* oocysts number and exhibited the anti-plasmodial property in the midguts in a way similar to its ortholog AgHPX2 in *A. gambiae*. In *A. gambiae* it is reported that HPX2 enhanced the clearance of *Plasmodium* ookinetes during midgut invasion. Thus, AsHPX2, a mosquito-specific gene may be targeted to design such strategy that can arrest *Plasmodium* development inside the mosquito.

Studies exploring the mosquito's innate immune defense mechanism against *Plasmodium* and detailing the importance of the midgut microbiota in vector competence may contribute towards the development of effective control strategies. In our study, we characterized Dual Oxidase (Duox) gene in Indian malaria vector and showed that it plays a crucial role in bacterial homeostasis and also in *Plasmodium* development in the mosquito midgut. Duox in *A. stephensi* is a transmembrane protein with N-terminal cytoplasmic heme peroxidase domain and a non-cytoplasmic NADPH oxidase domain at C-terminal. In addition, it also has a calcium binding domain and seven transmembrane domains. The AsDuox ortholog, AgDuox in *A. gambiae* performs tyrosine crosslinking of a mucin layer in the midgut. This mucin layer acts as a physical barrier and protects midgut commensal bacteria and *Plasmodium* against midgut immunity. Silencing of AsDuox in either sugar fed or blood fed midguts revealed increased endogenous bacterial growth. Thus, we assumed that AsDUOX gene plays a dual role, on one hand, it protects the bacteria from midgut immunity by creating low immunity zone through midgut barrier formation, and on the other hand, it controls their over-growth due to its anti-bacterial nature. The expression of this gene is induced in exogenous bacteria supplemented blood fed midguts and has a strong negative correlation with the growth of bacteria in these midguts. This indicates that Duox is one of the major molecules of midgut immunity.

AsDuox gene also plays important role in *Plasmodium* development and silencing of this gene suppressed *Plasmodium* oocysts number through activation of TEP1 (Thioester-containing proteins) molecules. Previously, in *A. gambiae,* it has been reported that Duox gene supports the development of *Plasmodium* and silencing of AgDuox reduced the number of *Plasmodium* oocysts. These findings explored that agonist role of Duox is conserved in both *A. stephensi* and *A. gambiae*. Thus, Duox gene is an important molecule of innate immunity against pathogens in *Anopheles*. Hence, this molecule might serve as a universal target to manipulate mosquito immunity

and midgut bacterial population that can be used to arrest *Plasmodium* development inside the vector host.

Heme peroxidases belong to multi-gene family in which gene duplication event is very common. So, we were interested in studying the gene duplication event in heme peroxidase family of *Anopheles*. We found that previously reported heme peroxidase HPX15, a transmission blocking vaccine candidate, in *A. gambiae* and *A. stephensi* has its tandemly duplicated paralog named HPX14. The duplicated genes are are flanked by the presence of boundary elements and might act as an independent domain of gene expression. We found that duplicated genes are under the purifying selection and hence, might maintain two distinct functional copies. We found that both the genes are functional and the mRNA levels of AsHPX15 gene are higher than AsHPX14 gene in the midguts. However, the spatial and temporal expression of AsHPX14 gene might be suppressed by CTCF (CCCTC-binding factor), an insulator protein. To reveal the function of AsHPX14 gene, we silenced the AsHPX15 gene in the midguts. We found that expression of AsHPX14 gene is induced in AsHPX15 silenced bacteria supplemented blood fed midguts but not in AsHPX15 silenced blood fed midguts against respective controls. The high bacterial load in absence of barrier formation (silencing of AsHPX15) cause some signal to displace CTCF protein and induce the expression of AsHPX14 gene. This data suggested that HPX14 gene may have a role in immunity against bacteria but not in physiology. Hence, we conclude that there is no redundancy in the function of duplicated gene. This strengthen the potentiality of AsHPX15 as a vaccine candidate to block *Plasmodium* transmission. This study suggests the potential functional roles of CTCF in the mosquito. This can be used to improve mosquito transgenesis. This can also provide a new model for the study of CTCF function in a species with medical significance. CTCF can be explored in managing and regulating genome-wide chromatin architecture and gene expression.

This thesis work contributes to a better understanding of the mosquito immune system to the malaria parasite, mosquito midgut microbiota interactions, mechanisms that maintain midgut homeostasis and chromatin organization of two duplicated heme peroxidases involved in the immunity. It would be of great interest to study the exact molecular mechanism that controls the expression of these heme peroxidases in the midgut. In future, functional study of these heme peroxidases in different anophelines may explore their use for widespread control of malaria.

ACKNOWLEDGEMENTS

Contents   Page No.

Rearing of mosquitoes

Retrieval of heme peroxidase genes from *A. stephensi* genome

Designing of *A. stephensi* heme peroxidases gene-specific primers

PCR amplification, cloning and full-length sequencing of heme peroxidases gene

Malaria parasite *Plasmodium berghei* maintenance and infection to mosquitoes

Mosquito tissue collection

Tissue collection of different developmental stages and body compartments of *A. stephensi*

Bacteria inoculation in 4th instar larvae of *A. stephensi*

A. stephensi bacterial feeding and tissue collection

P. berghei infection in *A. stephensi* and tissue collection

dsRNA synthesis

Analysing the effect of gene silencing on the growth of midgut bacteria

Analysing the effect of gene silencing on *Plasmodium* development

RNA isolation and cDNA preparation

Expression analysis of heme peroxidase and immune genes in different tissue samples using real time PCR

Analysis of conserved domains in heme peroxidases

Selection of heme peroxidases for phylogenetic analysis of AsHPX2 and AsDuox

***In silico* analysis of heme peroxidase proteins**

Retrieval, Identification and analysis of tandemly duplicated paralog (HPX15 and HPX14) in the genome of *Anopheles*

Analysis of gene sequence and structure of duplicated heme peroxidases in *A. stephensi* genome

Bioinformatics analysis of promoter, transcription factors, and regulatory elements for heme peroxidase genes

List of Tables

List of Figures

Abbreviations

HPX	Heme peroxidase
As	*Anopheles stephensi*
Ag	*Anopheles gambiae*
PBS	Phosphate buffer saline
FBS	Fetal bovine serum
PCR	Ploymerase chain reaction
qPCR	Quantative PCR
gDNA	Genomic DNA
cDNA	Complementary DNA
mRNA	Messenger RNA
rRNA	Ribosomal RNA
NaCl	Sodium chloride
Tris-Cl	Tris(hydroxylmethyl)aminomethane-Hydrochloride
kDa	Kilo dalton
AA	Amino acids
bp	Base pair
dsRNA	Double stranded RNA
RT	Room temperature
TF	Transcription factor
AMPs	Anti-microbial peptides
ROS	Reactive oxygen species
CTCF	CCCTC-binding factor
Duox	Dual oxidase
NOS	Nitric oxidase synthase
TEP1	Thioester-containing protein
nl	Nanolitre

Keywords

S.No.	Keywords
1	Mosquito
2	Malaria
3	*Anopheles gambiae*
4	*Anopheles stephensi*
5	Immunity
6	Midgut
7	Mucin barrier
8	Midgut epithelial cell
9	Heme peroxidase (HPX)
10	Peroxinectin
11	HPX2
12	Dual oxidase
13	HPX15
14	HPX14
15	Midgut bacteria homeostasis
16	Commensal bacteria
17	Exogenous bacteria
18	*Plasmodium*
19	Gene duplication
20	CTCF
21	Tolerance
22	Resistance

Chapter 1

Introduction and Review of Literature

1 Introduction

1.1 Malaria is a life-threatening vector-borne parasitic disease

Several major human diseases are caused by insects and malaria is one of the most important and life-threatening diseases. Malaria is caused by a protozoan parasite of the genus *Plasmodium,* which is transmitted to human by the female *Anopheles* mosquito. Out of hundred species of *Plasmodium,* six species can infect human. The four major species of *Plasmodium are P. falciparum, P. vivax, P. ovale* and *P. malariae* and the newly considered fifth species is *P. knowlesi,* a malaria parasite that normally infects macaques (William et al., 2013; White, 2008; Cox-Singh and Singh, 2008). Very rare cases of malaria in human that are caused by *P. cynomolgi* are reported (Ta et al., 2014). In endemic regions, several cases of malaria have been reported in the human having co-infection with more than one *Plasmodium* species (Mehlotra et al., 2000).

Malaria is reported over ninty-one countries, primarily in tropical and subtropical regions of sub-Saharan Africa, Central and South America, the Caribbean island of Hispaniola, the Middle East, South-East Asia, and Oceania (**Figure 1.1**). There is a high risk associated with malaria as almost about 3.2 billion people (almost half of the world's population) are at risk, making it a worldwide problem. On an average, more than 212 million people are infected and 0.429 million people die due to malaria every year worldwide (WHO, 2016). These data indicate the severity of malaria and its impact on human health.
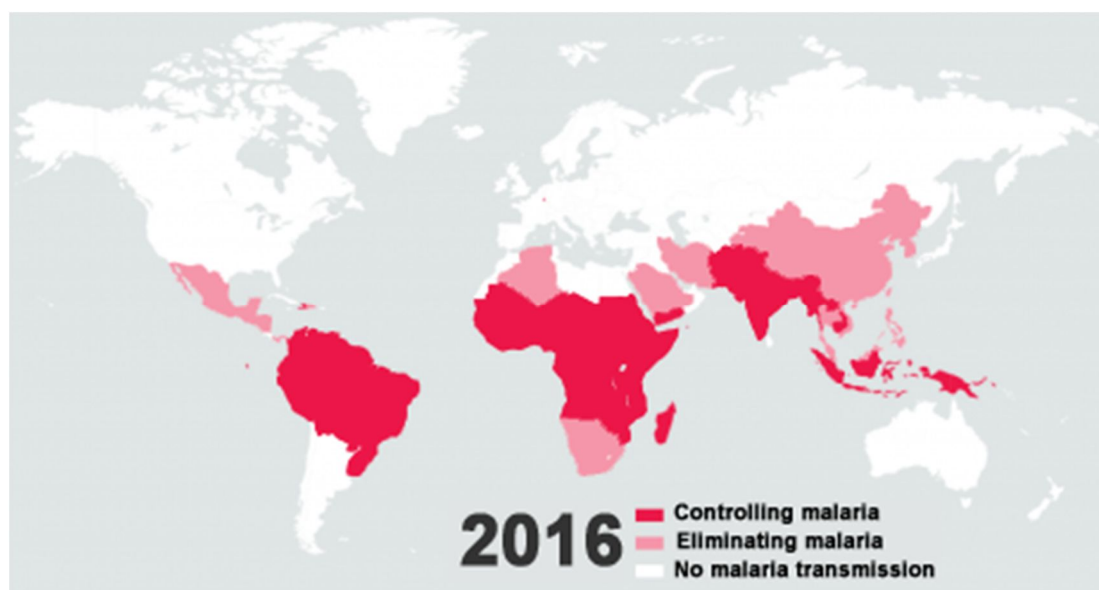
Figure 1.1: **Distribution of malaria worldwide**. The world map shows the malaria prevalent regions reported by UCSF Global Health Group's Malaria Elimination Initiative (MEI) (http://globalhealthsciences.ucsf.edu/our-work/diseases-and-conditions/malaria).

Tropical countries, including India, are at the risk of a number of vector-borne diseases including malaria. India reports approximately 61% of all confirmed malaria cases in the South-East Asia region (WHO, 2011). In the Indian subcontinent, for the past few years, severe complications (cerebral malaria, pulmonary edema, acute renal failure, etc.) are also reported in malaria patients (Kaushik et al. 2012; Patil, 2012) that make the situation worrisome. The high human population and suitable climate to support higher mosquito density might also accelerate the rate of malaria transmission throughout the country. Malaria is a major health problem in rural as well as tribal areas of 16 Indian states, including seven northeastern states and nine states of central India
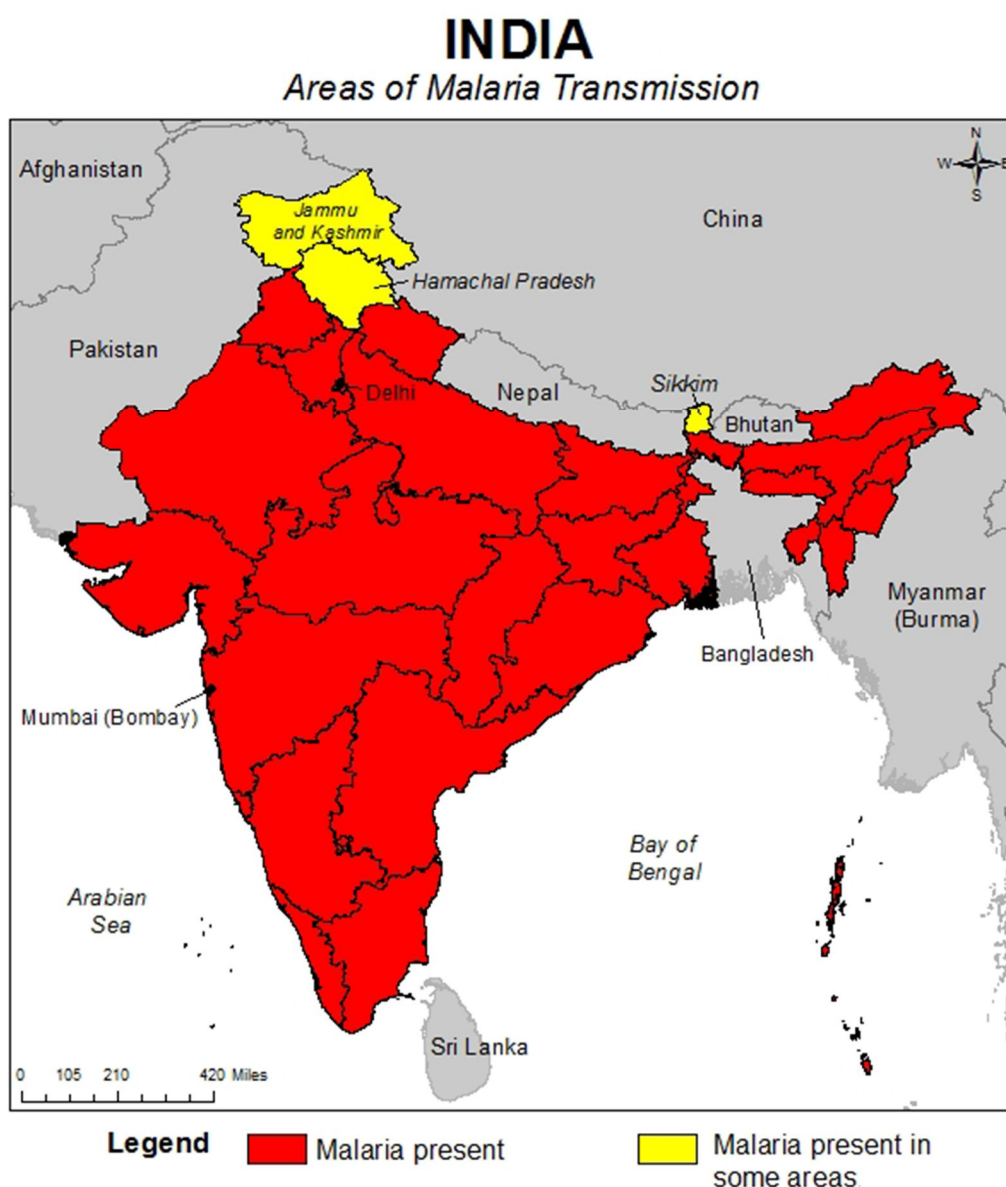
Figure 1.2: **Malaria affected areas in India**. In India, malaria is present throughout the country excluding areas >2000 m in Himachal Pradesh, Jammu and Kashmir, and Sikkim. (https://www.cdc.gov/malaria/travelers/country_table/i.html). Source: Global Health–Division of Parasitic Diseases and Malaria available at Centers for Disease Control and Prevention.

(National Vector Borne Disease Control Program, 2014). Malaria endemic areas in India are depicted in **Figure 1.2.**

1.2 Life cycle of malaria parasite

The malaria parasite, *Plasmodium* completes its life cycle in two evolutionary far hosts; an invertebrate host, the mosquito and a vertebrate host such as human. The sexual life cycle of *Plasmodium* is completed in female *Anopheles* mosquito (a definitive primary host) while asexual life cycle is completed in human (a secondary host). The life cycle of *Plasmodium* in *Anopheles* starts with the ingestion of *Plasmodium* gametocytes during blood feeding of malaria infected person. Soon after ingestion, *Plasmodium* gametocytes undergo gametogenesis and fertilization that form zygote inside the midgut lumen. In the midgut, the gametogenesis is activated by the drop in body temperature and mosquito derived Xanthurenic acid (Billker et al., 1998; Garcia et al., 1998; Kuehn and Pradel, 2010). Further, in the same compartment, the formation of ookinete from zygote takes place near about 15 h after ingestion of *Plasmodium* infected blood. Ookinete traverses the midgut epithelial cells around 24 h of ingestion and develops into an oocyst in the space between midgut epithelium and basal lamina (**Figure 1.3**). Oocyst matures approximately after 10 days and then releases thousands of sporozoites into the mosquito hemocoel. Sporozoites circulate throughout the hemolymph and some of them invade the salivary glands. Sporozoites in salivary gland further undergo maturation and during the subsequent feeding, they are ready to be injected into the new host (Zheng, 1997; Beier, 1998; Sinden, 2002).

Plasmodium life cycle in the human host begins with the infective bite of female *Anopheles* that releases sporozoites into circulating blood. These sporozoites migrate towards the liver and initiate a pre-erythrocytic cycle in hepatocytes that continue approximately for 6-15 days. During this phase, *Plasmodium* undergoes extensive growth and division. At the end, thousands of merozoites are released into the bloodstream. These merozoites further invade the red blood cells (RBCs) and initiate an erythrocytic cycle (Frevert, 2004). This asexual erythrocytic cycle further produces more merozoites and after 48 or 72 h of erythrocyte infection, depending on *Plasmodium* species, they are released from the RBCs and immediately invade new erythrocytes (**Figure 1.3**). In continuation of the erythrocytic cycle some merozoites differentiate into gametocytes (Cowman et al., 2012) and after ingestion by another female mosquito, they continue the sexual cycle as discussed above.
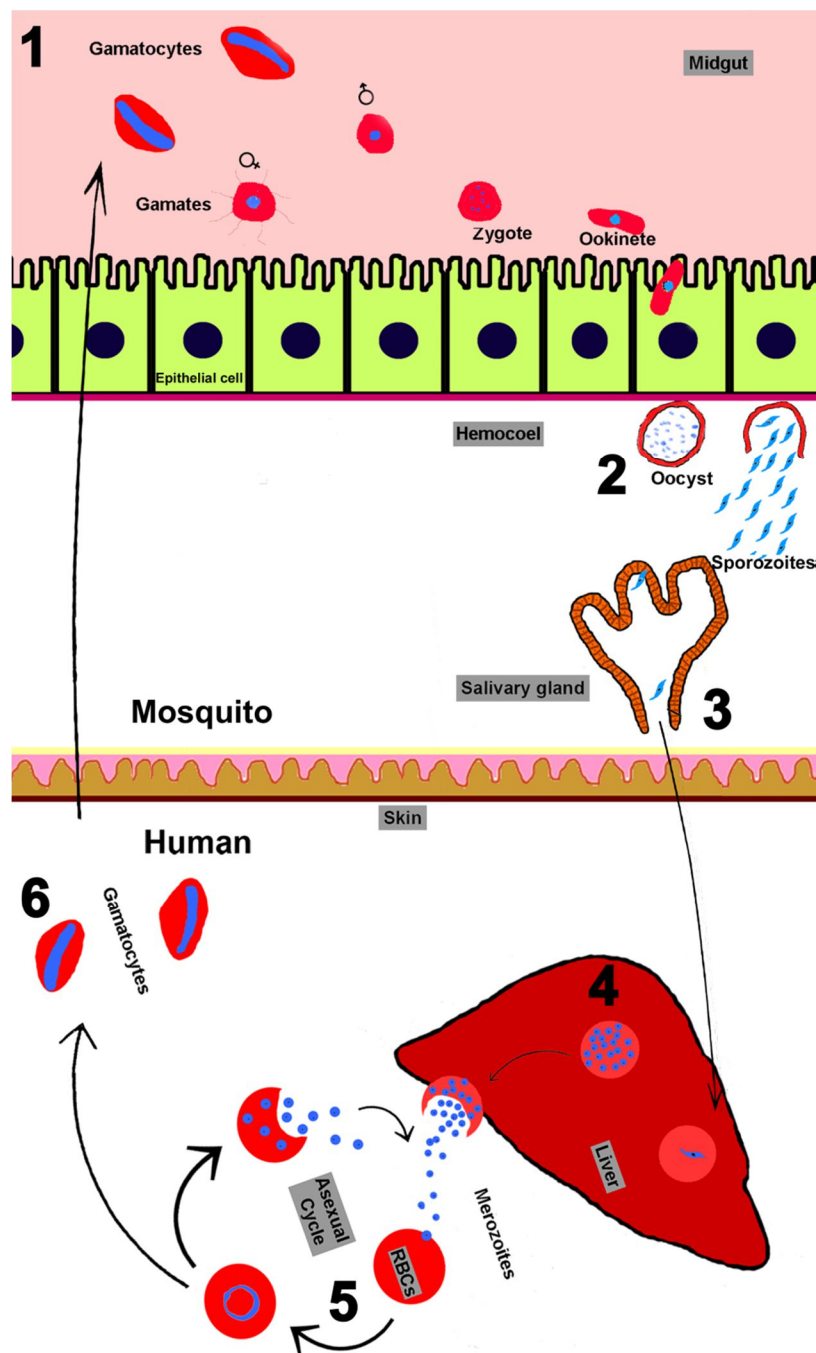
Figure 1.3: The life cycle of *Plasmodium*. The sexual life cycle of *Plasmodium* initiates when a mosquito ingests an infected blood meal containing gametocytes into the midgut lumen. **A) Mosquito stages: 1) Midgut stage**: Soon after the ingestion, gametocytes undergo gametogenesis and fertilization in the midgut that form zygotes. The zygotes differentiate into motile ookinetes that traverse the midgut and develop into oocysts. **2) Hemocoel stage**: After maturation, oocysts release thousands of sporozoites into the hemocoel. **3) Salivary gland stage:** The released sporozoites specifically invade the salivary glands. When an infected mosquito bites a human it releases sporozoites into the bloodstream. **B) Human stages**: **4) Liver stage**: These sporozoites travel towards the liver. In hepatocytes, *Plasmodium* schizont starts its exo-erythrocytic cycle and produces a number of merozoites. **5) Blood stage**: These merozoites are released into the bloodstream and invade, the RBCs that mark the beginning of the erythrocytic cycle. During the erythrocytic cycle, merozoites enter the ring stage and develop into early trophozoites. These trophozoites further develop into schizonts. The RBCs rupture and release new merozoites that immediately infect other RBCs. **6) Gametocyte stage:** The second fate of trophozoites during an erythrocytic cycle is to differentiate into female and male gametocytes. When female *Anopheles* takes blood from an infected human, its sexual life cycle starts again (Kakani et al., 2016).

1.3 Malaria vector distribution and its life cycle

The malaria parasite *Plasmodium* is transmitted by female mosquitoes of the genus *Anopheles*. Worldwide, there are approximately 465 *Anopheles* species, out of which, approximately 41 are dominant malaria species (DVS) that spread human malaria (Sinka et al., 2012). These anophelines are worldwide distributed as shown in **Figure 1.4** and transmit different species of *Plasmodium* depending on the region and the local environment.
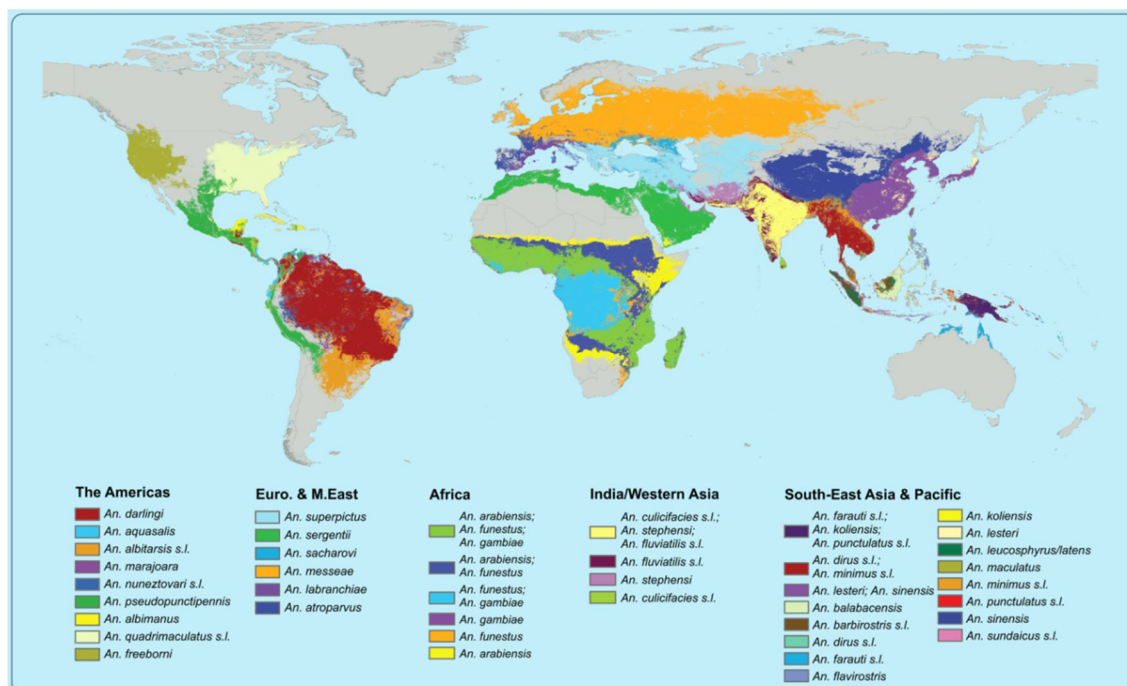
Figure 1.4: Dominant malaria vectors and their worldwide distribution: Graph shows the distribution of malaria vectors over the 5 major geographical regions (Sinka et al., 2012).

India is a tropical country and the epidemiology of malaria is complex because of geo-ecological diversity and climatic conditions which are suitable for wider distribution of anopheline vectors (Anvikar et al., 2016). In India, nine out of 58 *Anopheles* species are designated as the vector of the human malaria parasite (Nagpal and Sharma, 1995). Among these, six species (*A. minimus, A. fluviatilis, A. culicifacies, A. dirus, A. stephensi, A. sundaicus*) are considered as the primary vectors (commonly feed on humans and have measurable sporozoite rates), while three (*A. annularis, A. nivipes/ A. philippinensis* and *A. varuna*) are designated as secondary vectors (have low sporozoite rates) (Dash et al., 2007). Of these, *Anopheles stephensi* Liston 1901 considered as an important vector in India, mainly in the urban and industrial areas and prefers human blood (Adak et al., 1999; Sharma, 1999; Swami and Srivastava, 2012).

The mosquito life cycle is an example of complete metamorphosis. There are four distinct stages in the life cycle of a mosquito: egg, larvae, pupa, and adult. The life cycle starts with the tiny egg. The adult female *Anopheles* requires at least one blood meal before she can oviposit. Eggs are deposited singly on the surface of the water at various breeding sites, including shallow sunlit pools, borrow pits, drains, car tracks, hoof prints near water holes, rice fields, irrigation canals, pools left behind by receding rivers, and rainwater collecting in natural depressions (Gillies & DeMeillon, 1968). The eggs hatch into larvae 2-3 days after egg laying. The hatched larvae or "wrigglers" swims in the water and goes through four growth stages called instars, in which they grow in size. The larvae undergo three successive moults during their development. After the fourth instar larvae, the pupae form that
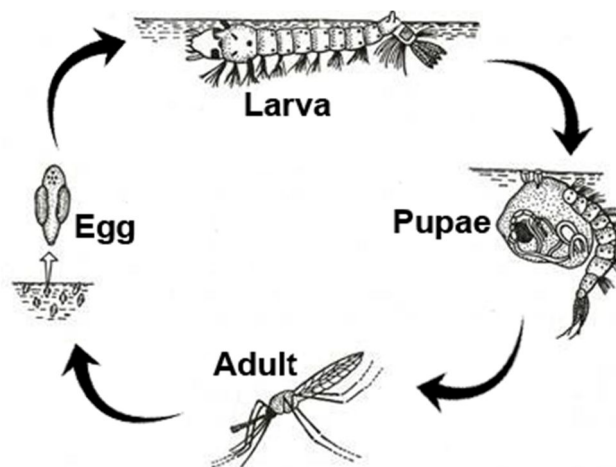
Figure 1.5: Mosquito life cycle. The life cycle of *Anopheles* starts with an egg that hatches into larva. The larva undergoes three consecutive moults and transforms into fourth instar larva. Larva finally develops into pupae, metamorphosis takes place and develops into an adult (WHO, 1997, Vector Control Methods for Use by Individuals and Communities).

often called "tumblers", do not feed and respire with the help of respiratory trumpets. After 2-3 days, the adult mosquitoes emerge from the pupae (Ramirez et al., 2009) (https://www.cdc.gov/malaria/about/biology/mosquitoes/index.html).

1.4 Current and past malaria control strategies

Malaria is a serious public health issue in many parts of the world with an enormous socioeconomic impact and people pay a heavy tax in terms of human health. This draws attention to initiate several prevention methods and development of control strategies. One such strategy is controlling the life cycle of *Plasmodium* in human using antimalarial drugs and vaccines. In addition, several other strategies involve reducing the contact with mosquitoes through the use of insecticide-treated bed nets, repellents, and controlling mosquitoes using insecticides, larvicides, and bacteria that produce toxins. Another strategy is to break the life cycle of *Plasmodium* by targeting important molecules of mosquito that are required for successful development of *Plasmodium* (transmission blocking strategies).

1.4.1 Treatment and prevention of malaria in human host

Early diagnosis of malaria and treatment of an infected person with the antimalarial drugs can also control transmission of *Plasmodium*. Synthetic drugs like chloroquine (CQ) and sulfadoxine/pyrimethamine (SP) are used extensively as antimalarial drugs. However, the extensive use of these drugs is the major cause of developing drug resistant *Plasmodium* (White, 2004; Klein, 2013). Presently, artemisinin-based combination therapies (ACTs) are used to treat malaria but some cases of resistance against them are also reported in South-East Asia (Dondorp et al., 2009; Noedl et al., 2010; Phyo et al., 2012). These reports warrant discovery of new drugs to treat malaria.

In addition to the drugs, vaccines can be used to prevent malaria. The purpose of vaccination strategies is to induce protective memory immune responses in advance of infection, to provide protection in the case of encountering the disease-causing agent again. Most current vaccine candidates target a single stage of the parasite's life cycle, at early pre-erythrocytic stages or later at blood stage. There are several malaria vaccines that are currently in clinical trials. The vaccines that target the pre-erythrocytic life-stage of *Plasmodium* are RTS,S/AS01E, RTS,S-AS01, PfCelTOS FMP012, Adenovirus (Ad35) vectored CS and RTS,S-AS01 in heterologous prime-boost regimen, ChAd63/MVA(CS; ME-TRAP), PfSPZ. The vaccines against Blood-stage of *Plasmodium* are ChAd63.AMA1/MVA.AMA1+Al/CPG7909, AMA1-C1-Alhydrogel+CPG 7909, BSAM-2-Alhydrogel+CPG7909, EBA175.R2, SE3, ChAd63/MVAAMA1, FMP2.1-AS01B (AMA1 3D7), NMRC.M3V.Ad.PfCA, MSP3[181–276], GMZ2 (Arama and Troye-Blomberg, 2014). The most advanced candidate for malaria vaccine is RTS,S/AS01 that is currently in phase 3 trials. This vaccine induces humoral and cellular immune responses against the circumsporozoite (CS) protein present on the surface of sporozoites or liver stage schizonts and block the pre-erythrocytic (PE) stage of *P. falciparum*. Recently in Africa, WHO will introduce Malaria Vaccine Implementation Programme (MVIP) in 2018 that will make the RTS,S vaccine (trade name Mosquirix™) available in selected areas of Ghana, Kenya, and Malawi (The RTS,S, 2014; RTS,S Clinical Trials Partnership, 2015). The RTS,S vaccine showed encouraging results, but low protection and high required amount of doses do not make the vaccine suitable for wide rollout in the areas that need it most (high disease burden, low resource settings). This calls for more research and development into a malaria vaccine that can give good protection and is easy to use in low resource contexts as a priority.

1.4.2 Vector control strategies

To reduce the risk of transmission, vector control is one of the most effective means to prevent malaria. Larvicides are used to control aquatic habitat of mosquito, larvae. The use of synthetic and non-biodegradable larvicidal is also causing some unfavorable effects such as killing of beneficial organisms and biological accumulation through the food chain that resulted in a negative impact on the ecological systems (Mrema et al., 2013). Secondary metabolites from plants (called natural phyto-mosquitocides) are better alternates for controlling mosquito populations. Several studies have documented the efficacy of plant extracts as a source of bioactive toxic agents against mosquito larvae (Ghosh et al., 2012; Kajla et al., 2016a). But only a few have been commercially produced and extensively used in vector control programs. To control the population of adult mosquito, adulticides are used. At the moment, insecticides belonging to different groups viz., organochlorines, organophosphates, carbamates and synthetic pyrethroids are widely used in various vector control programs. Insect resistance to insecticides of all classes of compounds mentioned above is observed (Smith Gueye et al., 2016). Well known mosquitocidal dichloro diphenyl trichloroethane (DDT) developed resistance in >50 species of anopheline mosquitoes and its adverse effect on human health as well as its non-biodegradability nature, makes it inappropriate as mosquitocidal (Hemingway and Ranson, 2000; Beard, 2006; Chen and Rogan, 2003; Cox et al., 2007; Eriksson and Talts, 2000). Currently, insecticide use still plays a significant role in malaria control programs involving the use of insecticide-treated bednets and indoor residual spraying.

Although, various methods are extensively used to cure and prevent malaria, such as anti-malarial drugs, bed nets, indoor residual spraying, larvicides and insectides but the emergence of drug resistant parasites and insecticide resistant mosquitoes have also played a significant role in the occurrence and severity of the malaria transmission worldwide. Also, insufficient financial support and health infrastructure in malaria-endemic areas have decreased the success of available control strategies to eradicate malaria. For these reasons, there is an urgent need to improve current malaria control methods. It is also important to develop new strategies to eliminate and eventually eradicate the disease. In recent years, there has been wider interest in the mosquito stages of *Plasmodium* life cycle as potential targets to develop new transmission-blocking strategies (TBS) that could help to control and ultimately eliminate the disease.

1.4.3 Transmission-blocking strategies

The eventual eradication of malaria implicates breaking the cycle of transmission of *Plasmodium* between human and mosquito hosts. One such strategy utilizes Transmission Blocking Vaccines (TBV) that can block the development of sexual stages of parasite in the mosquito midgut. A vaccine with the ability to prevent transmission of *Plasmodium* would be a critical tool for achieving the long-term goal of malaria eradication (Butler, 2009). TBV is designed to target *Plasmodium* antigens expressed during sexual stages of development or those *Anopheles* midgut proteins whose molecular interactions are necessary for the fertilization of gametes, ookinete invasion of midgut epithelium, or ookinete-to-oocyst transition. For that, *Plasmodium*- or mosquito target-specific antibodies are developed in the vertebrate host and these antibodies suppress parasite development inside the mosquito when ingested with the infected blood meal. The antigens such as Pfs230, Pfs48/45 present on the surface of *P. falciparum* gametocytes and gametes, and Pfs25/Pfs28 present on the surface of the zygote and ookinetes, Pvs25 and Pvs28 on ookinetes of *P. vivax* have been successfully targeted through transmission-blocking vaccines (Farrance et al., 2011; Pradel, 2007; Chaurio et al., 2016). Transmission-blocking vaccines (TBV) can also target those *Anopheles* molecules which are crucial for *Plasmodium* development. Midgut molecules are the prime vaccine candidates as *Plasmodium* infection to vertebrates is only possible, if there is successful development of malaria parasites in the mosquito midgut. Also, it is the first compartment of mosquito which is interacting with *Plasmodium*. In *Anopheles,* midgut molecule such as glycoproteins, several annexin-like protein isoforms, sulfated proteoglycans, Croquemort scavenger receptors (SCRBQ2), a secreted glycoconjugate of unknown function (SGU), enolase binding protein (EBP), alanyl aminopeptidase 1 (APN1), carboxypeptidases A and B, FREP1 (fibrinogen-like domain), and heme peroxidase HPX15 are reported as transmission blocking vaccine candidates (Raz et al., 2013; Armistead et al., 2014; Kajla et al., 2015a; Niu et al., 2017; Venkat Rao et al., 2017; Dinglasan et al., 2003; Kotsyfakis et al., 2005; Dinglasan et al., 2007; Gonzalez-Lazaro et al., 2009; Mathias et al., 2014; Vega-Rodriguez et al., 2014). Promising results have been obtained from the research conducted in this area but none have yet been utilized in the field. So, the screening of new such molecules will help in the development of more novel approaches. So, this area further needs understanding the details of immune interactions between *Plasmodium* and mosquito immunity.

Mosquito immune pathways and their effector genes have been implicated in limiting the development of *Plasmodium*. Mosquito's anti- *Plasmodium* immune repertoire is full of possibilities for the screening of new vaccine candidates. The immune responses offer a number of diverse mechanisms that could be manipulated to act synergistically, ensuring a more comprehensive elimination of malaria. So, the study of mosquito immunity is an important and interesting area, as it could be incorporated into a variety of control methods.

1.5 Mosquito immunity against *Plasmodium*

In order to continue the transmission and infection of malaria in the human host, the parasite engages in a series of complex interactions with the mosquito. Parasite suffers several major bottleneck situations (**Figure 1.6)** in the mosquito, such as, when the ookinete traverses the midgut epithelium prior to the development of the oocysts on the basal side and during the migration of sporozoites to the salivary glands (Sinden, 2002; Hillyer et al., 2007). The mosquito's innate immune system has been shown to
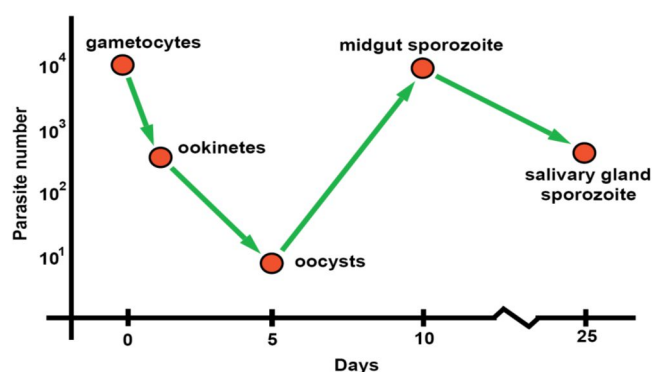
Figure 1.6: *Plasmodium* **bottleneck during development in the mosquito vector.** Malaria parasite faces bottleneck situation during sporogonic cycle that is considerably affected by the midgut environment of mosquito and is accredited to the mosquito defense mechanisms (Sinden, 1999).

play a key role in killing parasites and thereby affecting parasite development. Mosquitoes, like other insects, are known to mount potent immune responses against invading bacteria, fungi, viruses and parasites (Lehane et al., 2004). Several pieces of evidence suggested that mosquito is able to trigger a series of diverse defense reactions upon *Plasmodium* infection. These defense mechanisms are believed to account for the reduction in parasite number. The details of immune responses are discussed below.

1.5.1 Mosquito immune molecules and signaling pathways

Mosquitoes have innate immunity as a defense mechanism against parasite. Innate immune molecules are rapidly active after encountering pathogens and have specificity against a particular immune elicitor. Three major signaling pathways of mosquito

immunity contribute to defense mechanisms: the Toll, the Immune deficiency (Imd), and the Janus Kinase-Signal Transducers and Activators of Transcription (JAK-STAT) pathways (Christophides et al., 2002; Meister et al., 2005; Lemaitre et al., 1996). The Toll pathway is primarily activated by Gram-positive bacteria, *P. berghei*, fungi, and viruses (Michel et al., 2001; Ramirez and Dimopoulos, 2010; Frolet et al., 2006). The Imd (Immune deficiency) pathway is elicited by Gram-negative bacteria and *P. falciparum* (Lemaitre et al., 1995).

The recognition of pathogen-associated molecular patterns (PAMPs) is initiated by pattern recognition receptors (PRRs) such as Peptidoglycan Recognition Proteins (PGRPs) and Gram-Negative Binding Proteins (GNBPs). These receptors can either indirectly trigger the induction of intracellular signaling pathways that control transcription of effector genes or directly invoke microbial killing mechanisms such as encapsulation and phagocytosis (Hillyer, 2010). The PRRs activate the Toll and Imd pathways which lead to the nuclear translocation of the NF-κB family transcription factors Rel1 and Rel2, respectively. The activation of Toll and Imd pathways induce the transcription of numerous immune effector genes such as, Attacin, Cecropin, Defensin and Gambicin etc. that are collectively known as Anti-microbial Peptides (AMPs) (Luna et al., 2006). AMPs synthesis is considered as a hallmark of insect humoral responses to microbial infections such as Gram-positive as well as Gram-negative bacteria, yeast, fungi, and *Plasmodium*. AMPs interact with bacterial membrane or cell wall and mediate their killing. The net positive charge and a high ratio of hydrophobic amino acids let the AMPs to selectively bind a negatively charged bacterial membrane. The binding of AMPs to the bacterial membrane leads to non-enzymatic disruption of bacteria (Zhang and Gallo, 2016). Additionally, Imd pathway regulates a diverse set of anti-*Plasmodium* immune effectors such as *A. gambiae* Down syndrome cell adhesion molecule gene (AgDscam), leucine rich repeat containing proteins like LRRD19 (APL1), LRRD7 (APL2), LRIM1, FBN9 (Fibrinogen-Related Proteins), and TEP1 (Thioester containing proteins) (Meister et al., 2005; Riehle et al., 2008; Povelones et al., 2009; Blandin et al., 2004; Frolet et al., 2006; Garver et al., 2009 and 2012). LRIM1 and APLs are leucine rich repeat containing proteins that play an important role in melanization and killing of *Plasmodium* (Cirimotich et al., 2010). FBN9 is a member of Fibrinogen-Related Proteins (FREPs) family that binds ookinetes as they invade the midgut epithelium (Dong et al., 2006). Thioester-containing proteins (TEPs) are hemolymph proteins involved in the killing of bacteria and *Plasmodium* ookinetes by phagocytosis. TEP1 formed a complex with the leucine-rich repeat containing proteins LRIM1 and APL1C prior to bacteria

binding in hemocoel or ookinetes in the midgut, triggering their destruction (Fraiture et al. 2009).

The *A. gambiae* JAK/STAT-pathway and LPS-induced TNFα transcription factor (LITAF)-like 3 (LL3) are also reported to mediate late phase immunity. The JAK/STAT pathway regulates expression of nitric oxide synthase (NOS) that has a negative effect on developing oocysts (Gupta et al., 2009). In the *A. gambiae*, LL3 binds to the promoter region of an anti-plasmodial gene named serine protease inhibitor 6 (serpin 6) and modulates its expression. Silencing of LL3 gene restricts the differentiation of hemocytes and their response to parasite infection that results in the increased number of oocysts in silenced mosquitoes. Thus, LL3 implicates late-phase immunity against *Plasmodium* oocysts through hemocytes (Smith et al., 2012 and, 2015).

Although, gene silencing approaches identified several immune genes that effectively regulate *P. berghei* development in *Anopheles*, however, these genes are ineffective against human malaria parasites. For example, the silencing of leucine rich repeat containing proteins, LRIM1 gene in *A. gambiae* increased *P. berghei* infection and has no effect on *P. falciparum* infection. In addition, LRIM1 silencing has no effect in *A. stephensi-P. yoelii* system (Jaramillo-Gutierrez et al., 2009). TEP1 mediates *P. berghei* lysis but remained ineffective against human malaria parasite *P. falciparum* (strain 3D7, NF54 and GB4) (Molina-Cruz et al., 2012). This shows the different interaction of mosquitoes' immune system with different *Plasmodium* species. It is important to note that there are inconsistencies in the molecules' immune roles between mosquito and parasite strains that raised concern for widespread control of malaria using these molecules. Is this the ecological niche that influences the immunity and vector competence?. Understanding the dynamics of midgut endogenous bacteria, malaria parasites and midgut immune components may help us to improve vector-based control efforts and offer some unique control possibilities.

1.5.2 Role of gut microbiota in vector competence

Gut harbors numerous type of bacterial population often established the symbiotic relationship with the host. These bacteria perform important functions in supplying the nutrients, host development, food digestion, reproduction and also defend the host against colonization by opportunistic pathogens (Shin et al., 2011; Storelli et al., 2011; Venema, 2010; Tremaroli and Bäckhed, 2012; Gaio et al., 2011; Coon et al., 2016). The role of bacteria in blood digestion is evident by the secretion of hemolytic enzymes by *Serratia* and *Enterobacter* (Gaio et al., 2011). In *A. stephensi,* vitamin supplements are

provided by the bacterial species *Asaia bogorensis* (Crotti et al., 2010). The midgut bacteria of the insect have also been shown to modulate vectorial capacity by preventing the development of pathogens. This is an emerging area of interest since different *Anopheles* strains harbor quantitatively and qualitatively different microbial flora. The microbial environment can influence *Plasmodium* development within the mosquito (Pumpuni 1996). The symbiotic bacteria in the midgut activate the basal innate immune activity through induction of ROS producing gene, AMPs and other immune-specific genes that can also act against *Plasmodium* (Cirimotich et al., 2010). Moreover, immune responses that are raised by the mosquito against bacteria and *Plasmodium* overlap to a large degree. Additionally, host immune responses raised against gut bacteria also affect *Plasmodium* development (Dimopoulos et al., 2002). Researchers explored these findings to identify mosquito gut-specific symbionts that can regulate *Plasmodium* development (Boissière et al., 2012). Gram-negative bacteria (for example, *Serratia marcescens, Entobacter* species such as *E. amnigenus, E. cloacae, E. sakazaki*) have been identified in anophelines midgut that inhibit *P. falciparum or P. vivax* sporogonic development. It has been shown that co-feedings of either live or heat-killed bacteria cause decrease in *P. falciparum* oocysts whereas depletion of gut bacteria with antibiotics yields higher oocyst load (Dong et al., 2009a).

Symbiotic gut bacteria can also directly exert antiplasmodial activity. *Enterobacter* species, a resident of wild *A. arabiensis* can directly suppress *P. falciparum* survival through ROS production (Cirimotich et al., 2011). Presence of Gram-negative bacteria in *Anopheles* midgut has a negative effect on *Plasmodium* development (Pumpuni et al., 1993 and 1996; Beier et al., 1994). Another study demonstrated that more number of *P. falciparum* ookinetes transformed into oocysts when anti-bacterial antibodies were generated against total midgut lysate (Noden et al., 2011). So, it can be concluded that midgut bacteria play a role either directly through the production of various enzymes and toxins or indirectly by stimulating the mosquito's innate immune system to produce antimicrobial molecules that limit *Plasmodium* development (Pumpuni et al., 1993; Azambuja et al., 2005; Dong et al., 2009a). The mechanisms by which *Anopheles* midgut maintains bacteria homeostasis and process of bacteria mediated inhibition of *Plasmodium* infection are studied only to a limited extent. Thus, dissection of the tripartite interactions between mosquito midgut immunity, endogenous bacteria and *Plasmodium* and a fine-tuned manipulation of this system may provide an opportunity to devise an integrated approach. This may target multiple parasite stages and thereby increases the efficacy of the treatment.

1.6 Mechanisms maintaining midgut microbial homeostasis

Midgut of insect harbors natural bacteria (symbionts/commensal) that perform many functions and play a crucial role in the development of the host. It is noteworthy to mention that the gut epithelium tolerates the proliferation of endogenous bacteria to a certain extent to allow beneficial interactions with its host (Lemaitre and Hoffmann, 2007). So, the gut immune system must distinguish commensal and non-commensal microbes to maintain a fine balance between normal physiology and immunity. Then how does the gut epithelium tolerate commensal bacteria while retaining the ability to trigger an efficient immune response after infection with pathogenic bacteria?. It is reported in *A. gambiae* that Caudal is the negative regulator of IMD pathway and RNAi-mediated silencing of Caudal suppressed the midgut microflora (Clayton et al., 2013). A similar mechanism is also reported in *Drosophila* that in response to the commensal bacteria, there is activation of IMD signaling pathway but its negative regulator Caudal suppresses the production of AMPs to protect commensal bacteria (Lemaitre and Hoffmann, 2007; Ryu et al., 2008). The transcription of primary anti-bacterial immune molecules AMPs and Duox (Dual oxidase)-dependent ROS generation in gut epithelium is solely depended upon IMD pathway (Lemaitre and Hoffmann, 2007). However, infection with pathogenic bacteria increases the large quantities of PAMPs, which transiently increases the signaling strength and suppresses the activity of Caudal that triggers AMPs and ROS production. There are several studies that emphasized the antagonistic relationship between Caudal and Relish, a relationship that ensures a balance between immune defense and microbiota homeostasis in the gut (Charroux and Royet, 2010; Leulier and Royet, 2009; Muyskens and Guillemin, 2008; Royet, 2011; Ryu et al., 2008).

Mosquito being hematophagous in nature fed on sterile blood. It is noteworthy to mention that endogenous bacteria proliferate extensively after the blood feeding in the mosquito midgut (Kumar et al., 2010; Oliveira et al., 2011). However, midgut has a remarkable capacity to manage its environment to minimize the deleterious effects of midgut immunity on increased endogenous bacterial load as they are required for blood digestion (Kajla et al., 2015a; Dong et al., 2009a; Gupta et al., 2009; Warr et al., 2008).

Midgut synthesizes a peritrophic matrix (PM) in response to blood feeding, which shields direct contact of ingested food with midgut (Jacobs-Lorena and Oo, 1996). The peritrophic matrix in insects is a non-cellular sieve-like structure that lines the midgut epithelium and made up of chitin fibrils and chitin-binding proteins (Lehane, 1997). However, the soluble elicitors released by the microbes present in the food may interact

with the gut epithelium and can induce an immune response (Ahmed et al., 2002).

In *Anopheles*, it is reported that heme peroxidase HPX15/dual oxidase system catalyzes the cross-linking of mucin layer on the luminal side of the midgut in response to blood feeding. This mucin barrier decreases the permeability of the midgut PM to immune activators. In other words, this barrier creates a low immunity zone that protects the midgut microbiota from epithelial immunity (Kumar et al., 2010). RNAi-mediated silencing of HPX15 gene results in decreased bacterial load in the midgut and induced key antibacterial effectors such as heme peroxidase HPX8, cecropin, PGRPS3 and PGRP-LB. Hence, heme peroxidases modulate the midgut immunity and play important role in maintaining the bacterial growth in the blood fed midgut.

1.7 Heme peroxidases have a vital role in midgut immunity

Heme peroxidases are divided into two major families, namely the animal and non-animal heme peroxidases. Non-animal heme peroxidases include plant, bacterial, fungal and protist (Passardi et al., 2007). Animal heme peroxidases are present in vertebrates and invertebrates. Heme peroxidases catalyze the hydrogen peroxide-mediated one- and two-electron oxidation of thousands of molecules including aromatic molecules (e.g., coniferyl alcohol or tyrosine), cations (e.g., Mn^{2+}), anions (e.g., ascorbate or halides) or even proteins (e.g., cytochrome c). A heme peroxidase performs 2 different types of reactions. One is the formation of dityrosine covalent bonds via peroxidase cycle (dehydrogenation) in which H_2O_2 is reduced to water and one-electron donors such as oxygen is oxidized to the respective superoxide radical. The other reaction is catalyzed by chloro-peroxidase which involves the oxidation of two-electron donors like halides (X^-) to the corresponding hypohalous acids (HOX) (Zámocký et al., 2015; Kumar et al., 2010). In humans, heme peroxidases have a central role in defense mechanisms. For example, an oxidative killing of *Staphylococcus aureus* (Ramsey et al., 1982) or *Escherichia coli* by eosinophils or neutrophils are mediated through an azide-sensitive myeloperoxidase-dependent pathway (Persson et al., 2001). Some classical peroxidases, such as eosinophil peroxidase (EPO) and myeloperoxidases (MPO) are well-known effector molecules in vertebrate immunity. Several studies in *Drosophila* have shown that gut generates a dual oxidase (DUOX)-dependent reactive oxygen species (ROS) (Ha et al., 2005a, 2009a, 2009b) as the first line of defense against microbial population. Investigations in other insects like *Bactrocera dorsalis* and *Bombyx mori* also revealed the role of Duox generated ROS in maintaining the homeostasis of the gut bacterial community (Hu et al., 2013; Yao et al., 2016).

Several reports on mosquito immunity suggested the role of heme peroxidase in immunity against blood-borne antigens. There are several classes of peroxidases in *Anopheles* such as glutathione peroxidases, NADPH-dependent peroxidases, and heme peroxidases. Heme peroxidases have been found to associate with many immune responses directly or indirectly. The heme peroxidase HPX15 with dual oxidase forms mucin barrier and creates a low immunity zone (Kumar et al., 2010). *Plasmodium* takes advantage of this barrier and hides in the midgut. Another report in *A. gambiae* showed that heme peroxidase HPX2/NOX5 (NADPH oxidase) system has antiplasmodial response during *Plasmodium* development in the midgut. This system potentiate the toxicity of NO (nitric oxide) and enhanced the clearance of *Plasmodium* via TEP1 mediated lysis (Oliveira et al., 2012). So, mucin barrier formed by HPX15/Duox system protects *Plasmodium* from immune responses while at the time of midgut traversing, the midgut epithelial HPX2/NOX5 system enhanced their killing. In this way, heme peroxidases are modulating the gut immunity. Hence, studying the function of heme peroxidases in mosquito immunity may explore their potential role in limiting the *Plasmodium* development.

1.8   Evolution of heme peroxidase multi-gene family

A multigene family is a group of genes that have descended from a common ancestral gene and therefore have similar functions and similar DNA sequences. Heme peroxidases in eukaryotes belong to multigene family. The core event behind the expansion of multi-gene family is gene duplication that occurs in all organisms. The raw materials of evolution are mutations, such as gene gain, gene loss and single nucleotide changes that alter the protein-coding sequence. The divergence of these genes is driven by a dynamic and diversifying evolution. This process of evolution sometimes gives rise to loss or gain of lineage-specific genes or origin of multi-gene families (Zdobnov et al., 2002; Wu et al., 2009). In evolution, gene duplication is also a powerful driving force that plays significant roles leading to the formation of lineage-specific new duplicated genes (NGD) (Fan et al., 2008, Audemard et al., 2012; Pegueroles et al., 2013).

Gene duplication event generates duplicate copy of a gene through retroposition, unequal crossing over or chromosomal (or genome) duplication. Of all the types of gene duplication known, tandem duplication is one of the major mechanisms involved in the origin, expansion, and evolution of multigene families and also in the formation of lineage-specific tandem arrays. This tandem array or gene cluster shows a different

degree of divergence most probably the functional divergence. This may contribute to the novel function of the new gene (Tachida and Kuboyama, 1998; Walsh and Stephan, 2008; Jiang et al., 2013; Wang et al., 2013).

The evolutionary fate of duplicated gene can be broadly categorized into pseudogenization, function redundancy, subfunctionalization, and neofunctionalization. Pseudogenization leads to the loss in function due to the accumulation of degenerative or deleterious mutations over time like an introduction of a stop codon that causes the formation of immature mRNA or change in protein domain structure. Functional redundancy is a situation when newly duplicated genes perform the same function as of original gene and during evolution acquire symmetrical but not significant difference and rapidly fixed in the genome. It is also possible that the 2 copies can evolve symmetrically and accumulate different mutations leading to the splitting of roles (that had previously been performed by the original gene), a process is known as subfunctionalization (Force et al., 1999; Wagner, 2000; Gu et al., 2002). Gene duplication followed by purifying selection on one copy to retain its original function, whereas duplicated copy is allowed to evolve freely and can acquire new functions referred as neofunctionalization (Zhang et al., 1998). Freely evolving gene has accelerated rate of evolution leading to accumulation of asymmetric mutations in the direction of novel functions (Hurles, 2004; Zhang, 2003; Pegueroles et al., 2013). As stated above, one of the major consequences of gene duplication is the birth and evolution of multigene family and lineage-specific tandem array. Evolution of immune genes is dominated by expansions and contractions of gene families (Waterhouse et al., 2007). Immune genes like PGRPs, GNBPs, defensin and TEP are the most labile components of insect immunity showing rapid gene birth–death dynamics and lineage-specific gene families (Waterhouse et al., 2007). Hence, more studies are needed to better understand the evolutionary dynamics between hosts immunity and pathogens as pathogen exerts the strongest selective pressure on components of the immune system, for example, as seen in *Drosophila* Relish. In Relish, the positive selection sites were located in the caspase cleavage site that is required to activate Relish (Sackton et al., 2007; Viljakainen et al., 2015). Heme peroxidases belong to multigene family and genes belong to this family has very diverse roles ranging from development to immunity (Neafsey et al., 2015; Kumar et al., 2010; Oliveira et al., 2012; Ramsey et al., 1982; Kajla et al., 2016b, 2016c, 2017). The recent release of the genomes of 16 *Anopheles* species (Neafsey et al., 2015) offered the unique opportunity to get insights into the function and evolution of heme peroxidase gene family in anopheline mosquitoes. There

are number of heme peroxidases gene present in these 16 *Anopheles* species. These genes are phylogenetically related and have diverged gradually. It is reported that in *A. gambiae* this gene family has 18 heme peroxidases and out of these, 3 pairs of heme peroxidases are tandemly duplicated paralog (HPX15/HPX14, HPX10/HPX11, HPX7/HPX8) (Kajla et al., 2015a; Hughes, 2012).

Gaps in Existing Research

In spite of intensive research in this field, only particular *Anopheles-Plasmodium* interactions have been studied. Most of the interaction studies were carried out in *A. gambiae*. It is the major African malaria vector which predominantly transmits *P. falciparum*. In India, *A. stephensi* is one of the major malaria vectors which transmit both *P. falciparum* and *P. vivax*. It has been reported that the interaction of different *Plasmodium* varies with different *Anopheles* (Molina-Cruz et al., 2015)*.* The knowledge of gut immunity and immune active molecules (heme peroxidases) in Indian mosquitoes are least explored. Thus, parallel understanding of general mosquito gut immunity in major Indian malaria vectors is a prerequisite in developing malaria transmission-blocking strategies. This requires identification of potent mosquito gut molecules (heme peroxidases) which are specifically exhibiting a dynamic behavior (induction or suppression) in the presence of blood and/or microbial antigens. A novel vaccine target should have a conserved function in at least major malaria vectors so that it can be used for widespread control. Hence, it is required to study the function of already reported immune molecules of *A. gambiae* in the major Indian malaria vectors. Further, manipulation of these molecules by using modern methodology will explore their role in the regulation of *Plasmodium* development.

For exploring bacteria to inhibit *Plasmodium* development, it is required to study their interaction with midgut immune molecules. The mechanism that maintains microbial homeostasis inside the *Anopheles* midgut requires more intensive studies. The study of the *Anopheles* gut microbiota and efforts to characterize mosquito-parasite interactions in greater depth to develop new methods for controlling disease transmission has recently emerged as an important field. A deeper understanding of mosquito-microbiota interactions may highlight mechanisms by which bacteria can be utilized to block malaria transmission.

Research Objectives

The objectives of this research work are to isolate, characterize heme peroxidases (HPX2 and Duox) from Indian malaria vector and to understand their role in the regulation of *Plasmodium* and blood-borne antigens. Along with this, we also aimed to study the gene duplication event between heme peroxidases HPX15 and HPX14 in *Anopheles.* Studying and exploring the function of HPX14 might strengthen the potential of HPX15 as a vaccine candidate.

The proposed research objectives are:

Objective 1:   Molecular cloning and characterization of the heme peroxidases HPX2 and Duox from *Anopheles stephensi* genome.

Objective 2:   Functional characterization of heme peroxidase genes (HPX2 and Duox) and their role in bacterial homeostasis and *Plasmodium* development in the midgut of *Anopheles stephensi*.

Objective 3:   Studying the divergent expression of anopheline lineage-specific duplicated heme peroxidase genes (HPX15 and HPX14).

Chapter 2

Materials and Methods

2. Materials and Methods

2.1 Rearing of mosquitoes

Anopheles stephensi mosquitoes were reared in insectory at 28°C, 80% relative humidity (RH) and 12 h light: dark cycle as described before (Kovendan et al., 2012, Kajla et al., 2016a). Larvae were fed on a 1:1 mixture of dog food (Pet Lover`s crunch milk biscuit, India) and fish food (Gold Tokyo, India). Adult mosquitoes were regularly allowed to take 10% sucrose solution. For colony propagation, four to five days old females were fed on anesthetized mice and their eggs were collected in moist conditions (Day, 2016, Kajla et al., 2016b). The hatched larvae were floated in water to continue the life cycle. The mice were maintained in the central animal facility and all the procedures were approved by the Institutional Animal Ethical Committee, BITS-Pilani, Rajasthan under the protocol No: IAEC/RES/18/02.

2.2 Retrieval of heme peroxidase genes from *A. stephensi* genome

The heme peroxidase (HPX) genes that are reported as important molecules of midgut immunity against blood-borne antigens in *A. gambiae* (Kumar et al., 2004, 2010; Oliveira et al., 2012) were retrieved from *A. stephensi* genome (taxId: 30069) available at NCBI. In brief, *A. gambiae* heme peroxidase (AgHPX) genes listed in **Table 2.1** were individually blasted against *A. stephensi* genome to retrieve putative contig that contains the ortholog of respective AgHPX gene. Contig listed as best hit of BLAST result was further considered to retrieve the putative heme peroxidase gene in *A. stephensi* genome. For that, the obtained contig was analyzed using Augustus software (Stanke et al., 2008) to predict the HPX gene. *A. gambiae* heme peroxidase gene ortholog was identified in *A. stephensi* genome and respective contigs are listed in **Table 2.1** with SuperContig and Ensembl identifier in annotated genome.

Table 2.1: List of heme peroxidase genes. GeneID of *A. gambiae* heme peroxidases (HPX15, HPX14, HPX2 and Duox) and their respective orthologs in *A. stephensi* along with their contig number, supercontig and Ensemble identifier are listed in the table.

S. No.	*A. gambiae* Heme peroxidase (GeneID)	Heme peroxidase Ortholog in *A. stephensi*	Genomic location in *A. stephensi* genome retrieved from BLAST result		
			Contig number	Supercontig	Ensembl identifier
1	AgHPX2 (AGAP009033)	AsHPX2	7145	KB664566	ASTE003848
2	AgDuox (AGAP009978)	AsDuox	2339	KB664518	ASTE003295

3	AgHPX15 (AGAP013327)	AsHPX15	5284	KB665221	ASTE008179
4	AgHPX14 (AGAP010810)	AsHPX14	5285	KB665221	ASTE008179

2.3 Designing of *A. stephensi* heme peroxidases gene-specific primers

From the putative *A. stephensi* contigs obtained above, full-length heme peroxidase genes are predicted using Augustus software (Stanke et al., 2008). The nucleotide sequences of the predicted gene and respective AgHPX gene were aligned to design gene-specific primers (5′ to 3′) and are mentioned in **Table 2.2**. Other primers like immune genes, 16S rRNA universal primers and S7 are also mentioned.

Table 2.2: List of *A. stephensi* primers. The primers used to amplify and sequence the respective *A. stephensi* gene fragment are provided below.

Primer Set	Primer sequence (5′-3′)	Amplified PCR product size from template		Purpose	References
		cDNA (bp)	gDNA (bp)		
HPX15 F1 HPX15 R1	ACTATGAATGGTTGCCAATA GGCCCGGTGAATGTCGAT	428	499	Cloning, dsRNA preparation	Present study
HPX15 F2 HPX15 R2	GAGAAGCTTCGCACGAGATTA GAATGTCGATTGCTTTCAGGTC	329	400	Real time PCR	Kajla et al., 2016b
AsHPX14 F AsHPX14 R	CGAGGCCAAACACTTTCTCT GACCGTAGCGCGTGGAAT	181	181	Real time PCR	Present study
AsDUOX F1 AsDUOX R1	CGAGATCGAGAAGTGTGACG GTGCCAACGTAGGAACAGAA	398	466	Cloning, dsRNA preparation	Present study
AsDUOX F2 AsDUOX R2	TCGGTGGAGAAAATGGTAGC CGTAAGCCTCACGGAAGAAA	409	409	sequencing	Present study
AsDUOX F3 AsDUOX R3	TGCCATTCTGTTCCTGCGTT ACCATTTTCTCCACCGACAC	1180	1395	sequencing	Present study
AsDUOX F4 AsDUOX R4	GCGTTCCTGGACAAGGAGAT GGCGTCACGAATTCTGGTAA	624	624	Real time PCR	Present study
AsDUOX F5 AsDUOX R5	AAGTTCTTCTGGAACGCACA GTCCCTCGATGCGTATCTTC	448	448	sequencing	Present study
AsDUOX F7 AsDUOX R6	CGGTGTGTGGAGAGGGTTTT ATGAGCCACGTCGAAAAGCAG	4463	9479	sequencing	Present study
AsDUOX F8 AsDUOX R8	GAATCGCGACCTAATGCTGG GTAACGCCGGTGATGGTTTG	1179	1460	sequencing	Present study

AsHPX2 F1 AsHPX2 R1	CCGGTTGGATCTGTCGCAAC GCGGCCGTTGCGAACTCGTTGC	445	445	Cloning, dsRNA preparation	Present study
AsHPX2 F2 AsHPX2 R2	GCAGATCCTCGACGGCTA CGTCGTGCATCGTCTGGA	443	443	Real time PCR	Present study
AsHPX2 F3 AsHPX2 R3	CTGCTGCACACACTGTTCCT TGTCTAGCGCGAGCACATC	472	472	sequencing	Present study
AsHPX2 F4 AsHPX2 R4	ATGTGTAGGTCTGTGTTTGTGCA TACTCTTTCGAGGCGATCAA	482	482	sequencing	Present study
AsHPX2 F5 AsHPX2 R5	TTCGCTTCTACAGCACGATG TTAAACGAGTCGCGCAAGAT	706	706	sequencing	Present study
Defensin F Defensin R	AGTCGTGGTCCTGGCGGCTCT ACGAGCGATGCAATGCGCGGCA	298	298	Real time PCR	Dong et al., 2011
PGRP-LC F PGRP-LC R	ACCGTACAGGCTGTAGTTGGA ACTCGAGGAACTTTTCCGACAT	317	317	Real time PCR	Dong et al., 2011
SOCS F SOCS R	CGTCGTACGTCGTATTGCTC CGGAAGTACAATCGGTCGTT	241	306	Real time PCR	Dhawan et al., 2015
NOS F NOS R	ACATCAAGACGGAAATGGTTG ACAGACGTAGATGTGGGCCTT	250	382	Real time PCR	Luckhart et al., 1998.
GNBP F GNBP R	GAGTTCCAGTGGTACACCAACA CTTCGGCAGCAACCAGAT	333	493	Real time PCR	Kajla et al., 2016c
Toll prec F Toll prec R	ACCTGTCGGCGAATCCTTGG TCATCCTTGTGCGAGTACGA	358	358	Real time PCR	Kajla et al., 2016c
Gambicin F Gambicin R	GTGCTGCTCTGTACGGCAGCCG CTTGCAGTCCTCACAGCTATTGAT	344	344	Real time PCR	Kajla et al., 2016c
TEP1 F TEP1 R	GCTATCAAAATCAGATGCGCTATC ATCACAACCGCATGCTTCA	325	325	Real time PCR	Kajla et al., 2016c
HPX8 F HPX8 R	GATCCTTTGCCGATGCGCTCAAT CAGTTCGGGCAGTTTATGTCGCAC	381	381	Real time PCR	Kajla et al., 2016c
NOX5 F NOX5 R	TCATGCATCGCTACTGGAAG CCCGAACTGGTCACACTTGTA	473	473	Real time PCR	Present study
16S rRNA F 16S rRNA R (universal primers)	TCCTACGGGAGGCAGCAGT GGACTACCAGGGTATCTAATCCTGTT	467	467	Real time PCR	Kumar et al., 2010
Pb 28S rRNA F Pb 28S rRNA R	CGTGGCCTATCGATCCTTTA GCGTCCCAATGATAGGAAGA	168	168	Real Time PCR	Brandt et al., 2008
S7 F S7 R	GGTGTTCGGTTCCAAGGTGA GGTGGTCTGCTGGTTCTTATCC	487	600	PCR internal loading controls	Vijay et al., 2011

2.4 PCR amplification, cloning and full-length sequencing of heme peroxidase genes

To sequence the putative AsHPX2 and AsDuox genes, gene-specific primers were designed as discussed above. The PCR product was amplified and sequenced to confirm the presence of putative heme peroxidase gene in the retrieved contig of *A. stephensi*. This fragment was further cloned to prepare dsRNA for gene silencing using RNAi. In addition, full-length gene sequence was also amplified using PCR. For this, primers were designed and due to the longer size of cDNA, many sets of primers given in **Table 2.2** were also designed to get the full-length sequence of the gene.

PCR was done using Phusion High-Fidelity DNA Polymerase (Thermo scientific, #F-530S) and 24 h blood fed female midgut cDNA as a template with respective primer set mentioned in **Table 2.2** to get full-length AsHPX2 or AsDuox cDNA sequence. The PCR was initiated at $98^{o}C$ for 30 s and then followed by 35 cycles at $98^{o}C$ for 10 s, $62^{o}C$ for 30 s and $72^{o}C$ for 2 min. The final extension was carried out at $72^{o}C$ for 10 min. The PCR product was purified and sequenced commercially at Eurofins genomics (www.eurofinsgenomics.eu). Due to the longer size of AsHPX2 and AsDuox full-length cDNA, different sets of primers were used for sequencing purpose (as mentioned in **Table 2.2**). The sequence identity of full-length AsHPX2 and AsDuox cDNA was confirmed through BLAST result and sequence was submitted to NCBI Genbank. The sequence accession numbers GenBank: KY363390 for *A. stephensi* AsHPX2 and GenBank: KY386660 for AsDuox were obtained.

2.5 Malaria parasite *Plasmodium berghei* maintenance and infection to mosquitoes

P. berghei (ANKA strain) was provided by Prof. Asif Mohammad, Scientist, ICGEB, New Delhi, India), and a transgenic *P. berghei* PbGFP, expressing GFP at all developmental stages, was a gift from Dr. Agam Prasad Singh, National Institute of Immunology, New Delhi, India. Both the *Plasmodium* strains were maintained in Swiss albino mice following the protocols as described previously (Dong et al., 2006; Dhawan et al., 2015, Kajla et al., 2017). The parasitemia of the infected mice was determined from blood films stained with Giemsa under a light microscope (http://www.malariasite.com /tag/giemsa-stain/). For blood stage passages, 100-150 µl of blood from an infected mouse (parasitemia ~10%) was injected intraperitoneally into a healthy mouse. Parasitemia in this mouse blood was determined and potential infectivity to mosquitoes was established using exflagellation assays as described before (Billker et al., 1997). In

brief, 2 µl of tail blood was taken into heparinized syringe, immediately mixed with 10 µl of exflagellation buffer pH 8 (10mM Tris-Cl, 150mM NaCl and 10mM Glucose) and 10 µl FBS (10%) placed on a slide and covered with coverslip. Further, this slide was observed under microscope at 40X objective. In all the experiments, mice with 5-7% parasitemia and having exflagellation of 2-3 per field under 40X objectives were used for infecting mosquitoes.

2.6 Mosquito tissue collection

2.6.1 Tissue collection of different developmental stages and body compartments of *A. stephensi*

Different developmental stages of *A. stephensi* such as eggs, first to fourth instar larvae, pupae, adult males, and females were collected separately in RNAlater (Qiagen) and stored at -80°C. In some experiments, *A. stephensi* females were allowed to feed on an anesthetized mouse. After 24 h of blood feeding, the midgut (Mg) and the carcass (CC, rest of the body parts except midgut) were collected from the pool (n=20) of mosquitoes and stored at -80°C. In other experiment, midguts were collected at different time points post blood feeding. Sugar fed midguts and carcasses were also collected in a similar way and served as controls.

2.6.2 Bacteria inoculation in 4th instar larvae of *A. stephensi*

Escherichia coli (Gram negative bacteria, MTCC 40) and *Micrococcus luteus* (Gram positive bacteria, MTCC 106) were procured from MTCC (Microbial Type Culture Collection), Institute of Microbial Technology, Chandigarh, India. Bacterial cultures were grown to an optical density of A$_{600}$=0.5 in LB broth, and 500 µl of each culture was mixed and centrifuged for 5 min at 2300xg. The supernatant was discarded, and the pellet was washed twice with Ashburner's PBS (Phosphate Buffer Saline) pH 7.2 (3mM Sodium chloride, 7mM disodium hydrogen phosphate, and 3mM sodium dihydrogen phosphate). For bacterial injection, the pelleted cells were resuspended in 125 µl Ashburner's PBS and 69 nanoliters were injected into the thorax of 4th instar *A. stephensi* larvae using a nanoinjection system (Nanoject II microinjection system; Drummond).

2.6.3 *A. stephensi* bacterial feeding and tissue collection

The two bacterial strains (*E. coli* and *M. luteus*) were individually grown in LB media and pelleted in the same manner as described in section 2.6.2. The pellet was resuspended

either in saline or mouse blood to make the final count of 10^9 cells/ml (finally 0.5×10^9 cells of each bacterium per ml). *A. stephensi* adult female mosquitoes were allowed to feed on saline alone (control) or bacteria supplemented saline with the final concentration of 10^9 cells/ml through an artificial membrane feeder as before (Gupta et al., 2009). Similarly, *A. stephensi* adult female mosquitoes were allowed to feed on mouse blood alone or supplemented with a mixture of bacteria through an artificial membrane feeder as before (Gupta et al., 2009). After feeding the midguts were collected at different time points from fully engorged control and bacteria fed females in a similar way as mentioned above. Sugar fed midguts served as controls and were also collected in a similar way.

2.6.4 *P. berghei* infection in *A. stephensi* and tissue collection

Four to five days old and 12 h starved female mosquitoes were fed on anesthetized Swiss albino mouse infected with the wild-type or GFP expressing *P. berghei*. After feeding the females were maintained at 21°C, which is a permissive temperature for *P. berghei* development (Kumar et al., 2010; Dhawan et al., 2015). The mosquitoes fed on an uninfected mouse served as control. Sugar (SF), normal blood (control) or *P. berghei* infected blood (infected) fed midguts were dissected from a pool of mosquitoes (n=20) at different time points of feeding. The dissected midguts or carcasses (rest of the body except midgut) were kept in RNAlater solution (Qiagen) and stored at -80°C. Sugar fed midguts and carcasses served as controls and were also collected in a similar way.

2.7 dsRNA synthesis

A 218-bp fragment of the LacZ gene (control, to nullify the effect of injection and dsRNA) was amplified using the primers (5′ to 3′) F-GAGTCAGTGAGCGAGGAAGC and R-TATCCGCTCACAATTCCACA and cloned into the pCRII-TOPO vector (Gupta et al., 2010). The gene specific primers given in **Table 2.2** were used to amplify the PCR product of heme peroxidase gene. This was purified and processed to add 3′ poly-A overhang and inserted into pCR-II TOPO TA-Vector® (Invitrogen Cat No K46001-01) following the manufacturer instructions. The recombinant plasmid was used to transform high efficiency DH5-α TOPO10 *E. coli* chemi-competent cell provided with pCR-II TOPO TA-Vector® kit. Few white colonies (Blue/white screening) were screened through colony PCR with M13 universal primers present in the pCR-II TOPO TA-Vector® (**Figure 2.1**). The recombinant colony that contains insert was used for plasmid isolation. This recombinant plasmid already had a T7 promoter site at M13F end as

shown in **Figure 2.1**.

T7 promoter at the other end of the fragment was incorporated by amplifying the inserted fragment using the following primers: M13F-GTAAA ACGACGGCCAGT and T7-M13R-CTCG AGTAATACGACTCAC

Figure 2.1: The pCRII-TOPO-TA vector map. The vector map demonstrating the position of the PCR product insert, position of T7 promoter, and location of M13F and M13R primers (https: //tools.thermofisher.com/content/sfs/manuals/topota_man.pdf)

TATAGGGCAGGAAACAGCTATGAC. The PCR reaction was performed as follows: an initial step of 5 min at 94°C, followed by 40 cycles with denaturation at 94°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 30 s. After the final cycle, the product was extended for 10 min at 72°C using M13 F and T7-M13R primers. Amplicons were extracted from the gel with the QIAquick Gel Extraction Kit (cat no 28704, Qiagen, Valencia, CA, USA). PCR-purified amplicons tailed with T7 promoter sequences were used to synthesize dsRNAs (*In vitro* transcription) using the MEGAscript kit (Cat No. AM1626, Ambion, Austin). dsRNA was then purified with the help of Microcon YM-100 filter (Millipore) and finally concentrated to 3 µg/µl in DNase- and RNase-free water.

2.8 Analyzing the effect of gene silencing on the growth of midgut bacteria

One to two days old female mosquitoes were injected with 69 nl of 3 µg/µl dsAsHPX2 or dsAsDuox or dsAsHPX15 (207 ng/mosquito) RNA into their thorax using a nanojector (Drummond, Broomall, PA, USA). Control mosquitoes were injected with LacZ dsRNA in the same manner. In some experiments, these injected mosquitoes were maintained on continuous 10% sugar solution and after 4 days midguts were collected. In other experiments, four days after injection, mosquitoes were allowed to feed on a healthy mouse. Some of the injected mosquitoes were allowed to feed on mouse blood supplemented with a mixture of bacteria (*E. coli* and *M. luteus*) (final concentration of 10^9 bacteria/ml of blood) using an artificial feeder. Midguts were collected at 6 h and 24 h post feeding. The efficiency of RNAi silencing and its effect on the growth of midgut bacteria were analyzed through quantitative PCR.

2.9 Analysing the effect of gene silencing on *Plasmodium* development

One to two days old female mosquitoes were injected with 69 nl of 3µg/µl dsAsHPX2 or dsAsDuox (silenced) or dsLacZ (control) RNA into their thoraxes as mentioned above. Four days after injection, mosquitoes were allowed to feed on mice infected with *P. berghei* (Transgenic PbGFP). The gene silencing efficiency of RNAi method was analyzed in *P. berghei* infected midguts through qPCR against the respective controls. Further to evaluate the effect of gene silencing on *Plasmodium* development, some of the injected females (control and silenced) were allowed to feed on transgenic GFP-*P. berghei* infected mice and were continuously maintained on 10% sugar solution. Seven days post infection number of oocysts per midgut was determined. For that, midguts were dissected in Ashburner's PBS, fixed for 15 min with 4% paraformaldehyde, washed thrice in PBS and mounted on glass slides in Vectashield (Vector Laboratories) mounting medium. The numbers of green fluorescent oocysts were counted in each midgut under a fluorescent microscope (Olympus).

2.10 RNA isolation and cDNA preparation

Total RNA was isolated from above-mentioned tissue samples using RNAeasy mini kit from Qiagen (Cat no. 74104) with slight modification, adding 30 µl β-mercaptoethanol (2-ME) per 1 ml of RLT buffer given in the kit. First-strand cDNA was synthesized using Quantitect reverse transcription kit (Qiagen Cat no. 205311) following manufacturer's instructions.

2.11 Expression analysis of heme peroxidase and immune genes in different tissue samples using real time PCR

The mRNA expression analysis of heme peroxidases and other immune genes were carried out through real time PCR using SYBRgreen supermix in an IQ5 multicolor real-time PCR detection system (Bio-Rad), where ribosomal protein subunit S7 mRNA was used as internal loading control for normalization as described before (Kumar et al., 2010; Salazar et al., 1993; Kajla et al., 2016b). The primers set used for heme peroxidase and other immune genes for qPCR are mentioned in **Table 2.2**. PCR cycle parameters were as follows: an initial denaturation at 95°C for 5 min, 40 cycles of 20 s at 94°C, 30 s at 57°C, and 50 s at 72°C. Fluorescence readings were taken at 72°C after each cycle. For final extension, incubation at 72°C for 10 min was completed and then subjected for a melting curve, to confirm the identity of PCR product. Fold values of

mRNA expression were calculated against respective controls using $^{\Delta\Delta}$Ct method as described before (Livak and Schmittgen, 2001).

2.12 Analysis of conserved domains in heme peroxidases

Conserved domains in full-length sequences of AsHPX2 and AsDuox were identified using the Conserved Domain Database (CDD) search tool available online at NCBI (Marchler-Bauer et al., 2015) and SMART database (Letunic et al., 2014). The result was further compared with the respective ortholog of *A. gambiae* to compare the conserved domains present in the heme peroxidase gene.

2.13 Selection of heme peroxidases for phylogenetic analysis of AsHPX2 and AsDuox

To construct the evolutionary relationship, the sequences of heme peroxidase proteins from various species were downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/) and VectorBase database (https://www.vectorbase.org) and separate phylogenetic trees for HPX2 and Duox proteins were constructed. GenBank accession numbers of the protein sequences used for tree construction are given below in **Table 2.3**.

Table 2.3: List of peroxidases from diverse organisms selected for phylogenetic analysis. The heme peroxidases protein sequences of various organisms were obtained from NCBI and Vectorbase database.

List of peroxidase used to construct phylogeny with AsHPX2		
Organisms name (Abbreviation)	**Peroxidase nomenclature**	**Gene ID/ Accession Number**
A. stephensi	AsHPX2	KY363390
A. sinensis	AsiHPX2	ASIS008239
A. gambiae	HPX1	AGAP010734
	HPX2	AGAP009033
	HPX3	AGAP003714
	HPX4	AGAP007237
	HPX5	AGAP000051
	HPX6	AGAP003502
	HPX7	AGAP004036
	HPX8	AGAP004038
	HPX10	AGAP013282
	HPX11	AGAP010899
	HPX12	AGAP010735
	HPX14	AGAP010810
	HPX15	AGAP013327
	HPX16	AGAP011216

Aedes aegypti	Chorion Peroxidase	AAEL004386
	HPX2	AAEL013171
	Peroxinectin	AAEL003612
	Peroxidasin	AAEL000342
Aedes albopictus	HPX2	AALF004168
Culex quinquefasciatus	Chorion Peroxidase	CPIJ007711
	Chorion Peroxidase	CPIJ018105
	Chorion Peroxidase	CPIJ005949
	Peroxidase	CPIJ007579
	Peroxidase	CPIJ017588
	Peroxidase	CPIJ007710
	Thyroid Peroxidase	CPIJ016742
	HPX2	CPIJ001764
Drosophila melanogaster	Peroxidase	CG4009
	Peroxidase	CG5873
	Peroxidase	CG10211
	Peroxidase	CG42331
	Peroxidase (IRC)	CG8913
	Chorion Peroxidase	CG3477
	Peroxidasin	CG12002
	Peroxinectin	CG7660
Pediculus humanus corporis	Chorion Peroxidase	PHUM184790
	Peroxidase	PHUM103320
Tribolium castaneum	Peroxidase	GA19195
	Peroxinectin	CG7660
Caenorhabditis elegans	Peroxidase	CELE_K10B4.1
Arabidopsis thaliana	Peroxidase1	AAA32849
Homo sapiens	Eosinophil peroxidase	EPO
	Myeloperoxidase	MPO
	Thyroid peroxidase	TPO
	Lactoperoxidase	LPO
Ixodes scapularis	Peroxinectin	ISCW002680
Apis mellifera	Peroxidase	LOC724541
	Chorion Peroxidase	LOC412013
	Peroxidasin	LOC413025
	Peroxnectin	LOC551544
	Peroxidase	LOC408953
Sarcoptes scabiei	Peroxidase	KPM11066

List of orthologs of Duox used to construct phylogenetic tree		
Class	**Organisms name (Abbreviation)**	**Gene/sequence ID**

31

Diptera			
	A. stephensi		KY386660
	A. gambiae		AGAP009978
	Aedes aegypti		AAEL007563
	Culex quinquefasciatus		CPIJ003117
	Drosophila melanogaster		AAF51201.2
	Bactrocera dorsalis		AKS43593.1
	Lucilia cuprina		KNC33589.1
Bees	*Apis melifera*		FAA00352.1
	Melipona quadrifasciata		KOX68568.1
Lepidoptera	*Bombyx mori*		AFV61649.1
	Papilio machaon		KPJ08497.1
	Papilio xuthus		KPI94366.1
Ant	*Lasius niger*		KMQ97574.1
	Acromyrmex echinatior		EGI68387.1
Crustacea	*Daphnia magna*		KZS20913.1
	Daphnia pulex		EFX77976.1
	Marsupenaeus japonicus		BAM76968.1
	Pediculus humanus corporis		XP_002429864.1
	Ixodes scapularis		EEC10543.1
	Anasa tristis		AFK29281.1
Nematodes	*Caenorhabditis elegans*		NP_490684.1
	Meloidogyne incognita		AAY84711.2
Vertebrates	*Homo sapiens*	Duox1	AAI14939.1
		Duox2	EAW77288.1
	Sus scrofa	Duox1	AAN39338.1
		Duox2	AAN39339.2
	Pteropus alecto	Duox1	ELK05460.1
		Duox2	ELK05457.1
	Mus musculus	Duox1	NP_001092767.1
		Duox2	NP_808278.2
	Macaca fascicularis	Duox1	EHH63050.1
		Duox2	XP_005559477.1
	Cricetulus griseus	Duox1	ERE70908.1
		Duox2	ERE70901.1

To retrieve putative Duox from other anophelines, we performed BLAST of AsDuox full-length mRNA against their (anopheline species) whole genome shotgun (WGS) sequences available at NCBI. The contig with the best match was analyzed through Augustus software to obtain the predicted Duox gene and aligned with AsDuox by Clustal Omega. The predicted protein sequences are mentioned in **Table 2.4**. The phylogenetic tree was constructed using the full-length protein sequences of these putative *Anopheles* Duox as discussed below.

Table 2.4: List of putative Duoxes retrieved from different species of *Anopheles*. The putative Duox contig (or Ensembl identifier) from the genome of different anophelines were retrieved using nucleotide sequences of AsDuox clone (GenBank: KY386660) as a query.

Vectorial capacity	*Anopheles* species (abbreviation)	Retrieved Contig	Ensembl identifier or gene bank identity	Duox protein (amino acids)
Major vectors	*A. stephensi* (As)	cont 2339	KY386660	1475
	A. gambiae (Ag)	-	AGAP009078	1475
	A. arabiensis (Aarb)	cont1.662	AARA003798	1475
	A. atroparvus (Aatp)	cont1.8099	AATE007950	1475
	A. culicifacies (Acul)	Cont1.1826	ACUA025429	1475
	A. darlingi (Adar)	Cont8040	ADAC009476	1475
	A. dirus (Adir)	cont1.4135	ADIR009419	1475
	A. funestus (Afun)	cont1.1257	AFUN010730	1475
	A. maculatus (Amac)	cont 1.31810/1.6591	AMAM017333	1475
	A. sinensis (Asin)	cont016143	not available	1475
Minor vectors	*A. albimanus* (Aalb)	cont1.540	AALB000205	1475
	A. epiroticus (Aepi)	cont1.2810	AEPI007127	1475
	A. melas (Amel)	cont2.624	AMEC008242	1475
	A. merus (Amer)	cont2.3241	AMEM009260	1475
	A. minimus (Amin)	cont1.3730	AMIN006748	1475
Non-vectors	*A. quadriannulatus* (Aqua)	cont1.9713	AQUA009090	1475
	A. christyi (Achr)	cont1.905	ACHR003065	1475

Phylogenetic trees were constructed using the neighbor-joining (NJ) method implemented in MEGA 5.2 program as described before (Saitou and Nei, 1987). We aligned all selected protein sequences by Clustal W algorithm in the MEGA 5.2 program as before (Tamura et al., 2011). Branching pattern reliability was tested for NJ tree by 1000 bootstrap replicates. The resulting phylogenetic tree was analyzed to explore the evolutionary relationship with other heme peroxidases based on clusters and nodes formed.

2.14 *In silico* analysis of heme peroxidase proteins

The putative signal peptide sequence at the N-terminal of protein was predicted using SignalP software (Petersen et al., 2011). Protein 3D structure and putative substrate

binding sites were analyzed using TASSER (Yang et al., 2015), Phyre² (Kelly et al., 2015) and RaptorX (Källberg et al., 2012). Analysis of nature of protein (cytoplasmic, non-cytoplasmic or presence of transmembrane domain) was done using Phobius software that is based on hidden Markov model (HMM) (Käll et al., 2007). The molecular weight and isoelectric pH of heme peroxidase protein were analyzed online at http://web.expasy.org/compute_pi/ (Gasteiger et al., 2005).

2.15 Retrieval, Identification and analysis of tandemly duplicated paralogs (HPX15 and HPX14) in the genome of *Anopheles*

To retrieve the genomic, mRNA and protein sequences of heme peroxidases HPX15 and HPX14 genes from the *A. gambiae* genome, we have used NCBI (http://www.ncbi.nlm.nih.gov/) and Vectorbase database (http://www.vectorbase.org). The 5′ upstream and 3′ downstream regions of genes in *A. gambiae* were also retrieved from the NCBI. To find out the lineage-specific AgHPX15 duplicate gene, protein BLAST against *A. gambiae* database was performed and gene satisfying ≥50% alignment identity and ≥70% alignment coverage at the amino acid levels was considered as paralog generated through gene duplication as discussed previously (Audemard et al., 2012). We also analyzed the genome location of paralogs to understand the mechanism behind their duplication. We defined paralogs as generated by tandem duplication mechanism if both copies are adjacent to each other. To further confirm their tandem duplication, we used two-step strategy, first at genome level and second at the protein level. The paralogs are said to be tandemly duplicated if they are less than T kb (150 kb) apart and each tandem unit (gene) has a minimum length of l bp (500 bp) with alignable units separated by less than L kb (40 kb) apart (Audemard et al., 2012). To further confirm gene duplication at protein level we computed the identity I'= I x min (n_1/L_1, n_2/L_2) where I is the proportion of identical amino acids in the aligned region (including gaps) between the sequence 1 and sequence 2 by the alignment program FASTA, L is the length of sequence, and n is the number of amino acids in the aligned region in sequence and I' > 30% as proposed by Li et al., 2001.

2.16 Analysis of sequence and structure of duplicated heme peroxidase genes in *A. stephensi* genome

The *A. gambiae* heme peroxidase showed duplication (HPX15) and was used as query and BLAST against the whole genome sequence of *A. stephensi* (taxid: 30069) available at NCBI. The retrieved contig (5285) (Supercontig KB665221, Ensembl

identifier ASTE008179 in the annotated genome) was then analyzed to explore the duplicated gene and duplication event in the *A. stephensi* genome. For this, genome sequence was analyzed to understand the exon-intron boundaries, untranslated regions (UTRs) of the duplicated genes with the help software such as Genscan and Augustus as before (Burge and Karlin, 1997; Stanke et al., 2008). The retrieved genes were then analyzed to reveal the tandem duplication event as described above. Post-translational modifications in the proteins were analyzed by programs available at CBU prediction server (http://www.cbs.dtu.dk/services/).

2.17 Bioinformatics analysis of promoter, transcription factors, and regulatory elements for heme peroxidase genes

The 5′ region of HPX15 and HPX14 in both *A. gambiae* and *A. stephensi* was analyzed to locate the promoter sequence in these genes. The 5′ upstream sequences of heme peroxidases gene were analyzed by Genscan (http://genes.mit.edu/GENSCAN.html) (Organism vertebrate and Suboptimal exon cutoff 1), Augustus (http://bioinf.uni-greifswald.de/augustus/submission.php) (organism *D. melanogaster*) and NNPP2.2 (http://www.fruitfly.org/seq_tools/promoter.html) (organism selected Eukaryotes and Minimum promoter score 0.8) software to decipher the promoter region as mentioned before (Reese et al., 2001; Burge and Karlin, 1997; Stanke et al., 2008). In order to decipher putative regulatory motifs within the 5′ upstream region of HPX15 and HPX14 genes of both *A. gambiae* and *A. stephensi* transcription factor binding sites in the 5′ upstream region of genes were predicted by MatInspector (https://www.genomatix. de/matinspector.html) and JASPAR (http://jaspar.genereg.net/cgibin/jaspar_db.pl?rm= browse&db= core&tax_group=insects) software (Cartharius et al., 2005; Mathelier et al., 2013). For this analysis in both the software, the transcription binding sites from insect family were selected with 80% of similarity as threshold value. The promoter region was also analyzed by MEME suits (Bailey et al., 2009) and the obtained motifs were then analyzed by the TOMTOM feature (Gupta et al., 2007) available in the MEME suit by selecting JASPAR insect core database as a reference.

2.18 Analysis of sequence divergence in duplicated genes

Sequence divergence and phylogenetic relationships of *A. gambiae* and *A. stephensi* duplicated heme peroxidases proteins were constructed using Neighbor-Joining (NJ) methods implemented in MEGA 5.2 software with 1000 bootstrap replicate to assess support (Tamura et al., 2011). In addition, we also calculated ka/ks ratio using tools

available at http://services.cbu.uib.no/tools/kaks keeping all the parameters default. The default parameter for Ka/Ks were set at none, means no codon bias is used in the algorithm. The window option is also set at none means no windows are used and the whole sequence is averaged together. To analyze selective forces on each amino acid we used Selecton server version 2.4 and using Null model as Evolutionary model with no positive selection (M8a, beta + w =1) (Stern et al., 2007; Doron-Faigenboim et al., 2005).

2.19 Analysis of chromatin boundary elements and insulators in 5′ and 3′ region of duplicated genes

We used cdBEST software (Srinivasan and Mishra, 2012) to predict boundary elements in the domain that contains the duplicated gene in both *A. gambiae* and *A. stephensi*. The tool was run with a set window size of 1000 bp and window slide of 10 bp across the duplicated domain. The image output width is based on sequence length where 1 pixel is equal to 100 bases.

2.20 Statistical analysis of the data

All the data were expressed as a mean ± standard deviation. Statistical significance between test and respective controls was analyzed by Student's t-test or one-way ANOVA post-Tukey's Multiple Comparison Test using GraphPad Prism 5.0 software (Motulsky et al., 1999). The data with $p < 0.05$ was considered significant. The sigma plot (SigmaPlot 10.0 Systat Software, San Jose, CA) and GraphPad Prism 5.0 software (Motulsky et al., 1999) were used to prepare graphs of data.

Chapter 3

Molecular cloning and functional characterization of the heme peroxidase HPX2 from *Anopheles stephensi* and its role in immunity against blood-borne antigens

3.1. Abstract

In this chapter, we carried out the molecular characterization of heme peroxidase HPX2 from Indian malaria vectors *A. stephensi*. This gene encodes for a secreted peroxinectin-like protein of 692 amino acids. Phylogenetic analysis of AsHPX2 with heme peroxidases of other organisms revealed that its orthologs are present only in mosquitoes. In AsHPX2 silenced sugar fed midguts, increased bacterial growth is observed against control. Blood feeding suppressed the expression of AsHPX2 gene, however, the bacteria supplemented blood feeding induced its expression in the midguts. In AsHPX2 silenced bacteria supplemented blood fed midguts, increased bacterial growth against control is observed. Collectively, AsHPX2 gene is one of the molecules that maintain bacterial homeostasis in the sugar fed midguts and its reduced expression in blood fed midguts may create a physiological condition that allow proliferation of gut bacteria as they are required for the digestion of blood. Silencing of AsHPX2 gene increased the *Plasmodium* oocysts number and hence, is an anti-plasmodial molecule. This is in agreement with anti-plasmodial role of AgHPX2 in the *A. gambiae*. Thus, AsHPX2, a mosquito-specific gene can be targeted to arrest *Plasmodium* development inside the mosquito.

3.2. Introduction

The completion of a sexual cycle of malaria parasite inside the mosquito is a complex process. During this process, parasite interacts with a large number host molecules that are beneficial for its development called agonists. On the other hand, the successful development of the parasite in the mosquito vector is also determined by its interaction with host immunity. *Plasmodium* overcomes the several immune responses to complete its life cycle. Identification of these immune mechanisms will be helpful in arresting *Plasmodium* development inside the vector. This is the basic strategy to develop transmission-blocking vaccines (TBV).

In this aspect, mosquito-derived molecules that support the development of parasite such as carboxypeptidase A/B, alanyl aminopeptidase N, Fibrinogen-related protein 1 (FREP1) and heme peroxidase HPX15 have been proposed as transmission-blocking vaccine (TBV) candidates (Raz et al., 2013; Armistead et al., 2014; Kajla et al., 2015a; Niu et al., 2017; Venkat Rao et al., 2017). Another way for developing transmission blocking strategies is to target those molecules which are the antagonist to the *Plasmodium*. We mainly focus on those molecules which are active in antiplasmodial immunity during midgut invasion by ookinetes.

Therefore, in this study, we explored the role of heme peroxidase HPX2 in *A. stephensi* because in *A. gambiae* its ortholog AgHPX2 acts as an anti-plasmodial molecule (Oliveira et al., 2012). Thus, exploring the role of *A. stephensi* HPX2 in the migdut immunity against bacteria and *Plasmodium* may open up news ways to target this molecule for the widespread control of malaria.

3.3. Results

3.3.1. Cloning and characterization of AsHPX2 gene from *A. stephensi*

3.3.1.1. PCR amplification and full-length sequencing of AsHPX2 gene

To identify AsHPX2 gene in the genome of *A. stephensi,* the PCR-based approach was used. Gene-specific primers were designed as described in the Material and Methods. *A. stephensi* genomic DNA (gDNA) or cDNA (midgut and carcass) was used as a template to amplify AsHPX2 gene. The results presented in **Figure 3.1A** showed the amplification of a single 450 bp fragment from each template. This fragment was sequenced and BLAST result confirmed that it has similarity with *A. gambiae* HPX2 gene (E value 7e-54, identity 84%). The partial clone of 450 bp confirmed the presence of HPX2 in *A. stephensi* genome. Then the full-length AsHPX2 gene was predicted using bioinformatics approaches. In brief, the AgHPX2 gene was BLAST against *A.*
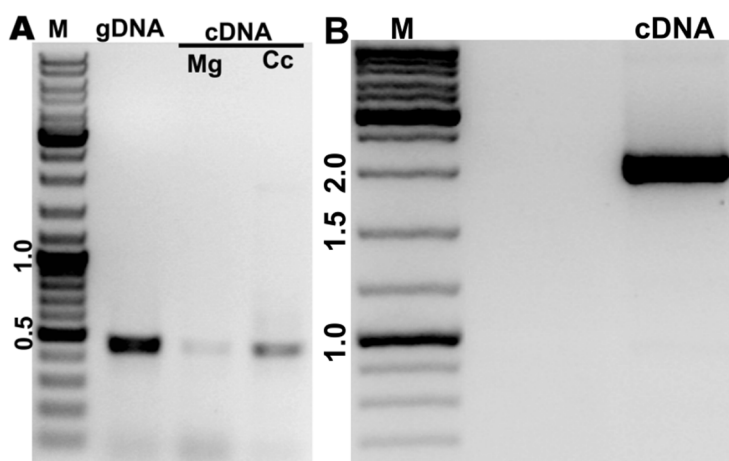
Figure 3.1: PCR amplification of *A. stephensi* AsHPX2 gene. **(A)** PCR amplification of *A. stephensi* partial AsHPX2 gene segment using F1R1 primers and genomic DNA or cDNA (450 bp) as a template. **(B)** PCR amplification of full-length AsHPX2 using F4R5 primers and cDNA of 24 h blood fed midguts as a template. M represents the DNA ladder in kb used as a reference for identifying the product size.

stephensi genome to retrieve the putative Contig that contains AgHPX2 ortholog. The AsHPX2 gene is present in contig 7145 (SuperContig KB664566, Ensembl identifier ASTE003848 in the annotated genome of *A. stephensi*).

From the *A. stephensi* contig, full-length AsHPX2 is predicted using Augustus software to reveal the gene structure. The predicted AsHPX2 gene is 2868 bp long. It contains an open reading frame (ORF) of 2079 bp that encodes for a protein of 692 amino acids. The predicted 5′-untranslated region (5′-UTR) is 438 bp and 3′-UTR is 351

bp. The signal for polyadenylation AATAAA was found at the 331 bp downstream from stop codon in 3′-UTR **(Figure 3.2)**. Further, primers for full-length cloning were designed based on the alignment of AgHPX2 and AsHPX2 genes mentioned in **Table 2.2** and PCR product was amplified using cDNA of 24 h blood fed midguts as a template (**Figure 3.1B**). The F4R5 primers amplified the expected PCR product of ~2 kb. The product was sequenced and BLAST against the NCBI nucleotide database. It showed 80% identity with its ortholog AgHPX2 and was submitted to the NCBI (GenBank

```
                                          AP-1                                              GATA
-461 CTCCAACTAACTGGCGGAATCCTAGTGACTGGAGCTGCAAATATGTCAAAGGGTCTGACGATTCGCCAATCTGAACTCAATCGATGGGCAA
                                                    Rel          -310  -54
-373 CTGCACCACACGAGCCAAGCAATTTATTCCTCCGTTTTGGGACGGGTTTCTTTCTTTAAAT // CCATCATCCATCCCGCGCGTACACAAA
                                    TSS          +16  +175                     Broad Complex
-26  AGCGCGAACCGATCCCCGCGGGGTTTAACGCATCGCGCAAGGG // GGAAAACTGCGCGATAATGTTTGGAGTAAACAAAAACATCCATG
                                     +1
+222 CGAAGCGAGAAG // ACTAAAAACCGTTTGAAAAGAGTGAACTTAACCACA
                    +401                                  +438
439  ATGTGTAGGTCTGTGTTTGTGCAGTGCATCGCCGTTGCACTGCTTGCACTGGCAGTGCAGGGCCGTACCACGTCCGGCATAGCGTCCGAT
  1  M  C  R  S  V  F  V  Q  C  I  A  V  A  L  L  A  L  A  V  Q  G  R  T  T  S  G  I  A  S  D
     CAAGTAGATCGATACGAGCTAGATCCACCGCTCGCTCTTCGATTGTCCCGGTTTGCAAGCACATCTAACCTCGAGGAGGCGCTCGATTAC
     Q  V  D  R  Y  E  L  D  P  P  L  A  L  R  L  S  R  F  A  S  T  S  N  L  E  E  A  L  D  Y
619  AGTGAGACAATATTTGAAAAATCCAAACGTCTCGAAGGTTCCCTGGCCGCCGGGTCGCGGGGAAAGGTTGTCAAGGGTGGCTCGACGTAC
 61  S  E  T  I  F  E  K  S  K  R  L  E  G  S  L  A  A  G  S  R  G  K  V  V  K  G  G  S  T  Y
     GCGCAGATCCTCGACGGCTATCCAACGCCGAGCACCCAGCAGGAAGATCTGCTTGCCCGCAAGGTCCTCCGAGGTCTCCTTCTTCGTC
     A  Q  I  L  D  G  Y  P  T  P  S  T  Q  Q  Q  D  L  L  A  R  K  V  L  R  A  S  S  F  F  V
799  AACCGCCACTGCATCCCGAATGGCATTTCTTCGTACGAGTGTGGCCTGTATCTCAGCGGAACGCAGCTTCCCTCCTCGACGATCGGAGAT
121  N  R  H  C  I  P  N  G  I  S  S  Y  E  C  G  L  Y  L  S  G  T  Q  L  P  S  S  T  I  G  D
     CAGTGTCGAAGTTTGATCGCCTCGAAAGAGTACAACGATGAGTATCGCCGACTGTTGCCCGCACTGTACGACGATGGAGTGTACGCGTTC
     Q  C  R  S  L  I  A  S  K  E  Y  N  D  E  Y  R  R  L  L  P  A  L  Y  D  D  G  V  Y  A  F
979  CGGCGCAGTACGACCGGGGCTGACGGGGCTGAGTTGCCACTCGCGCGCCAAGTGTCCAGCCGGTTCCATACGGCCTGGAAGAACCAGCTGAC
181  R  R  S  T  T  G  A  E  L  P  L  A  R  Q  V  S  S  R  F  H  T  A  W  K  N  Q  E  P  L  D
     AGTACGCGCTCCGTTGCCCTCGTCCAGTGGACGCAGTTCATCGAGCACGATCTCGCTAAAACCACGGTCCAGACGATGCACGACGGGACG
     S  T  R  S  V  A  L  V  Q  W  T  Q  F  I  E  H  D  L  A  K  T  T  V  Q  T  M  H  D  G  T
1159 GACATTGAATGCTGCTCCGGCGACCATGGTCAGCTGTTGCCCCGGTACCGGCATCCTTCCTGCAAACCGATGCAGGTCCCCGAAGACGAC
241  D  I  E  C  C  S  G  D  H  G  Q  L  L  P  R  Y  R  H  P  S  C  K  P  M  Q  V  P  E  D  D
     CCCTACTACCAGACGTACCGGGCAACGTGCTTGAACTACGTGCGCAGTGCCTGTCGCTCGGTGGCGACGGTTGCCATCTGGGCCCGGCA
     P  Y  Y  Q  T  Y  R  A  T  C  L  N  Y  V  R  S  A  L  S  L  G  G  D  G  C  H  L  G  P  A
1339 AACCAGCTCAACCAGGCCACCAACCGGTTGGATCTGTCGCAACTGTACGGCAGTCTGGCGAACGACACGAAGCTGCTGCGCGCAGGCAAA
301  N  Q  L  N  Q  A  T  N  R  L  D  L  S  Q  L  Y  G  S  L  A  N  D  T  K  L  L  R  A  G  K
     GGTGGCCGACTGCGCGCCCAACTGTACGACTCGGCCGAATATCTGCAGCAGGCGGACAACGATCGGCTGTGTGTGACCGATCCCAACCTC
     G  G  R  L  R  A  Q  L  Y  D  S  A  E  Y  L  Q  Q  A  D  N  D  R  L  C  V  T  D  P  N  L
1519 GAAACCGTCTGCTACGATTCCGGCGACACACGGGTAAACGTGAACCCGTACATCACCCTGCTGCACACACTGTTCCTGCGGTCCCACAAT
361  E  T  V  C  Y  D  S  G  D  T  R  V  N  V  N  P  Y  I  T  L  L  H  T  L  F  L  R  S  H  N
     CGGCTAGCGAAACATTTGGCCCTGCTGCGACCGAGCTGGACCGACGATCAGCTGTTTGCCGCGGCCCGGACCGTCAACACGAAACTGTAC
     R  L  A  K  H  L  A  L  L  R  P  S  W  T  D  D  Q  L  F  A  A  A  R  T  V  N  T  K  L  Y
1699 CAGAAGATTGTCCGCGAATGGTTGACGGCGGTGGCAGGCGAACGGGCGACGCGAACCTCGCAACCGCGCCCAGGACGGTCGGGTT
421  Q  K  I  V  R  E  W  L  T  A  V  A  G  E  R  A  T  R  T  S  Q  P  T  A  P  Q  D  G  R  V
     AGCAACGAGTTCGCAACGGCCGCCATTCGCTTCTACAGCACGATGATGCCGGGCGAAATCTCGAACGCGCTCGGATCGTACGATCTGGAG
     S  N  E  F  A  T  A  A  I  R  F  Y  S  T  M  M  P  G  E  I  S  N  A  L  G  S  Y  D  L  E
1879 CATCTCTTCTATCGACCGAAGGATTTGCGCAAACGGGACTATTTCGCCCATCTGGTCGGTTCCGTGCTCGGGCAGAATGCGATGTCGCTC
481  H  L  F  Y  R  P  K  D  L  R  K  R  D  Y  F  A  H  L  V  G  S  V  L  G  Q  N  A  M  S  L
     GACACGGCCTACGTGGACGATATGGCCCATCTCCTGTTCGGCGTGCGCAACGTCGGGCTGGATGTGCTCGCGCTAGACATTCAGCGCGGT
     D  T  A  Y  V  D  D  M  A  H  L  L  F  G  V  R  N  V  G [L  D  V] L  A  L  D  I  Q  R  G
2059 CGCGACCATGGTCTCGCCCGCTACACCGACTACTACACGGTTGTGCAACGGCAAACCGGTCACCGGGTGGGCCGACCTCGAGCGGGCGCTC
541  R  D  H  G  L  A  R  Y  T  D  Y  Y  T  L  C  N  G  K  P  V  T  G  W  A  D  L  E  R  A  L
     AAACCGGACGATTTGGAAATGATACGCGATTCGTACGCAACCGTGCACGATGTCGATCTGATCGTCGGTGCCATCGCCGAATGGCCCGTT
     K  P  D  D  L  E  M  I  R  A  S  Y  A  T  V  H  D  V  D  L  I  V  G  A  I  A  E  W  P  V
2239 GCGGGTGCTACCGTTGGACCTACGCGTGTCCTGTCTTATCCGAGAACAAATCGACAACTCGCTGGACAAGGCGCACGAGCCACAAGACAGT
601  A  G  A  T  V  G  P  T  L  S  C  L  I  R  E  Q  I  D  N  S  L  D  K  A  H  E  P  Q  D  S
     GCGACACCCGTAGCACTGCGGGCGGACCGCGTGGAGGCGGTGTTGGCGGAGTACAGTGCAGCAAGGTTTATGTGCGATACGGCACAGGTC
     A  T  P  V  A  L  R  A  D  R  V  E  A  V  L  A  E  Y  S  A  A  R  F  M  C  D  T  A  Q  V
2419 GATCGCGTTCAGCGTGATATTTTCCGCGTCCATCGGACGATAACCCGCAGATCCGGTGCGCACAGTTGCCAGCGCTCGATCTTGCGCGA
661  D  R  V  Q  R  D  I  F  R  V  P  S  D  D  N  P  Q  I  R  C  A  Q  L  P  A  L  D  L  A  R
2509 CTCGTTTAA
691  L  V  *                    2537 // 2834
2518 ACCGCTACTCGGCGACGGC // CGTTCTGATGCATTTAGTGGAAATAAAGTAAGTTAATGAAT 2868
                                      Poly-A                          TTS
```

Figure 3.2: The nucleotide and deduced amino acid sequences of AsHPX2 gene. The cDNA (2868 bp) encodes a protein of 692 amino acid residues (residue number mentioned on the left). The figure depicts the amplified CDS and predicted UTRs. The predicted promoter, start codon (ATG) and stop codon (TAA) are highlighted in gray. Transcription start site (TSS) is denoted by +1. Polyadenylation (Poly-A) and transcription termination site (TTS) are indicated with an underline. Binding sites for various transcription factors such as GATA/Rel, Broad complex and AP-1 are depicted in the regulatory region of AsHPX2 gene as predicted by JASPAR software. The integrin binding motif, LDV is highlighted.

database accession number: KY363390). The AsHPX2 cDNA was aligned with the *A. stephensi* genomic sequence (contig 7145), to determine the exon-intron boundaries and revealed no intron is present in this gene. Sequence comparison with its ortholog AgHPX2 of *A. gambiae* revealed similar gene structure (**Figure 3.3**). Further analysis of AsHPX2 and AgHPX2 cDNA, revealed that AsHPX2 encodes for a polypeptide protein that is 20 amino acids longer than AgHPX2.
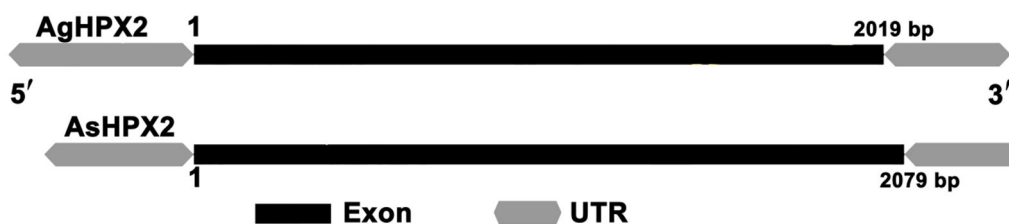
Figure 3.3: Genomic organization of HPX2 gene. HPX2 gene of both *A. gambiae* and *A. stephensi* have single exon of 2019 bp and 2079 bp, respectively. The 5′ and 3′ UTR of AgHPX2 gene are 372 bp and 351 bp while in AsHPX2 gene they are 438 bp and 351 bp, respectively.

3.3.1.2. Sequence and domain analysis of AsHPX2 protein

The complete domain structure of AsHPX2 protein was analyzed by the SMART program (**Figure 3.4**) and it was categorized as an animal heme peroxidase. The deduced amino acid sequence has a signal peptide of 21 amino acid residues in the N-terminus with the signal peptidase cleavage site located between Gly_{21}-Arg_{22} residues. This suggested that AsHPX2 is a secreted protein. The deduced amino acid sequence had a predicted molecular mass of 76.9 kDa and an isoelectric point of 6.09. The secondary structure prediction by Phyre^2 software revealed that it is composed of 47% of alpha helix (**Figure 3.5**). Since, the three-dimensional structure can be a valuable indicator of protein structure and function (Edwards and Cottage, 2003), thus, we built the three-dimensional structure of AsHPX2 using TASSER server which selected lactoperoxidase of buffalo (PDB ID 3FAQ) as a template. The quality of the predicted model is assessed by TM-score and normalized Z-score and are 0.70±0.12 and 2.62, respectively (**Figure 3.6**). This suggested that the predicted model is likely to have correct fold (Xu and Zhang, 2010). The active site of the protein includes amino acid positioned at Thr_{212}, Ala_{430}, Gly_{602}, and Arg_{669}. The conserved domain database (CDD) analysis of AsHPX2 protein revealed the presence of LDV, an integrin binding motif, indicating that it is a peroxinectin like protein. The 5' region near promoter was analyzed using JASPAR to mark putative transcription factor binding motifs (TFBM) for understanding the transcriptional regulation of AsHPX2. The 5′ regulatory region contains the binding site for GATA, Rel, AP-1 and Broad complex (Br-C) (**Figure 3.2**).
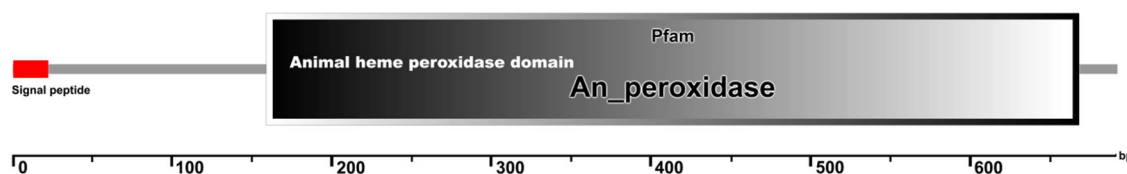
Figure 3.4: Domain organization of AsHPX2 protein. The protein sequence of AsHPX2 is analyzed with the help of SMART server (http://smart.embl-heidelberg.de/). The AsHPX2 protein has an animal heme peroxidase domain and a signal peptide.
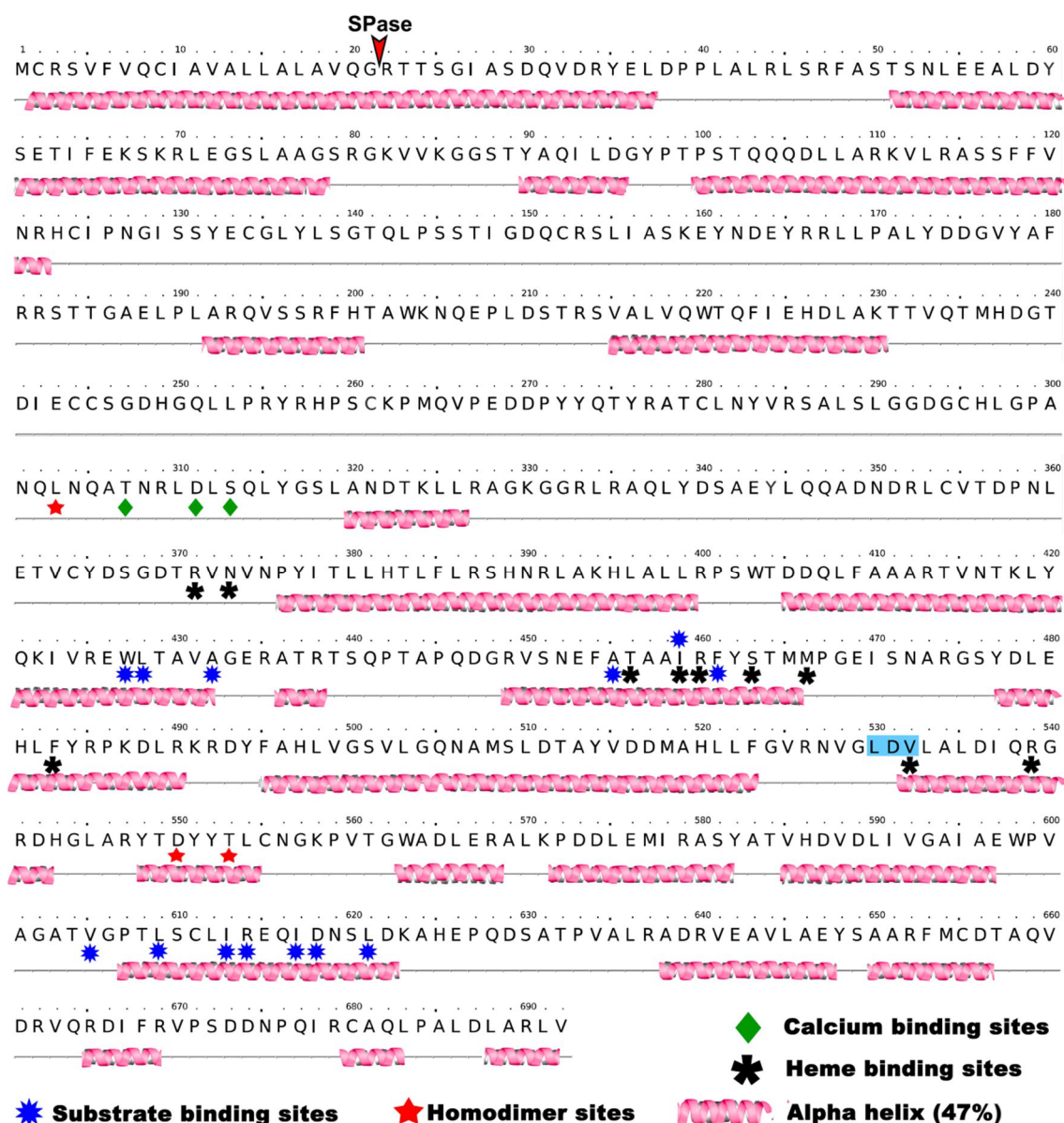
Figure 3.5: The secondary structure of AsHPX2 protein. The AsHPX2 protein contains several alpha helixes (represented by spirals) as predicted by Phyre[2] software. The cleavage site for signal peptidase (SPase) is indicated by an arrowhead as predicted by the SignalP software. Conserved binding sites in the protein are represented by characteristic symbols.
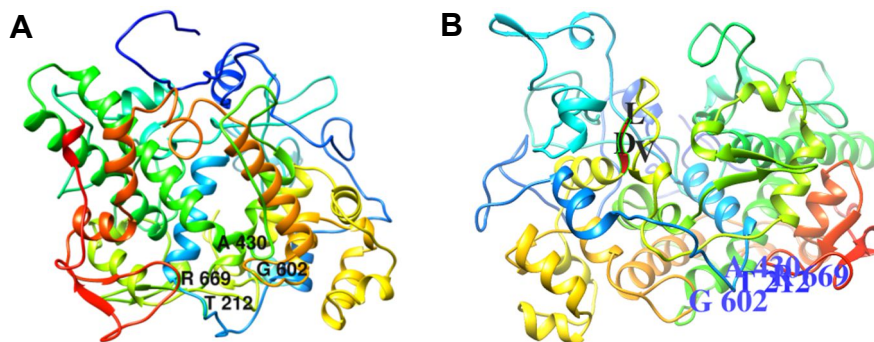
Figure 3.6: The Tertiary structure of AsHPX2. **A)** The 3D structure of AsHPX2 protein and the amino acids at active site are represented by one letter symbol and number indicate their position. **B)** The integrin binding motif, LDV is present on the surface of the protein.

3.3.1.3. Sequence homology of AsHPX2 protein with other heme peroxidases

The AsHPX2 protein sequence was aligned with the protein sequence of heme peroxidases from species such as *Anopheles, Aedes, Culex, Drosophila* and human to understand the sequence conservation. The Clustal Omega alignment result is used to determine the percentage identity with the other heme peroxidases. AsHPX2 shared 79% and 73% similarity with *A. gambiae* HPX2 and *A. sinensis* HPX2, respectively. AsHPX2 shared 57%, 59% and 58% similarity with HPX2 of *A. aegypti*, *A. albopictus* and *C. quinquefasciatus,* respectively. Interestingly, AsHPX2 shared only 35% similarity with immune-related catalase (IRC, it is a heme peroxidase which performs catalase cycle (Ha et al., 2005a, 2005b)) of *D. melanogaster*. The sequence homology with human peroxidases like MPO, TPO, EPO and LPO is less than 38%. The above analysis showed that AsHPX2 shared high sequence homology with other anophelines.

3.3.1.4. Phylogenetic analysis of AsHPX2

To better explain the evolutionary relationships between AsHPX2 and other heme peroxidases, a phylogenetic tree was constructed. AsHPX2 protein sequence was BLAST to access the protein sequence of peroxidases from other organisms. The accession numbers, symbols, and nomenclature of sequences used in the phylogenetic analysis are given in **Table 2.3** (on page 31)**.** The phylogenetic tree showed that HPX2 of different mosquitoes appeared in a single cluster and get diverged from *D. melanogaster* IRC (**Figure 3.7**)*.* Human peroxidases appeared in the separate cluster from the rest of the insect's peroxidases. It is noteworthy to mention that AsHPX2 shared >55% similarity with mosquito HPX2 while <40% similarity with other heme peroxidases. This indicated that its orthologs are present only in the family Culicidae (Mier et al., 2015).
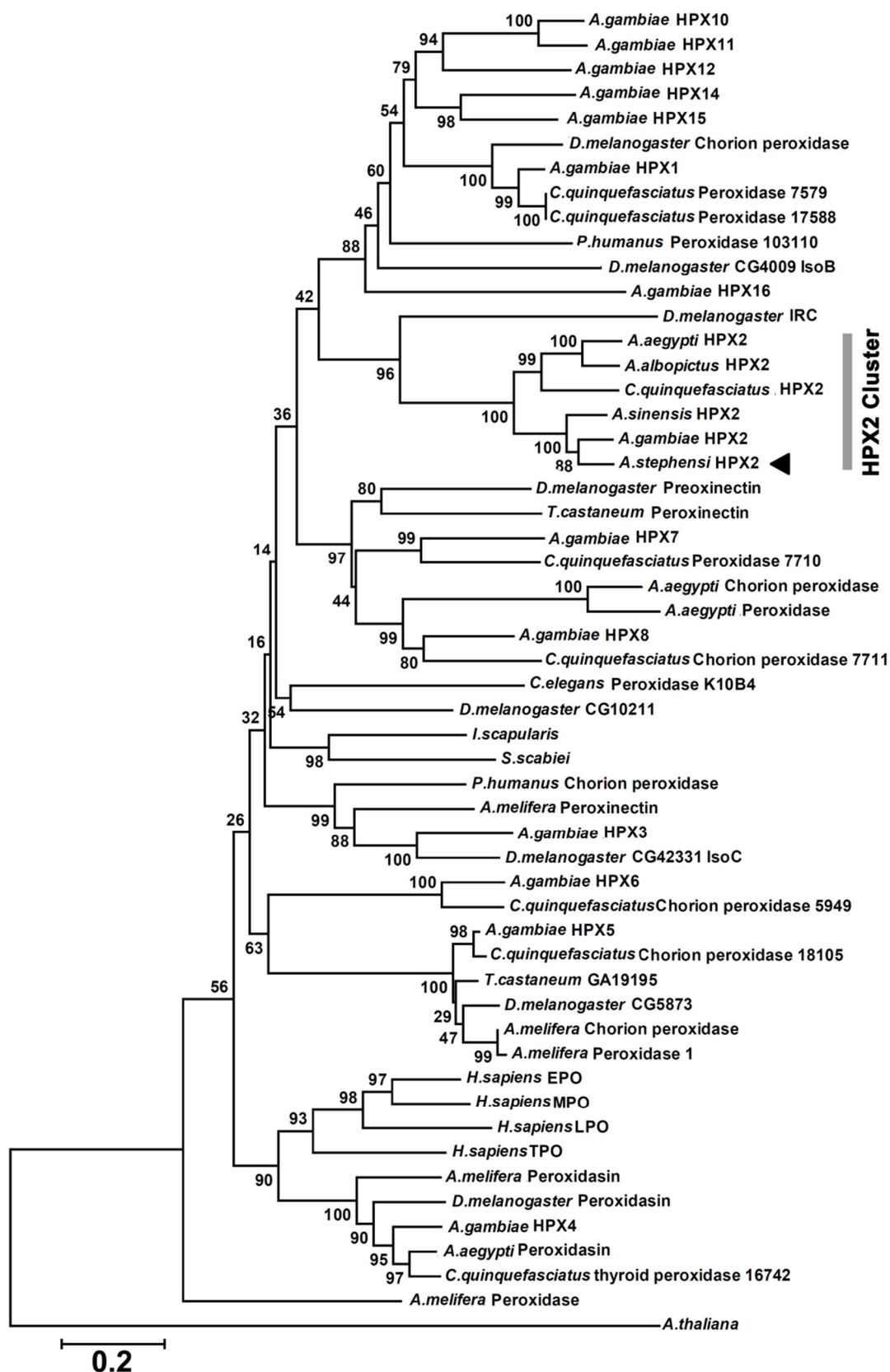
Figure 3.7: Phylogenetic analysis of AsHPX2 protein. The Neighbour-joining (NJ) method was used to construct the phylogenetic tree of AsHPX2 with heme peroxidases (HPX) protein selected from different organisms as mentioned in **Table 2.3** (on page 31). Arrowhead indicates AsHPX2 protein. The vertical gray line indicates the mosquito-specific cluster of HPX2. The scale bar represents base substitutions per site. The numbers on the branches represent the % of 1000 bootstraps.

3.3.2. Expression analysis of AsHPX2 gene

3.3.2.1. Expression profile of AsHPX2 in different developmental stages of *A. stephensi*

The expression of AsHPX2 in the different developmental stages of *A. stephensi* was determined using semi-quantitative real-time PCR. The AsHPX2 gene was expressed throughout all developmental stages of *A. stephensi*; namely, egg, first to fourth instar larvae, pupa, male and female adults. The results presented in **Figure 3.8** showed that the relative mRNA levels of AsHPX2 are 6-fold in first instar larvae, 4-fold in second instar larvae and 2.5-fold in third instar larvae, respectively when compared to the eggs. The mRNA levels are 0.65-fold in the fourth instar larvae and 2-fold in pupae in comparison to the eggs. The mRNA levels of AsHPX2 showed high induction in the adult stage with the highest expression in the adult females. Our analysis showed that expression of AsHPX2 gene increased 26-fold in the males while 84-fold in the females against eggs. Statistical analysis revealed that the relative levels of AsHPX2 mRNA in males and females are highly significant in comparison to other developmental stages (**Figure 3.8**).
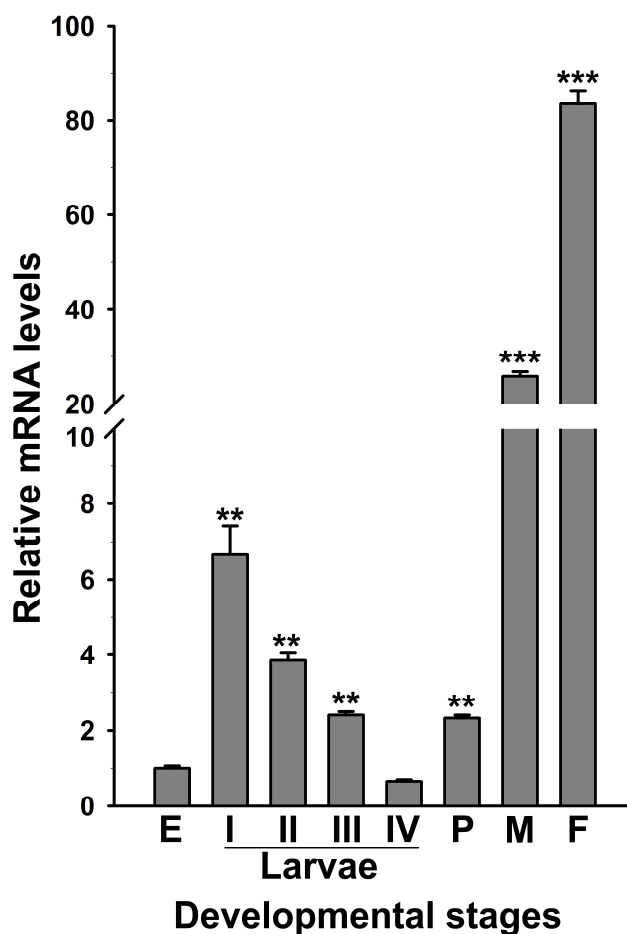
Figure 3.8: Relative mRNA levels of AsHPX2 gene in different developmental stages of *A. stephensi*. Results represented the mean ± SD of relative mRNA levels of AsHPX2 gene in different developmental stages. The mRNA levels in eggs were considered as control or 1.0. Significant differences ($p < 0.001$ or 0.01) among relative mRNA levels of different stages against control are indicated by three or two asterisks (*), respectively. E, eggs; I, II, III, IV, various stages of instar larvae; P, pupae; M, males; F, females.

3.3.2.2. Expression of AsHPX2 in mosquito body compartments

The relative mRNA levels of AsHPX2 gene were analyzed in sugar fed and 24 h post blood fed midguts and carcasses to decipher its tissue-specific expression. The result presented in **Figure 3.9** showed that AsHPX2 expression levels are 33-fold ($p<0.001$) higher in sugar fed carcasses than midguts. However, its expression is reduced in 24 h post blood fed midguts and carcasses. The result shown in **Figure 3.9** revealed that expression of AsHPX2 in the midgut is reduced by 2-fold ($p=0.0367$) while in the carcass it is decreased by 10-fold ($p<0.0001$) when compared to the sugar fed midguts and carcasses, respectively. The downregulation of AsHPX2 gene in blood fed midguts can be explained by the presence of Broad complex binding motifs. The broad complex (Br-C) is a transcriptional regulator (Zhu et al., 2007; Chen et al., 2004) and might be active in the transcriptional regulation of AsHPX2 upon blood feeding as depicted described in other genes such as vitellogenin (Vg) in *A. aegypti* (Chen et al., 2004).
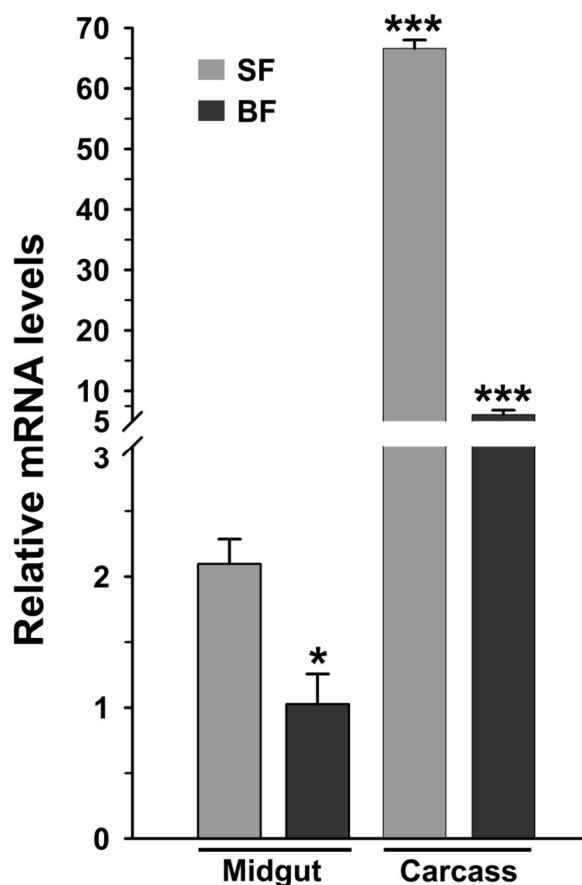
Figure 3.9: AsHPX2 gene expression in different mosquito body compartments. Relative mRNA levels of AsHPX2 in midguts and carcasses of sugar fed or 24 h post blood fed mosquito. Significant differences $p< 0.001$ or 0.05 are indicated by three or one asterisk (*), respectively.

The CDD analysis of AsHPX2 protein revealed its peroxinectin like nature. Peroxinectin is involved in the invertebrate immunity and is a homolog of vertebrate myeloperoxidase. This is involved in the antibacterial responses by selectively binding to the bacteria. This protein kills the bacteria by the enzymatic reaction with H_2O_2 and halide which leads to the formation of an anti-bacterial molecule, hypochlorous acid (HOCl) (Klebanoff, 1968; Allen and Stephens, 2011). Hence, the presence of peroxinectin like domain in AsHPX2 gene might suggest its involvement in anti-bacterial immunity. In conclusion, downregulation of this gene in blood fed midgut may create a physiological condition that allows the proliferation of bacteria as they are required for

the proper digestion of blood meal (Kajla et al., 2015b; Kumar et al., 2010).

3.3.2.3. Blood feeding downregulates AsHPX2 gene in the midguts

To study the expression kinetics of AsHPX2 gene in the midgut after blood feeding, we analyzed its relative mRNA levels using qPCR. The results presented in **Figure 3.10** revealed that AsHPX2 mRNA levels are downregulated 3-fold at 3 h post blood feeding against the sugar fed controls. The mRNA levels of AsHPX2 further reduced with time and showed 2-fold and 20-fold downregulation at 6 h and 9 h post blood meal, respectively. In blood fed midguts, the maximum reduction of 22-fold in the expression of AsHPX2 is observed at 12 h post feeding. The mRNA levels of AsHPX2 showed 10-fold reduction at 18 h and 2-fold reduction at 24 h was observed in blood fed midguts in comparison to the sugar fed midguts. This result indicated that expression of AsHPX2 is more in sugar fed midguts and its mRNA levels reduced after the blood meal. It is noteworthy to mention that bacterial growth is more dominant post 6 h of blood feeding in midguts (Kumar et al., 2010), hence, reduced expression of AsHPX2 gene in blood fed midguts post 6 h correlates with the bacterial growth (relative levels of 16S rRNA) as shown in **Figure 3.10**.
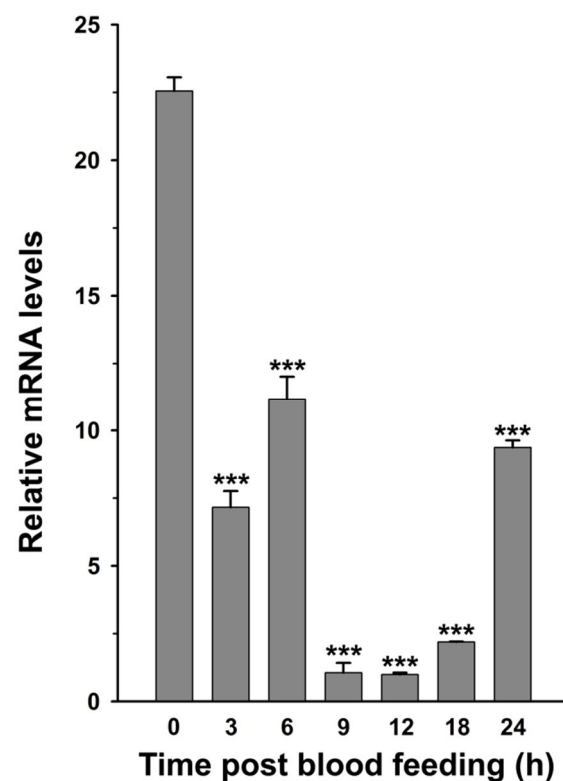
Figure 3.10: Temporal expression of AsHPX2 mRNA and 16S rRNA in blood fed midguts. Relative mRNA levels of AsHPX2 gene and 16S rRNA were analyzed in midguts collected at different time points after the blood feeding. The mRNA levels of sugar fed midguts were considered as controls or 1. Significant differences ($p < 0.001$) between each time point and control are indicated by three asterisks (*).
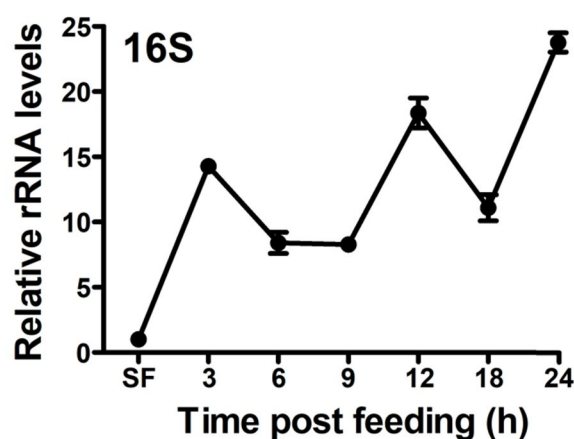
3.3.3. AsHPX2 has role in antibacterial immunity

3.3.3.1. Expression of AsHPX2 gene is induced in bacteria supplemented saline fed midguts

Since the expression of AsHPX2 is higher in the sugar fed midgut than blood fed midguts; we were interested to know whether this gene has any role in immunity against exogenous bacteria. In this context, we allowed mosquito to artificially feed either on saline alone (control) or supplemented with a mixture of *E. coli* and *M. luteus* (total 10^9 bacterial/ml). To confirm that the exogenous bacteria which were fed to the mosquito act as immune elicitors, we checked the expression of GNBP (Gram-Negative Bacteria-
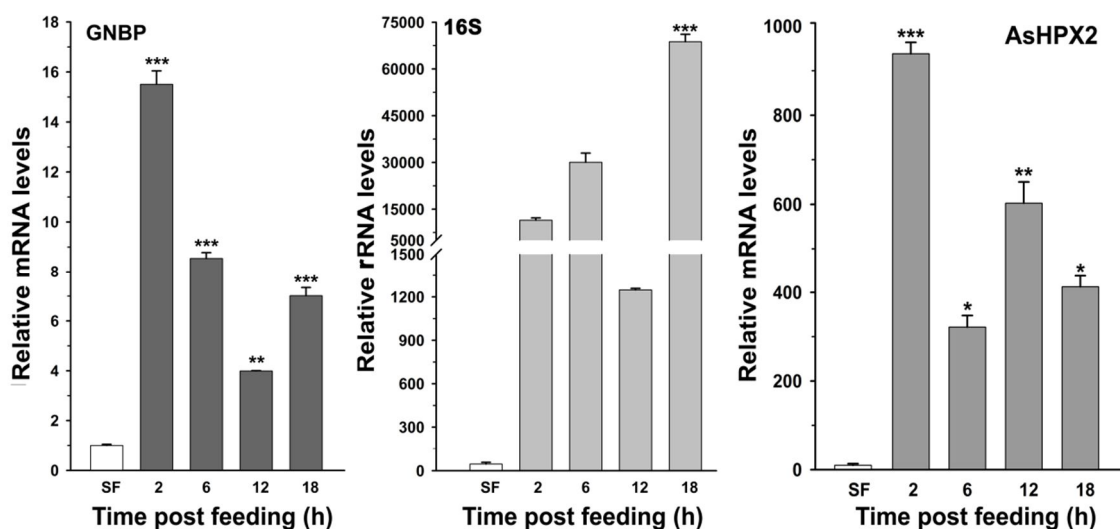
Figure 3.11: Relative mRNA levels of various genes in bacteria supplemented saline fed midguts. Mosquitoes were fed either on saline alone (control) or bacteria (*M. luteus* and *E. coli*) supplemented saline and midguts were dissected at different time points post feeding. The expression kinetics of 16S rRNA, GNBP and AsHPX2 genes was analyzed in their midguts. SF represents the sugar fed midguts and taken as 1. Relative levels are presented against respective control at each time points. Significant differences *p*< 0.001, 0.01 or 0.05 between sugar fed and bacteria fed are indicated by three, two or one asterisk (*) respectively.

Binding Protein). The GNBP is a PRR which is transcriptionally upregulated in the presence of bacterial elicitors like ß-1,3-glucan and lipopolysaccharide (Dimopoulos et al., 1997; Lemaitre et al., 1995; Michel et al., 2001; Waterhouse et al., 2007; Osta et al., 2004). The expression kinetics of GNBP shown in **Figure 3.11** revealed that this gene was significantly upregulated in bacteria supplemented saline fed midguts by 15-fold at 2 h, 8-fold at 6 h, 4-fold at 12 h, and 7-fold at 18 h against sugar fed midguts. The 16S rRNA levels presented in **Figure 3.11** showed that there is increased bacterial growth by 11000-fold at 2 h, 30000-fold at 6 h, 1200-fold at 12 h and 69000-fold at 18 h against sugar fed midguts. The mRNA levels of AsHPX2 gene in the presence of exogenous bacteria is induced by 938-fold at 2 h, 321-fold at 6 h, 602-fold at 12 h, and 412-fold at 18 h in bacteria supplemented saline fed midguts against sugar fed midguts (**Figure 3.11**). This result indicated that AsHPX2 gene is induced due to heavy load of

exogenous bacteria in the midgut. The induced AsHPX2 mRNA levels and its peroxinectin like nature might indicate that this gene is participating in anti-bacterial immunity.

3.3.3.2. Expression of AsHPX2 gene is induced in bacteria supplemented blood fed

We also investigated the expression kinetics of AsHPX2 gene in exogenous bacteria supplemented blood fed midguts. For this, female mosquitoes were allowed to feed on either blood alone (control) or bacteria supplemented blood (test) and midguts were collected at different time points as described in Material and Methods. The bacterial growth was analyzed in controls and exogenous bacterial fed midguts. The 16S rRNA levels presented in **Figure 3.12** revealed the growth of exogenous bacteria in the bacteria supplemented blood fed midguts against respective blood fed midguts. In these samples, expression of gambicin, an antimicrobial peptide was studied to ensure that the bacteria that we had
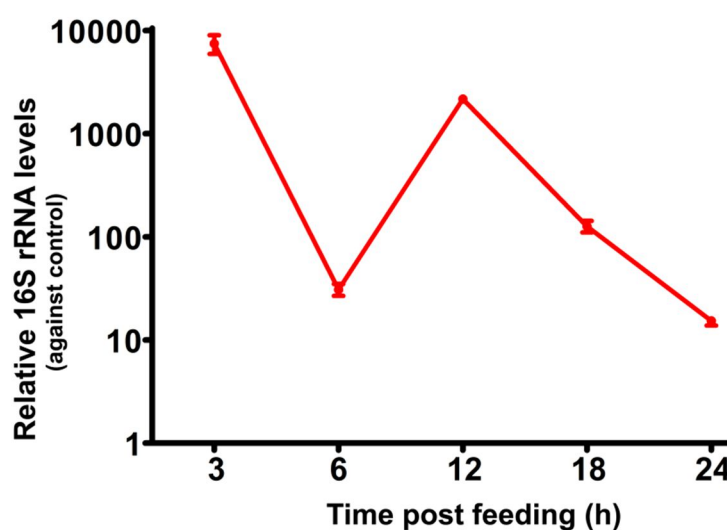
Figure 3.12: The expression kinetics of 16S rRNA levels in bacteria supplemented blood fed midguts. Mosquitoes were fed either on blood alone (controls) or supplemented with a mixture of *M. luteus* and *E. coli* bacteria (total 10^9 cells/ml blood). The relative levels of 16S rRNA were analyzed in their midguts collected at different time points post feeding and represented here in the $\log_{10}$ scale against the control at each time points.

given to the mosquito act as immune elicitors. The result showed in **Figure 3.13** revealed that gambicin is induced 43-fold at 3 h, 10-fold at 6 h, 8-fold at 12 h, and 2.5-fold at 18 h in bacteria supplemented blood fed midguts in comparison to their respective blood fed controls. We then analyzed the mRNA levels of AsHPX2 gene in these midguts to understand its role against exogenous bacteria. We found that expression of AsHPX2 is downregulated after blood feeding when compared to the sugar fed controls (**Figure 3.13**). The mRNA levels of AsHPX2 were unaffected by the bacterial feeding at 3 h, 12 h and 18 h when compared to their respective blood fed controls. However, in the bacteria supplemented blood fed midguts, AsHPX2 gene showed significant induction of 11-fold at 6 h and 2-fold at 24 h. Analysis of relative

levels of 16S rRNA in bacteria supplemented blood fed midguts against blood fed controls revealed reduced bacterial growth this two time points (**Figure 3.12**). So, we conclude that mRNA levels of AsHPX2 are induced in the midgut in the presence of exogenous bacteria and might participate in the anti-bacterial immunity. This was further studied by AsHPX2 silencing using RNAi approach.
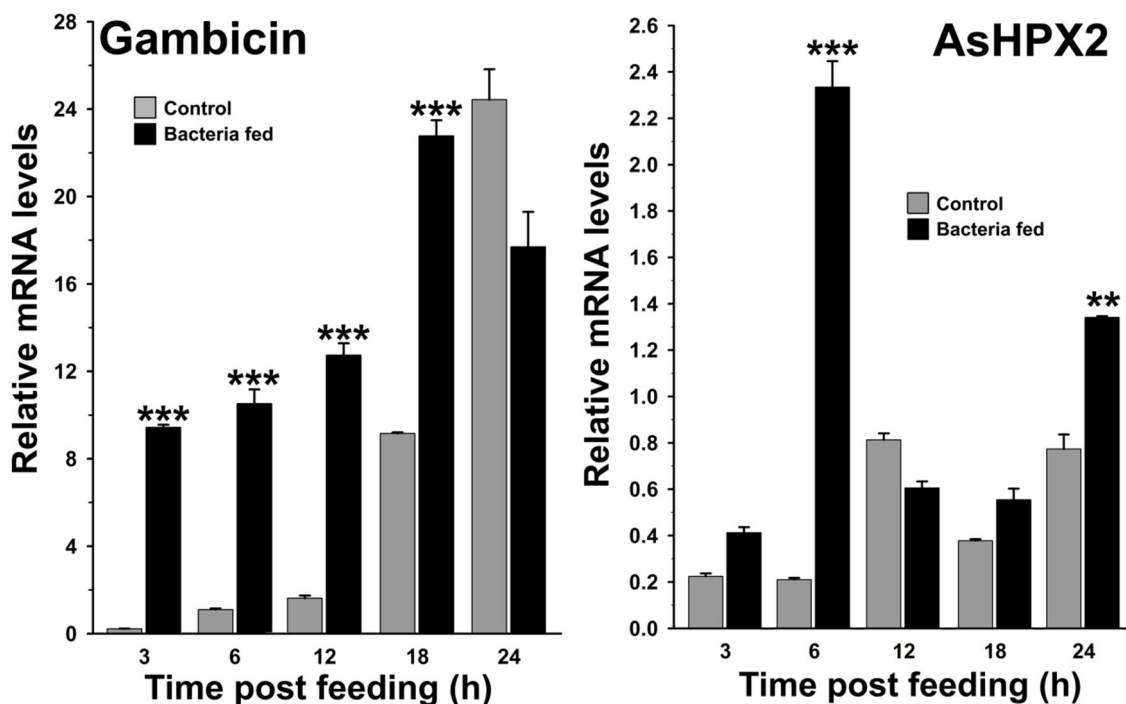
Figure 3.13: **Expression kinetics of gambicin and AsHPX2 genes in bacteria supplemented blood fed midguts**. Relative mRNA levels of gambicin and AsHPX2 genes were analyzed in the blood alone (controls) or bacteria supplemented blood fed midguts. Significant differences *p*<0.001 or 0.01 between controls and bacteria fed are indicated by three or two asterisks (*), respectively.

3.3.3.3. Silencing of AsHPX2 gene induces the growth of midgut endogenous bacteria

The expression of AsHPX2 gene is reduced after blood feeding when compared against sugar fed midguts (**Figure 3.9** and **3.10**). The peroxinectin like nature of AsHPX2 gene and its induction after bacterial feeding prompted us to hypothesize that this gene might be involved in maintaining the bacterial homeostasis in the sugar fed midguts. To further explore antibacterial role of AsHPX2 gene, we carried out the silencing of the gene and evaluated its effect on the growth of endogenous bacteria in the sugar fed midguts. In particular, to achieve this goal, a group of mosquitoes was injected with the LacZ (control) or AsHPX2 (silenced midguts) dsRNA and after 4 days midguts were collected. We analyzed the percentage of gene silencing by comparing the relative mRNA levels of AsHPX2 gene in control and silenced midguts which revealed that we could achieve

90% silencing of this gene (**Figure 3.14**). In addition, to explore the effect of AsHPX2 silencing on bacteria growth, we analyzed 16S rRNA levels of endogenous bacteria in the above-mentioned midguts. Results presented in **Figure 3.14** revealed that bacteria levels were increased ~4-fold in the silenced midguts against control (p=0.0012). Thus, AsHPX2 silencing increased the overall growth of bacteria and hence, this gene participates in maintaining the bacterial homeostasis in sugar fed midguts.
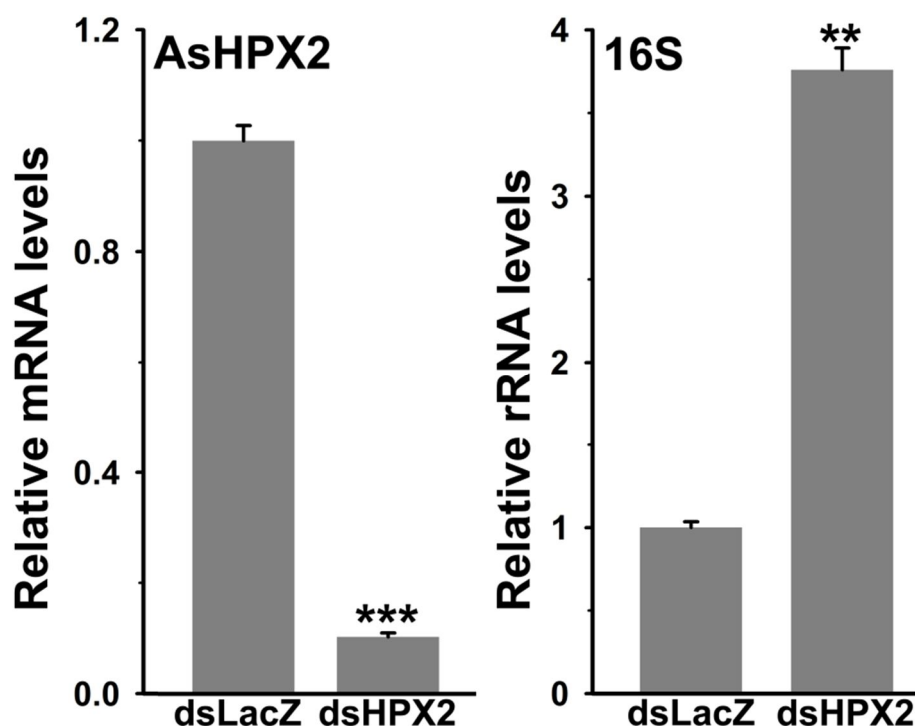
Figure 3.14: Relative levels of AsHPX2 mRNA and 16S rRNA in AsHPX2 silenced sugar fed midguts. Mosquitoes injected with dsLacZ (control) or dsAsHPX2 (silenced) RNA were kept for 4 days and continuously fed on 10% sugar. After 4 days relative levels of AsHPX2 mRNA and 16S rRNA were analyzed in their midguts. Significant differences ($p < 0.001$ or 0.01) are shown by three or two asterisks (*), respectively.

3.3.3.4. AsHPX2 silencing promotes the growth of bacteria in exogenous bacteria fed midgut

We were then interested in exploring the role of AsHPX2 against the high load of the exogenous bacteria in the midguts. To achieve this, we silenced AsHPX2 gene through dsRNA-mediated interference. Silenced mosquitoes were fed on bacteria supplemented blood as mentioned in the Materials and Methods. The result presented in **Figure 3.12** revealed that the levels of 16S rRNA in exogenous bacteria supplemented blood fed midguts showed reduced bacterial growth at 6 h and 24 h when compared against their respective blood fed midguts. Moreover, expression of AsHPX2 in exogenous bacteria supplemented blood fed midguts showed biphasic induction at 6 h and 24 h post feeding in comparison to their respective blood fed midguts (**Figure 3.13**). Thus, to

explore the immune regulatory mechanism that balances the mosquito midgut immunity against exogenous bacteria, these two time points were further analyzed in detail. The mRNA levels of AsHPX2 in control and silenced midguts revealed that this gene was silenced 74% and 63% at 6 h and 24 h, respectively (**Figure 3.15**). The growth of bacteria in the As-HPX2 silenced midguts was analyzed by studying the 16S rRNA levels in exogenous bacteria fed midguts of control and silenced mosquitoes. Results presented in **Figure 3.15** revealed that there is a significant increase in the bacterial growth in silenced midguts by 1.75-fold at 6 h and 2-fold at 24 h post
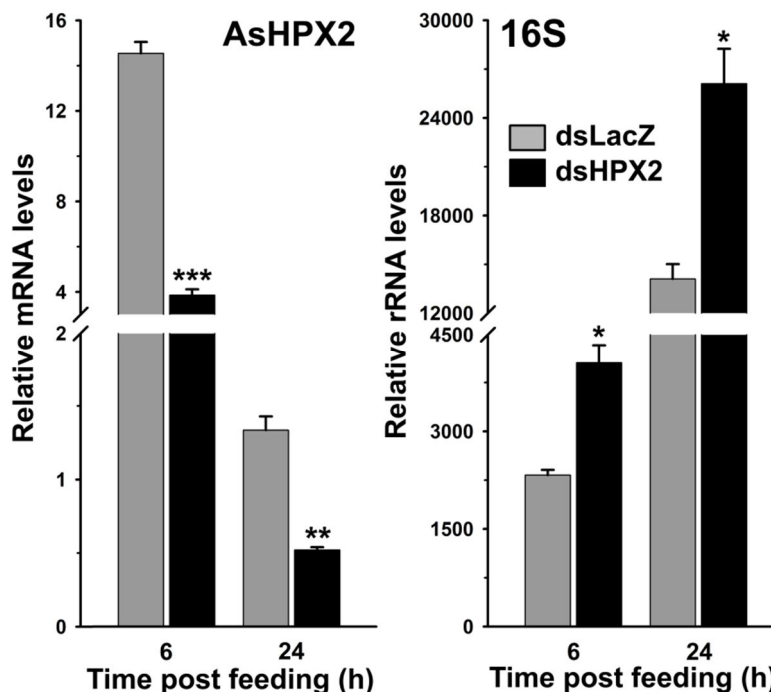
Figure 3.15: Relative levels of AsHPX2 mRNA and 16S rRNA in AsHPX2 silenced bacteria supplemented blood fed midguts. Mosquitoes injected with dsLacZ (controls) or dsAsHPX2 (silenced) RNA were fed on bacteria supplemented blood and relative levels of AsHPX2 mRNA and 16S rRNA were analyzed in their midguts at 6 h or 24 h post feeding. Relative levels are presented against the sugar fed midguts. Significant differences $p<0.001$, 0.01 or 0.05 between control and silenced midguts are indicated by three, two or one asterisk (*), respectively

feeding in comparison to the controls (p=0.0140 and p=0.0416, respectively). Thus, AsHPX2 silencing increases proliferation of bacteria in the midgut.

We then analyzed the mRNA levels of antibacterial immune genes to study the regulation of bacteria by mosquito immunity in the AsHPX2 silenced midguts. The mRNA levels of GNBP and gambicin were analyzed in these samples to understand the involvement of classical immune pathways in maintaining the midgut bacterial homeostasis (Lemaitre et al., 1995; Michel et al., 2001; Waterhouse et al., 2007; Osta et al., 2004; Tanji et al., 2007; Vizioli et al., 2001). It was found that pattern recognition receptor GNBP induces 46-fold and 40-fold at 6 h and 24 h, respectively in bacteria supplemented blood fed controls **(Figure 3.16).** However, in the silenced midguts, GNBP expressions was reduced significantly by 36% at early time point (6 h) (p=0.0022) and remained similar at later time point (24 h) (p= 0.5228) when compared

to the non-silenced controls (**Figure 3.16**). We also studied the expression of gambicin in the AsHPX2 silenced bacteria supplemented blood fed midguts. We found that the mRNA levels of gambicin were 4.5-fold and 7-fold at 6 h and 24 h, respectively in non-silenced bacteria supplemented blood fed midguts against sugar fed controls (**Figure 3.16**). However, gambicin mRNA levels were similar at 6 h but induced 1.5-fold at 24 h (*p*=0.0057) in AsHPX2 silenced midgut when compared to the control. We also analyzed the expression of one of the AMPs, defensin which is controlled by both IMD and TOLL pathways (Tanji et al., 2007). Data presented in **Figure 3.17** showed that mRNA levels of defensin have no difference at 6 h or 24 h post feeding in AsHPX2 silenced midguts against controls.

Previously, our laboratory reported that *A. stephensi* heme peroxidase AsHPX15 modulates the immunity in exogenous bacteria supplemented blood fed midgut. When we silenced the AsHPX15 gene, the heavy load of bacteria was reduced by the induction of NOS (Kajla et al., 2016c).
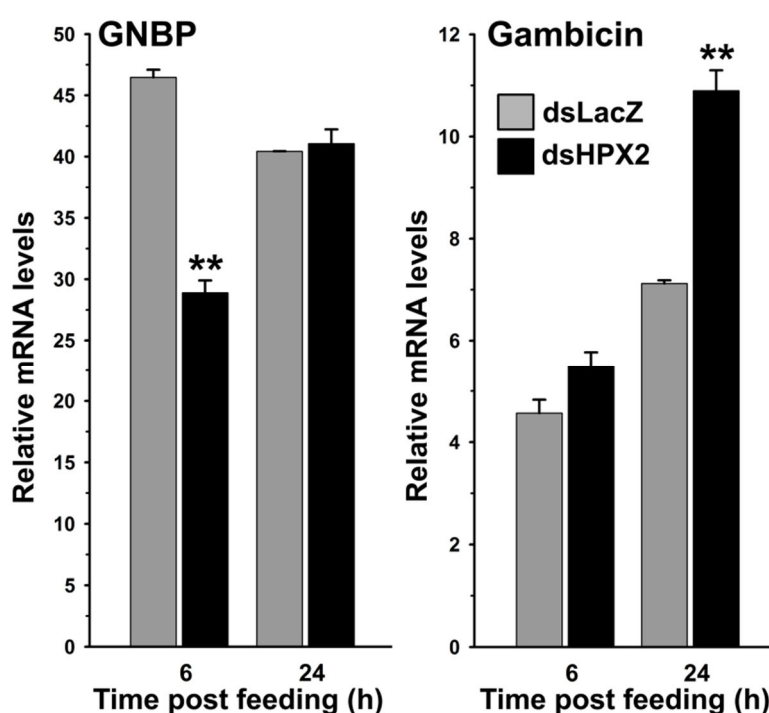
Figure 3.16: Expression of immune genes in AsHPX2 silenced bacteria supplemented blood fed midguts. Mosquitoes injected with dsLacZ (controls) or dsAsHPX2 (silenced) were fed on bacteria supplemented blood and relative mRNA levels of GNBP and gambicin genes were analyzed in their midguts at 6 h or 24 h post feeding. Relative levels of mRNA are presented against the sugar fed midguts. Significant differences *p* < 0.01 between control and silenced midgut are shown by two asterisks (*).

Thus, we were interested to find out the effect of AsHPX2 silencing on the expression pattern of AsHPX15 in the presence of exogenous bacteria. The result shown in the **Figure 3.17** revealed that AsHPX15 is induced 50-fold and 30-fold at 6 h and 24 h post bacterial supplemented blood feeding in non-silenced control midguts, respectively, against sugar fed controls. Interestingly, its mRNA levels were 82-fold and 43-fold at 6 h and 24 h in AsHPX2 silenced bacteria supplemented blood fed midguts, respectively. Hence, AsHPX15 is induced in AsHPX2 silenced bacteria supplemented

blood fed midguts against control. HPX15 gene mediates the crosslinking of mucin layer after blood feeding that reduces the permeability between midgut bolus and epithelial cells and thus, creates a low immunity zone. This will, in turn, protect the bacteria from the midgut immune responses. Thus, the induced levels of AsHPX15 might reflect the important function of barrier formation (Kumar et al., 2010) during such a heavy bacterial load, so that mosquito immunity does not compromise the survival of the mosquito as we observed similar rate of survival in control and AsHPX2 silenced mosquitoes groups.

So, above findings suggested that AsHPX2 silencing increased bacterial growth in the midgut. The formation of gut barrier protects the bacteria from being re-

Figure 3.17: Expression of defensin and AsHPX15 genes in AsHPX2 silenced bacteria supplemented blood fed midguts. Mosquitoes injected with dsLacZ (controls) or dsAsHPX2 (silenced) were fed on bacteria supplemented blood and relative mRNA levels of defensin and HPX15 genes were analyzed in their midguts at 6 h or 24 h post feeding. Relative levels of mRNA are presented against the sugar fed midguts. Significant differences $p < 0.01$ between control and silenced midguts are shown by two asterisks (*).

cognized by epithelial immunity and so, we could not observe induction of mosquito immunity (immune genes like defensin and GNBP) against the increased bacterial load. In addition, we also found that rate of survival of mosquitoes was indifferent in AsHPX2 silenced bacteria supplemented blood fed midgut when compared to the controls. This suggested the advancement in mosquito immunity as compared to the other insects like *Bactrocera dorsalis* and *Drosophila*. In these insects, when they were fed on bacteria supplemented food, their mortality increases (Yao et al., 2016; Ha et al., 2016).

3.3.4. AsHPX2 gene is antagonist to *Plasmodium* development

3.3.4.1. Expression analysis of AsHPX2 gene in *P. berghei* infected mosquitoes

To decipher the regulation of AsHPX2 gene during malaria infection, we analyzed its

mRNA levels in blood fed (control) and *P. berghei* infected midguts and carcasses that were collected at different time points after feeding. Results shown in **Figure 3.18** revealed induced expression of AsHPX2 gene in the infected midguts in comparison to the controls. The mRNA levels of AsHPX2 gene were
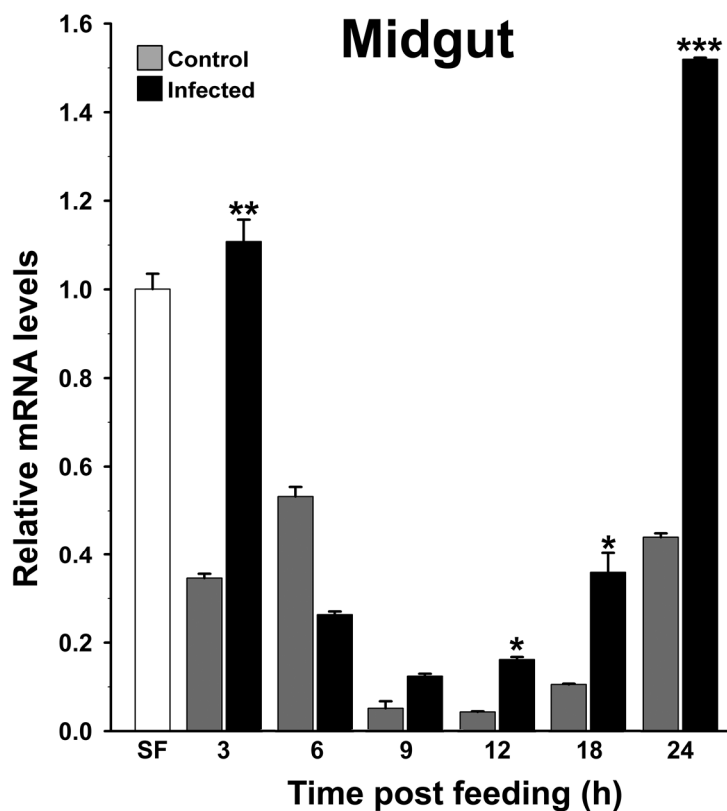
Figure 3.18: Expression kinetics of AsHPX2 gene in the midguts and carcasses during *P. berghei* infection. Relative mRNA levels of AsHPX2 gene were analyzed in *A. stephensi* midguts and carcasses at different time points after feeding on an uninfected (Control) or *P. berghei*-infected mouse. Relative fold induction was calculated against sugar fed midguts (open bar). Significant differences *p* <0.001, 0.01 or 0.05 between control and infected midguts and carcasses are indicated by three, two or one asterisk (*), respectively.
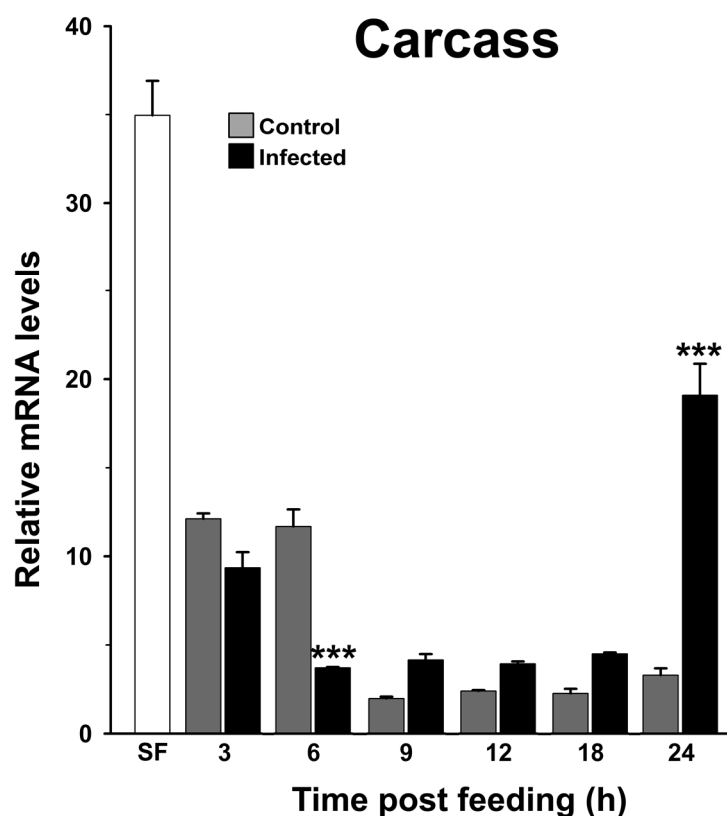
significantly induced by 4-fold at 3 h, 3-fold at 12 h, 3-fold at 18 h and 3-fold at 24 h (p=0.0022, p=0.0035, p=0.0126, p=0.0001, respectively) post infection against their uninfected controls.

These results indicated that the expression of AsHPX2 gene is induced in infected midguts when the pre-ookinete stages of *Plasmodium* development predominate in the blood bolus (3 h) as well as when the mature ookinetes start invading the midgut epithelial cell (~24 h) (Smith et al., 2014) (**Figure 3.18**). These results are in accordance with the previous findings where *A. gambiae* AgHPX2 gene, an ortholog of AsHPX2, is induced in *Plasmodium* infected midguts at 24 h post feeding (Kumar et al., 2004; Oliveira et al., 2012).

The mRNA levels of AsHPX2 gene were also analyzed in *P. berghei* infected carcasses as shown in **Figure 3.18**. The result revealed that unlike midgut, there is no induction of AsHPX2 gene in the carcass up to 18 h post infected blood feeding. The expression of AsHPX2 is induced 6-fold at 24 h when compared to its respective blood fed control. However, in comparison to sugar fed carcasses, there is no induction of AsHPX2 mRNA levels in the *Plasmodium* infected carcass.

These findings are in accordance with the previous data where it has been shown that HPX2 in *A. gambiae* is induced at the time of midgut invasion by the ookinetes (Kumar at al., 2004). In these mosquitoes, the HPX2/ NOX5 system potentiates the toxicity of nitric oxide by mediating epithelial nitration which modifies the ookine-
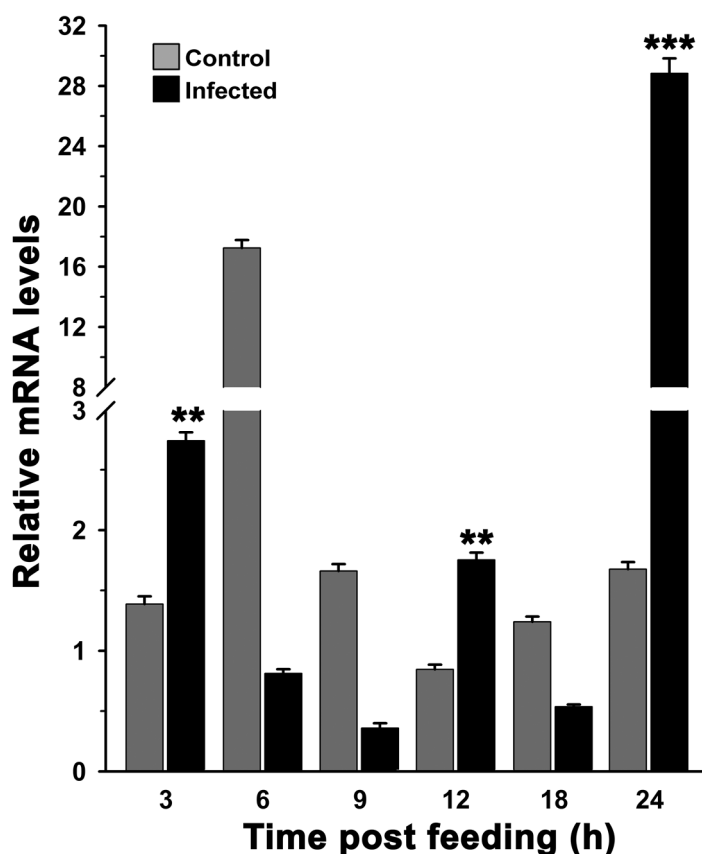
Figure 3.19: Expression kinetics of NOX5 gene in the midguts during *P. berghei* infection. Relative mRNA expression levels of NOX5 were analyzed in *A. stephensi* midguts at different time points after feeding on an uninfected (Control) or *P. berghei*-infected mouse. Relative fold induction was calculated against sugar fed midguts. Significant differences p <0.001 or 0.01 between control and infected midguts are indicated by three or two asterisks (*), respectively.

tes and makes them visible to the TEP1 (Thioester-containing protein 1), a molecule of mosquito complement system that mediates their lysis while traversing the midguts (Oliveira et al., 2012).

To better understand the similar mechanism in the *A. stephensi*, we analyzed the expression of NOX5 and TEP1 genes in these *P. berghei* infected midguts. The mRNA levels of NOX5 gene induced 2-fold at 3 h (*p*=0.0025), 2-fold at 12 h (*p*=0.0031) and 17-fold at 24 h (*p*=0.0007) in *P. berghei* infected midguts as compared to their respective controls (**Figure 3.19**). The mRNA levels of TEP1 also induced 11-fold at 3 h (*p*=0.0004), 2-fold at 6 h (*p*=0.0324), 2-fold at 18 h (*p*=0.0041) and 2-fold at 24 h (*p*=0.0409) in infected midguts against respective controls (**Figure 3.20**). These observations revealed that all the genes, AsHPX2, NOX5, and TEP1 showed induced expression at 3 h and 24 h post *P. berghei* infection in the midguts in comparison to the respective blood fed control midguts. The previous report showed the induction of these genes in *A. gambiae* at 24 h post infection in the midguts, the time when ookinetes are traversing the midgut epithelial cells (Oliveira et al., 2012). Based on these findings, we can speculate that in *A. stephensi* AsHPX2/NOX5 system works in a similar manner and the lysis of ookinetes is mediated by TEP1.

Figure 3.20: Expression kinetics of TEP1 gene in the midguts during *P. berghei* infection. Relative mRNA expression levels of TEP1 were analyzed at different time points in *A. stephensi* midguts after feeding on an uninfected (Control) or *P. berghei*-infected mouse. Relative fold induction was calculated against sugar fed midguts. Significant differences *p* <0.001, 0.01 or 0.05 between control and infected midgut are indicated by three, two or one asterisks (*), respectively.
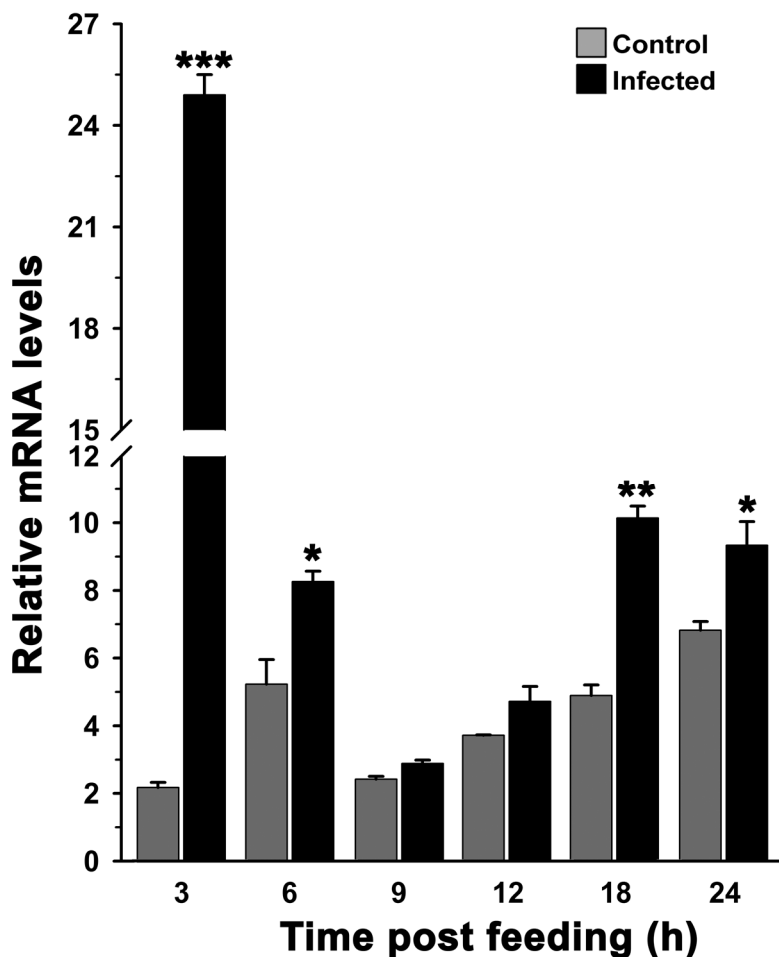
In the previous study, HPX15 gene from *A. gambiae* is involved in the formation of midgut mucin barrier which creates a low immunity zone and supports the growth of bacteria as well as *Plasmodium* (Kumar et al., 2010). However, this barrier can only protect ookinetes which are present in the midgut and ookinetes traversing the midgut are modified by HPX2/NOX5 system. This modification makes them visible to the mosquito complement-like system and mediates their killing (Oliveira et al., 2012).

3.3.4.2. AsHPX2 gene silencing increased *Plasmodium* development

The mRNA levels of AsHPX2 gene were induced post *P. berghei* infection in the mosquito midguts as compared to the respective controls (**Figure 3.18**). To decipher the role of AsHPX2 in the regulation of *P. berghei* development, we performed silencing of this gene. For this, we injected mosquitoes with LacZ (controls) or AsHPX2 (silenced) dsRNA and after 4 days these mosquitoes were fed on *P. berghei* infected mouse. The percentage of gene silencing and number of developing oocysts was analyzed as described in Material and Methods. The relative AsHPX2 mRNA levels in controls and silenced midguts revealed 80% silencing of this gene (**Figure 3.21A**).
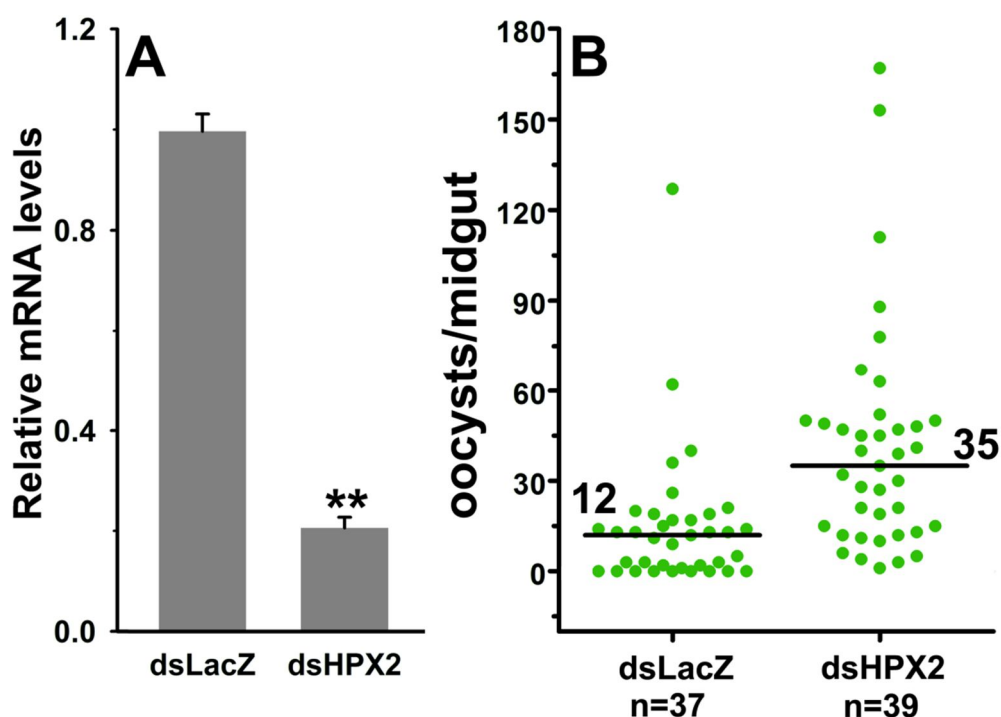
Figure 3.21: Effect of AsHPX2 gene silencing on *P. berghei* development. A) Relative mRNA levels of AsHPX2 gene in dsLacZ (unsilenced) or dsAsHPX2 (silenced) RNA injected and 24 h post *P. berghei* infected midguts. Two asterisks shows significant difference ($p < 0.01$). **B)** Effect of AsHPX2 silencing on the number of live oocysts (green dots) in midguts analyzed seven days post infection. Horizontal line indicates the medians. Distributions are compared using the Kolmogorov-Smirnov test ($p=0.0003$); n = number of mosquitoes.

Further, we counted the number of developing oocysts in control and silenced midguts after 7 days of *Plasmodium* infected blood feeding. Results presented in **Figure 3.21B** revealed a variable number of oocysts in control and silenced midguts. However, the median value for the oocysts numbers in controls and AsHPX2 silenced midguts was 12 and 35, respectively. This indicated that the numbers of developing oocysts are increased significantly by 3-fold in the silenced midguts against controls (**Figure 3.21B**, *p*= 0.0003). Our data showed that AsHPX2 is an antagonist of *P. berghei* development in the mosquito midgut. These findings are in agreement with the previous report where silencing of the AgHPX2 gene in *A. gambiae* also increased *Plasmodium* survival (Oliveira et al., 2012).

3.4. Discussion

In the present study, we have characterized and studied the functional role of heme peroxidase HPX2 in *A. stephensi*, an equivalent ortholog of *A. gambiae* HPX2. To understand the transcriptional regulation of AsHPX2, we have cloned the full-length gene and analyzed its sequence (**Figure 3.1 and 3.2**). BLAST result of AsHPX2 gene revealed that it has closest match to *A. gambiae* and *A. sinensis* HPX2. Our result showed that AsHPX2 shared 55-60% of identity with its ortholog present in *A. aegypti*, *A. albopictus* and *C. quinquefasciatus*. Furthermore, general BLAST result showed that its orthologs are not present in species other than Culicinae. Hence, this gene is present only in the mosquito and shared a high percentage of similarity (> 70%) within the genus *Anopheles*.

The expression analysis of AsHPX2 gene showed that it is expressed in all the developmental stages of *A. stephensi*. This gene showed induced expression in 1st instar larvae in comparison to the eggs and the highest expression is seen in the adult females (**Figure 3.8**). This suggested the important role of AsHPX2 in mosquito development. The similar pattern of expression of peroxidase is seen in *Mayetiola destructor* across developmental stages (Chen et al., 2016). The 1st instar larvae encounter various pathogens and should mount immune responses against them. Failure in this will result in larval death. Hence, AsHPX2 induction in 1st instar larvae might be due to immune responses against pathogens. The reduced expression of AsHPX2 gene in 4th instar larvae and pupae might be because of transition stage from larvae to pupae and pupal stage has overall acquiescent metabolic activity. This further correlates with the earlier study where reduced peroxidase expression is reported during larvae to pupal transition (Chen et al., 2016).

The mRNA levels of AsHPX2 gene decreased after the blood feeding in the mosquito midguts (**Figure 3.10**). In other words, the mRNA levels of this gene were highest in sugar fed midguts. Thus, we silenced the gene in the sugar fed midguts and analyzed its effect on bacterial load. Surprisingly, the levels of 16S rRNA in the silenced midguts were significantly higher in comparison to the control midguts (**Figure 3.14**). This data might suggest the anti-bacterial nature of AsHPX2 gene which maintains midgut bacterial homeostasis in the sugar fed midguts. In this study, we have also shown that mRNA levels of AsHPX2 gene are induced in the presence of exogenous bacteria in the midguts (**Figure 3.13**). The presence of TFBM for GATA/Rel suggested the tissue-specific expression of this gene in immunity (Senger et al., 2006).

Previous reports have shown that *Plasmodium* invaded midgut cells produced a high level of NOS (nitric oxide synthase) and peroxidase that cause an increase in the level of midgut nitration (Kumar et al., 2004). Invasion of Midgut epithelial cells by parasite activates the HPX2/NOX5 system which causes the nitration of epithelial cells and is regulated by the c-Jun N-terminal kinase (JNK) pathway. This epithelial nitration enhanced the thioester-containing protein 1 (TEP1)-mediated lysis of ookinetes (Oliveira et al., 2012; Garver et al., 2013). Silencing of AgHPX2 greatly increased the number of oocysts. Our study also showed that mRNA levels of AsHPX2 are significantly induced post 24 h of *P. berghei* infection in the midguts the time when motile ookinetes traverse the midguts (**Figure 3.18**). The presence of AP-1 binding motifs in the 5′ region of AsHPX2 gene suggested its expression through JNK pathway during immunity against *Plasmodium* (Garver et al., 2013). Silencing of AsHPX2 gene increased the *Plasmodium* infection by 3-fold (**Figure 3.21**). This result indicated the anti-plasmodial property of AsHPX2 in a way similar to AgHPX2.

Collectively, based on the present finding, it is clear that AsHPX2 gene plays important role in mosquito physiology and immunity. AsHPX2 is one of the molecules that maintain the bacterial homeostasis in the sugar fed midgut and its mRNA levels are reduced in blood fed midgut may be for the proliferation of the endogenous bacteria, as they are required for the various processes like digestion and reproduction (Kajla et al., 2015b). Also, in the blood fed midgut there is the induction of another heme peroxidase HPX15 which catalyzes the crosslinking of mucin layer that acts as a barrier and blocks the recognition of gut proliferating bacteria by the immunoreactive epithelium. This mechanism creates a "Low Immunity Zone" in this body compartment to promote the bacterial growth that supports blood digestion (Kajla et al., 2015b; Kumar et al., 2010). The AsHPX2 along with NOX5 showed upregulation in mRNA levels in *P. berghei*

infected midgut is involved in anti-plasmodial responses when ookinetes are traversing the midgut epithelial cells that greatly enhance their clearance by TEP1 mediated lysis, a similar mechanism reported in *A. gambiae* (Oliveira et al., 2012).

3.5. Conclusion

The identification of *Anopheles* genes that are involved in immune responses against both bacteria and *Plasmodium* is of great value. Here, we found that AsHPX2 gene is exclusively present in the genome of mosquito species. It has an important role in the mosquito development and mosquito immunity against both bacteria and *Plasmodium*. Our results showed that the anti-plasmodial role of HPX2 is conserved among both *A. gambiae* and *A. stephensi*. Hence, this heme peroxidase can be targeted for arresting *Plasmodium* development.

Chapter 4

Functional characterization of *Anopheles stephensi* midgut Dual Oxidase gene and its role in bacterial homeostasis and *Plasmodium* development

4.1 Abstract

Our study characterized the Dual Oxidase (Duox) gene from Indian malaria vectors *A. stephensi.* Here, we showed that silencing of this gene in either sugar fed or blood fed midguts increased the bacterial population in comparison to the controls. Moreover, expression of this gene has a strong negative correlation with bacterial growth in exogenous bacteria supplemented blood fed midguts. Hence, Duox is one of the key molecules of mosquito immunity that maintains the bacterial homeostasis in the midguts. AsDuox gene is also induced in the midguts post-*P. berghei* infection and silencing of this gene suppressed *Plasmodium* development. This finding corroborates with the previous report where AsDuox ortholog AgDuox in *A. gambiae* mediates the formation of mucin barrier on the luminal side of the midgut. This creates a low-immunity zone and suppresses the activation of mosquito immunity against the *Plasmodium*. Thus, AsDuox gene is an important molecule of innate immunity against pathogens in the mosquito. Hence, this molecule might be targeted to manipulate mosquito immunity and arrest *Plasmodium* development inside the vector host.

4.2 Introduction

The insect gut is in direct and constant contact with microbes and harbors naturally occurring bacteria. The symbiotic bacteria perform important functions in nutrient supply, host development, food digestion, reproduction, defense against colonization by opportunistic pathogens (Ley et al., 2006; Weiss et al., 2011; Broderick et al., 2014; Shin et al., 2011; Storelli et al., 2011; Venema, 2010; Tremaroli and Bäckhed, 2012, Gaio et al., 2011; Coon et al., 2016). Thus, gut epithelium must be able to tolerate the optimal proliferation of natural gut bacteria but still capable of eliminating the pathogenic or opportunistic bacterial community.

Till date, few investigations have focused on this phenomenon to understand the interaction mechanism between bacteria and gut epithelial cells. In *Drosophila* anti-microbial peptides (AMPs) are secreted by gut epithelial cells against the pathogenic bacteria (Lemaitre and Hoffmann, 2007), additionally dual oxidase (DUOX) pathway also activates to maintain microbial homeostasis (Ha et al., 2009a; 2009b; 2005a). In mosquitoes, maintenance of bacterial homeostasis is more challenging as endogenous bacteria are present in the midgut and proliferate extensively after blood feeding (Dong et al., 2009; Kumar et al., 2010; Oliveira et al., 2011).

In *Anopheles* mosquito, bacterial homeostasis is one of the major components that determine the *Plasmodium* development (Boissière et al., 2012; Dong et al., 2009;

Bahia et al., 2014). The molecular mechanism of midgut immunity that tolerates endogenous bacteria while triggering immune responses against exogenous bacteria remains to be elucidated in *Anopheles*. So in this study, we investigated the role of Duox in maintaining midgut bacterial homeostasis and in the development of malaria parasite, *Plasmodium* in the midguts of *A. stephensi*.

4.3 Results

4.3.1 Molecular cloning and sequence analysis of AsDuox gene

We have retrieved and analyzed AsDuox gene in *A. stephensi* genome using AgDuox gene (AGAP0099778) of *A. gambiae* as a reference. In brief, the AgDuox gene (AGAP0099778) was BLAST against *A. stephensi* genome to retrieve the contig that contains putative AgDuox ortholog. We found that putative AsDuox gene is located in contig 2339 (SuperContig KB664518, Ensembl identifier ASTE003295 in the annotated genome of *A. stephensi*). From this contig, the full-length putative AsDuox gene (9.45 kb) is predicted using Augustus software. The PCR based approach was used to confirm the presence of *A. stephensi* AsDuox gene in contig 2339. For that, gene-specific primers were designed based on the alignment of AgDuox and Augustus program predicted AsDuox gene as described in the Material and Methods and primers are mentioned in **Table 2.2.**

 A. stephensi genomic DNA (gDNA) or cDNA was used as a template to amplify the AsDuox gene segment. The result presented in **Figure 4.1A** showed that the primer set F1R1 amplified 466 bp and 398 bp of PCR product from gDNA and cDNA template, respectively. This partial AsDuox clone was sequenced and BLAST result revealed that it has the closest match to *A. gambiae* AgDuox (AGAP0099778) (identity 89% and E value 2e-134).
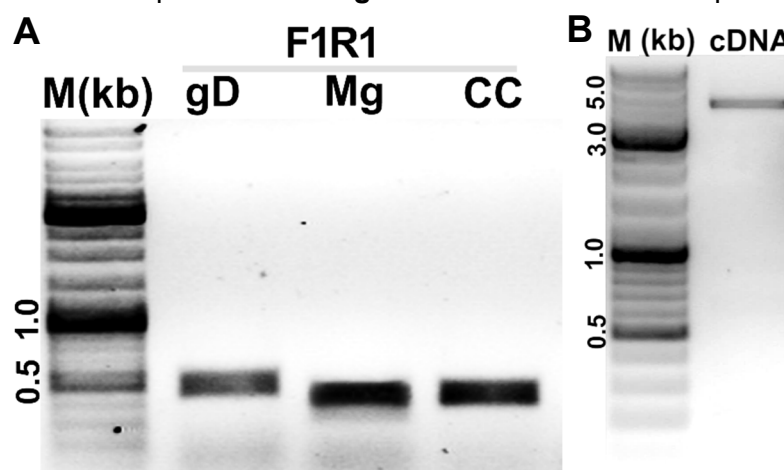
Figure 4.1: PCR amplification of *A. stephensi* Duox gene. A) amplificantion of partial AsDuox gene. PCR product was amplified from genomic DNA (466 bp) and cDNA (398 bp) using F1R1 primers. cDNA of midguts (Mg) and carcasses (CC) post 24 h of blood feeding was used as a template. **B) PCR amplification of full-length AsDuox gene.** Full-length AsDuox cDNA is amplified using F7R6 primers and cDNA of 24 h blood fed midgut is used as a template. The left lane (M) represents the DNA ladder in kb and used as a reference for identifying the product size.

Ecdysone

-908 GGTAGCTTTTGGTTTTTAGGGAAAATATTAAAACTTTAATTACCATACAAAAAAGGGAATTTGTTTCGCCACCACTGCATCGCTTAATGGGAT

ATF-2

-818 TCGCCATTTTGGACAACCATTCTTCTACCATGGTCTGAGAT /-776 /-414 AAGACACTCTCGCGAAATCTGTCATGTTTCATCTGACTAGGAAAAAA

-367 CGAAAAACAATTGTTCCACCTCCAACGGAAAGAATAATTCGTGTGCTC /-320 /-48 ACTGCTCGCGAAAAGATATAAACGTTAGCGTCGAAGATTT

TSS   **GATA**

-7 ACGATCAGGTAGTATGGAACATTAATCAACCGATATCGGTGACATGCAACAGTAAATTGTTCCTTTTGTGGAGCCATTGCCAAAATAAAG
  +1

+84 TCCATTCTCGGCGAACCTTGCCTGCGTACGTTCTGCCCGGCGTTCGGCGATAAGATCCTTTTCCTTACACGAATCCAGCTTG

```
   1 ATGAGCCACGTCGAAAAGCAGCGCTACGACGGATGGTACAACAATCTGGCCCATCCGGACTGGGGAGCTGTCGATAATCACCTGACAAGGAAAGCTCCGTCGGCGTATTCGGACGGGGTG
   1  M  S  H  V  E  K  Q  R  Y  D  G  W  Y  N  N  L  A  H  P  D  W  G  A  V  D  N  H  L  T  R  K  A  P  S  A  Y  S  D  G  V
     TATGTGTTGGCCGGTGCCAATCGACCCTCGCCCCGCAAGCTTCAGCCGACTGTTCATGCGTGGCACCGATGGGCTACCATCGATGGAAAACCGTACCGCTCTGCTGGCCTTCTTCGGTCAG
      Y  V  L  A  G  A  N  R  P  S  P  R  K  L  S  R  L  F  M  R  G  T  D  G  L  P  S  M  E  N  R  T  A  L  L  A  F  F  G  Q
 241 GTGGTGACGAACGAGATAGTCATGGCGTCCGAATCGGGCTGTCCGATCGAGATGCACCGCATCGAGATCGAGAAGTGTGACGAAATGTATGATCGAGAGTGTCGCGGCGATCGCTACATA
  81  V  V  T  N  E  I  V  M  A  S  E  S  G  C  P  I  E  M  H  R  I  E  I  E  K  C  D  E  M  Y  D  R  E  C  R  G  D  R  Y  I
     CCGTTCCACCGTGCGGCCTACGATCGTAATACGGGCCAGAGCCCGAATGCACCACGCGAACAAATCAATCAGATGACGGCTTGGATCGATGGTAGCTTCATCTACAGTACGTCGGAGGCA
      P  F  H  R  A  A  Y  D  R  N  T  G  Q  S  P  N  A  P  R  E  Q  I  N  Q  M  T  A  W  I  D  G  S  F  I  Y  S  T  S  E  A
 481 TGGCTGAACGCGATGCGATCGTTCCAGGACGGTGCACTGCTGACGGACAAGCAGGGCACGATGCCGATAAAGAACACGATGCGGGTCCCACTGTTCAACAATCCTGTGCCGCATGTGATG
 161  W  L  N  A  M  R  S  F  Q  D  G  A  L  L  T  D  K  Q  G  T  M  P  I  K  N  T  M  R  V  P  L  F  N  N  P  V  P  H  V  M
     CGGATGCTTAGCCCGGAACGGCTGTACCTTCTCGGTGATCCTAGAACTAACCAAAATCCAGCCCTACTGTCGTTCGCCATTCTGTTCCTACGTTGGCACAATGTGGTAGCGAAGCGCGTC
      R  M  L  S  P  E  R  L  Y  L  L  G  D  P  R  T  N  Q  N  P  A  L  L  S  F  A  I  L  F  L  R  W  H  N  V  V  A  K  R  V
 721 CGACGACAGCACCGAGATTGGAGTGATGAAGAAATCTTCCAACGCGCACGCCGTGTAGTGATCGCTAGTCTGCAAAACGTGATCGCTACGAGTATCTGCCAGCGTTCCTGGACAAGGAG
 241  R  R  Q  H  R  D  W  S  D  E  E  I  F  Q  R  A  R  R  V  V  I  A  S  L  Q  N  V  I  A  Y  E  Y  L  P  A  F  L  D  K  E
     ATACCTCCGTACGATGGTTATAAGGCCGACACCCATCCGGGTGTAAGCCATATGTTCCAGGCTGCTGCTTTCCGATTTGGACATTCTTTGATTCCACCGGGACTGTTTCGTCGCGATGGT
      I  P  P  Y  D  G  Y  K  A  D  T  H  P  G  V  S  H  M  F  Q  A  A  A  F  R  F  G  H  S  L  I  P  P  G  L  F  R  R  D  G
 961 CAGTGCAACTTCCGTCGTACTAATATGGACTTTCCTGCGCTTCGACTATGTTCAACATGGTGGAATTCTAACGACGTGCTTGACAACAACCACCGGTAGAAGATTCATCATGGGCATGGCA
 321  Q  C  N  F  R  R  T  N  M  D  F  P  A  L  R  L  C  S  T  W  W  N  S  N  D  V  L  D  N  T  P  V  E  E  F  I  M  G  M  A
     TCGCAGATCGCCGGAAAGGAAGACCCCCTGCTGGTGCTCGGACGTGCGCGATAAACTTGTTCGGACCGATGGAGTTTACACGCCGGGATCTTGGTGCTTTGAATATTATGCGCGGTCGTGAC
      S  Q  I  A  E  K  E  D  P  L  L  C  S  D  V  R  D  K  L  F  G  M  E  F  T  R  R  D  L  G  A  L  N  I  M  R  G  R  D
1201 AACGGACTTCCAGACTACAATACAGCGCGTGCGGCCGTACCGGCTGCCGAGAAAGAAAATTGGCGTGACATCAACCCGCAGGTGTTTGAGCGGCAGCCTGAACTACTGGAGTTGCTAATC
 401  N  G  L  P  D  Y  N  T  A  R  A  A  Y  R  L  P  R  K  K  N  W  R  D  I  N  P  Q  V  F  E  R  Q  P  E  L  L  E  L  L  I
     AATACGTACGATAACCAGCTGGACAACGTAGACGTTTATGTCGGTGGCATGCTGGAGTCCGATGGTAGACCTGGGGAACTGTTCTCGGCTGTCATCATTGATCAGTTTACCAGAATTCGT
      N  T  Y  D  N  Q  L  D  N  V  D  V  Y  V  G  G  M  L  E  S  D  G  R  P  G  E  L  F  S  A  V  I  I  D  Q  F  T  R  I  R
1441 GACGCCGATCGGTTTTGGTTCGAGAATGAAGATAACGGCATATTCACCAAAGAAGAAATAGCCGAAATACGCAAGTTCACGCTCTGGGACATCATCGTCAACAGTACGGACATCGAAGCG
 481  D  A  D  R  F  W  F  E  N  E  D  N  G  I  F  T  K  E  E  I  A  E  I  R  K  F  T  L  W  D  I  I  V  N  S  T  D  I  E  A
     GACGAAATTCAGCGTGATGTTTTCCACTGGAAGCAAGGCGATCCCTGTCCGCAACCGGAACAGCTCAACGCAACATTGCTCGAACCGTGCAACTATCTCGAGGGCTACGATTATTTCTCC
      D  E  I  Q  R  D  V  F  H  W  K  Q  G  D  P  C  P  Q  P  E  Q  L  N  A  T  L  L  E  P  C  N  Y  L  E  G  Y  D  Y  F  S
1681 GGCTCGGAGCTGGCGTACATCTACTCGTGCGTCTTTTTGGGCTTCGTGCCCATTCTATGCGCTGGTGCGGGATACTGCGTGATCAAGCTACGAACAGTCGACGCCGCAAGCTGAAAATC
 561  G  S  E  L  A  Y  I  Y  S  C  V  F  L  G  F  V  P  I  L  C  A  G  A  G  Y  C  V  I  K  L  Q  N  S  R  R  R  K  L  K  I
     AAACAAGAAGCCATGAAAAACACTGCCAACAGTAAAGTGTCGGTGGAGAAAATGGTAGCTCGCGAGTGGTTGCACGCGAACCACAAACGCTGGTGACAGTCAAATTCGGACCGGAAGCG
      K  Q  E  A  M  K  N  T  A  N  S  K  V  S  V  E  K  M  V  A  R  E  W  L  H  A  N  H  K  R  L  V  T  V  K  F  G  P  E  A
1921 GCTATCTACACGGTCGATCGCAAGGGCGAGAAACTTCGCACCTTCAATCTGAAGCACGTCGATGTGGTGACGGTGGAACAATCGCAAGAAAACTACAAGGCAAAAAAACCGTACATATTG
 641  A  I  Y  T  V  D  R  K  G  E  K  L  R  T  F  N  L  K  H  V  D  V  V  T  V  E  Q  S  Q  E  N  Y  K  A  K  K  P  Y  I  L
     CTGCGCGTCCGAGCGATCACGATCTGGTGCTGGAACTGGAATCGAACTCGGCTCGTCGGAAGTTTGTGAAAAAGCTGGAGGACTTCCTAGTGGTGCACAAAAAGTCTATGACATTCGTT
      L  R  V  P  S  D  H  D  L  V  L  E  L  E  S  N  S  A  R  R  K  F  V  K  K  L  E  D  F  L  V  V  H  K  K  S  M  T  F  V
2161 GAGTCGAATCGCGACCTAATGCTGGCCAAGGCAGAAACTCGTGAGCGTCGTCAGAAGCGGCTGGAACATTTCTTCCGTGAGGCTTACGCACTTACGTTCGGATTACGTCCCGGCGAGCGC
 721  E  S  N  R  D  L  M  L  A  K  A  E  T  R  E  R  R  Q  K  R  L  E  H  F  F  R  E  A  Y  A  L  T  F  G  L  R  P  G  E  R
     CGTCGTCGCTCGGATGCATCGCTGGAGGTAATGACGGTGATGCGCACCAGTCGTGTCAAAGTCGGAGTTTGCGGCAGCACTCGGCATGAAACCGGACGACATGTTCGTTCGAAAG
      R  R  R  S  D  A  S  L  D  G  V  M  T  V  M  R  T  S  L  S  K  S  E  F  A  A  A  L  G  M  K  P  D  D  M  F  V  R  K
2401 ATGTTCAATATCGTCGATAAGGACAAAGACGGAGAATATCCTTCCAGGAATTCTTGGAGACCGTGGTGTGTTTTCGCGCGGAAAAACTGACGATAAACTGAGGATCATCTTCGATATG
 801  M  F  N  I  V  D  K  D  K  D  G  R  I  S  F  Q  E  F  L  E  T  V  V  L  F  S  R  G  K  T  D  D  K  L  R  I  I  F  D  M
     TGTGACAACGATCGAAACGGAGTGATCGATAAGGGAGAGCTGAGTGAGATCGATCTTTGGTGGAGATCGCTCGAACCACAAGCGTGACTGATGAACAAGTTAACGAGTTGATCGAC
      C  D  N  D  R  N  G  V  I  D  K  G  E  L  S  E  M  M  R  S  L  V  E  I  A  R  T  T  S  V  T  D  E  Q  V  N  E  L  I  D
2641 GGCATGTTCCAGGACGTTGGTCTCGAGCATAAAAAATCATCTAACCTACGAGGATTTCAAGCTGATGATGAAGGAATACAAGGGAGACTTTGTCGCGATTGGCTTGGATTGTAAGGGTGCT
 881  G  M  F  Q  D  V  G  L  E  H  K  N  H  L  T  Y  E  D  F  K  L  M  M  K  E  Y  K  G  F  D  V  A  I  G  L  D  C  K  G  A
     AAGCAGAACTTCCTCGACACATCCACCAACGTGGCTCGCATGACCTCGTTCCATATTGAACCGATCTCGGACTCACGCCGTCACTGGATGCAGGAAAAGTGGGACTGCTACACGACCTTC
      K  Q  N  F  L  D  T  S  T  N  V  A  R  M  T  S  F  H  I  E  P  I  S  D  S  R  R  H  W  M  Q  E  K  W  D  C  Y  T  T  F
2881 CTGGAGGAAAATCGTCAAAACATCTTCTATCTTTTCCTGTTCTATGTTATCACGATCGTCTTGTTTGTCGAGCGATTCATTCATTATTCCTTTATGGCTGAGCACACTGATCTACGGCAC
 961  L  E  E  N  R  Q  N  I  F  Y  L  F  L  F  Y  V  I  T  I  V  L  F  V  E  R  F  I  H  Y  S  F  M  A  E  H  T  D  L  R  H
     ATCATGGGCGTCGGTATTGCGATAACACGCGGGTCCGCTGCTTCATCGTGTCATTCTGCTATTCACTCTTTTGCTGCTCACAATGTCCAGAAACCTGCTGAAAGCTGAAAGAGTTTCCCAATC
      I  M  G  V  G  I  A  I  T  R  G  S  A  A  S  L  S  F  C  Y  S  L  L  L  L  T  M  S  R  N  L  L  T  K  L  K  E  F  P  I
3121 CAACAGTACATCCCGCTTGATTCGCACATTCAATTCCACAAGATCGCCGCCTGTACGGCGCTGTTCTTTCGCTGCTCCACACCGTGGGACACATCGTGAACTTCTACCACGTTTCCACA
1041  Q  Q  Y  I  P  L  D  S  H  I  Q  F  H  K  I  A  A  C  T  A  L  F  S  L  L  H  T  V  G  H  I  V  N  F  Y  H  V  S  T
     CAATCGATCGAGAATCTGAAATGTCTCACGAAAGAAGTGCACTTTACGTCCGACTACCGGCCCGACATCACGTACTGGCTGTTCCAAACCATCACCGGCGTTACCGGCGTCATGCTGTTC
      O  S  I  E  N  L  K  C  L  T  K  E  V  H  F  T  S  D  Y  R  P  D  I  T  Y  W  L  F  Q  T  I  T  G  V  T  G  V  M  L  F
3361 GTCACAATGTGCATCATATTCGCCTTCGCCCATCCGACCATTCGCAAGAAGGCGTACAAGTTCTTCTGGAACGCACACTCCCTGTACGTGGTGCTGTACGCATTGTGTCTCGTGCACGGG
1121  V  T  M  C  I  I  F  A  F  A  H  P  T  I  R  K  K  A  Y  K  F  F  W  N  A  H  S  L  Y  V  V  L  Y  A  L  C  L  V  H  G
     CTTGCCCGGCTGACCGGTGCGCCCCAGGTTCTGGCTGTTTTTCATCGGGCCCGGCATTGTTATACACGCTCGATAAGATCGTCTCGCTTCGCACAAAGTACATGGCTGGATGTCATTGAA
      L  A  R  L  T  G  A  P  R  F  W  L  F  I  G  P  G  I  V  Y  T  L  D  K  I  V  S  L  R  T  K  Y  M  A  L  D  V  I  E
3601 ACCGATCTGCTACCATCGGATGTCATCAAGATCCCGGTTCTACCGGCCGCGCGAATCTAAAGTATCTGTCCGGCCAGTGGGTACGGCTATCGTGCACAGAGATAAAGCCGGAGGAAATGCAC
1201  T  D  L  L  P  S  D  V  K  I  R  F  Y  R  P  N  L  K  Y  L  S  G  Q  W  V  R  L  S  C  T  E  I  K  P  E  E  M  H
     AGCTTTACGCTCACGTCGGCACCGCACGAGAACTTCCTCAGCTGTCACATCAAGGCACAGGGTCCGTGGACTTGGAAGCTGCGCAACTACTTCGACCCGTGCAACTACAACCCGGACGAT
      S  F  T  L  T  S  A  P  H  E  N  F  L  S  C  H  I  K  A  Q  G  P  W  L  W  K  L  R  N  Y  F  D  P  C  N  Y  N  P  D  D
3841 CAGCCGAAGATACGCATCGAGGGACCGTTCGGTGGGGGCAATCAGGATTGGTACAAGTTCGAGGTCGCCGTTATGGTTGGTGGCGGAATCGGTGTAACGCCGTACGCATCGATCTTGAAC
1281  Q  P  K  I  R  I  E  G  P  F  G  G  G  N  Q  D  W  Y  K  F  E  V  A  V  M  V  G  G  G  I  G  V  T  P  Y  A  S  I  L  N
     GATCTCGTGTTTGGTACGAGCACCAACAGGTACTCGGGAGTGGCCTGTAAAAAGGTGTTACTTCCTCTGGCACCGTCCCCACAAGACACTTTGAGTGGTTTATCGATGTGTTGCGAGAT
      D  L  V  F  G  T  S  T  N  R  Y  S  G  V  A  C  K  K  V  Y  F  L  W  I  C  P  S  H  K  H  F  E  W  F  I  D  V  L  R  D
4081 GTGGAAGAAGGGACGTTTCGAACGTGCTGGAGATTCACATTTTCATCACACAATTCTTCCACAAATTCGACCTACGAACTACCATGCTGTACATTTGCGAGAACCACTTCCAACGACTC
1361  V  E  K  K  D  V  S  N  V  L  E  I  H  I  T  Q  F  F  H  K  F  D  L  R  T  T  M  L  Y  I  C  E  N  H  F  Q  R  L
     TCGAAAAGTCCATGTTCACCGGGCTCAAAGCAGTGAACCACTTCGGACGTCCAGCATGTCCAGTTTCTCAAATTCGTACAGAAGAAGCATTCTTACGTTTCCAAGATCGGAGTCTTC
      S  K  T  S  M  F  T  G  L  K  A  V  N  H  F  G  R  P  D  M  S  S  F  K  F  V  Q  K  K  H  S  Y  V  S  K  I  G  V  F
4321 TCGTGCGGTCCACGACCACTGACTAAGAGTGTCATGTCGGCCGTCGGATGAAGTGAACAAGAGCCGCAAATGGCCATACTTCATACATCATTTTGAAACCTTTGGATGA
1441  S  C  G  P  R  P  L  T  K  S  V  M  S  A  C  D  E  V  N  K  S  R  K  W  P  Y  F  I  H  H  F  E  T  F  G  *
```

4429 GCAGCAGCGGGCAGAAAAACCCTCTCCACACACCGTCATCGTCCTTTTTCA /4579 /5409 TATAAGATTAATTATTAATAAAAAATATTAATATTTT
                                                         Poly-A   **TTS**

Figure 4.2: The nucleotide and deduced amino acid sequences of *A. stephensi* dual oxidase (AsDuox). The AsDuox cDNA (4,428 bp) contains a complete ORF that encodes a protein of 1,475 amino acids (residue number indicated on the left). The promoter, start codon (ATG) and the stop codon (TGA) are shaded in gray color. Polyadenylation and transcription termination site (TTS) is indicated with the underline. Various transcription factors were also marked in the promoter region of AsDuox gene as predicted by JASPAR.

65

This confirmed the presence of Duox gene in *A. stephensi* genome. Further, this 398 bp long partial clone was used to prepare dsRNA to silence the AsDuox gene.

After the confirmation of AsDuox gene in *A. stephensi* genome, we further analyzed the full-length gene using Augustus software. The predicted AsDuox cDNA is of 5709 bp with an open reading frame (ORF) of 4428 bp. This encodes for a protein of 1475 amino acids. It has 5′-untranslated region (5′-UTR) of 165 bp and 3′-UTR of 1116 bp. The signal for polyadenylation, ATTTA, is found at 964 bp downstream to the stop codon (**Figure 4.2**). The deduced AsDuox protein has a predicted molecular mass of 171 kDa and an isoelectric point of 8.47. Further, to amplify the full-length cDNA of AsDuox, the primer set F7R6 are used and PCR product was amplified using cDNA of 24 h blood fed midgut as a template (**Figure 4.1B**). The primer set F7R6 amplified the expected PCR product of ~5 kb from the cDNA template. This PCR product was sequenced and BLAST to confirm the identity. The sequence was submitted to NCBI (GenBank database accession number: KY386660 and Ensembl identifier ASTE003295 in the annotated genome of *A. stephensi*). The full-length AsDuox cDNA was aligned with *A. stephensi* genomic contig 2339 sequence to determine exon-intron boundaries. We found that AsDuox gene is organized into 12 exons and 11 introns in a way similar to *A. gambiae* AgDuox gene (**Figure 4.3**). To understand the transcriptional regulation of AsDuox gene, the 5′ upstream sequence adjacent to the promoter or start codon was analyzed using JASPAR software to mark putative transcription factor binding motifs (TFBM). The 5′ sequence of AsDuox has binding site for transcription factors such as Ecdysone, GATA, and ATF-2 (**Figure 4.2**).

4.3.2 Domain Structure of AsDuox

The complete domain structure of AsDuox protein was compared with other Duoxes such as *A. gambiae* and *D. melanogaster* by SMART program (Letunic et al., 2014) (**Figure 4.4**). AsDuox has an N-terminal heme peroxidase domain of total 522 amino acids (at positions 7–528), calcium-binding domain containing 3 EF-hand (a helix-loop-helix topology) motifs of 87 amino acids (at positions 797-825, 833-861 and 878-906). A ferric–reductase domain of total 148 amino acids (at positions 1010–1157), a FAD binding domain of 122 amino acids (at positions 1173-1294) and a NAD binding domain of 157 amino acids (at positions 1300–1456) are also identified. Transmembrane domain analysis of was carried out using TMHMM Phobius (Käll et al., 2007) and seven transmembrane domains of total 158 amino acid are identified (at positions 566-588, 968-988, 1000-1025, 1055-1076, 1106-1129, 1141-1162 and 1168-1187) (**Figure 4.5**)
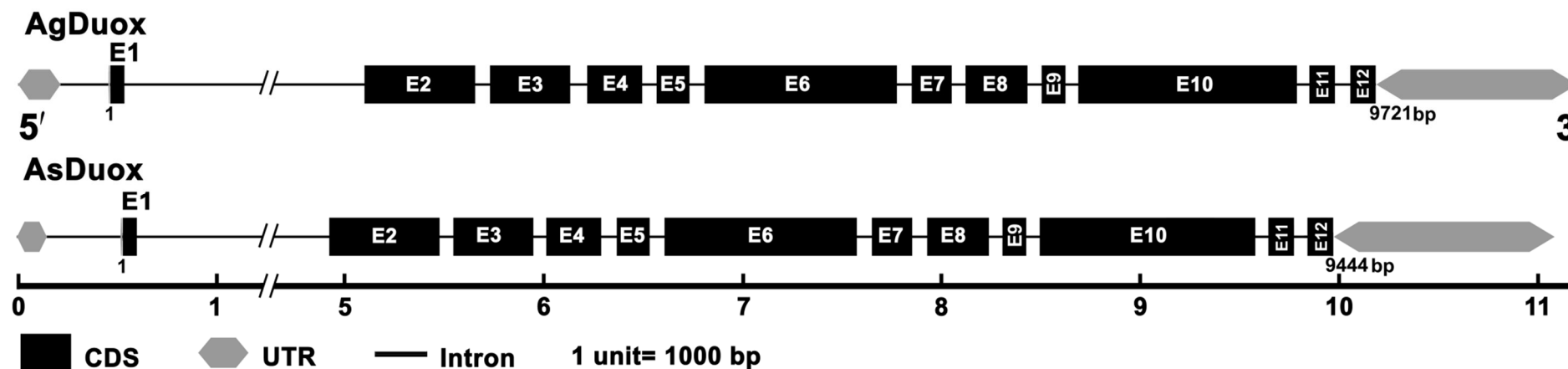
Figure 4.3: Genomic organization of Duox gene in *A. gambiae* and *A. stephensi*. The genomic structure of AgDuox and AsDuox genes revealed the presence of 12 exons. AgDuox gene is 9.72 kb long while AsDuox gene is 9.45 kb and each of the genes encode for the ORF of 4428 bp.
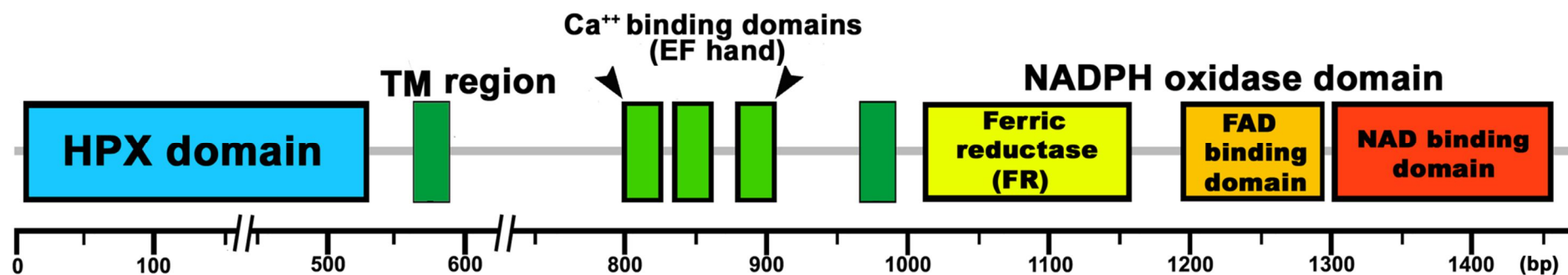
Figure 4.4: Domains organization of AsDuox protein. Schematic view of different domains present in AsDuox protein. AsDuox protein contains an N-terminal heme peroxidase domain (blue) (HPX), Transmembrane (TM) domains (dark green), EFh or EF hand (calcium binding sites) (green) and an NADPH oxidase domain that contains ferric reductase domain (yellow), FAD binding domain (golden yellow) and NAD binding domain (brisk).

The similar domain structure was found in Duoxes of many species; e.g. *A. gambiae* AgDuox, *D. melanogaster* DmDuox, *Homo sapiens* Duox1 and Duox2 (Hu et al., 2013; Sumimoto, 2008; Donkó et al., 2005). The AsDuox protein has 12 heme binding sites, 12 substrate binding sites, 19 calcium-binding sites and 25 active sites in NADPH-binding domain. The N-terminal heme peroxidase domain of the protein is non-cytoplasmic and the C-terminal NADPH oxidase domain is cytoplasmic (**Figure 4.5**). The 3-dimensional structure of AsDuox was modeled using RaptorX as shown in **Figure 4.6**. The Best template chosen by the program for peroxidase domain (1-557) was goat peroxidase (2E9E) and residues from 692-1370 were modeled based on the structure of 4IL1 (Rat calcineurin) which contains several EF-hand motifs. The residues from 1371-1475 were modeled based on the template 3A1F (NADPH-binding domain of gp91 (phox)). The software modeled the 100% (1475 amino acids) residues of AsDuox protein. The quality of predicted model is assessed by un-normalized Global Distance Test (GDT, a measure of similarity between two protein structures) and the overall GDT for AsDuox model is 620. The three-dimensional structure of AsDuox revealed domain organization similar to the 3D structure of Duox previously described in kuruma shrimp, *Marsupenaeus japonicus* (Yang et al., 2016).
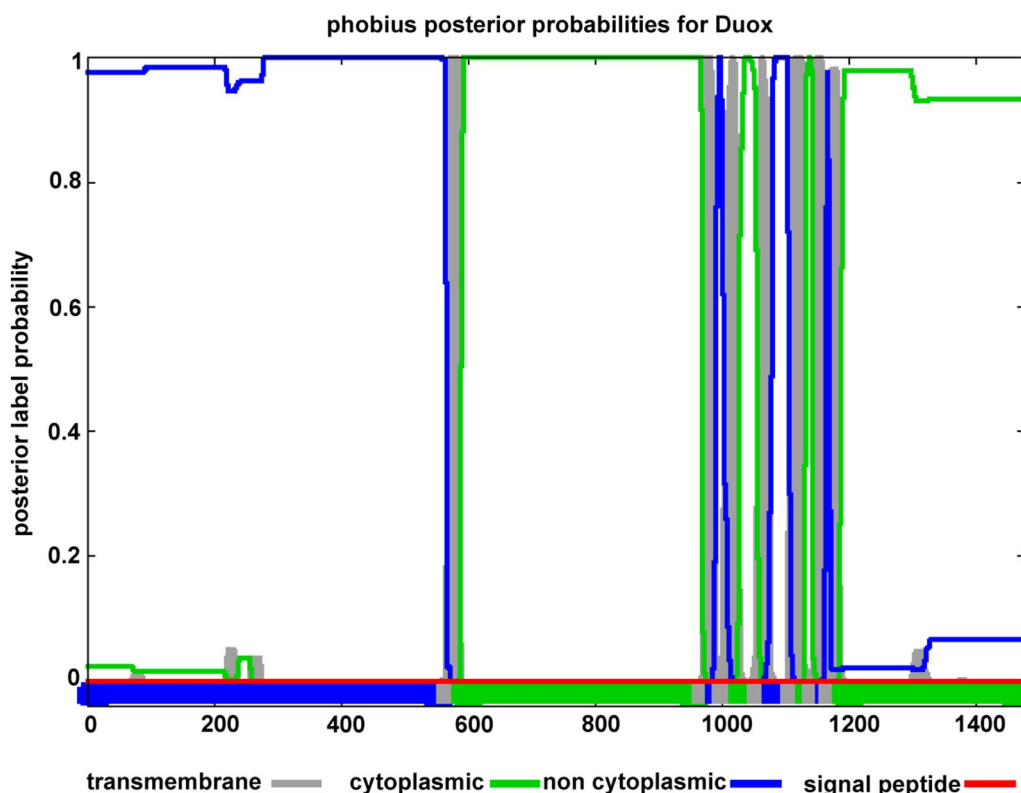
Figure 4.5: Transmembrane domains in AsDuox protein. AsDuox protein contains seven putative transmembrane domains as shown in gray color. The result showed that N-terminal is non-cytoplasmic indicated by blue color while C-terminal is cytoplasmic as shown by green color.
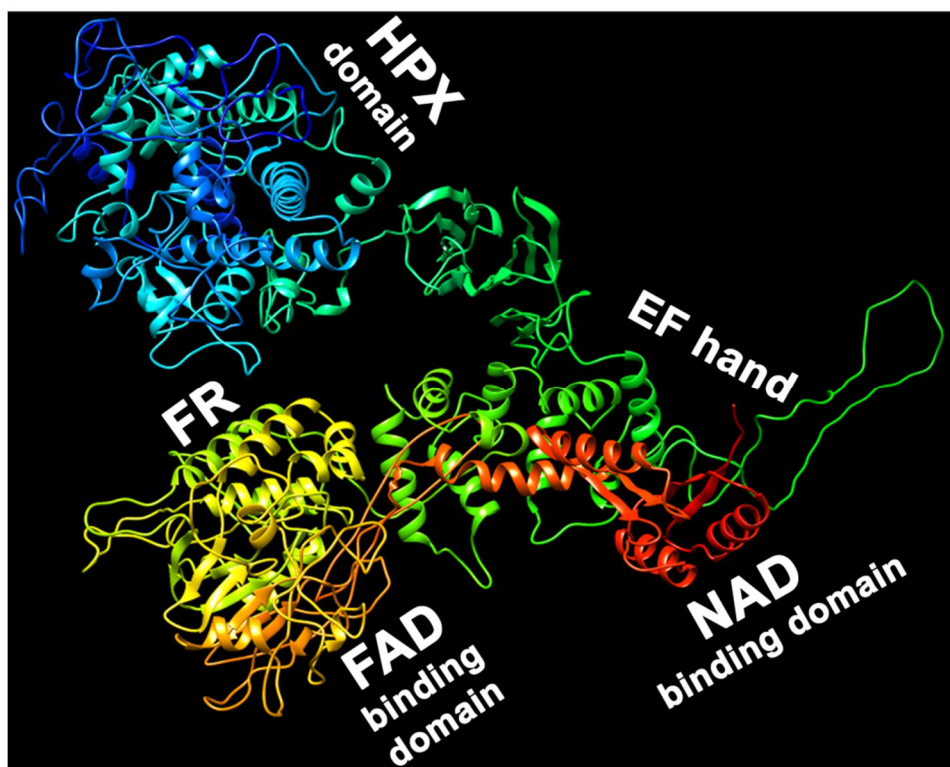
Figure 4.6: Model of three-dimensional structure of AsDUOX. The heme peroxidase (HPX) domain is highlighted in blue color and calcium binding domains in green color. NADPH oxidase domain contains Ferric reductase (FR), FAD and NAD-binding domains. The 3D structure was predicted using RaptorX software.

4.3.3 Sequence homology and phylogenetic analysis of AsDuox protein

The protein sequence of AsDuox was aligned with Duoxes of other organisms to determine the percentage of protein similarity. AsDuox protein shared sequence homology of 99%, 94% and 93% with Duox of *A. gambiae, A. aegypti* and *Culex quinquefasciatus,* respectively. The AsDuox shared more than 70% similarity with Duoxes of other insects and less than 40% similarity with Duoxes of human. To better understand the evolutionary relationships between AsDuox and several Duox from other organisms, a phylogenetic tree was constructed using MEGA 5.2 as described in Material and Methods. The accession numbers, symbols and nomenclature of sequences used in the phylogenetic analysis are given in **Table 2.3** (on page 31). The phylogenetic tree is divided into two major clusters; one is composed of Duoxes from Vertebrates and another of Duoxes from Arthropods and Nematodes as shown in **Figure 4.7**. The cluster of Arthropods was further divided into 4 major sub-clusters that include the Duoxes from Crustacean, Hymenoptera, Lepidoptera, and Diptera. The Duoxes of *Drosophila* and mosquitoes (*Anopheles*, *Aedes, and Culex*) appeared in a single cluster of order Diptera (**Figure 4.7**). The major Vertebrate Duoxes cluster is further divided into two sub-clusters of Duox1 and Duox2. Hence, it is clear from the

phylogenetic analysis that Duoxes although present in lower species to higher species animals but form different cluster with their closest match. The AsDuox formed cluster with Dipteran Duoxes and is far diverged from the vertebrate Duoxes.
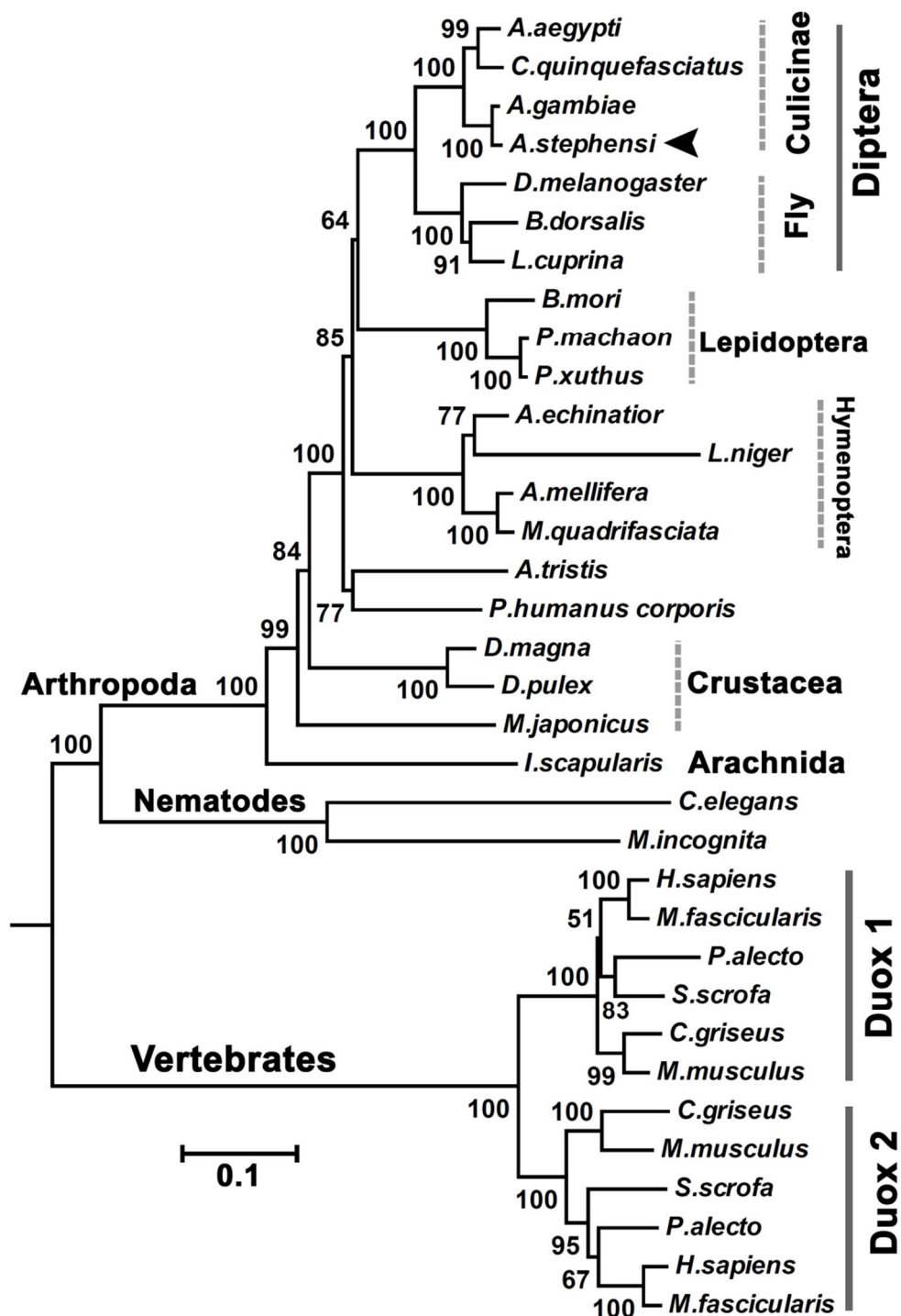
Figure 4.7: Phylogenetic analysis of *A. stephensi* Duox. Phylogenetic relationship of Duoxes from various insects, vertebrates, and nematodes was built using Neighbor joining (NJ) method implemented in MEGA 5.2. Details regarding Duox nomenclature are mentioned in **Table 2.3** (on page 31). The scale bar represents base substitutions per site. The numbers on the branches represent the % of 1000 bootstrap.

To elucidate the presence of Duox protein in the 17 worldwide distributed anophelines and similarity among them, we searched the putative ortholog of AsDuox in their genome as described in the Material and Methods. The list of these anophelines Duoxes is mentioned in **Table 2.4** (on page 33). To better understand the evolutionary relationship between Duox proteins of anophelines, we constructed the phylogenetic tree. The phylogeny of Duox proteins within the genus *Anopheles* revealed that it followed the classical classification of molecular taxonomy of *Anopheles* (**Figure 4.8**) (Harbach, 2004). Alignment of different anophelines Duox proteins using Clustal Omega revealed that Duox is highly conserved and shared 97-100% sequence identity within the genus *Anopheles* **(Table 4.1)**. Interestingly, Duox genes of anophelines listed in **Table 4.1** encode for a protein 1475 amino acids. The high conservation in Duox proteins across different species of genus *Anopheles* suggested its essential role in the fundamental cellular processes. It has been established through various studies that essential genes play a central role in the viability of the organism and should be highly conserved during evolution (Bergmiller et al., 2012).
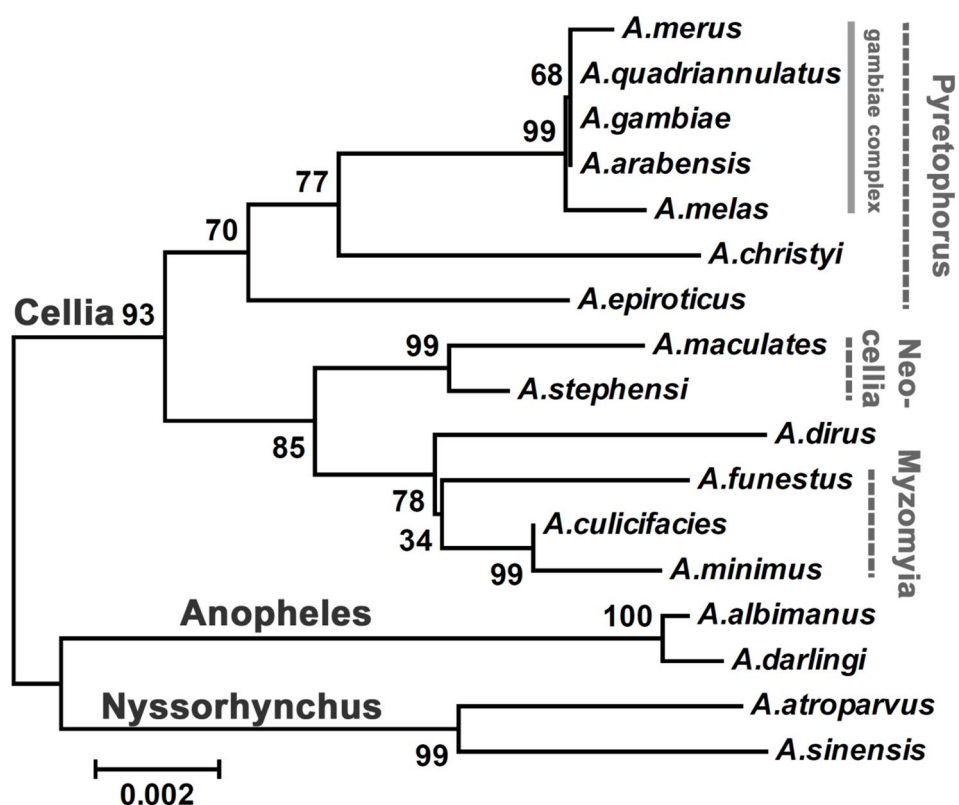
Figure 4.8: Phylogenetic analysis of putative Duoxes from different species of *Anopheles* mosquitoes. Neighbor joining (NJ) tree was constructed using Duox protein sequences from different species of *Anopheles* mosquitoes. Details regarding Duoxes from anopheline species are mentioned in **Table 2.4** (on page 33). The scale bar represents base substitutions per site. The numbers on the branches represent the % of 1000 bootstrap.

71

Table 4.1: Percentage amino acids identity among full-length Duoxes obtained from different anophelines. The identity among different Duox proteins was analyzed through their alignment in Clustal Omega. The total amino acids (AA) used for analysis and the nomenclature of individual Duox are shown in the table given below. The Duox gene ID and abbreviations of *Anopheles* species are adopted from the **Table 2.4** (on page 33).

		AA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	AatpDUOX	1475		99.05	97.90	97.90	97.76	98.17	97.76	97.97	98.10	97.90	97.76	97.83	97.83	97.90	97.90	97.90	97.83
2	AsinDUOX	1475	99.05		97.90	97.76	97.63	97.90	97.90	97.83	98.10	97.90	97.76	97.76	97.83	97.90	97.90	97.90	97.83
3	AalbDUOX	1475	97.90	97.90		99.86	97.63	97.83	98.03	97.90	98.17	97.97	98.03	97.83	97.97	98.03	98.03	98.03	97.97
4	AdarDUOX	1475	97.90	97.76	99.86		97.63	97.83	97.90	97.76	98.03	97.83	97.90	97.83	97.97	98.03	98.03	98.03	97.97
5	AmacDUOX	1475	97.76	97.63	97.63	97.63		99.59	98.71	98.98	99.12	98.92	98.71	98.44	98.51	98.58	98.58	98.58	98.51
6	AsDUOX	1475	98.17	97.90	97.83	97.83	99.59		98.92	99.19	99.32	99.12	98.85	98.71	98.71	98.78	98.78	98.78	98.71
7	AdirDUOX	1475	97.76	97.90	98.03	97.90	98.71	98.92		99.05	99.32	99.12	98.58	98.17	98.24	98.31	98.31	98.31	98.24
8	AfunDUOX	1475	97.97	97.83	97.90	97.76	98.98	99.19	99.05		99.46	99.25	98.58	98.37	98.37	98.44	98.44	98.44	98.37
9	AminDUOX	1475	98.10	98.10	98.17	98.03	99.12	99.32	99.32	99.46		99.80	98.85	98.58	98.64	98.71	98.71	98.71	98.64
10	AculDUOX	1475	97.90	97.90	97.97	97.83	98.92	99.12	99.12	99.25	99.80		98.64	98.37	98.44	98.51	98.51	98.51	98.44
11	AepiDUOX	1475	97.76	97.76	98.03	97.90	98.71	98.85	98.58	98.58	98.85	98.64		98.71	98.85	98.98	98.98	98.98	98.92
12	AchrDUOX	1475	97.83	97.76	97.83	97.83	98.44	98.71	98.17	98.37	98.58	98.37	98.71		98.92	99.05	99.05	99.05	98.98
13	AmelDUOX	1475	97.83	97.83	97.97	97.97	98.51	98.71	98.24	98.37	98.64	98.44	98.85	98.92		99.86	99.86	99.86	99.80
14	AarbDUOX	1475	97.90	97.90	98.03	98.03	98.58	98.78	98.31	98.44	98.71	98.51	98.98	99.05	99.86		100	100	99.93
15	AgDUOX	1475	97.90	97.90	98.03	98.03	98.58	98.78	98.31	98.44	98.71	98.51	98.98	99.05	99.86	100		100	99.93
16	AquaDUOX	1475	97.90	97.90	98.03	98.03	98.58	98.78	98.31	98.44	98.71	98.51	98.98	99.05	99.86	100	100		99.93
17	AmerDUOX	1475	97.83	97.83	97.97	97.97	98.51	98.71	98.24	98.37	98.64	98.44	98.92	98.98	99.80	99.93	99.93	99.93	

4.3.4 AsDuox is highly expressed in pupal stage of developmental

The mRNA levels of AsDuox gene were analyzed in different developmental stages of *A. stephensi* namely egg, larvae, pupae and adults (**Figure 4.9**). The result shown in **Figure 4.9** revealed that AsDuox is expressed in all the developmental stages of mosquito. The mRNA levels of AsDuox remained similar from egg to 3rd instar larvae but reduced significantly in 4th instar larvae by 10-fold in comparison to the eggs. The gene then showed the highest induction of 7-fold in the pupal stage in comparison to the eggs. In the adult stages, AsDuox gene expression is significantly higher in the adult females by 2-fold in comparison to the eggs. In *A. stephensi*, during development from 1st instar to 3rd instar larvae there is molting, change in cuticle layer and growth in size occur. The extracellular cuticle should be strong enough to protect larvae from the external environment (temperature, moisture and

Figure 4.9: Relative mRNA levels of AsDuox gene in different developmental stages of *A. stephensi.* Results presented the mean ± SD of relative mRNA levels of AsDuox in different developmental stages of *A. stephensi*. The mRNA levels of eggs were considered as 1. Significant differences ($p < 0.001$ or < 0.01) in the relative mRNA levels of different stages against eggs are indicated by three or two asterisks (*), respectively.

pathogens, etc.), or it will die (Tajiri et al., 2017; Page and Johnstone, 2007). In *A. stephensi,* AsDuox gene might be responsible for the proper formation of extracellular cuticle layer as reported earlier in *Caenorhabditis elegans* that its ortholog stabilizes the extracellular cuticle layer. In *C. elegans,* the extracellular cuticle matrix of larvae is stabilized by the cross-linking of tyrosine residues of cuticle protein and is catalyzed by the peroxidase MLT-7 and the dual oxidase BLI-3. The collagen cross-linking and proper extracellular matrix formation are critical for post-embryonic viability in *C. elegans.* Silencing of either gene causes body morphology defects most notably molt,
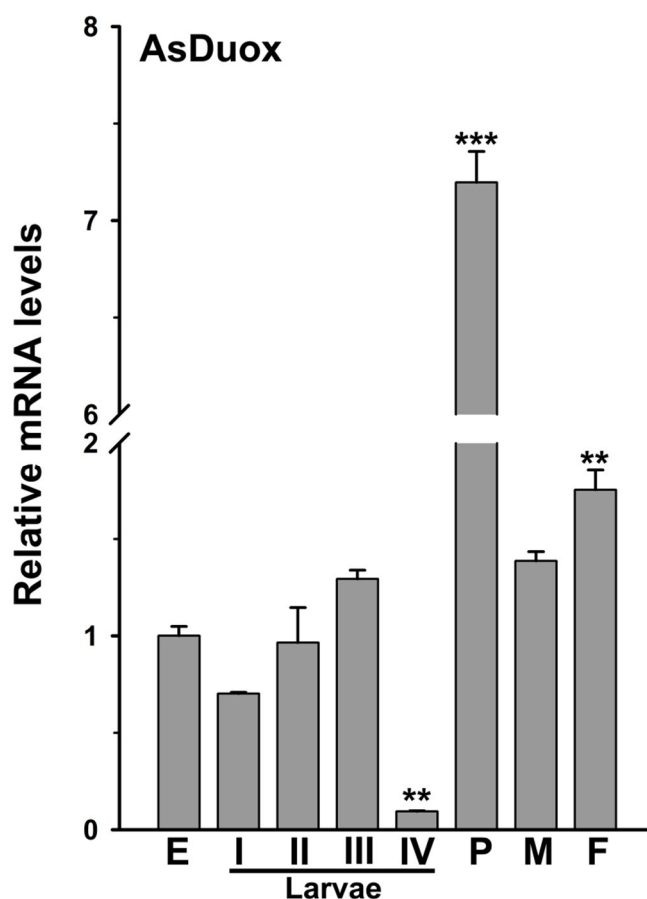
dumpy and arrest at early larval stage (Edens et al., 2001; Thein et al., 2009). The presence of ecdysone binding site is in the promoter region of AsDuox gene might explain the reduced expression of this gene in 4th instar larvae. It has been reported in *Drosophila* that during larval growth the high amount of ecdysone hormone induces molting and its levels are reduced in the last stage of larvae that causes larvae to enter the non-feeding or wandering stage (Nijhout et al., 2014). Further, the induced expression of AsDuox gene in pupal stage might indicate an important role of this gene in pupae development mainly for the stabilization of the cuticle structure present like in the wings through tyrosine cross-linking processes. This assumption was further supported by a study carried out in *Drosophila* that revealed the role of Duox in the development of normal wings via stabilization of the cuticle structure (Anh et al., 2011). Another study in *Drosophila* reported that protein named Curly Su (Cysu) which acts as Duox stabilizes the wing on the last day of pupal development (Hurd et al., 2015). Moreover, the high expression of AsDuox in pupae might be due to the presence of putative binding motif of ecdysone hormone in the promoter region of this gene **(Figure 4.2)**, which regulates the larval molting and metamorphosis in other insects like *Drosophila* and *Aedes aegypti* (Akagi and Ueda, 2011; Telang et al., 2007).

4.3.5 Expression of AsDuox gene is induced in bacteria inoculated 4th instar larvae of *A. stephensi*

The expression of AsDuox in bacteria (*E. coli* and *M. luteus*) challenged 4th instar larvae was evaluated using qPCR as described in Material and Methods to decipher its role in immunity. The result presented in **Figure 4.10** revealed that expression of 16S rRNA is increased gradually post bacteria challenge. To confirm that bacteria given act as immune elicitors, the expression of the well-known anti-bacterial peptide defensin was analyzed (Tanji et al., 2007). The mRNA levels of defensin showed 363-fold induction post 30 minutes of bacterial injection in the larvae against non-injected larvae. The mRNA levels were induced 45-fold, 91-fold and 286-fold at 1 h, 2 h and 4 h, respectively post bacterial challenge against non-injected larvae. The mRNA levels of defensin showed induction of 2170-fold at 6 h, 770-fold and 1824-fold at 9 h and 12 h, respectively post bacterial challenge against non-injected larvae (**Figure 4.10**). The induced expression of defensin gene is also reported in bacteria infected larvae of *Spodoptera littoralis* (Seufi et al., 2011). Expression analysis of AsDuox gene showed upregulation of 2-fold and 4-fold at 6 h and 12 h post bacterial challenge, respectively as shown in **Figure 4.10**. The defensin responds rapidly to the bacterial challenge while
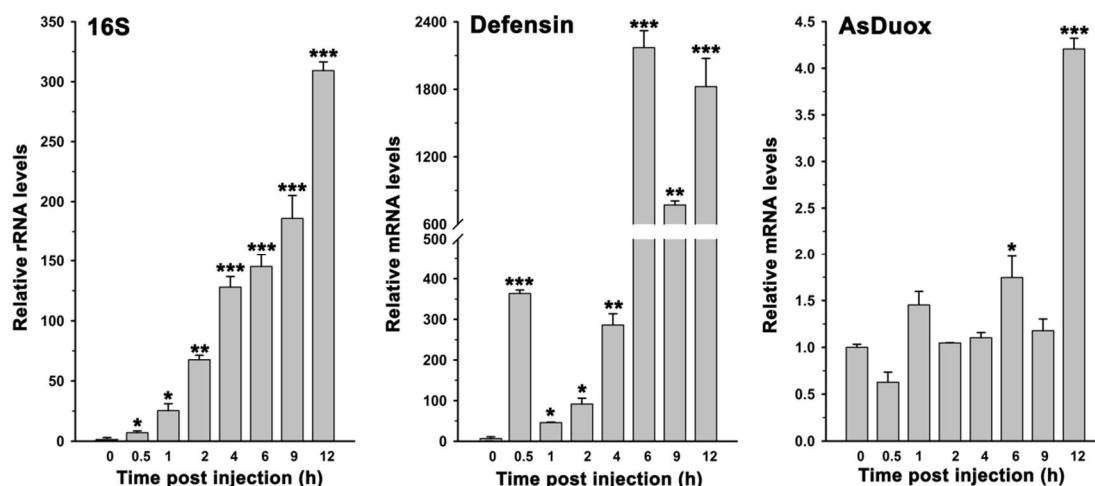
Figure 4.10: Expression analysis of 16S rRNA, defensin and AsDuox genes in bacteria challenged 4th instar *A. stephensi* larvae. Relative levels of 16S rRNA, defensin and AsDuox mRNA were analyzed at different time point after bacterial challenge. 0 h represents the non-injected larvae and is taken as 1. Relative levels at each time points are presented against their respective PBS injected controls. Significant differences ($p<0.001$, 0.01 or 0.05) against control are indicated by three, two or one asterisk (*), respectively.

Duox showed upregulation at the later time points. The similar pattern of expression of AMP genes like cecropin, attacin, defensin, diptericin, muscin, and domesticin were observed in the *Musca domestica* larvae during bacterial infection (Tang et al., 2014). In *C. elegans* similar type of study revealed that although there is no induction of Duox (BLI-3) in response to infection with *Enterococcus faecalis* or *Candida albicans* but Duox mutant *C. elegans* is more susceptible to infection (Van der Hoeven et al., 2015). Similarly, Duox in Zebrafish larvae conferred immunity against enteric pathogen *Salmonella* and in *Bombyx mori*, Duox showed significant upregulation in the larval midgut upon challenge by *E. coli* and *B. mori* nucleopolyhedrovirus (BmNPV) (Flores et al., 2010; Hu et al., 2013). This suggested that like other organisms, in *A. stephensi,* AsDuox is also actively involved in immune responses against bacteria in larvae.

4.3.6 Blood feeding induces expression of AsDuox gene in midguts

The expression of AsDuox gene was analyzed in different body compartments of *A. stephensi* to decipher its spatial-temporal expression. For that, we compared the relative mRNA levels of AsDuox gene in sugar fed or blood fed female midguts and carcasses post 24 h blood feeding. The relative mRNA levels of AsDuox were 100-fold higher in the carcasses than midguts of sugar fed mosquitoes. The AsDUOX mRNA levels post 24 h blood feeding are significantly upregulated by 7-fold in the midguts (**Figure 4.11**). However, there is no significant difference in the expression of AsDuox in the 24 h blood-fed carcasses in comparison to the sugar fed carcasses (**Figure 4.11**). Thus, AsDuox is induced in blood fed midguts only. We found two transcription factors binding

motifs in the regulatory region of the gene (**Figure 4.2**) that might explain its transcriptional regulation in the blood fed mosquitoes. One of the transcription factors is

ecdysone, which might be responsible for the induction of AsDuox gene in response to blood meal, as the ecdysone levels increase soon after the blood feeding in the mosquito and in other insect body (Swevers and Latrou, 2003; Hagedorn et al., 1975). The other transcription factor is GATA that might be responsible for the midgut-specific induction of AsDuox gene. It drives the tissue-specific expression of genes as reported earlier in *Drosophila* (diptericin) and mosquito (hexamerin-1.2) (Jinwal et al., 2006; Senger et al., 2006).
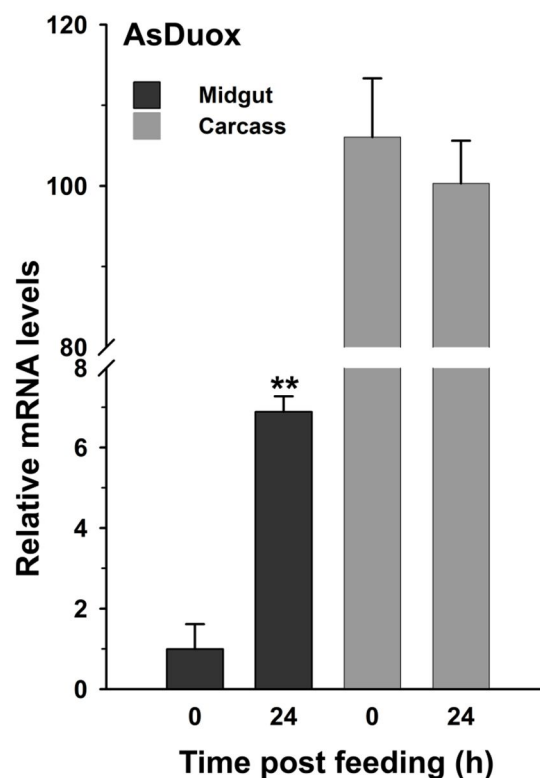
Figure 4.11: Relative mRNA levels of AsDuox in different body compartments of mosquito. Relative mRNA expression levels were analyzed in midgut and carcass of sugar fed (0 h) and 24 h post blood fed female mosquitoes. The mRNA levels of sugar fed midgut were considered as controls. Significant differences ($p<0.01$) are indicated by two asterisks (*).

4.3.7 AsDuox maintains microbial homeostasis in sugar fed midguts

The life cycle of *Drosophila* and *Anopheles* is much more alike, having stages of eggs, larvae, pupae and adults. One of the major differences between *Drosophila* and *Anopheles* is in the nutritional behavior of female adults. *Drosophila* is saprophagous in nature (Anagnostou et al., 2010) while for egg development *Anopheles* female relies on blood meal (Logue et al., 2016). Besides taking the blood meal, female *Anopheles* also feed on the nectar of plants. Hence, the sugar fed midgut of female *Anopheles* might resemble the midgut of female *Drosophila* adult as both reflect the physiology of non-hematophagous insect. Since *Drosophila* is saprophagous in nature, it encounters more food-borne micro-organisms (Broderick and Lemaitre. 2012). It is reported that in the midguts of *Drosophila* the bacterial homeostasis is maintained by Duox (Ha et al., 2005a). So, it was hypothesized that bacterial homeostasis is maintained by Duox in the sugar fed midguts of the female *Anopheles*.

For this, silencing of AsDuox gene was done using RNAi. Two groups of 60 mosquitoes were taken, one group is injected with dsLacZ (control) and another group is injected with dsAsDuox (silenced) RNA after that they were continuously fed on 10% sugar solution and survival rate of mosquitoes was recorded. The sugar fed midguts were collected after 4th day of dsRNA injection as described in Materials and Methods. The mRNA levels of different genes were then analyzed in these midguts. The mRNA levels of AsDuox in control and silenced midguts revealed that we could achieve 70% silencing of this gene in sugar fed midguts (**Figure 4.12**). The 16S rRNA levels of endogenous bacteria were increased by 32-fold in AsDuox silenced midguts as compared to the non-silenced controls (**Figure 4.12**). The data shown in
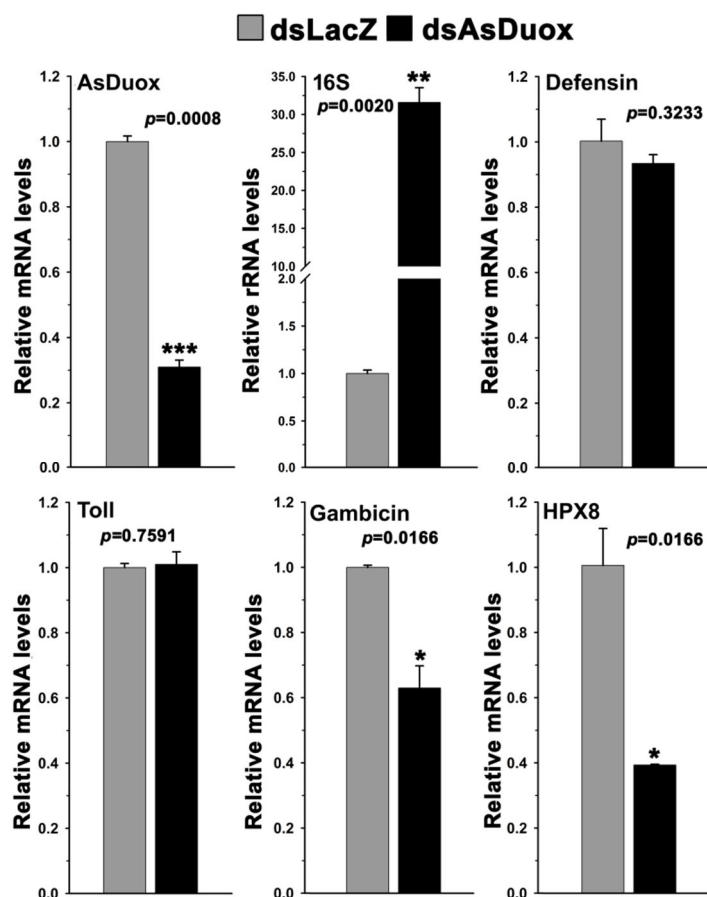
Figure 4.12: Relative mRNA levels of different genes in AsDuox silenced sugar fed midguts. Mosquitoes injected with dsLacZ (controls) or dsAsDuox (silenced) RNA were fed on sugar and expression of AsDuox mRNA, 16S rRNA and other immune genes was analyzed. Relative mRNA levels are presented against control midguts. Significant differences *p<* 0.001, 0.01 or 0.05 are shown by three, two or one asterisk (*), respectively.

Figure 4.12 are in agreement with the previous reports where Duox from other insects such as *A. aegypti*, *D. melanogaster*, *B. dorsalis* and *Marsupenaeus japonicus* regulates the homeostasis of gut bacterial community and silencing of Duox increases the bacterial load in the gut (Yao et al., 2016; Inada et al., 2013; Oliveira et al., 2011; Ha et al., 2005a).

In spite of the increased bacterial growth, we observed indifferent changes in the survival of AsDuox silenced mosquitoes as compared to the controls. The increased endogenous bacterial load can easily trigger the immune responses in the midgut. Hence, we decided to study the expression of known anti-bacterial immune genes in

these control and silenced midguts. The expression analysis of known anti-bacterial genes showed either non-significant changes or downregulation in their mRNA levels in control and silenced midguts. The mRNA levels of defensin and Toll showed non-significant differences in silenced mosquito midguts against controls. On the other hand, the mRNA levels of gambicin and HPX8 showed downregulation in silenced midguts as compared to the controls (p=0.0166 and p=0.0166, respectively). The downregulation of gambicin and HPX8 may be controlled by a single mechanism similar to the *Drosophila*. In *Drosophila,* although there is activation of IMD pathway in response to the endogenous bacteria, however, its negative regulator Caudal suppresses the production of AMPs (Lemaitre and Hoffmann, 2007). So, a similar mechanism might be working that regulates the expression of AMPs against endogenous bacteria in the midgut of *A. stephensi*.

This data showed that like *Drosophila*, *Anopheles* midgut epithelial does not mount an effective immune response against endogenous bacteria (Capo et al., 2016). Thus, finely tuned regulation of signaling mechanism developed immune tolerant environment against endogenous bacteria in the sugar fed midguts. From the above study, it can be concluded that Duox is one of the key molecules that play a major role in maintaining the bacterial homeostasis in sugar fed midguts of *A. stephensi*.

4.3.8 Blood feeding induces bacterial growth and expression of AsDuox gene in *A. stephensi* midgut

To decipher the expression kinetics of AsDuox gene after blood feeding, we analyzed its relative mRNA levels in the blood fed midguts at different time points. The result presented in **Figure 4.13** revealed that AsDuox is induced by 1.5-fold at 3 h post blood feeding in midguts against the sugar fed midguts (p= 0.0101). The relative mRNA levels of AsDuox gene in blood fed midguts further increased with time and induced 20-fold at 6 h (p= 0.0007) against sugar fed midguts. At 24 h, the mRNA levels of AsDuox in the blood fed midgut is 4-fold in comparison to the sugar fed midguts (p= 0.0058). Hence, AsDuox gene is induced by blood feeding in the midguts (**Figure 4.13**). The bacterial growth in the blood fed midguts was analyzed using 16S rRNA levels. The 16S rRNA levels showed 14-fold induction at 3 h in blood fed midguts (p< 0.0001) against sugar fed midguts. The 16S rRNA levels increased by 9-fold at 6 h (p= 0.0061), 18-fold at 12 h (p= 0.0022), 11-fold at 18 h (p= 0.0049) and 24-fold at 24 h (p= 0.0005) in the blood fed midguts against sugar fed midguts. This data is in accordance with the previous findings where bacteria proliferate after blood feeding and is necessary for the proper digestion

of blood (Kumar et al., 2010; Kajla et al., 2015b; Minard et al., 2013). These bacteria are protected from midgut epithelial immunity by the formation of midgut barrier by HPX15/Duox system as reported in *A. gambiae* (Kumar et al., 2010). But at the same

time bacteria should not over-proliferate to cause any deleterious effect on the survival of the host. Analysis of rRNA levels of 16S and mRNA levels of AsDuox in blood fed midguts at different time points revealed a weak negative correlation (r=-0.529, *p*= 0.0425) between bacterial growth and AsDuox expres-sion. Hence, we hypothesized that *A. stephensi* Duox, an ortholog of *A. gambiae* Duox, not only protects the gut

Figure 4.13: Relative levels of 16S rRNA and AsDuox mRNA in blood fed midguts. Mosquitoes were fed on blood and relative levels of AsDuox mRNA and 16S rRNA were analyzed in their midguts at different time points after feeding. Relative levels are presented against the sugar fed midguts.

bacteria through barrier formation but might also be responsible for balancing their population after the blood meal (Kumar et al., 2010).

4.3.9 AsDuox maintains bacterial homeostasis in blood fed midguts

Female *Anopheles* requires a blood meal for the development of eggs. The mosquito feeds on sterile blood and endogenous bacteria present in the mosquito midgut proliferate (Kumar et al., 2010; Oliveira et al., 2011). Hence, it is required for the mosquito to develop some strategy so that at the same time it allows the proliferation of the bacteria but not over proliferation.

It is evident from **Figure 4.13** that bacteria grow post blood feeding in the midgut and a weak negative correlation between bacterial growth and AsDuox expression is observed. This suggested that AsDuox maintains bacterial growth in blood fed midguts. To further support our hypothesis, we silenced AsDuox gene in the midguts and allowed the female *Anopheles* to feed on a healthy mouse. The time points post feeding early at 6 h and later at 24 h were selected based on the expression kinetics of AsDuox gene in blood fed midguts for further analysis.

The mRNA levels of AsDuox gene in control and silenced midguts post 6 h and 24 h blood meal revealed that this gene was silenced 80% and 79%, respectively as shown in **Figure 4.14**. The relative 16S rRNA levels in controls and silenced midguts are analyzed and revealed that in AsDuox silenced blood fed midguts there is a significant increase in the bacterial load **(Figure 4.14)**. The levels of 16S rRNA are induced 10-fold and 5-fold at 6 h and 24 h post blood feeding,

Figure 4.14: Relative levels of 16S rRNA in AsDuox silenced blood fed midguts. Mosquitoes injected with dsLacZ (control) or dsAsDuox (silenced) RNA were fed on a mouse and relative levels of AsDuox mRNA and 16S rRNA were analyzed in their midguts at 6 h or 24 h post blood feeding. Relative levels are presented against the control midguts. Significant differences ($p < 0.01$ or 0.05) between control and silenced midguts are shown by two or one asterisk (*), respectively.

respectively in the silenced midguts against controls. This data supported our hypothesis that AsDuox is responsible for maintaining the midgut microbial homeostasis in the blood fed midguts. This result is similar as described in the previous section where AsDuox silencing in sugar fed midguts also increased bacterial growth (**Figure 4.12**). It might be possible that in AsDuox silenced blood fed midgut of *A. stephensi*, only one or certain types of bacteria are proliferating, which needs further analysis of bacterial communities. Reports from other insects like *A. aegypti*, *D. melanogaster* and *B. dorsalis* revealed that there is a shift in midgut commensal bacterial population in Duox silenced midguts (Yao et al., 2016; Xiao et al., 2017).

Further to explore the immune responses against an increased bacterial load in AsDuox silenced blood fed midgut, we analyzed the expression of AMPs such as, defensin and gambicin that are controlled by both IMD and TOLL pathways (Tanji et al., 2007; Vizioli et al., 2001). Data presented in **Figure 4.15** showed that expression of defensin is similar at 6 h but reduced significantly at 24 h in AsDuox silenced blood fed midguts as compared to the controls. The expression of gambicin is induced at 6 h but reduced significantly at 24 h in AsDuox silenced blood fed midguts as compared to the controls (**Figure 4.15**).

The crosslinking of mucin layer upon blood feeding by HPX15 and Duox creates a low immunity zone and hence leads to an immune tolerant environment in the midgut (Kumar et al., 2010). However, in AsDuox silenced midguts where mucin barrier has been broken, induction of immunity is not pronounced against the increased
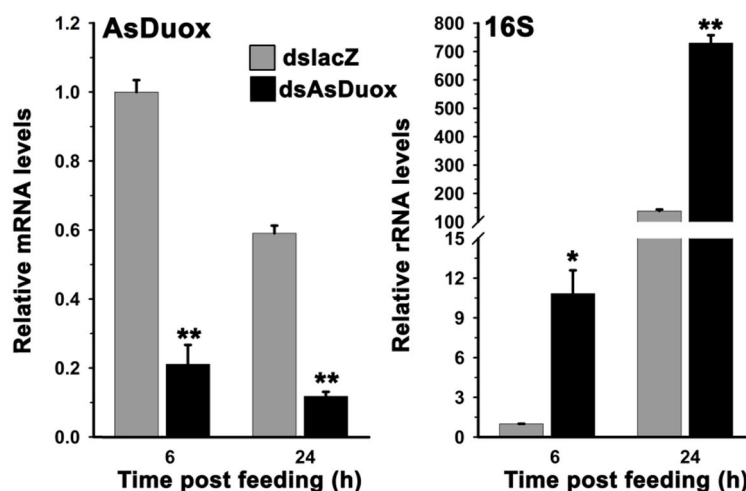

Figure 4.15: Relative mRNA levels of immune genes in AsDuox silenced blood fed midguts. Mosquitoes injected with dsLacZ (control) or dsAsDuox (silenced) were fed on a mouse and relative levels of defensin and gambicin genes were analyzed in their midguts at different time points after feeding. Relative levels are presented against control midguts. Significant differences ($p < 0.05$) between control and silenced midguts are shown by one asterisk (*).

sed bacterial load. In AsDuox silenced blood fed midguts increased bacterial load did not induce the expression of known anti-bacterial peptide like defensin and gambicin. Hence, mosquito gut can tolerate the endogenous bacteria without mounting effective immune responses against them. This process might be similar to the *Drosophila* where Caudal, a negative regulator of IMD pathway suppresses the activation of pathway against endogenous bacteria. The similar result was observed in *B. dorsalis* where feeding with dominant communities of endogenous bacteria does not evoke induction of Duox and hence midgut is tolerant to commensal bacteria (Yao et al., 2016).

4.3.9.1 Induction of anti-plasmodial genes in AsDuox silenced blood fed midguts

It is noteworthy to mention that bacteria proliferate extensively after blood feeding in the midguts and so, levels of bacteria are very high in AsDuox silenced blood fed midguts than AsDuox silenced sugar fed midguts. This high bacterial load can easily trigger the death of mosquitoes but we did not observe significant change in the rate of survival. So, we also analyzed the expression of other immune genes that might be involved in regulating the bacterial growth. Expression analysis of well known anti-plasmodial genes, NOS and TEP1 showed induced expression of these genes in AsDuox silenced blood fed midguts against controls (**Figure 4.16**). We also analyzed the expression of an anti-plasmodial HPX2/NOX5 system (Oliveira et al., 2012) and found that it is

significantly induced in the AsDuox silenced blood fed midguts against controls as shown in **Figure 4.16**. As reported in *A. gambiae* that Duox gene catalyzes the crosslinking of mucin layer in the midgut and silencing of this gene causes disruption of mucin barrier (Kumar et al., 2010). In AsDuox silenced blood fed midguts, the disruption of barrier and midgut bre-aching by bacteria might induce the expression of antiplasmodial immune molecules. It is note-worthy to mention that midgut epithelial cells can distinguish between endo-genous and pathogenic bacteria. So, it might have some mechanism that regulates the increased endogenous bacterial growth to avoid any delet-erious effect on survival of host and also it should not completely kill the bac-teria, as they are required
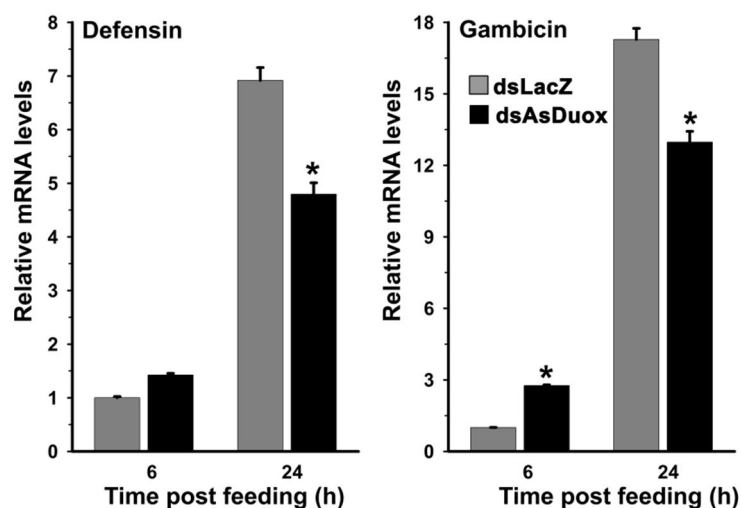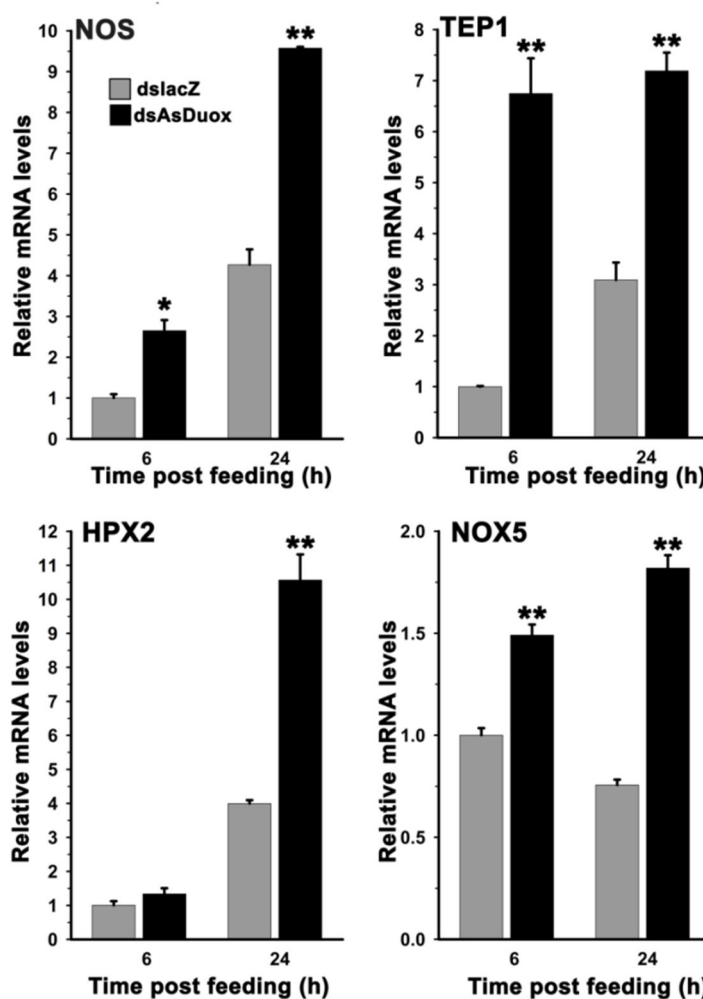

Figure 4.16: Relative mRNA levels of various anti-plasmodial genes in AsDuox silenced blood fed midguts. Mosquitoes injected with dsLacZ (control) or dsAsDuox (silenced) were fed on a mouse and relative levels of anti-plasmodial genes such as NOS, TEP1 and HPX2/NOX5 were analyzed in their midguts at different time points after feeding. Relative levels are presented against the dsLacZ injected midguts. Significant differences ($p < 0.01$) between control and silenced midguts are shown by two asterisks (*).

for the proper physiology of mosquito (Kalja et al., 2015a). Thus, the induced expression of anti-plasmodial gene suggested their involvement in regulating the bacterial growth as described earlier (Kajla et al., 2016c; Blandin et al., 2004).

4.3.10 AsDuox is induced in *Plasmodium berghei* infected midguts

A previous study has revealed that Duox in *A. gambiae* is upregulated in *P. berghei* infected midgut and silencing of this gene negatively affects the development of *Plasmodium*. So, we want to understand the role of AsDUOX during midgut stages of *Plasmodium* development. For this, the control (uninfected mouse fed) or *P. berghei* infected mouse fed mosquito midguts were collected at different time points and
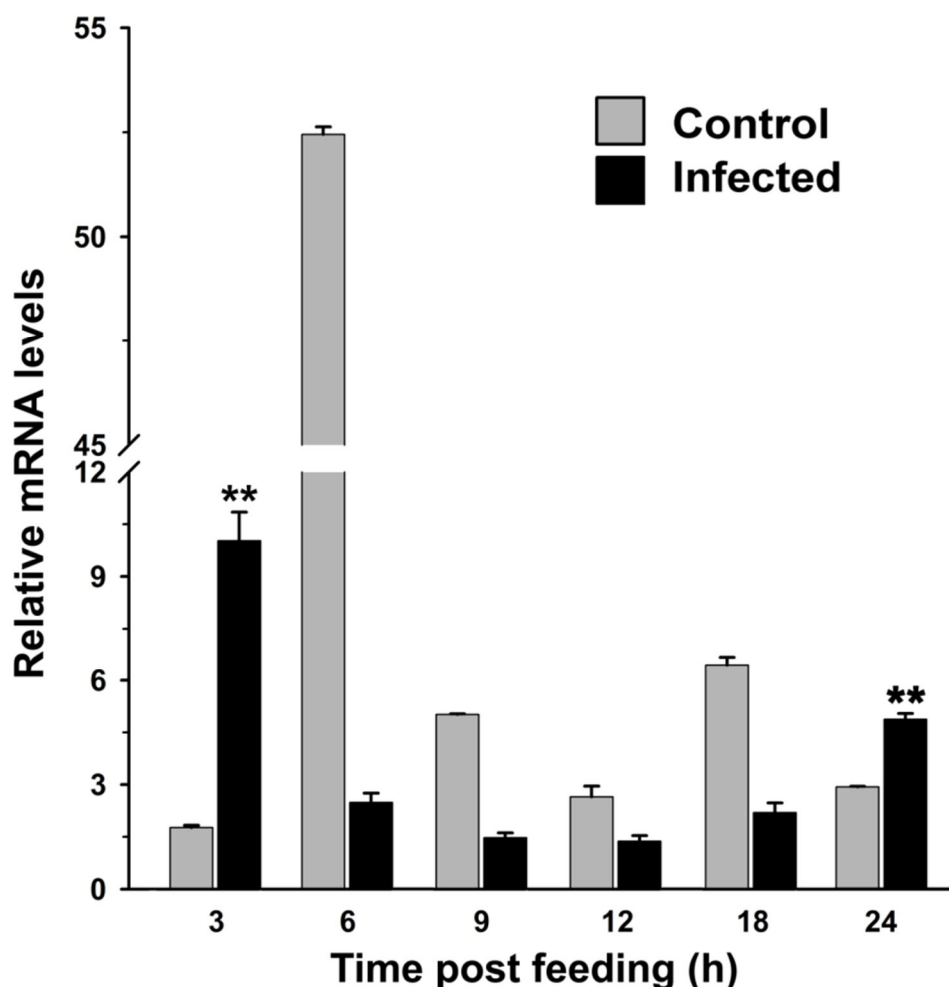
Figure 4.17: Expression kinetics of AsDuox gene during *P. berghei* infection in the midguts. Relative mRNA levels of AsDuox gene were analyzed in *A. stephensi* midguts at different time points after normal (Control) or *P. berghei*-infected blood feeding. Relative fold induction was calculated against sugar fed midguts. Significant differences $p < 0.01$ between control and infected midguts are indicated by two asterisks (*).

AsDuox gene expression was analyzed. The result presented in **Figure 4.17** showed that mRNA levels of this gene are induced in *P. berghei* infected blood fed midguts against controls. The expression of AsDuox gene is induced 7-fold at 3 h in *P. berghei* infected midgut in comparison to the respective non-infected controls. Gamete fertilization and zygote formation predominate at 3 h post infected blood feeding in the midgut. The significant reduction in the mRNA levels of AsDuox was observed from 6 h

to 18 h in *Plasmodium* infected midguts in comparison to their respective controls. The reduced expression of AsDuox gene might be regulated by *Plasmodium* to avoid oxidative burst caused by increased ROS level. At 24 h, this gene showed significant upregulation in the *Plasmodium* infected midguts against blood fed controls as shown in **Figure 4.17**. The similar pattern of expression was observed in the mRNA levels of the AgDuox at 24 h in *P. berghei* infected midguts as compared to the blood fed controls (Kumar et al., 2004). At 24 h, *Plasmodium* is motile ookinete and is invading the midgut epithelial cells. The induced mRNA levels of AsDuox in *P. berghei* infected midguts might suggest its role in *Plasmodium* development which was further confirmed using RNAi.

4.3.10.1   AsDuox silencing decreases *Plasmodium* oocysts number

To decipher the role of AsDuox in the development of *Plasmodium*, RNAi-mediated gene silencing approach was used. For this, we compared the number of *Plasmodium* oocysts in dsLacZ (control) and dsAsDuox (silenced) RNA injected midguts as discussed in Materials and Methods. The relative AsDuox mRNA levels in control and AsDuox silenced midguts revealed that we could achieve 80% silencing of this gene (**Figure 4.18**). In these control and AsDuox silenced midguts, we analyzed the expression of 28S rRNA of *P. berghei*, to study the effect on its growth.
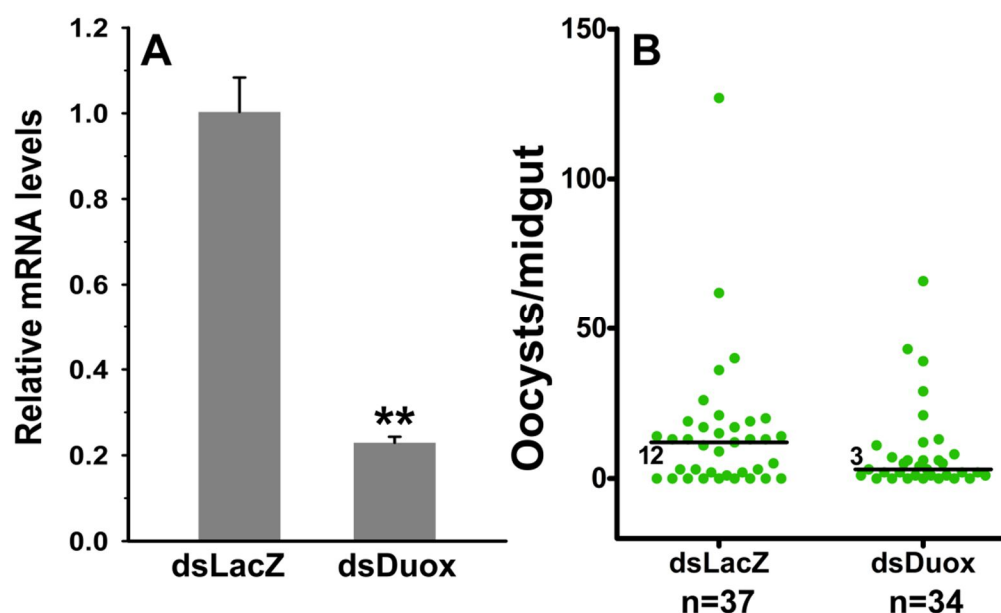
Figure 4.18: AsDuox silencing in the midgut suppresses *P. berghei* development. A) Relative mRNA abundance of AsDuox gene in dsLacZ (unsilenced) or dsAsDuox (silenced) RNA injected and 24 h post *P. berghei* infected midguts. Asterisks showed the significant difference ($p < 0.01$) between unsilenced and silenced midguts. **B)** The effect of midgut AsDuox silencing on the number of live oocysts (green dots) seven days post *P. berghei* infection. The horizontal line indicates median number of parasites with $p=0.017$ calculated by the Kolmogorov-Smirnov test. n = number of mosquitoes.

The result shown in **Figure 4.19** revealed reduced expression of 28S rRNA in the AsDuox silenced *P. berghei* infected midguts 24 h post feeding against controls. Further, we compared the number of developing oocysts in control and silenced midguts post 7 days of *Plasmodium*-infected blood feeding. Results presented in **Figure 4.18** revealed that the variable numbers of oocysts were observed in control and silenced midguts. However, the median value of the oocysts numbers in control and AsDuox silenced midguts are 12 and 3, respectively (p= 0.017). These results revealed that the oocysts numbers are reduced significantly in the silenced midguts when compared to the controls (**Figure 4.18**). Reduction in the oocysts number by 4-fold in the AsDuox silenced midguts indicated the positive role of this gene in *Plasmodium* development. Thus, AsDuox gene is *Plasmodium* agonist that positively regulates *Plasmodium* development inside the mosquito midgut. These results are in accordance with the previous findings where silencing of AgDuox gene in *A. gambiae* also reduced *Plasmodium* oocysts numbers (Kumar et al., 2010).
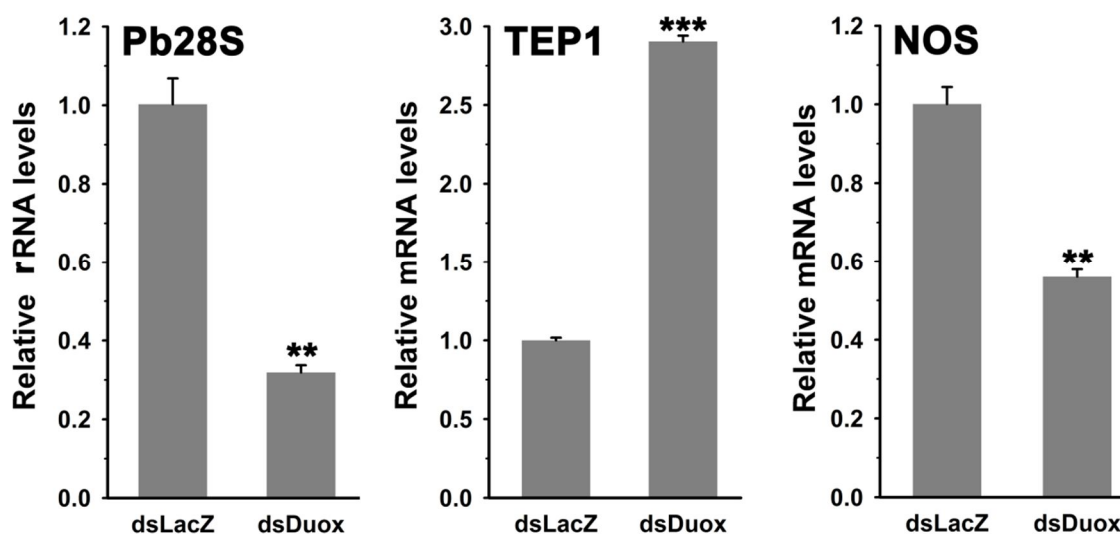
Figure 4.19: Expression of *P. berghei* 28S rRNA and immune genes in AsDuox silenced *P. berghei* infected midguts. Relative mRNA expression levels of various immune genes Thioester-containing Protein 1 (TEP1) and Nitric Oxide Synthase (NOS) in 24 h post *P. berghei* infected blood fed *A. stephensi* mosquitoes midguts injected with dsLacZ (control) or dsAsDuox (silenced) RNA. Significant difference (p < 0.001 or 0.01) between control and silenced midguts are shown by three or two asterisks (*), respectively.

In the previous study, it has been shown that silencing of Duox in *A. gambiae* disrupt the crosslinking of mucin barrier. This makes immune system to recognize the parasite and activates the anti-plasmodial immunity (Kumar et al., 2010). Thus, we assumed that the silencing of AsDuox gene and *Plasmodium* infection might induce anti-plasmodial immunity and these assumptions were confirmed by analyzing the mRNA levels of various known anti-plasmodial genes (Blandin et al., 2004; Kumar et al.,

2010) in the AsDuox silenced *P. berghei* infected blood fed midguts. Results presented in **Figure 4.19** revealed that the relative mRNA levels of an antiplasmodial immune gene thioester-containing protein 1 (TEP1) are induced 3-fold in AsDuox silenced mosquito midguts post 24 h of *P. berghei* infected blood feeding as compared to the controls (*p*=0.0002). Furthermore, expression analysis of another antiplasmodial immune gene Nitric Oxide Synthase (NOS) in the above samples revealed that its mRNA levels reduced 2-fold in the silenced midguts against the controls (**Figure 4.19**, *p*=0.005). These findings suggested that the elevated TEP1 might play an antiplasmodial role as reported before in *A. gambiae* (Blandin et al., 2004; Ramphul et al., 2015; Fraiture et al., 2009). In both *A. gambiae* and *A. stephensi* silencing of Duox gene causes reduction in *P. berghei* oocysts number. In *A. stephensi* the upregulated TEP1 expression in AsDuox silenced *P. berghei* infected midguts against control might explain the reduction of *Plasmodium* oocysts number. However, in *A. gambiae*, reduction in *Plasmodium* oocysts number is mediated by the induced expression of NOS gene (Kumar et al., 2010). Hence, there is a subtle difference in the immune mechanism of two different *Anopheles* species with the same outcome.

4.3.11 AsDuox gene is induced in exogenous bacteria supplemented blood fed midguts

The gut of insects is in contact with different types of microbes like symbionts, non-symbionts, food-borne microbes and life-threatening pathogens. The gut should be able to differentiate both endogenous and exogenous bacteria. In *Drosophila*, the finely tuned IMD pathway regulates the gut immunity against bacteria. The induction of IMD pathway against endogenous bacteria is inhibited by its negative regulator Caudal while in presence of exogenous bacteria signaling strength driven by PAMPs to activate pathway increases and overcome the suppressive effect of Caudal and activates the transcription of several AMPs (Lemaitre and Hoffmann, 2007). In *Drosophila,* several reports have shown that in response to bacterial infection there is induction of AMPs. The infection with *Erwinia carotovora* induced the expression of AMPs like attacin, diptericin, drosomycin, metchnikowin, and defensin (Tzou et al., 2000). The infection with *M. luteus/E. coli, Pseudomonas entomophila* also induced the expression of diptericin, drosomycin in *Drosophila* (Vodovar et al., 2005). Hence, infection triggers the activation of a number of immune molecules. Additionally, in response to natural infection of *Erwinia carotovora carotovora* 15 (Ecc15), induction of Duox is also reported in *Drosophila* (Ha et al., 2005a). In other insects like *A. aegypti* and *B. dorsalis*

involvement of Duox is also reported in immunity (Yao et al., 2016; Xiao et al., 2017).

In AsDuox silenced sugar fed midguts and blood fed midguts increased bacterial growth is observed (**Figure 4.12 and 4.14**). So, in *A. stephensi*, AsDuox gene is one of the molecules that maintain the bacterial homeostasis in the midguts. Moreover, we observed induced expression of anti-plasmodial gene in AsDuox silenced blood fed midguts against increased bacterial load. So, we wanted to study the gut immune responses against exogenous bacteria as it has been shown in *A. gambiae* that co-feedings of bacteria and *Plasmodium* cause a decrease in *Plasmodium* oocysts number (Dong et al., 2009; Cirimotich et al., 2010; Bahia et al., 2014). Thus, exploring the role of AsDuox in midgut immunity against exogenous bacteria and possible manipulation of the gut environment with the help of exogenous bacteria may open new ways to limit *Plasmodium* development inside the mosquito.

For this, *A. stephensi* females were allowed to feed on blood alone or supplemented with a mixture of Gram$^+$ (*M. luteus*) and Gram$^-$ (*E. coli*) bacteria. We investigated the expression of AMPs to check the activation of immunity against exogenous bacteria as their synthesis is considered a hallmark of insect humoral responses to microbial infections. Thus, we analyzed the expression of anti-microbial peptide gambicin and defensin as both can be regulated by Toll and IMD pathways (Tanji et al., 2007). The expression of gambicin is induced in response to exogenous bacteria in the midguts (**Figure 3.13**). We also analyzed the expression of defensin in these exogenous bacterial fed midguts. The result presented in **Figure 4.20** showed induced expression of
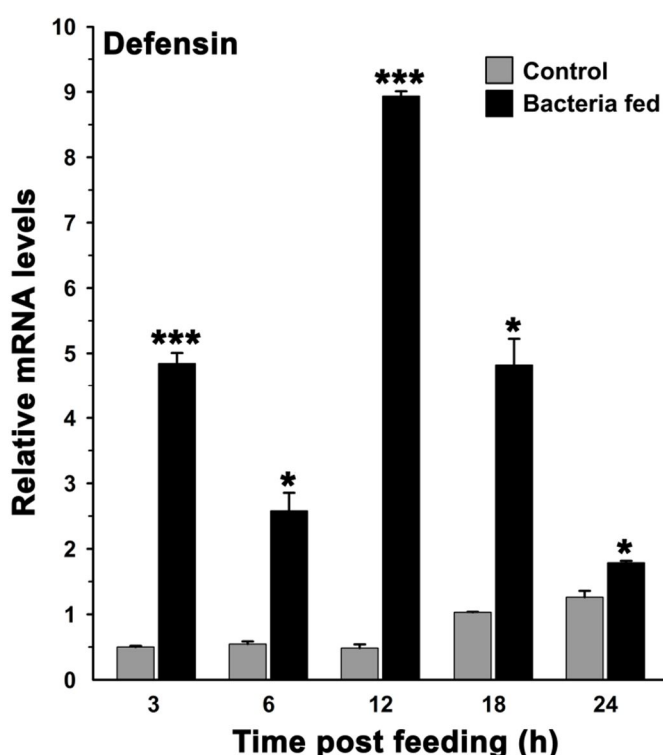
Figure 4.20: Expression kinetics of defensin gene in bacteria supplemented blood fed midguts. The kinetics of relative mRNA levels of defensin gene were analyzed in the blood alone (controls) or bacteria supplemented blood fed midguts. Relative levels are presented against the sugar fed midguts. Significant differences ($p < 0.001$ or 0.05) between control and bacteria fed midguts are shown by three or one asterisk (*), respectively.

defensin gene in response to bacterial feeding. The mRNA levels of defensin showed 9-fold induction at 3 h and 4-fold at 6 h in bacteria supplemented blood fed midguts against the blood fed controls (p=0.0007, p=0.0092), respectively. The mRNA levels of defensin showed the induction of 18-fold at 12 h in bacteria supplemented blood fed midguts against the blood fed midguts (p=0.0001). The mRNA levels of defensin showed 4-fold induction at 18 h and 1.5-fold at 24 h post bacteria supplemented blood feeding in midguts against controls (p=0.0206, p=0.0177, respectively). This data suggested the active involvement of mosquito immunity against exogenous bacteria.

Further, to analyzed the role of AsDuox in immunity against exogenous bacteria, the relative mRNA levels of AsDuox were analyzed through qPCR and are presented in **Figure 4.21**. The mRNA levels of AsDuox were similar at 3 h in bacteria supplemented blood fed midguts but showed 12-fold induction at 6 h in bacteria supplemented blood fed midguts in comparison to the controls (p=0.0007). Again the mRNA levels of AsDuox is reduced 3-fold at 12 h in bacterial fed midguts than control (p=0.0043). At 18 h and 24 h post feeding, AsDuox showed significant upregu-lation of 8-fold and 2-fold in bacteria supple-

Figure 4.21: Expression kinetics of AsDuox gene in bacteria supplemented blood fed midguts. The relative mRNA levels of AsDuox gene were analyzed in the blood alone (controls) or bacteria supplemented blood fed mosquito midguts. Relative levels are presented against the sugar fed midguts. Significant differences ($p < 0.001$ or 0.01) between control and bacteria fed midguts are shown by three or two asterisks (*), respectively.

mented blood fed midguts as compared to controls (p=0.0185, p=0.0031, respectively). Therefore, we conclude that AsDuox is induced in a biphasic manner, early at 6 h and later at 18 h and 24 h. The similar report is observed in *B. dorsalis* that BdDuox gene regulates the homeostasis of the gut bacterial community and induced only in the presence of opportunistic pathogens like *E. coli* DH5α and *Staphylococcus aureus* (Yao et al., 2016).

To understand the biphasic expression of AsDuox and its role in bacterial growth, we have analyzed expression of 16S rRNA in bacterial supplemented blood fed midguts. The result presented in **Figure 4.22** showed that mRNA levels of AsDuox are induced in bacteria fed midguts at 6 h and there is reduced bacterial load at 6 h in bacteria fed midguts when compared to 3 h bacteria fed midguts. At 12 h in bacteria fed midguts, expression of AsDuox gene is reduced and increased bacterial growth is observed in comparison to the 6 h. Further, at 18 h mRNA levels of AsDuox is again induced and at the same time point bacterial growth is reduced. At 24 h in bacteria fed midgut, expression of AsDuox is reduced and increased bacterial growth is observed. The expression of AsDuox and growth of bacteria showed a strong negative correlation (r=-0.7948, *p*= 0.029) in bacteria supplemented blood fed midguts at different time points.
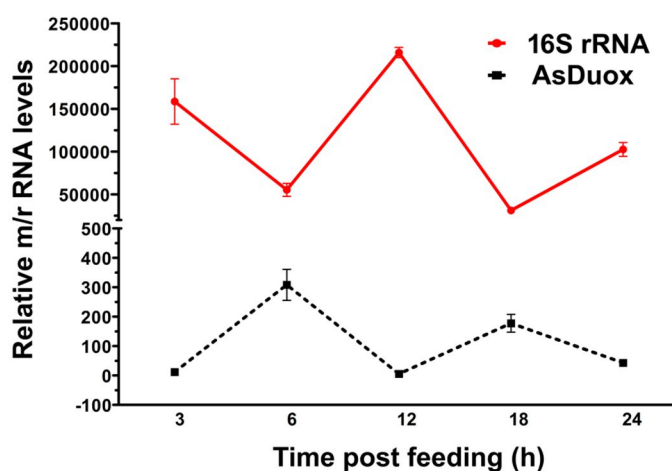
Figure 4.22: Relative levels of 16S rRNA and AsDuox mRNA in bacteria supplemented blood fed midguts. Mosquitoes were fed on bacteria supplemented blood and relative levels of AsDuox mRNA and 16S rRNA were analyzed in the midguts at different time points after feeding. Relative levels are presented against the sugar fed midguts.
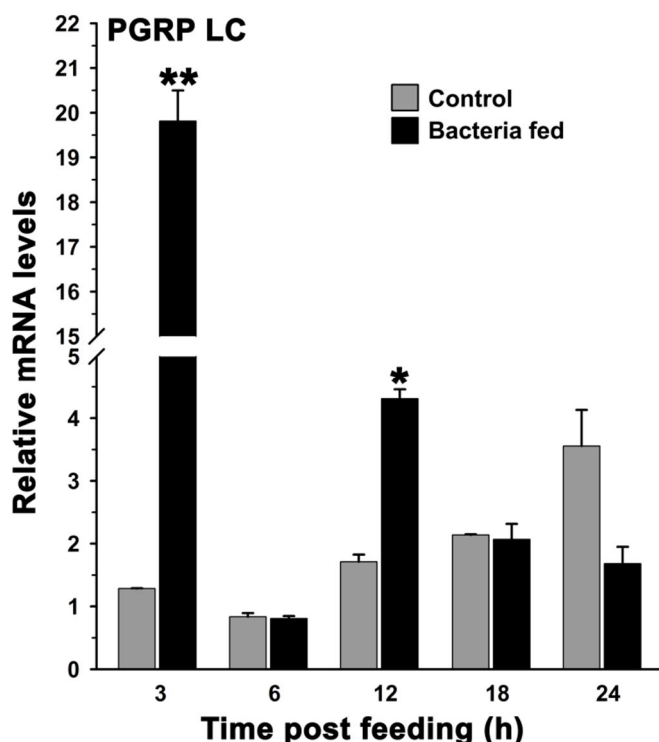
Figure 4.23: Expression kinetics of PGRP-LC gene in bacteria supplemented blood fed midguts. The relative mRNA levels of PGRP-LC were analyzed in the blood alone (controls) or bacteria supplemented blood fed midguts. Relative levels are presented against the sugar fed midguts. Significant differences *p*< 0.01 or 0.05 in the gene expressions between control and bacteria fed midguts are denoted by two or one asterisk (*), respectively.

This data suggested that AsDuox is one of the major molecules of immunity that maintains the bacterial homeostasis. In *Drosophila*, PGRP-LC activates IMD pathway that induces Duox (Leulier and Royet, 2009). So, to decipher the mechanism behind the activation of AsDuox in *A. stephensi*, we analyzed the expression of PGRP-LC. The mRNA levels of PGRP-LC in **Figure 4.23** showed 16-fold induction at 3 h in exogenous bacteria supplemented blood fed midguts against controls (p=0.0007). At 6 h, 18 h, and 24 h the expression of PGRP-LC is indifferent in bacteria supplemented blood fed midguts as compared to their controls. However, its expression levels were induced 3-fold at 12 h in bacteria supplemented blood fed midguts against controls (p=0.0026). Hence, PGRP-LC mRNA levels showed 16-fold and 3-fold induction at 3 h and 12 h post bacteria supplemented blood fed controls against the blood fed midguts, respectively (**Figure 4.23**).

In *Drosophila*, IMD pathway is strongly activated in response to the high microbial burden due to a large amount of microbe-associated molecular pattern (MAMP) especially PGN (peptidoglycan). PGRP-LC, the canonical membrane-associated receptor for peptidoglycan in flies, is an activator of IMD pathway which mediates the peptidoglycan-dependent activation of the MEKK1-MKK3-p38-ATF2 pathway (Leulier and Royet, 2009). In addition, Duox transcription is upregulated via p38-mediated activation of ATF2. A large quantity of DUOX produces the abundant ROS and is needed to fight against infectious microbes (Leulier and Royet, 2009).

Collectively, we can speculate from the expression data of 16S rRNA, PGRP-LC and AsDuox that upon high bacterial load induced expression of PGRP-LC is observed due to increased PAMP/PGN. This, in turn, activated the IMD pathway and caused induction of downstream gene, Duox. The AsDuox mediated generation of ROS kills the bacterial cells. Hence, we observed the strong negative correlation between bacterial growth and AsDuox expression in the exogenous bacteria fed midguts.

4.3.12 AsDuox silencing reduced the growth of exogenous bacteria in the midgut

To understand the role of AsDuox in maintaining the homeostasis in midguts upon exogenous bacterial feeding, we used RNAi mechanism to silence this gene. The control and AsDuox silenced mosquitoes were then fed on bacteria supplemented blood as described in Materials and Methods. The 16S rRNA levels in bacteria supplemented blood fed midguts in comparison to their respective controls, revealed that there is reduced bacteria growth at 6 h and 24 h (**Figure 4.22**) when compared to the other time

points. Also, the expression of AsDuox at different time points revealed that it is induced at these two time points (6 h and 24 h) in bacteria supplemented blood fed midguts against control (**Figure 4.21**). Hence, the two time points 6 h and 24 h were further analyzed in detail to explore the immune response against the exogenous bacteria in AsDuox silenced midguts similarly as described earlier (Kajla et al., 2016c).

The result presented in **Figure 4.24**, showed that we could achieve 80% (*p*= 0.0024) and 65% (*p*= 0.0024) silencing of AsDuox gene at 6 h and 24 h, respectively in bacteria supplemented blood fed midguts against control. We compared the rate of survival in the control and silenced group of mosquitoes post bacterial feeding and surprisingly, we found no significant difference in the rate of survival between two groups.
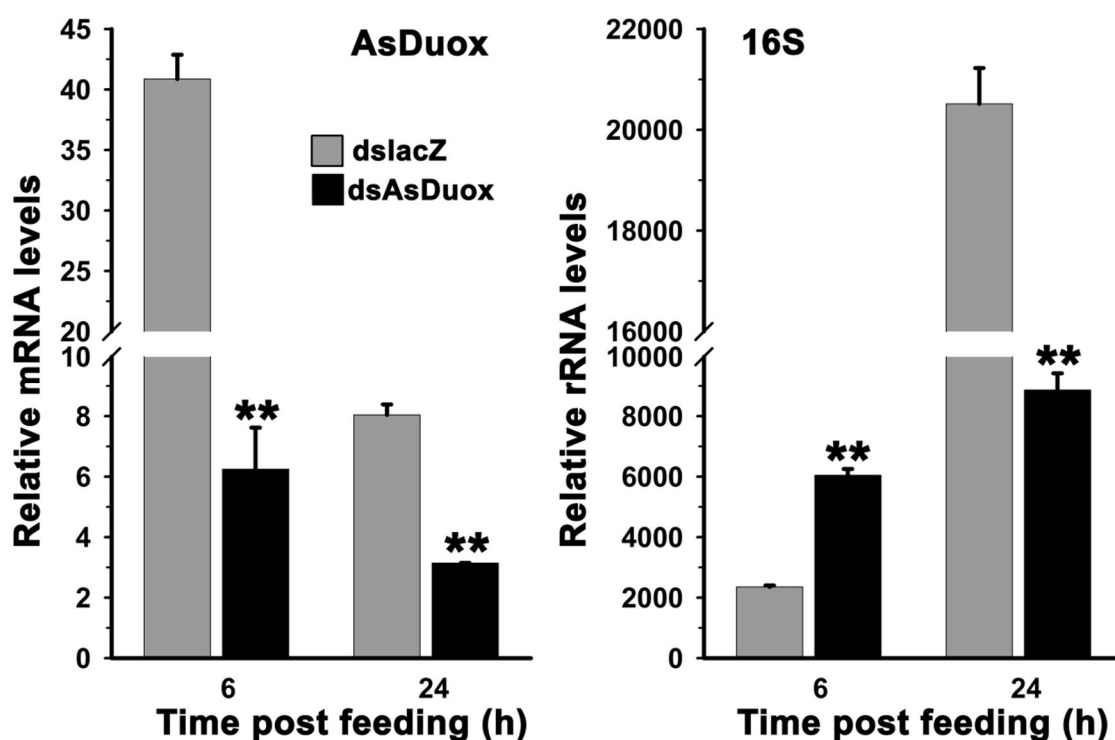
Figure 4.24: Relative levels of 16S rRNA in AsDuox silenced bacteria supplemented blood fed midguts. Mosquitoes injected with dsLacZ (controls) or dsAsDuox (silenced) RNA were fed on bacteria supplemented blood and relative rRNA levels of 16S and mRNA levels of AsDuox were analyzed in their midguts at different time points post feeding. Relative levels are presented against the sugar fed midguts. Significant differences (*p*< 0.01) between control and silenced bacteria fed midguts are shown by two asterisks (*).

Thus, we analyzed the expression of 16S rRNA in the two groups to monitor the bacterial growth. The result showed in **Figure 4.24** revealed that expression of 16S rRNA at 6 h was 2.5-fold higher in silenced midguts than controls (*p*= 0.0017). However, in the AsDuox silenced bacterial supplemented blood fed midguts, the expression of 16S rRNA is reduced by 56% at 24 h (*p*= 0.0031) against controls (**Figure 4.24**).

To decipher the mechanism behind the reduction of bacterial population, we analyzed the expression of various immune genes in these samples. As stated previously in *Drosophila*, the immune responses of gut epithelial cells are solely under the control of IMD pathway (Lemaitre and Hoffmann, 2007). So, we analyzed the expression of defensin in the controls and AsDuox silenced bacteria supplemented blood fed midguts. The expression of defensin induced 2-fold at 24 h (*p*= 0.0040) in AsDuox silenced midguts as compared to the controls (**Figure 4.25**). The mRNA levels of GNBP and Toll showed significant upregulation of 3-fold and 2-fold in the AsDuox
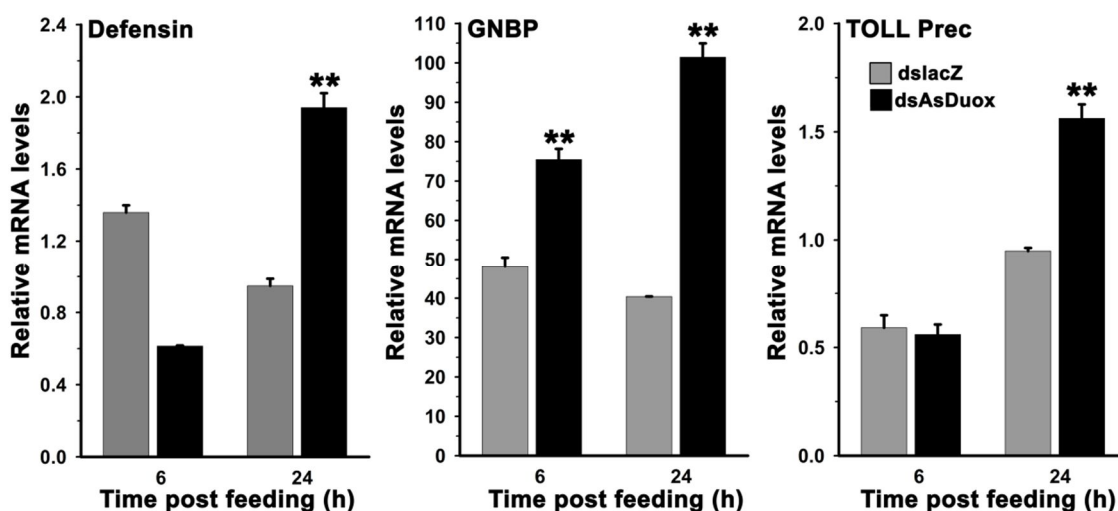
Figure 4.25: Expression of classical immune genes in AsDuox silenced and bacteria fed midguts. Mosquitoes injected with dsLacZ (controls) or dsAsDuox (silenced) RNA were fed on bacteria supplemented blood and relative mRNA levels of defensin, GNBP or Toll gene were analyzed in these midguts after 6 h or 24 h post feeding. Relative levels of mRNA are presented against the sugar fed midguts. Significant differences (*p*< 0.01) between control and silenced are shown by two asterisks (*).
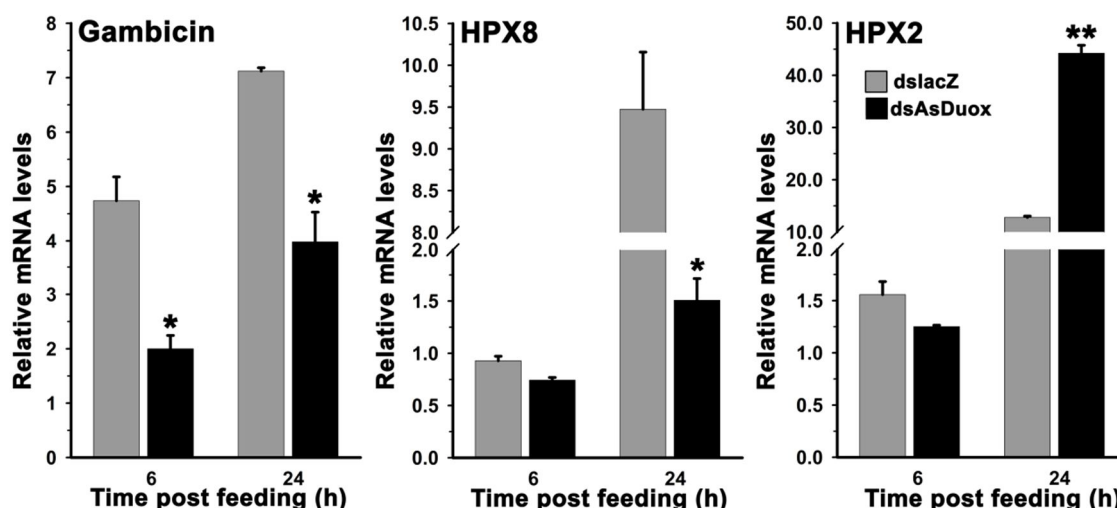
Figure 4.26: Expression of immune genes in AsDuox silenced and bacteria fed midguts. Mosquitoes injected with dsLacZ (controls) or dsAsDuox (silenced) RNA were fed on bacteria supplemented blood and relative mRNA levels of gambicin, HPX8, and HPX2 were analyzed in these midguts after 6 h or 24 h post feeding. Relative levels of mRNA are presented against the sugar fed midguts. Significant differences (*p*< 0.01 or 0.05) between control and silenced are shown by two or one asterisk (*), respectively.

silenced exogenous bacteria fed midguts at 24 h against controls (p= 0.0017, p= 0.0049), respectively. The induced expression of immune genes might be responsible for reduced bacterial growth in AsDuox silenced bacterial supplemented blood fed midguts against controls.

We also analyzed the expression of other anti-bacterial genes like gambicin and HPX8 (Vizioli et al., 2001; Kumar et al., 2010) in the silenced midguts to explore their role in the reduction of bacteria. We found that both the immune genes showed reduced expression in the AsDuox silenced bacteria supplemented blood fed midguts as compared to the controls as shown in **Figure 4.26**. Additionally, we also analyzed the expression of AsHPX2 in AsDuox silenced bacterial fed midguts. Results presented in the **Figure 4.26** showed that AsHPX2 is induced 4-fold at 24 h (p= 0.0012) in AsDuox silenced bacteria supplemented blood fed midguts against controls. This induced expression of AsHPX2 suggested that this might be one of the molecules that control the bacterial growth.

We also explored the involvement of JAK/STAT pathway in the reduction of bacterial load (Gupta et al., 2009; Kajla et al., 2016c). The expression of NOS, an effector molecule of JAK/STAT pathway and SOCS (a negative regulator of JAK/STAT pathway) were ana-lyzed in AsDuox silenced bacteria supplemented blood fed midguts. Data shown in **Figure 4.27** revealed reduced expression of NOS at 6 h and indifferent expression at 24 h in the silenced midguts as compare to the controls. The expression of SOCS remained similar at 6 h but induced at 24 h
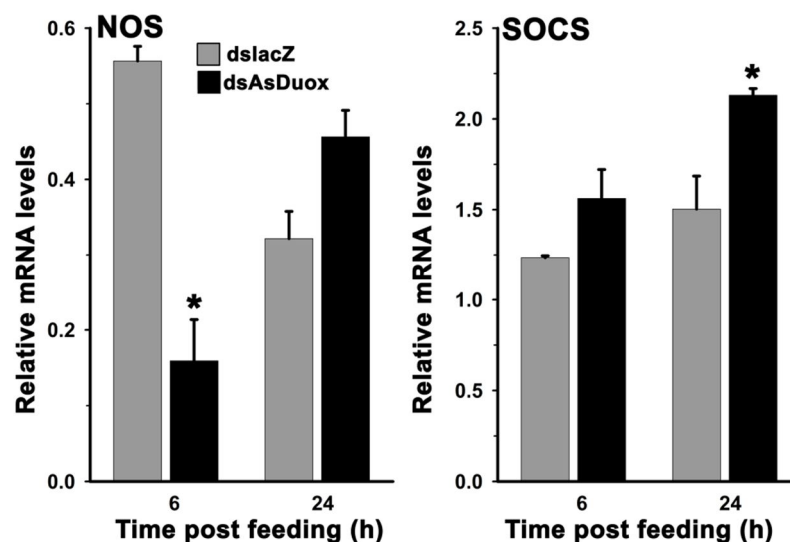
Figure 4.27: Relative mRNA levels of NOS and SOCS genes in AsDuox silenced bacteria fed midguts. Mosquitoes injected with dsLacZ (controls) or dsAsDuox (silenced) RNA were fed on bacteria supplemented blood and relative levels of NOS or SOCS mRNA were analyzed in their midguts at different time points after feeding. Relative levels of mRNA are presented against the sugar fed midguts. Significance differences (p< 0.05) are denoted by an asterisk (*).

post feeding in the silenced midguts against controls. From the above data, it can be concluded that IMD and TOLL pathways are responsible for the reduced bacterial load

and not JAK/STAT pathway in the AsDuox silenced bacterial fed midguts. This data is in partial agreement with our previous finding where silencing of AsHPX15 suppressed the growth of exogenous bacteria and is mediated by JAK/STAT pathway (Kajla et al., 2016c). In conclusion, HPX15 and Duox are involved in midgut barrier formation and disruption of barrier by silencing of these genes causes reduction in the bacterial load but the immune mechanism of bacterial killing is different. In AsHPX15 silenced exogenous bacteria supplemented blood fed midgut, the reduction in the bacterial load is mediated by the upregulation of the NOS via JAK/STAT pathway (Kajla et al., 2016c) while in AsDuox silenced exogenous bacteria supplemented blood fed midgut, the bacterial reduction is caused by the upregulation of Toll and IMD pathway.

This reduction in the bacterial growth in the AsDuox silenced midgut could be the possible explanation for the survival of the mosquito in spite of heavy bacterial load. This suggested that mosquito midgut is much more evolved and highly efficient and equipped to handle the excessive bacterial load. This showed that in mosquito if one mechanism fails to control the bacterial growth another pathway(s) activates and maintains the microbial homeostasis. While in case of *Drosophila* silencing of Duox causes great mortality (Ha et al., 2005a).

4.4 Discussion

The most important mechanism of innate immunity of insect midgut is the production of free radicals by dual oxidases (Duox), a class of enzymes from the NOX family of proteins (Ha et al, 2005a; Ha et al, 2009). This transmembrane enzyme is capable of producing H_2O_2, a ROS molecule and has N-terminal peroxidase domain. This event is also found in mosquitoes and affects their vector competence to transmit human diseases such as malaria (Kumar et al., 2003; Kumar et al., 2010; Molina-Cruz et al., 2008).

The ORF of AsDuox gene is 4428 bp long and encodes for a protein of 1475 amino acids and is highly conserved among anophelines (sequence identity 97-99%). Analysis of the AsDuox protein has revealed that it is a membrane-bound protein with a non-cytoplasmic heme peroxidase domain and a calcium-binding domain at the N-terminus and cytoplasmic conserved FAD- and NAD-binding domains at the C-terminus (**Figure 4.5**). This showed that structure of AsDuox is similar to other Duoxes from invertebrates (Hu et al., 2013; Sumimoto, 2008) and vertebrates (Donkó et al., 2005).

The AsDuox protein contained amino acid sequences SGQWVR, FTLTSAPHEN and GIGVTPYAS in the NAD-binding domain which are conserved in both invertebrates

and vertebrates. The NAD-binding sites have an important role in transporting electrons from NADPH to FAD in Nox2 of Human (Torres and Dangl, 2005). Therefore, the presence of these conserved sites in the NAD-binding domain of AsDuox suggests a crucial role of this protein in the electron delivery system. The AsDuox active sites within the peroxidase domain shared 9% identity and substrate binding domain shared 50% identity with human Duoxes. The calcium-binding domain shared 57% identity and NADPH oxidase domain shared 76% identity for with human Duoxes.

The *A. stephensi* Duox gene is expressed in all the developmental stages. The expression of Duox is similar from egg to 3rd instar larvae to ensure the proper formation of extracellular cuticle layer as described previously in *C. elegance.* The CeDuox is required for post-embryonic development. The presence of ecdysone hormone binding motif in the promoter region of AsDuox gene might explain the highest expression of this gene in pupae as ecdysone regulates the larval molting and metamorphosis in other insects like *Drosophila* and *Aedes* (Akagi and Ueda, 2011; Telang et al., 2007). Also, in *Drosophila* recent study showed that Duox is required to stabilize the wing on the last day of pupal development (Hurd et al., 2015). The bacterial challenge to 4th instar larvae induced the expression of AsDuox gene (**Figure 4.10**). This suggested the role of AsDuox in immunity against bacterial infection in larvae similarly as previously characterized in *C. elegans*, Zebrafish larvae and in *Bombyx mori* (Van der Hoeven et al., 2015; Flores et al., 2010; Hu et al., 2013).

In innate immunity, the role of ROS, such as O^-_2 and H_2O_2, is to eliminate microbes. In *Aedes aegypti,* ROS is described to produce in sugar fed midguts and silencing of Duox increases the bacterial load (Oliveira et al., 2011). There is increased bacterial growth in the AsDuox silenced midguts as compared to control (**Figure 4.12**). It is well reported in *Drosophila* that Duox is expressed in the gut and involved in immune responses. This gene has been shown to exhibits microbicidal activity that prevents over proliferation of dietary bacteria and yeast during natural infection (Ha et al., 2005a). Thus, our results suggested that AsDuox plays an important role in maintaining the bacterial homeostasis in sugar fed midguts similar to that of the Duoxes of other insects (Yao et al., 2016; Ha et al., 2005a).

To elucidate the mechanism underlying in maintaining the bacterial homeostasis after blood feeding in the midgut, we compared the expression pattern of AsDuox and 16S rRNA. The transcriptional analysis revealed that although mRNA of AsDuox is highly expressed in carcass but induced only in the midgut after blood feeding (**Figure 4.11** and **4.13**). Our analysis showed that bacterial growth and AsDuox expression at

various time points post blood meal in midgut has a weak negative correlation. Hence, *A. stephensi* dual oxidase (AsDuox) has a dual role it is not only protecting the gut bacteria through barrier formation but is also responsible for balancing their population after the blood meal. In AsDuox silenced blood fed midguts, we observed increased bacterial population against control midguts and no induction of immunity against increased bacterial load (**Figure 4.14**). Thus, like *Drosophila*, mosquito gut can also tolerate the endogenous bacteria and hence there is no upregulation of AMPs in AsDuox silenced blood fed midguts. This data suggested that AsDuox is involved in maintaining the bacterial homeostasis. The similar mechanism is reported in both *A. aegypti* and *D. melanogaster* where dual oxidase (Duox)-regulating pathway maintains the bacterial homeostasis in the gut (Xiao et al., 2017). In *A. aegypti,* in response to the blood feeding expression of AaDuox gene is induced and proliferation of bacteria takes place in midguts. Silencing of AaDuox gene increased the bacterial load. This study has revealed that Duox is the major molecule in maintaining the gut homeostasis and manage healthy gut–microorganism interactions in insects (Xiao et al., 2017).

The upregulation of AsDuox gene in bacteria supplemented blood fed midguts suggested that *A. stephensi* also uses AsDuox dependent immune response against exogenous bacteria in the gut in a way similar to *Drosophila* (Ha et al., 2005a; Bae et al., 2010). In this study, we also explored the molecular pathway behind that might regulate the expression of AsDuox during bacterial infection. We have shown that AsDuox expression is under the control of the IMD pathway as previously described for *Drosophila* Duox (Lemaitre and Hoffmann, 2007). During bacterial feeding, induced expression of AMPs is observed. So, in *A. stephensi*, during bacterial infection, there is activation of IMD pathway, which upregulates the expression of AMPs and Duox. Thus, bacterial feeding manipulates the immunity of mosquito midgut and this can be explored for arresting *Plasmodium* development since high levels of PAMPs activated the immune pathways this, in turn, culminating in an increased abundance of effector molecules that may defend against infection with *Plasmodium.*

The reduced bacterial growth is observed in the AsDuox silenced bacteria supplemented blood fed midguts than controls and is mediated by induced mosquito midgut immunity. In the previous study, it has been shown that Duox is involved in the formation of the midgut barrier that protects bacteria from the mosquito immunity (Kumar et al., 2010). So, silencing of Duox disrupts the formation of mucin barrier and this activates the mosquito immunity against exogenous bacteria. The similar report was published earlier where disruption of the gut barrier by silencing of HPX15 leads to

reduction of bacteria in silenced bacteria fed midguts via activation of NOS molecule (Kumar et al., 2010, Kajla et al., 2016c).

We also investigated the role of *A. stephensi* Duox gene in the regulation of *Plasmodium* development. We found that mRNA levels of AsDuox gene were induced in the midguts post *Plasmodium*-infected blood feeding. Interestingly, the expression of AsDuox is significantly increased during ookinete invasion of the midgut epithelium (24 h post infected blood feeding) (**Figure 4.17**). These findings are in agreement with the previous reports where AgDuox gene is induced post 24 h of *Plasmodium* infection in midguts (Kumar et al., 2004).

Furthermore, our data revealed that AsDuox plays important role in the regulation of *Plasmodium* development in midguts. The silencing of AsDuox gene greatly reduces the number of developing oocysts in the silenced midgut as compared to the control (**Figure 4.18**). The disruption in the formation of mucin barrier in the Duox silenced midguts, in turn, induced antiplasmodial immunity and we found that TEP1, is highly induced in these samples (**Figure 4.19**) and might be one of the important negative regulators of *Plasmodium* development similarly as reported in *A. gambiae* (Kumar et al., 2010). In *A. gambiae* the reduction in *Plasmodium* oocysts number is mediated by the activation of NOS molecule.

4.5 Conclusion

A. stephensi Duox gene has an important role in the innate immunity. This gene is responsible for ROS generation and maintenance of bacteria homeostasis in the midgut. The AsDuox gene has a weak negative correlation with the endogenous bacterial growth while a strong negative correlation with the exogenous bacterial growth. AsDuox gene also has an important role in *Plasmodium* development. Hence, manipulation of this gene will open up new frontiers in blocking the *Plasmodium* development as bacteria are the antagonist of *Plasmodium* development.

Chapter 5

CTCF, an insulator protein, regulates the divergent expression patterns of lineage-specific duplicated *Anopheles* heme peroxidases HPX15 and HPX14

5.1 Abstract

In this study, we have investigated the gene duplication event in the heme peroxidase multi-gene family. It is found that heme peroxidase HPX15 in *A. gambiae* and *A. stephensi* has its tandemly duplicated paralog gene termed HPX14. The two genes although shared similar exonic structures but differed in their intronic structures. These genes are under purifying selection and phylogenetic analysis showed that the duplication event occurred prior to the speciation. The two genes are flanked by the presence of boundary element and might act as an independent domain of gene expression. The genes of the duplicated cluster are highly divergent in their spatial-temporal expression. The expression of AsHPX15 gene is higher in all the developmental stages and was thousand fold in blood fed and bacteria supplemented blood fed midguts of *A. stephensi* in comparison to the expression of AsHPX14. The presence of binding motifs of an insulator protein, CTCF in the regulatory region of AsHPX14 suggested that it might regulate the transcription of AsHPX14 and hence, expression of AsHPX14 remained suppressed. AsHPX15 silencing and blood feeding did not induce the expression of AsHPX14 in the midgut but its expression is induced in the AsHPX15 silenced bacteria supplemented blood fed midguts in comparison to the respective controls. This data suggested that there is no functional redundancy between the two duplicated genes and AsHPX14 might evolve to have a role in immune responses against bacteria but not in physiology. Silencing of HPX15 causes disruption of mucin barrier and we assume that the high bacterial load might cause some signal to displace CTCF protein and activates the expression of HPX14 gene. This study provided insight regarding the potential functional role of CTCF in mosquito and may be used to improve mosquito transgenesis.

5.2 Introduction

Mosquitoes exhibit several aspects of innate immune responses to withstand *Plasmodium* infection. *Plasmodium* interaction with mosquito draws major attention to understand both the conservative and the dynamic evolutionary nature of immune responses interfering with malaria pathogen development *(*Waterhouse et al., 2007). Thus, understanding mosquito immunity in the regulation of *Plasmodium* development is a hot topic because these findings will be helpful in designing strategies in the future that will control *Plasmodium* transmission.

Previous studies from our laboratory found that heme peroxidase HPX15 is induced upon blood feeding in the midguts and catalyzed the crosslinking of tyrosine

residues of mucin layer to form a gut barrier (Kumar et al., 2010). This gut barrier creates a low immunity zone that supports the growth of bacteria as well as *Plasmodium*. This is a lineage-specific gene that is present in all 19 worldwide distributed anopheline species and thus, we proposed that mosquito heme peroxidase HPX15 may be targeted to block the *Plasmodium* development (Kumar et al., 2010 and Kajla et al., 2016b, 2016c, 2017).

In previous reports, it has been suggested that heme peroxidase family belongs to the multi-gene family and has several gene duplication events (Loughran 2008; Neafsey et al., 2015). Gene duplication event generates a duplicate copy of a gene and is very eminent in eukaryotes. Thus, we analyzed the duplication event of HPX15 in the *Anopheles* genome to understand the relationship with its duplicate copy HPX14 (Kajla et al., 2015a). Studying the gene duplication event and expression pattern of duplicated copy is necessary to claim HPX15 as a transmission-blocking candidate. Therefore, in the present study, we have identified and characterized the duplicated paralog of HPX15 gene in *Anopheles* and studied the gene structure and expression pattern of the duplicated genes. We also studied the sequence divergence, characterized spatial-temporal expression and involvement of putative transcription factors that might regulate their expression to decipher the probable evolutionary divergence between these two duplicated genes.

5.3 Results

5.3.1 *Anopheles gambiae* HPX15 (AgHPX15) gene is evolved with its tandemly duplicated paralog HPX14 (AgHPX14)

The gene structure of the HPX15 in *A. gambiae* genome was retrieved and further, we performed the paralog search to identify whether a copy of this gene was generated through duplication event. Our criterion of gene satisfying ≥50% alignment identity and ≥70% alignment coverage (query coverage) at the amino acid level is taken into further consideration as discussed before (Audemard et al., 2012). We have identified that AgHPX14 is the paralog of AgHPX15 generated through gene duplication (**Figure 5.1**).

The paralogs are said to be tandemly duplicated if they are less than 150 kb apart from each other. In addition, each tandem unit (gene) has a minimum length of 500 bp and separated by less than 40 kb (Audemard et al., 2012; Li et al., 2001). Our results revealed that tandemly duplicated gene cluster of AgHPX15 and AgHPX14 (43363 bp, Chromosome 3L**:** 10755137-10798500) is situated adjacent to each other.
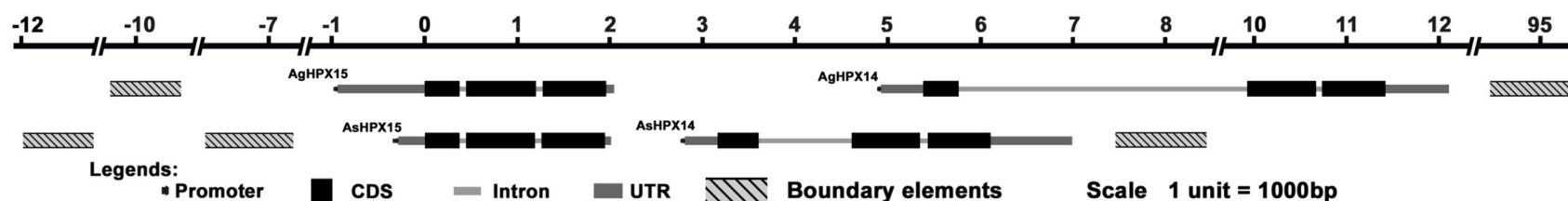
Figure 5.1: Location and distribution of the boundary elements across the duplicated genes of mosquitoes, *A. gambiae* (Ag) and *A. stephensi* (As). The gene cluster of *A. gambiae* and *A. stephensi* are shown with tested boundary elements as lining shaded bars present on both the side of the duplicated gene cluster. Exonic structures in genes are shown as a black rectangle box and the gray lines within the genes indicate intronic regions. The UTRs are shown in gray rectangle. The scale on the upper line corresponds to 1000 bp for 1 unit.

Table 5.1: Comparison of the genomic features between *A. gambiae* and *A. stephensi* duplicated paralogs: The various genomic features of HPX15 and HPX14 from *A. gambiae* and *A. stephensi* are given in the table below. The genomic sequence of *A. gambiae* and *A. stephensi* was retrieved from NCBI and gene structure was then predicted using Augustus software.

Features	AgHPX15	AgHPX14	AsHPX15	AsHPX14
Gene length	1961	6042	1941	2913
Intergenic region (bp) between paralogs	-	2885	-	754
Number of Exons in gene	3	3	3	3
1st exon (bp)	372	387	369	397
2nd exon (bp)	750	741	743	736
3rd exon (bp)	684	714	682	706
1st intron (bp)	82	4159	76	993
2nd intron (bp)	72	66	71	81
5' UTR (bp)	940	449	344	410
3' UTR (bp)	82	688	72	879
Gene location Chromosome/contig	Chromosome 3L	Chromosome: 3L	SuperContig KB665221	SuperContig KB665221
Gene span or start- end positions	10,786,057-10,788,017	10,791,477-10,797,515	711,425-718,063	711,425-718,063
Protein length (amino acids)	602	614	597	612
Signal Peptidase cleavage signal position	GLS_{21}-Q_{22}T (D=0.764)	VLA_{27}-V_{28}C (D=0.768)	VLS_{21}-Q_{22}T (D=0.838)	VLA_{30}-V_{31}C (D=0.697)

AgHPX14 gene is situated 2885 bp downstream of AgHPX15 gene on same DNA strand in head to tail orientation (**Figure 5.1** and **Table 5.1**). The protein length of AgHPX15 and AgHPX14 is 602 and 614 amino acids, respectively. The BLASTP results revealed that the alignable region between these two proteins is of 584 amino acids with 326 identical amino acids. Further, we confirmed the gene duplication between AgHPX15 and AgHPX14 gene at protein level through computed identity [I'= I x min (n$_1$/L$_1$, n$_2$/L$_2$)] as discussed before (Li et al., 2001). In general, I' > 30% satisfies the tandemly duplicated paralogs. The calculated I' value for AgHP15 and AgHP14 is 55.87% with the query coverage of 95% and 56% identity. In conclusion, AgHPX15 and AgHPX14 are true tandemly duplicated paralogs.

5.3.2 *A. stephensi* AsHPX15 also has tandemly duplicated paralog

The nucleotide BLAST of AgHPX15 gene against partially annotated *A. stephensi* genome revealed that its ortholog is present in contig 5285 (SuperContig KB665221: 711,425-718,063) and situated between nucleotide ranges from 14000 to 16000 and termed it as AsHPX15 (Kajla et al., 2016b). Both genes, AgHPX15 and AsHPX15 have 75% sequence identity with the query coverage of 96%. Further, BLAST of AsHPX15 against *A. stephensi* genome revealed that its paralog is also located in the same contig between nucleotide 17000 and 21500. Interestingly, this paralog is located 754 nucleotides downstream to AsHPX15, on the same DNA strand in head to tail orientation, similarly as AgHPX15 and AgHPX14 (**Figure 5.1** and **Table 5.1**). It is noteworthy to mention that the genomic location of AsHPX15 paralog is equivalent to AgHPX14 thus, we named it as AsHPX14. AgHPX14 gene shared 81% sequence identity with AsHPX14 gene and hence, the two are also orthologs.

5.3.3 Structural divergence in duplicated genes of *Anopheles*

The genomic sequences retrieved in section 5.3.1 and section 5.3.2 that has HPX15 and HPX14 genes were analyzed using Augustus software to elucidate the gene structure for further comparison. The comparative analysis of *A. gambiae* (AgHPX15 and AgHPX14) and *A. stephensi* (AsHPX15 and AsHPX14) gene clusters revealed that both share several common features (**Figure 5.1** and **Table 5.1**). *A. gambiae* AgHPX15 gene is of 1961 bp with 3 exons and 2 introns. Interestingly, its paralog AgHPX14 gene is of 6042 bp with the same number of exons and introns. However, the first intron is comparatively larger in case

of AgHPX14 than AgHPX15 (4159 bp versus 82 bp) (**Table 5.1**). This may indicate the insertion of an extra segment of DNA in the first intron of AgHPX14. This situation is in agreement with other tandemly duplicated genes such as F-box, AP2, and cyclin where one of the paralogs has longer intron (Xu et al., 2012).

We found that all the four genes AgHPX15, AgHPX14, AsHPX15 and AsHPX14 showed structural conservation. We also found that gene structure of AsHPX15 and AgHPX15 showed more conserved structure, whereas AgHPX14 and AsHPX14 diverged structurally in the intronic region. The first intron of HPX14 showed significant divergence in length when compared to the first intron of HPX15 in both *A. gambiae* and *A. stephensi*. In *A. gambiae* the first intron of AgHPX14 is 4135 bp and almost 50 times in length (bp) in comparison to the first intron of AgHPX15 and in *A. stephensi* the first intron of AsHPX14 is 993 bp and 10 times of the first intron of AsHPX15 (**Figure 5.1**). Collectively, the gene organization of these two duplicated genes showed intronic divergence in case of both paralogs as well as orthologs while showed conservation in exonic structures.

The presence of structural divergence in the intronic region of paralogs suggested that different selective pressure might act on introns during evolution as also reported in the growth hormone GH1 and GH2 in the eight species of *Salvelinus* genus (Pankova et al., 2013). The previous study showed that the structural divergence found in the exons-introns of duplicated pairs showed a positive correlation with the divergence in expression. This data is further supported by the findings reported in *Arabidopsis* duplicated genes such as F-box, AP2, and cyclin (Xu et al., 2012).

5.3.4 Duplicated heme peroxidases are secreted globular proteins

It has been reported earlier that AgHPX15 and AsHPX15 proteins are secreted into the gut lumen (Kumar et al., 2010, Kajla et al., 2016b). Analysis of presence of signal peptide sequence by SignalP (Petersen et al., 2011) in AgHPX14 and AsHPX14 revealed that both these proteins also contained the cleavage site for the signal peptidase (**Figure 5.2**). This data suggested that HPX15 and HPX14 both are secreted proteins. The alpha helixes present in the proteins of AsHPX15 and AsHPX14 are 44% and 41%, respectively (**Figure 5.2**). This indicated that these are globular proteins, as this category of proteins contains more alpha helixes than beta sheets (Pace and Scholtz, 1998). The result presented in **Figure 5.2** revealed that both AsHPX15 and AsHPX14 proteins share conserved heme, substrate, calcium and homodimer interface binding sites.
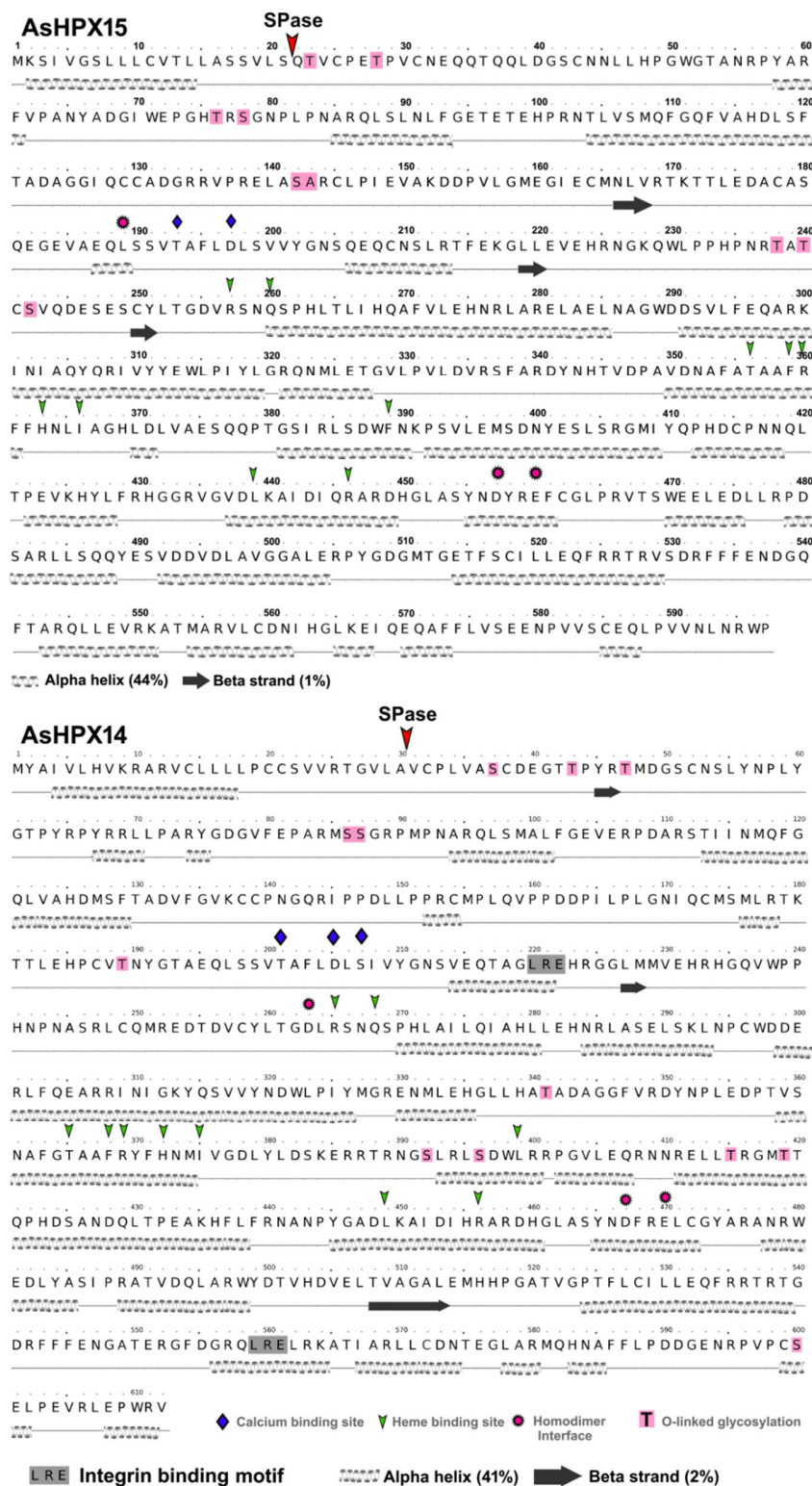
Figure 5.2: Deduced secondary structure of duplicated paralog. Both the proteins contain alpha helixes (spirals) and beta sheets (arrows) as predicted by Phyre[2] software. The red arrowhead indicates the cleavage site for signal peptidase (SPase) and various conserved binding sites in the protein are represented by characteristic symbols. (AsHPX15 secondary structure modified from Kajla et al., 2016b)

The deduced AsHPX15 and AsHPX14 polypeptide were also subjected to a three-dimensional structural prediction by Phyre[2] software. Interestingly, 93% and 91% amino acid residues (554 and 558 amino acids) of the full-length AsHPX15 or AsHPX14 protein (total 597 and 612 amino acids) were modeled with buffalo lactoperoxidase available in protein data bank (PDB ID:2GJM) with 100% level of confidence, respectively (**Figure 5.3**). The active site of AsHPX15 consists of ARG_{102}, ILE_{317}, GLN_{571} and PHE_{574} amino acids and in AsHPX14 it is consists of ARG_{111}, ILE_{325}, ASN_{584} and PHE_{587} amino acids as predicted by 3DLigandSite software using NAG (N-acetyl-D-glucosamine) as heterogen.
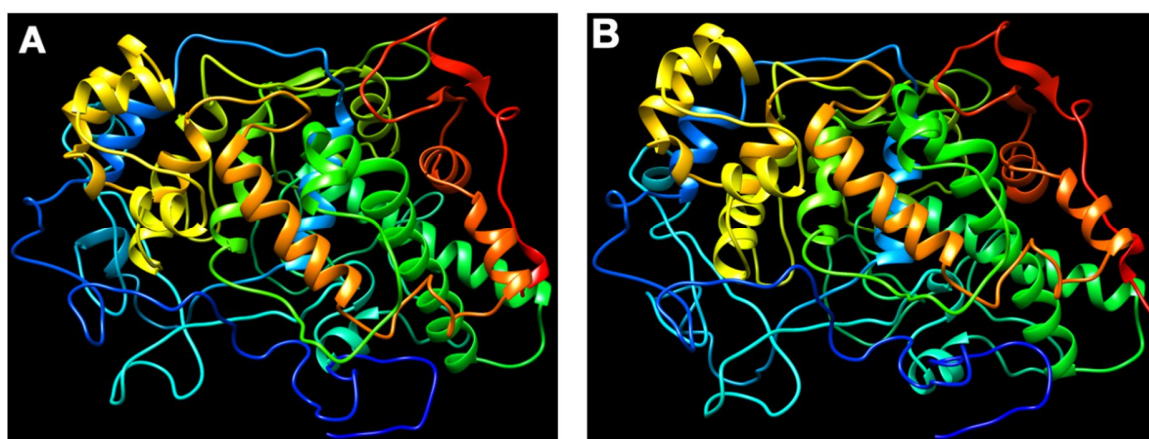
Figure 5.3: 3D structure of AsHPX15 and AsHPX14 proteins. 3D structure of **A)** AsHPX15 and **B)** AsHPX14 proteins were modeled by Phyre[2] software that used lactoperoxidase (PDB ID: 2GJM) as a template.

5.3.5 Tandemly duplicated heme peroxidase paralogs of *A. gambiae* and *A. stephensi* are functional

Fixation and conservation of newly duplicated genes in genome require conquering over ample hurdles. Once fixed, the evolutionary fate of duplicated gene may end up in pseudogenization. This process leads to the loss in function due to the accumulation of degenerative or deleterious mutations over time in the regulatory region or the gene itself (Roth et al., 2007). In order to understand the functional nature of the tandemly duplicated AgHPX14 and AsHPX14 genes, we analyze its regulatory elements such as a promoter, TATA protein binding site, transcription start site (TSS) as they play a central role in the functional nature of a gene.

We analyzed the 5' upstream region of AgHPX14 and AsHPX14 genes with the help of various computational programs like Augustus, GenScan, and NNPP to predict the putative promoter as described in Material and Methods. We found that both the genes

have promoter region with TATA-binding motifs and transcription start site (TSS). The promoter of AgHPX14 is 449 bp upstream of start codon while in AsHPX14 it is 410 bp upstream of start codon as analyzed by NNPP and Augustus software (**Figure 5.1**).

5.3.6 Phylogeny revealed gene duplication of heme peroxidase HPX15 and HPX14 occurred before speciation

We used the phylogeny based algorithm to infer duplication and speciation events from the phylogenetic tree as described earlier (Huerta-Cepas et al., 2007). The Phylogenetic analysis of protein sequences was performed using the neighbor-joining (NJ) methods implemented in MEGA 5.2, with 1000 bootstrap replicates to assess support as described in Material and Methods. The HPX15 from both *A. gambiae* and *A. stephensi* formed one cluster while HPX14 formed the other cluster. This data showed that there is a gene duplication event followed by speciation, which might be accelerated the rate of evolution (**Figure 5.4**) (Pegueroles et al., 2013).
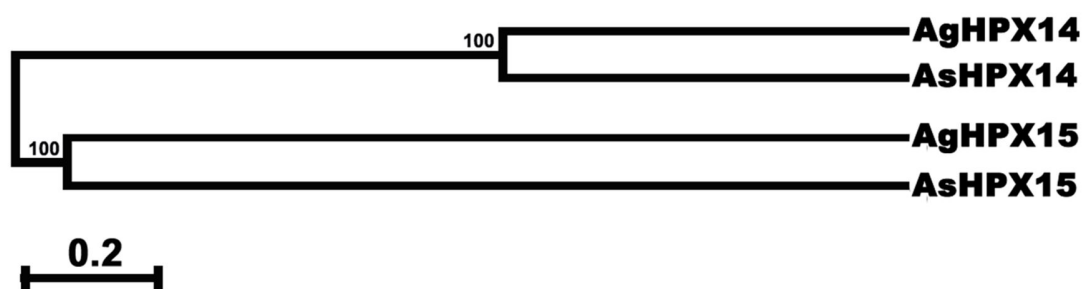
Figure 5.4: Phylogeny of duplicated genes HPX15 and HPX14 of *A. gambiae* and *A. stephensi*. The neighbor-joining (NJ) method was used to construct phylogeny of duplicated genes. The scale bar represents base substitutions per site. The numbers on the branches represent the % of 1000 bootstrap.

5.3.7 HPX15 and HPX14 are under purifying selection

Nucleotide substitutions in protein-coding regions are divided into two classes, ones that change amino acid (nonsynonymous) and those that do not (silent or synonymous). For most protein-coding regions the ratio Ka/Ks has been found to be significantly smaller than 1 (Nekrutenko et al., 2002; Makalowski and Boguski 1998). The Ka/Ks ratio > 1 indicates positive selection while < 1 indicates the purifying selection on the duplicated genes. Predictably, Ka/Ks ratio < 0.5 revealed functional constraints on both paralogs (Li, 1997). Subsequently, purifying selection maintains two functionally distinct copies of duplicated genes in the genome. We further calculated the non-synonymous (Ka) (amino-acid-

changing) and synonymous (Ks) (silent) substitution rate of coding sequences for duplicate pairs. The Ka/Ks ratio of the duplicated genes HPX15 and HPX14 of *A. gambiae* and *A. stephensi* revealed that the values at each node < 0.5 and is shown in **Figure 5.5** and hence, the duplicated genes are under purifying selection **(Figure 5.6)**. Therefore, we hypothesized that the functional constraints on HPX15 and HPX14 genes might contribute in the evolution of the duplicate paralog in *Anopheles*.
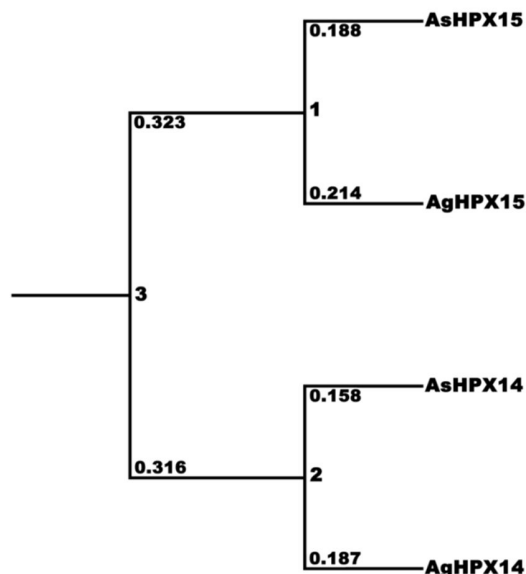
Figure 5.5: Ka/Ks annotated evolutionary tree. Ka/Ks values for each node in the tree. Numbers indicated the nodes of the tree (1, 2 and 3). Ka/Ks ratio values are indicated above each branch.
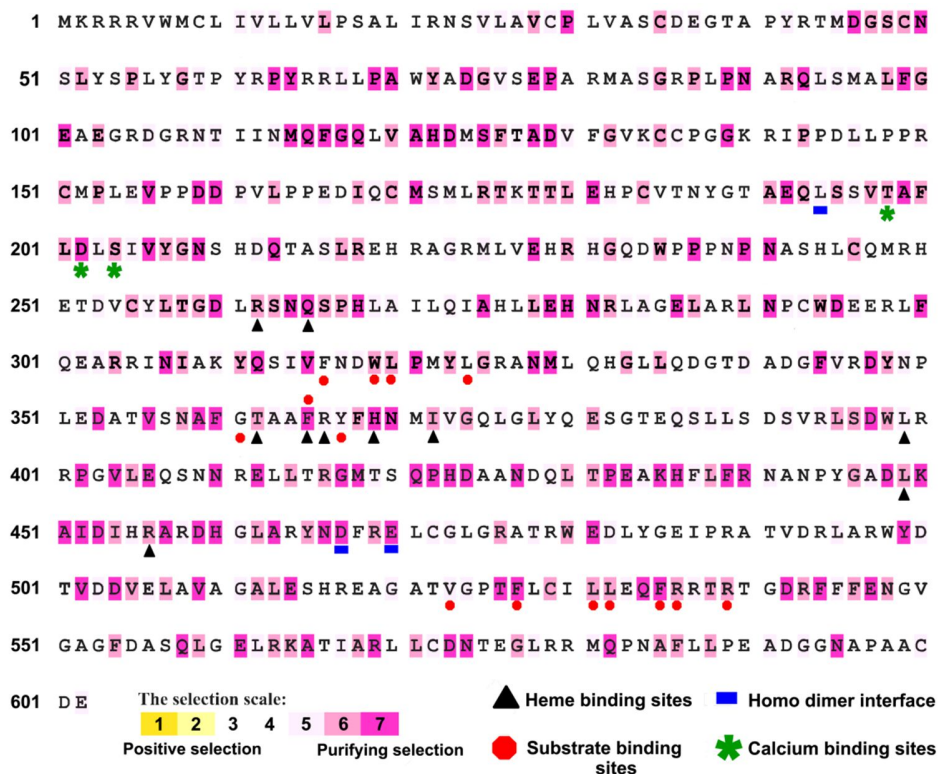
Figure 5.6: Purifying selection on HPX15 and HPX14 of *A. gambiae* and *A. stephensi*. Analysis of natural selection on four genes was carried out using Selecton server. Purifying selection is colored in shades of magenta in the consensus sequence of four genes. The heme binding, calcium binding, putative substrate binding and homodimer interface sites as represented by the characteristic symbols. All four genes showed a significantly high level of purifying selection.

5.3.8 Identification of boundary elements in the duplicated cluster

Boundary elements are able to contribute in the organization of the chromatin structure by supporting the formation of functional domains on the chromosome and might be responsible for gene clustering. Chromatin loops created by boundary elements causes regulation through inter-insulator contacts, thus, provide the chance of rigorous regulation inside this domain (Ma et al., 2016).

Boundary elements share functional properties from flies to human despite their poor sequence conservation. However, they contain small sequence motifs like other cis-regulatory elements that serve as binding sites for proteins involved in boundary function. *Drosophila* uses the binding sites of multiple proteins, dCTCF, Su(Hw), BEAF, GAF, Zw5, Mod(mdg4) and CP190, to derive its boundary function. Due to conserved nature of the *Drosophila* boundary interacting proteins such as CTCF with mosquitoes and other insects also, might make them recognize the similar DNA motifs and contribute towards boundary function in *A. gambiae* as well in *A. stephensi* (Schoborg and Labrador, 2010; Gray and Coates,2005).

Based on these facts, we have used cdBEST software (Srinivasan and Mishra, 2012), to identify potential boundary elements from the duplicated domain of HPX15 and HPX14 in *A. gambiae* and *A. stephensi*. Using cdBEST that was developed originally for *Drosophila*, we predicted putative boundary elements across the duplicated cluster of HPX15 and HPX14 of *A. gambiae* as well as in *A. stephensi* **(Figure 5.1)**. The predicted boundaries fall in the intergenic regions and have multiple binding sites for GAGA Associating Factor (GAF). To understand how boundary elements may define domains in *A. gambiae* and *A. stephensi* cluster, we mapped putative transcriptional enhancers identified by different TFBs prediction tool.

Interestingly, the marked boundary elements flank the mapped TFB motifs across the cluster both in *A. gambiae* as well as in *A. stephensi*. This arrangement of boundaries flanking the cis-regulatory region is in accordance with their suggested role in the formation of chromatin domains. Comparison of cis-regulatory elements revealed that in both the species these clusters contain an array of transcriptional enhancers organized into domains flanked by boundary elements. It is interesting to note that in *A. gambiae* the size of duplicated gene cluster is 40 Mb, which is roughly 2 times the size of duplicated cluster in *A. stephensi.* These differences in size and domain organization may be specific to the evolution of these genes in *A. gambiae* and in *A. stephensi*. Overall, the special localization

of boundaries do not have protein-coding sequences and is in agreement with a role for these elements in organizing the genome into independent domains of gene expression (Ahanger et al., 2013).

5.3.9 *A. stephensi* AsHPX15 and AsHPX14 exhibit a differential expression pattern

To date, studying functional genomics by comparing the gene expression of duplicated genes is one of the principal ways to decipher rapid acquisition of divergent tissue-expression patterns after gene duplication (Huminiecki and Wolfe, 2004; Farre and Alba, 2010).
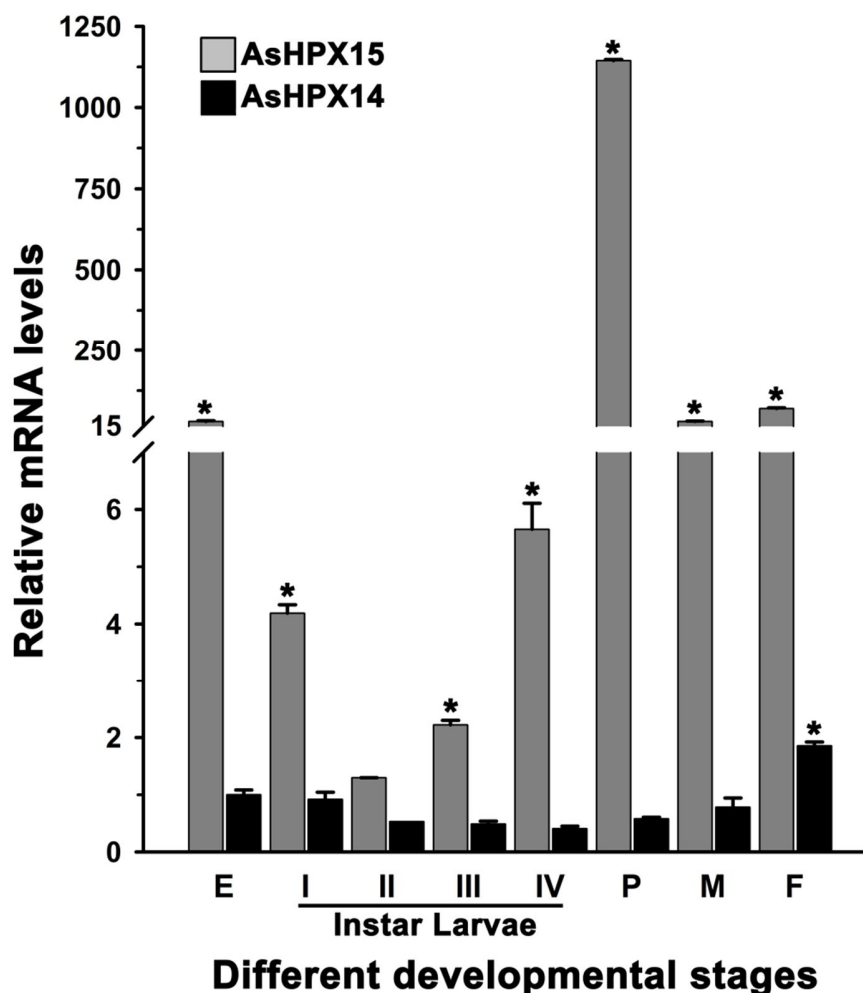
Figure 5.7: Relative mRNA levels of AsHPX15 and AsHPX14 genes in different developmental stages of *A. stephensi*. Results present the mean ± SD of relative mRNA levels of AsHPX15 and AsHPX14 in different developmental stages of *A. stephensi*. The mRNA levels of eggs of AsHPX14 were considered as controls or 1. Significant differences ($p<0.05$) among different stages are indicated by an asterisk (*).E, eggs; I, II, III, IV, various stages of instar larvae; P, pupae; M, males; F, females.

To analyze AsHPX15 and AsHPX14 gene expression in *A. stephensi,* gene-specific primers for both these genes were designed as mentioned in Material and Methods. The qPCR analysis was used to study the expression of AsHPX15 and AsHPX14 genes in different developmental stages. We have investigated the expression of AsHPX15 and AsHPX14 genes in the developmental stages of *A. stephensi* viz. eggs, I to IV instar larvae, pupae, males and females. The gene expression profiles of AsHPX15 and AsHPX14 were highly divergent in all the stages of mosquito development (**Figure 5.7**). AsHPX14 mRNA was found in all the stages of development and showed significant upregulation by 2-fold (p= 0.0067) in adult females in comparison to the eggs. Comparing the expression profiling of AsHPX15 with respect to the mRNA levels of AsHPX14 gene in eggs, AsHPX15 mRNA levels remained high in all the stages (**Figure 5.7**). It has been found that the mRNA levels of AsHPX15 gene were the highest in the pupae (**Figure 5.7**). The mRNA levels of AsHPX15 were 30-fold in eggs (p=0.0030), 4-fold, 2-fold and 6-fold in first instar, third instar and fourth instar larvae (p=0.0014, p=0.0043, p=0.0092), respectively in comparison to the mRNA levels of AsHPX14 in the eggs. The expression of AsHPX15 was 1150-fold in pupae (p<0.0001), 30-fold in adult males (p=0.0006) and 70-fold in the adult females (p=0.0067) when compared to the mRNA levels of AsHPX14 in the eggs. This data showed that mRNA levels of AsHPX15 were higher in all the developmental stages when compared to the mRNA levels AsHPX14.

Mosquitoes are hematophagous in nature and therefore blood feeding is one of the major signals that activate the number of genes participating in the digestion (e.g. trypsin) (Müller et al., 1995) and immunity (e.g. HPX15) (Kajla et al., 2016b, Kumar et al., 2010). We have investigated the expression of AsHPX15 and AsHPX14 in the blood fed midgut of *A. stephensi* at different time points post blood meal. The result presented in **Figure 5.8** revealed that midgut mRNA levels of AsHPX14 showed significant upregulation of 4-fold and 4-fold at 6 h and 12 h post blood feeding (p=0.0070, p=0.0003), respectively. The mRNA levels of AsHPX14 were 2.5-fold at 36 h post blood feeding (p=0.0026). The mRNA levels of AHPX15 also showed upregulation and were thousand folds when compared against the mRNA levels of AsHPX14 in the sugar fed midguts (**Figure 5.8**). The mRNA levels of AsHPX15 was 140-fold higher (p=0.0110), and increases to 19890-fold at 6 h (p=0.0015) and 21500-fold at 12 h (p=0.0007) post blood meal in midguts, in comparison to the mRNA levels of AsHPX14 in sugar fed midguts. AsHPX15 showed induction of 12500-fold at 24 h (p=0.0001) and 4781-fold at 36 h (p=0.0013) post blood feeding in midguts

against mRNA levels of AsHPX14 in sugar fed midgut. This suggested that AsHPX15 is highly inducible gene upon blood feeding in comparison to AsHPX14 which showed induction but its mRNA levels are very less in comparison to the AsHPX15 (**Figure 5.8**).

In both *A. gambiae* and *A. stephensi,* the HPX15 is highly induced by blood feeding in midguts and mediates the cross-linking of mucin layer to form a gut barrier (Kumar et al., 2010, Kajla et al., 2016b). The midgut barrier blocks the interaction of midgut bacteria and epithelium immune responses thereby, supporting the growth of natural gut flora. Our previous study showed that mRNA levels of AsHPX15 are induced in response to the exogenous bacterial supplemented blood feeding (Kajla et al., 2016c) and its silencing suppressed the growth of gut bacteria due to disruption of gut barrier and activation of epithelial immunity (Kajla et al., 2016c; Kumar et al., 2010).
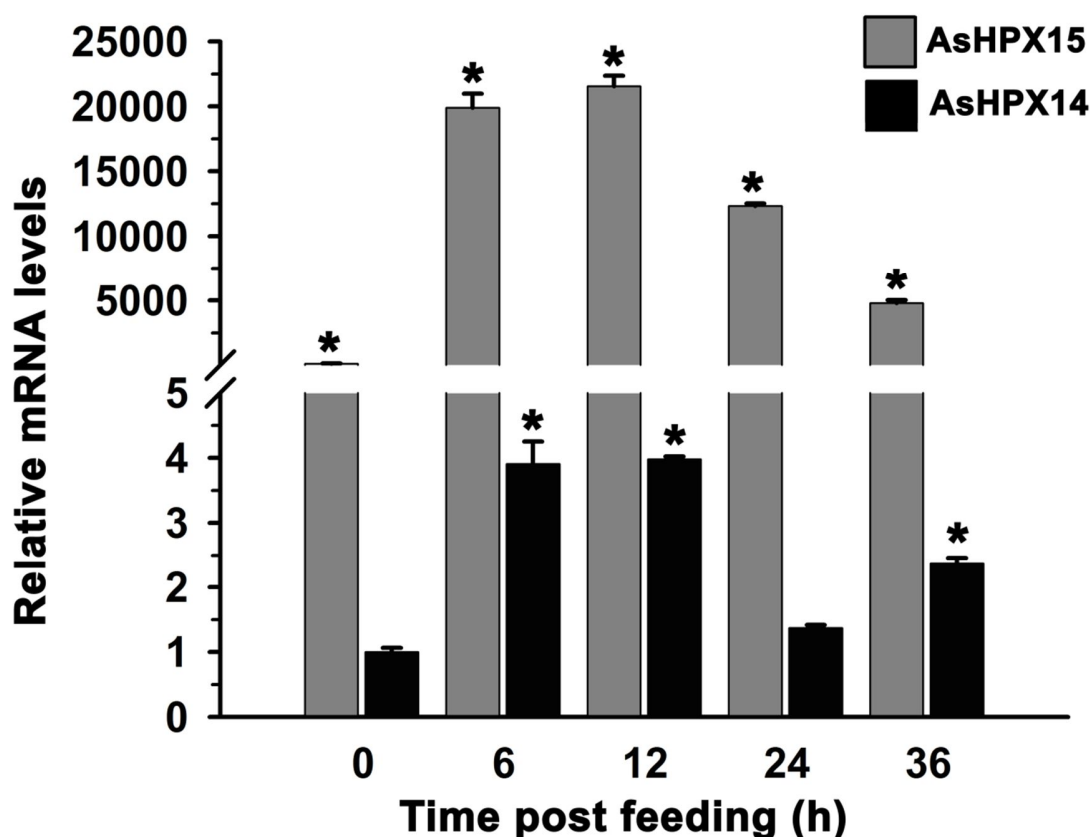
Figure 5.8: Expression kinetics of AsHPX15 and AsHPX14 genes in blood fed midguts. Relative mRNA levels of AsHPX15 and AsHPX14 genes were analyzed in blood fed midguts collected at various time points post blood feeding. The mRNA levels of AsHPX14 sugar fed (0 h) were considered as control or 1. Significant differences ($p<0.05$) against control are indicated by an asterisk (*).

To explore the role of AsHPX14 and immune response associated with this gene cluster we have studied the expression of AsHPX15 and AsHPX14 in the blood fed (BF) or bacteria supplemented blood fed (BF+Bac) midgut of *A. stephensi* as described in Materials and Methods. The result presented in **Figure 5.9** showed that both AsHPX15 and AsHPX14 were induced in bacteria supplemented blood fed midguts at 24 h post feeding. However, their mRNA levels were non-significantly different at 6 h against their respective blood fed midguts. The mRNA levels of AsHPX14 showed significant induction of 2-fold at 24 h (p=0.0317). The induced mRNA levels of AsHPX15 were thousand folds in comparison to mRNA levels of AsHPX14 and induced 1.5-folds against its respective blood fed control (p= 0.0164) (**Figure 5.9**).

So, the above expression analysis revealed that although AsHPX14 showed spatial-temporal upregulation in its mRNA levels in blood fed midguts or bacteria supplemented blood fed midguts but its mRNA levels are very less when compared with the mRNA levels of AsHPX15 that were thousand folds in
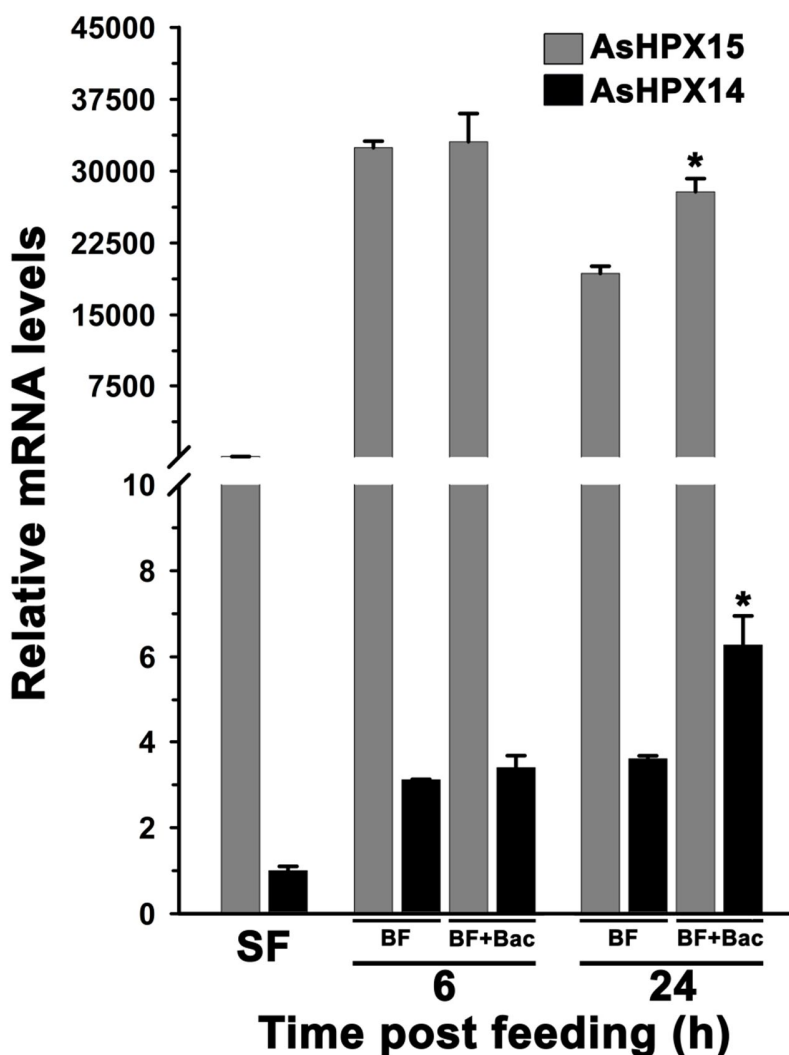
Figure 5.9: Expression kinetics of AsHPX15 and AsHPX14 in bacteria supplemented blood fed midguts. The kinetics of relative mRNA levels of AsHPX15 and AsHPX14 genes were analyzed in the blood alone (control) (BF) or supplemented with a mixture of *M. luteus* and *E. coli* bacteria fed midguts (BF+Bac). The relative mRNA levels were calculated against AsHPX14 levels in sugar fed controls (0 h). The significant differences (p<0.05) in the gene expressions against control are denoted by an asterisk (*).

comparison to the mRNA levels of AsHPX14 in midguts (**Figure 5.9**). So we were interested in exploring the function of AsHPX14 gained after gene duplication.

We further investigated the expression of AsHPX14 in the AsHPX15 silenced *A. stephensi* female to decipher the AsHPX14 function. The control and AsHPX15 silenced female mosquitoes were allowed to feed on mouse or bacteria supplemented blood as described in the Material and Methods. We could achieve the significant silencing of

Figure 5.10: Relative mRNA levels of AsHPX15 and AsHPX14 genes in AsHPX15 silenced midguts. Mosquitoes injected with dsLacZ (controls) or dsAsHPX15 (silenced) RNA were fed on a **A)** mouse (blood) or **B)** and **C)** bacteria supplemented blood and relative mRNA levels of AsHPX15 and AsHPX14 were analyzed at 6 h or 24 h post feeding in **A)** blood fed midguts, **B)** bacteria supplemented blood fed midguts and **C)** bacteria supplemented blood fed carcass. Relative mRNA levels are presented against the mRNA levels of sugar fed midguts of AsHPX14. Significant differences ($p<0.05$) between control and silenced midguts are shown by an asterisk (*).
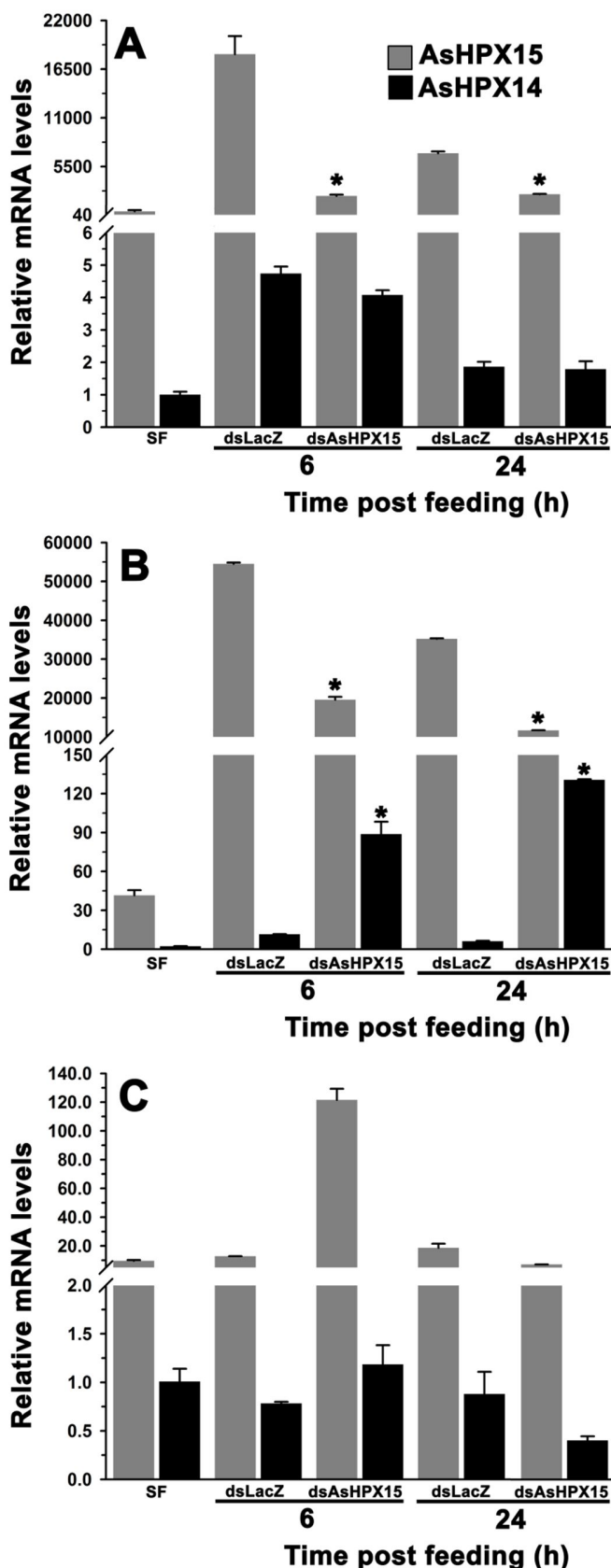
AsHPX15 gene in both the experiments in the range of 60-80% in mRNA levels when compared against control midguts (**Figure 5.10**) and expression of AsHPX14 was then analyzed in these samples. It is found that when we silenced the AsHPX15 gene and allowed the female to feed on a mouse, there is no significant difference in the expression of AsHPX14 in the control and silenced blood fed midguts (**Figure 5.10A**). This might suggest that AsHPX14 has no role during blood feeding in comparison to the AsHPX15 which is induced in midgut in response to the blood feeding and involved in the gut barrier formation (Kajla et al., 2016b, Kumar et al., 2010). We have also analyzed the expression of AsHPX14 in AsHPX15 silenced mosquitoes which were allowed to feed on bacteria supplemented blood. Expression analysis of AsHPX14 gene showed that it induced 8-fold and 21-fold at 6 h and 24 h in the AsHPX15 silenced bacteria supplemented blood fed midguts (p= 0.0076, p< 0.0001), respectively against controls (**Figure 5.10B**).

To know the spatial expression of AsHPX14, we also analyzed its expression in the AsHPX15 silenced bacteria supplemented blood fed mosquito carcasses. The data presented in **Figure 5**.**10C** showed that its mRNA levels were indifferent in the carcass tissue. Hence, it can be concluded that AsHPX14 mRNA levels were induced only in the midguts of AsHPX15 silenced bacterial supplemented blood fed midguts.

The induced expression of AsHPX14 gene in the midgut where gut barrier is disrupted and high bacterial load is present suggested its involvement in the anti-bacterial immune response and this is further strengthened by the presence of peroxinectin-like domain in the protein of AsHPX14. The protein of AsHPX14 contains the LRE motif as an integrin binding domain (Ruoslahti, 1996) as shown in **Figure 5.2**. The peroxinectin is a homolog of vertebrate myeloperoxidase and involved in the invertebrate immunity. It selectively binds to the bacteria and kills them by the enzymatic reaction with H_2O_2 and halide which leads to the formation of hypochlorous acid (HOCl), an anti-bacterial molecule (Klebanoff, 1968; Allen and Stephens, 2011).

Analysis of post-translational modifications in HPX15 and HPX14 proteins revealed that although both the proteins have similar protein sequence but expected post-translational modifications sites in the proteins are different as given in **Table 5.2**. The number of phosphorylation sites is similar in AgHPX15 and AsHPX15 proteins as shown in **Table 5.2**. The phosphorylation sites in AgHPX14 and AsHPX14 also similar. The number of N-linked glycosylation sites is 2 in AgHPX15 and AsHPX15 proteins and only one site is present in AgHPX14 protein while two sites are present in AsHPX14 protein. Hence, the

number of N-linked glycosylation is similar in all the four proteins. The most diverged type of modification is O-linked glycosylation which is mostly present at N-terminal domain of both AgHPX15 and AsHPX15 proteins while in AgHPX14 and AsHPX14 proteins it is also situated in the C-terminal region as highlighted in pink color in **Figure 5.2** (**Table 5.2**).

Table 5.2: Predicted post-translational modifications in HPX15 and HPX14 protein of *A. gambiae* and *A. stephensi*. Post-translational modifications were analyzed by program available at CBU prediction server (http://www.cbs.dtu.dk/services/) and numbers indicate the occurrence of modification, respectively.

Type of modification		Genes			
(numbers indicate the total sites of modification)		AgHPX15	AgHPX14	AsHPX15	AsHPX14
Phosphorylation		81	66	82	72
N-linked glycosylation		2	1	2	2
O-linked glycosylation	Total number of sites in the protein	8	7	9	12
	At N-terminal	5	1	9	6
	In middle	3	3	0	0
	At C-terminal	0	3	0	6

In the previous studies, it has been shown that heme peroxidase can perform either peroxidase cycle or halogenation cycle (Sundaramoorthy et al., 1995; Davies, 2011). One of the factors that decide the reaction catalyzed by heme peroxidase is post-translational glycosylation such as N-linked oligosaccharide chains and O-linked oligosaccharide chains. The peroxidase that performs halogenation cycle has more O-glycosylation sites in the C-terminal domain of the protein, which contains numerous serine and threonine residues (Sundaramoorthy et al., 1995). Hence, it suggests that due to the presence of O-glycosylation site in the C-terminal of HPX14 protein and not in HPX15, HPX14 might be able to perform the halogenation cycle. Thus, we can postulate that AsHPX14 is a peroxinectin that has both cell adhesion and peroxidase activity. The peroxinectin protein might produce hypochlorous acid in the vicinity of the pathogen and mediates their killing (Cerenius et al., 2008; Jiravanichpaisal et al., 2006; Dong et al., 2009b; Johansson, 1999).

The spatial-temporal expression analysis of AsHPX14 deciphered the possible role of this gene in immunity against bacteria. In conclusion, this gene might be evolved to combat bacterial infection in the absence of midgut barrier. Our study showed differential spatial and temporal expression pattern of AsHPX15 and AsHPX14 genes. So, to understand the regulatory mechanisms behind these differential expressions of the

duplicated genes, we further analyzed the structure and position of the promoter with other cis-regulatory elements of these genes.

5.3.10 Analysis of transcription factors binding motifs

The promoter is required for the basal expression of a gene, but its differential expression is contributed by the presence of different type of cis-regulatory elements near or in the promoter region. To explore the mechanism of differential expression of HPX15 and HPX14, we have analyzed the regulatory elements or transcription factor (TF) binding motifs in the 5′ upstream region of these genes in both *A. gambiae* and *A. stephensi*. Various TF databases like JASPAR, MEME suit, MatInspector and Consite have been used to predict putative transcription factor binding sites as described in Material and Methods. List of predicted transcription factors obtained from MatInspector is given in **Table 5.3.** The transcription factors that regulate the expression of genes involved in the developmental process, immunity, physiology and transcription factors that mediate transcriptional repression of the gene are found in the regulatory region of these four genes.

Table 5.3: The list of transcription factors binding motifs in 5′ regulatory region of duplicated gene. Transcription factors in the promoter region of *A. gambiae* and *A. stephensi* HPX15 and HPX14 genes are analyzed by MetInspector and JASPAR database.

Transcription factors	Frequency of TF in 5′ regulatory region of Genes			
	AgHPX15	AgHPX14	AsHPX15	AsHPX14
Activator of alcohol dehydrogenase gene, ADF-1 protein	2	1	-	-
Activator-, mediator- and TBP-dependent core promoter element for RNA polymerase II transcription from TATA-less promoters	1	-	1	-
Bicoid-like homeodomain transcription factors	1	3	1	-
Boundary element associated factor	10	13	6	3
BTB/POZ proteins Bric a brac	4	5	-	-
Cis-acting silencer sequences binding Drosophila polycomp group proteins	-	1	-	-
Core promoter initiator elements	8	14	10	3
Dead ringer factor	2	6	2	
DNA replication-related element factor	11	17	9	2
Drosophila Abd-B group	5	7	4	-
Drosophila basic helix-loop-helix transcription factors	3	3	4	-
Drosophila broad-complex for ecdysone steroid response	6	4	2	1
Drosophila C/EBP like bZIP transcription factors	6	6	-	-
Drosophila Chorion Factor 1 /Ultraspiracle	-	1	1	-

Drosophila Chorion Factor 2	4	13	2	1
Drosophila Dorsal Ventral Factor	7	11	6	1
Drosophila ecdysone induced protein E74A	4	7	3	1
Drosophila fork head factors	1	1	-	-
Drosophila gap gene hunchback	1	2	2	
Drosophila gap gene Krueppel	2	10	6	1
Drosophila GLI ortholog factors	-	1	-	-
Drosophila Glia Cell Missing factors	-	3	-	-
Drosophila heat shock factors	3	1	4	-
Drosophila homeobox transcription factor with CUT domain	2	11	2	-
Drosophila homeodomain protein caudal	1	3	2	-
Drosophila homeoproteins	22	42	15	4
Drosophila neuronal cis element binding factor	3	7	3	1
Drosophila odd-skipped	-	2	3	-
Drosophila proneural repressor	2	5	2	1
Drosophila segmentation gene knirps	4	4	7	1
Drosophila segmentation gene tailless	4	3	2	-
Drosophila sex determinating transcription factor doublesex	1	3	-	1
Drosophila Six factors	1	2	1	1
Drosophila snail protein	4	3	3	-
Drosophila STAT	4	4	6	1
Drosophila supressor of Hairless	2	2	2	1
Drosophila T-box transcription factors	2	6	1	-
Drosophila T-cell factor	4	7	2	-
Drosophila tramtrack protein	1	-	-	-
Drosophila winged-helix nude	4	1	3	-
Drosophila OVO transcription factor	4	5	5	1
Drosophila giant transcription factor	7	9	4	2
GAGA element, binding sites for proteins of the trithorax group (trxG)	1	-	-	-
General transcription factor IID, GTF2D	2	1	-	-
General transcription factor IIIC, GTF3C	1	-	-	-
Iroquois group of transcription factors	4	10	3	1
NK-family of homeodomain transcription factors	1	-	-	-
Paired homeodomain factors	8	19	7	-
Plant TATA binding protein factor	1	15	2	-
Proximal proneural response elements	-	1	-	-
RNA polymerase II transcription factor II B	-	1	-	1
Runt-domain transcription factors	3	1	1	
TGIF (TG-interacting factor)-Exd (extradenticle) group	2	3	2	1
Transcription factors with POU-domain - N-terminal to homeobox domain	6	4	5	1
Vertebrate TATA binding protein factor	13	38	9	3
Yeast TATA binding protein factor	4	8	-	-
Zeste transvection gene product	-	-	-	1

The frequency of various transcription factors in the regulatory region of four genes has been analyzed. The TFs have been categorized into conserved and specific transcription factors among paralogs and orthologs based on their presence in the promoter region of the duplicated gene in *A. gambiae* and *A. stephensi* and are presented in **Figure 5.11**. The frequency of transcription factors that might explain the spatial-temporal expression of the duplicated gene was analyzed and presented in the **Figure 5.12**.

In natural condition, AgHPX15 showed high induction only in blood fed midguts (Kumar et al., 2010). The presence of ecdysone binding sites with GATA in the 5' region of AgHPX15 may explain its specific spatial and temporal expression pattern in

Figure 5.11: Transcription factors distribution in paralogs and orthologs of *A. gambiae* and *A. stephensi*. Distribution of conserved and specific transcription factors in the regulatory region of paralogs and orthologs of *A. gambiae* and *A. stephensi*.

blood fed midguts (**Figure 5.13**). Similar expression pattern is obtained in AsHPX15 gene in response to blood feeding as shown in **Figure 5.8**. So, the presence of putative binding sites for transcription factors ecdysone and GATA in the 5' upstream region of AsHPX15 gene suggested the plausible involvement of these two TFs in the tissue-specific differential expression.
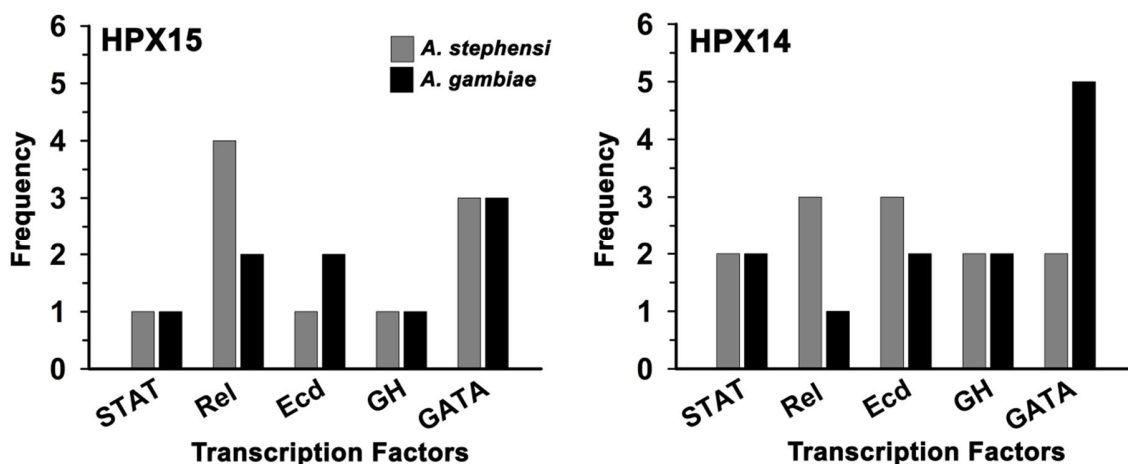
Figure 5.12: The frequency of transcription factors in orthologs. The frequency of transcription factors in the regulatory region of orthologs of HPX15 and HPX14 of *A. gambiae* and *A. stephensi*.

The previous study also showed that AgHPX15 causes the cross-linking of mucin layer and forms the gut barrier to provide low immunity zone (Kumar et al., 2010). Genes involved in the cross-linking mechanism of the matrix are regulated by TF grainy head (Narasimha et al., 2008). Our finding showed the presence of putative binding sites for TF grainy head in the vicinity of AgHPX15 and AsHPX15 promoter, correlates with its known biological function that is cross-linking of mucin layer on the luminal side of the midgut.

5.3.11 CTCF, an enhancer-blocking element (insulator), regulates expression of duplicated paralog

To decipher the putative mechanism which regulates the spatial and temporal expression of the duplicated genes AsHPX15 and AsHPX14, we searched for the regulatory element in the 5' upstream region of both the genes in *A. gambiae* and *A. stephensi*. Our findings revealed the presence of putative binding sites for evolutionary conserved insulator element CTCF in the promoter region of both AsHPX14 and AgHPX14 genes. CTCF has enhancer-blocking activity and it ensured that only appropriate promoter will be active in a gene cluster domain under particular enhancer. CTCF protein consists of 11 zinc-finger domains and for CTCF binding, a central ~12-20 bp core (DNA sequence common to most CTCF sites) DNA motif is required (Kim et al., 2007; Renda et al., 2007). In addition, 10 bp of upstream and downstream motifs may also affect the binding affinity (Nakahashi et al., 2013). We found 3 CTCF binding sites in the regulatory region of HPX14 gene (the longest motif of 21 bp (CACCACATCGAGTGCGCCGTA) is present in the regulatory region of AsHPX14 gene at -21 from start codon). We also found CTCF binding site in the upstream region of AsHPX15 (-308 bp) and AgHPX15 (-1150 bp) (**Figure 5.13** and **5.14**), but motifs predicted were only 6 bp long and hence, CTCF binding may not be strong to drive gene repression as described earlier (Phillips and Corces, 2009; Nakahashi et al., 2013). Thus, CTCF might regulate the spatial and temporal expression of HPX14 genes (Ishii and Caemmli, 2003).

```
-1366 CTAAGAGAACATCGATGGTGAAGTGGGTGGTGGGAGGAAAAAATATCAGGTGTAAACAAACG

-1304 CTTTCTGCACTTCTTTAACATCGTTTTTTTCATAATTTTGTTGGAAATGGGGTAAATATGAG

-1242 TAATGAGAGTGTTGATAAACATATGTTTTATTACACACGTAGGCAGTGGGTTAAATTTTTGT

-1180 GAGGTAAAATTTGCGCGATTCCGACGGCTTGAGCTGTCGGCAGAAAATCTCTCGAATGCACC
                            *******

-1118 TTCCTGTTCGAGCAATACGGTACCGATTTTTACTTCGTCGACTCGAATCTGCAGATGAGTTA

-1056 CTCGAAGCGGAATGTCGCTAGCAAAACGATACGAAAATTAACATAATTAACGTGCGAATGAA

                                              ................
-994  TATTCTCGAATGCCGAACGCAATATAATAAGCCTGTATGTATGTAACAACAGCAAACATTT

                                        .............
-933  GGGGTTTGCTAAAAAATCGTTGTAGAGGTTTATGTGTCGCTTTTGCTTGATCGGATCAAACG

-871  ATATTTTTGAACTGACTGTAGAAAAAAAGAATGAAAATTGAATGCGAAGCTTTAAATCGTGT

-809  CAAACTAGGCGATTTCATGTCAAAGAACGTTAGCTAACAACACTCGTCGAGAGAGAAAGAGG

-747  AACTGGGAACTATTCAGTGAAATAGCCCAAGATCTCAAAATTACAGTATTGTTGTTCGATAG
                                                           -275
-682  TTTGACAAA//TGACCTGCCTTCAGAAAAACAAATCACTGATTAAGAATCATTATCTTAATT

-224  AGTAATTGTGTATTTTTGGTGAAACATCCAACCTTTTGAGAAAAATTTCTTGGATTAGATTT

-162  CAGTAGCATAGCGTTGAGCAATCATAGTGAAGATTAGACTGCTAGCAATATCCATTTGTCAA

-100  GTGAAAAACAATTAAAATATTGGTATTGCTTGCATCACTCAAGCTCCTCTCTCGTCCAAAGC

-38   TCTTGAAATAAAAATGAATTTCTCTCACTTTCCTTAGG

1 ATGTACACCGTCCCAGGTTTATGGATTGCCATTACGCTCCTTTCA//GAGAGCAAT 1764
1 M  Y  T  V  P  G  L  W  I  A  I  T  L  L  S    E  S  N   588

1765 CCGGTGATTCCGTGTGAGCAAATTTCAAAGGTGAATTTAACGCGATGGCGTTGA    1809
589 P  V  I  P  C  E  Q  I  S  K  V  N  L  T  R  W  R  *    602
```

Ecdysone   GATA   Rel   GrainyHead   STAT   CTCF

Figure 5.13: Analysis of transcription factors in the regulatory region of AgHPX15 gene. The putative transcription factor binding sites for various factors such as Ecdysone, Rel, STAT, GATA and Grainy head are depicted in the regulatory region of AgHPX15 gene as predicted by MatInspector and JASPAR software. The binding of CTCF is highlighted by the gray box. The promoter region is shown by double zigzag underline with TATA Box and transcription start site (A) in bold letters.

-1751 GGACGTTGCAGGAACTTTGCCGGATCGAATATTTCCGGGTACATGTGCAAACCGGAACATTC

-1689 CACCAGCAGCACTGCAACCAGAATGAAGTTTTCCGCACGACACATCGCCACTAGTCTTGGAC

-1627 ACTTGAAACTTTCGGTATTTCACTGAATGCTTTTGTGTTTTGTTGTTAATGTCTTTCGCTGG

-1565 CGTTGTTTTACTGTCGCACTCGACGCACCGTATGCAAATGTATGCTGTTCGGATCGGTGAGG

-1503 TGTGATTTATAGCAGGAATGTGGTTAGGAGCGTTCCCTACCTTAGCGGTCCTACGCCATCGA

-1441 AAAATTGTGCAATGACCACTCGACCCATTGTCCCAGAAATATCGGC \\ GACTGCCGGTGACG
　　　　　　　　　　　　　　　　　　　　　　　　　　　-1394 \\ -1223

-1209 GAATTCCCTGGGGAGGAAATGTTTCTTCATCTTCGGTTTGTGGAACGATCGGGAGGGAATTC

-1147 GCGGTGAAATGGACGAGAAAAACTAACCCAGTTCAATGATTGGTTTTGTGTTTGAAATGAGA

-1085 TTGAA \\ CTAAGCTCGGCGGAACTATTTATAGCAGCA \\ TGGTTTCATTTTTCGCAAGCTCG
　　　　-747　　　　　　　　　　　　　　　　　　-717 \\ -420

-397 CGGCGGCCCA**GTATAAATGA**GCCGGGCGTGAAGTGTAACGAATTCAGTAACCG**A**TTCCTG

-337 GTTTTCTGGCTGTCAAGAAGTTGCGCCAGGTACGGGCTTTTGATTTGAAATTTGGTTCGGTT
　　　　　　　　　　　　　　　*******

-275 CGATTGGTTCAGTTTTGGTGATTCTGTGGTGAATCAGAGTGACGGGAAAAGTGTGTTTATTG

-213 CTTGCTTAGGATTGTGCCTTATCATGTGATATTGATATGGATTTAGCGCTCTTGAACGGGTC

-151 AATAAGTGTTCAGTTATCGATCGAAATTACGGTCTTTGAAGTTTTACACCTCTCAAAGCACA

-89 GTTATTTGAACAGACTTTTTTTTGGAAAGGTGTAAATATAAATCATTTTCTAACTGAAATAT

-27 TCCAACAAACATCTCACAGCTATCAGG

1 ATGAAATCTATCGTCGGTTCACTGCTGCTTTGCGTGACTCTGCTGGCCAGT \\ GTCAGTGAG 1534
1 M K S I V G S L L L C V T L L A S   V S E 578

1735 GAGAATCCGGTGGTTTCGTGTGAGCAGCTTCCGGTGGTGAATTTAAACCGATGGCCTTGA 1785
579 E N P V V S C E Q L P V V N L N R W P * 595

Ecdysone　　GATA　　Rel　　GrainyHead　　STAT　　CTCF
　　　　　　　　　　　　　　..............　　　　　*****

Figure 5.14: Analysis of transcription factors in the regulatory region of AsHPX15 gene. The putative transcription factor binding sites for various factors such as Ecdysone, Rel, STAT, GATA and Grainy head are depicted in the regulatory region of AsHPX15 gene as predicted by MatInspector and JASPAR software. The binding of CTCF is highlighted by the gray box. The promoter region is shown by double zigzag underline with TATA Box and transcription start site (A) in bold letters.

121

-1282 AACTTTTACCAGCATAAGCGTAAGTTAAGCCAGCAATTATGATAATGCCAGGAAAACTAATA

-1220 ATGTA**//**CCACTGAACCCAGCATTGTGCTGTTACTAAATAA**//**GAGGTGCAATATTCGCTGCA
　　　　　　　-1001　　　　　　　　　　　　　　　　　　　　-966　-794

-774 ATCTCGACCATCTAAACCCTACCAGGCAAAAGAAAAGTCGGGTCGGGGGGTGTCTCGTCGGA

-712 AGTGGTGCACTAATGGGTGCCTGATAGGGTTTAATGTGCTGCTTGATATGGTACATATTGTA

-650 GCATAAAAGAGAGGTGGAGAGCATGGTGCAGCTAATCGAGCCAAACCGGGGGCATTGTTTGA

-588 TCCGCCAGTAACGAGAACGGCATACAAACGCCATTCTAATAAGGGGCAGGCTTGCGTTGCAT

-526 TGAATGCAAAACCGCCCGTGACAAAAAGGTCGAC**TATAA**ACGGTTTTAAAAAACGCGTGGCA

-464 CGAGCGCGCAGCGC**A**AACACAAAGCATTTATCGCATCAGCGTCGCACACATTAGCCAGCATT

-402 GTGTTGCTGCCGATGAAAGCGATTGGTATCAATTACAATCCTTAGTGCCCCCCTCCCCTCAC

-340 CAGATCGACCTGACCTGGTGGGTTTGTGGTATAAATTATAACGTTGTAAATATTGGCGCTGC

-278 CATAATTTGTGCTACGTTTAGCGTTTACTCTTCTCACTTGCCACCGGTTGCATCGTCTGGGG

-216 GTCGTGTTTAGTGGTTTGCAAAATTGTGCCTTCCGCTGGTTGTGGTTTTATCCGCCCGTTTC

-154 GATTCCCCGTTTTTTCCCACCTAATTTGCAATGCTAGTTTTGGAGTGATTTTGTGTGTTTCT

-92 TTTTTTGTGTTAGTGAAAAAGAAAAACGAGAAACTCACCACACGCGCGTGGTGAACGGTGAC

-30 GCGCCCGAAATGTATGCAATCGTGCGGAAA

1 ATGAAGCGCAGGCGAGTTTGGATGTGTGTTTAATAGTGCTGCTG **//** GCGCCGGCA 1795
1 M　K　R　R　R　V　W　M　C　L　I　V　L　L　　　 A　P　A　 598

1796 GCGTGCGACGAGCTGCCGAAGGTATTGCTGGAACCGTGGCGTGGACGATAG 1845
599 A　C　D　E　L　P　K　V　L　L　E　P　W　R　G　R　*　 614

Ecdysone　GATA　Rel　GrainyHead　STAT　CTCF

Figure 5.15: Analysis of transcription factors in the regulatory region of AgHPX14 gene. The putative transcription factor binding sites for various factors such as Ecdysone, Rel, STAT, GATA and Grainy head are depicted in the regulatory region of AgHPX14 gene as predicted by MatInspector and JASPAR software. The binding of CTCF is highlighted by the gray box. The promoter region is shown by double zigzag underline with TATA Box and transcription start site (A) in bold letters.

-977 GGAAGCGTAAATCAGCTCACTGGCACCGGTGTTT<u>TGTATATTCC</u>CACGCTTCCTAAACTATAC

-915 TGTCAAAGCATACCACGAATCCGAAACGTTTGTGCAGTTACAAAACCGTCGCTACAAACCCA

-853 TCCATCGTTGTAAACCAGCTCAGAACCGCGACCTGTGTGCGAGCCAGTGCATCAATCAGCGG
⁣ ⁣ ⁣ ⁣ ⁣ ⁣ ⁣ ⁣ ⁣ ******************

-791 TACATTAGCCGTGT<u>CATCGATATGC</u>ACAAACAGTGCCCAAACCCCCGAAAACAATGGCATGC

-729 AAAATGAGGTGCAATATTTTCTAAACAATCTAAACAGTTGCTGCTCGATCCGCCACCATCAC

-667 TACTGCGG<u>CAATCGGAAGTGT</u>GCGCTAATGGTGCGCCTGATAGGGTTTAATTTGCAGCTTAC

-605 GGCATAATGTACGCGCAGCACAAACAGCCAACAAACCGCCCCGTAATGGAGGGCACACTGTA

-543 TTGTTTGTTCCGGCAGTAACGCGAACTGCATACAAC<u>GGGATATTGC</u>AACCTGTCCCGTGGTG

-481 AGAGGCTGGTGCCTGGCTGTGGTTAACGTTCGGGTTAGGG**TGAGTGTTTT**TAAAGCGCCCC

-420 GGCACGCGC**A**CACAAAGCATTTATCGCATCAGCATCACACATTAGCTGCTTTGTTGGCAGCG

-358 ATGAAAG<u>CAAACGGG</u><u>TATCAATTAC</u>ATTCTTATTGCCCGCGCCCCATTGTGCCCACGATTTC
⁣ ⁣ ⁣ ⁣ ⁣ ⁣ ⁣ **************************

-296 CCGCGACATCAAACAATAACTCGTCGTTGACGATACGCCC<u>GGCGCTTCC</u>GAGCAGGCCGTAT

-274 AAATTATGGCGCTTGTAAACATTCGGTAGCATAATTTTGTGCTGCGATT<u>GGCATATTCC</u>ACT

-172 CGGGGTGGTTGGTAATTTGCCACCAGTTGCGTCAGGGGTCGGATGTGTCGAAAAATGTGTGT

-110 TTTTATTTCGCAGATATTAGTTGTGTGAAGTGATTTTTGGTACGATTTTTAGTGTTTCGTTT

-48 CGCGCCAGAAAAAAACGAACTG<u>GAAATT</u><u>CACCA</u><u>CATCGAGTGCGCCGTA</u>
⁣ ⁣ ⁣ ⁣ ⁣ ⁣ ⁣ ⁣ **************************

```
   1 ATGTATGCAATCGTGCTGCACGTGAAGCGAGCTAGGGTGTGT \\ CCAGTG  1797
   1  M  Y  A  I  V  L  H  V  K  R  A  R  V  C        P  V   599

1798 CCGTGCTCGGAATTGCCCGAGGTACGGTTGGAACCGTGGCGAGTGTGA       1839
 600  P  C  S  E  L  P  E  V  R  L  E  P  W  R  V  *        612
```

| Ecdysone | GATA | Rel | GrainyHead | STAT | CTCF |

Figure 5.16: Analysis of transcription factors in the regulatory region of AsHPX14 gene. The putative transcription factor binding sites for various factors such as Ecdysone, Rel, STAT, GATA and Grainy head are depicted in the regulatory region of AsHPX14 gene as predicted by MatInspector and JASPAR software. The binding of CTCF is highlighted by the gray box. The promoter region is shown by double zigzag underline with TATA binding protein motif and transcription start site (A) in bold letters.

123

The presence of putative CTCF binding site in the 5′ upstream region (promoter) of AgHPX14 gene as shown in **Figure 5.15** is also supported by the previous finding reported in a study carried out by Johanson, 2013. Further, the presence of CTCF motifs in 5′ upstream region of AsHPX14 gene (**Figure 5.16**) might demonstrate that although similar TF binding motifs are present in the promoter region of both AsHPX15 and AsHPX14 genes but CTCF controls the expression of AsHPX14 (non-significant differential expression in comparison to AsHPX15) during the natural condition (presence of AsHPX15 and formation of gut barrier), Hence, basal expression of AsHPX14 gene is observed and its mRNA levels are very less in all the conditions in comparison to the mRNA levels of AsHPX15 gene.

The expression of AsHPX14 gene showed non-significant difference in AsHPX15 silenced blood fed midguts against controls. However, its expression is significantly induced in AsHPX15 silenced bacteria supplemented blood fed midguts against controls. We have suggested that change in the natural condition because of disrupted gut barrier and challenge with bacteria causes some signal to displace CTCF from the promoter region of AsHPX14 and recruitment of immunity associated TFs (putative binding sites are shown in **Figure 5.15** and **5.16**) in the promoter of AsHPX14 causes its induction. Thus, the function of AsHPX14 might be against exogenous bacteria when midgut barrier was disrupted. In the previous study, it has been shown that lysozyme gene repression is mediated by the enhancer-blocking protein CTCF/cohesin complex and pro-inflammatory stimuli, such as lipopolysaccharide (LPS) treatment, causes transcriptional activation of this gene. The treatment with LPS for 30 min to chicken cell lines HD11 and HD37 causes binding of C/EBPb and AP1 to a newly activated cis-regulatory element and recruitment of RNA Pol II causes nucleosome remodeling. This chromatin remodeling eventually leads to the removal of CTCF and its partner cohesin, allow transcription factor to act on the promoter and cancel the effect of relief CTCF mediated gene repression (Lefevre et al., 2008).

Hence, from the above data, we conclude that transcription of AsHPX14 gene is under the control of CTCF and AsHPX14 is a peroxinectin, might perform halogenation cycle, as summarized in the **Figure 5.17**. In AsHPX15 silenced bacteria supplemented blood fed midgut, mucin barrier disrupts and exogenous bacteria releases PAMPs. These processes cause immune signaling pathway to recruit transcription factors in the regulatory region of AsHPX14 which initiate displacement of CTCF. The recruited transcription factors induce the transcription of AsHPX14 gene.
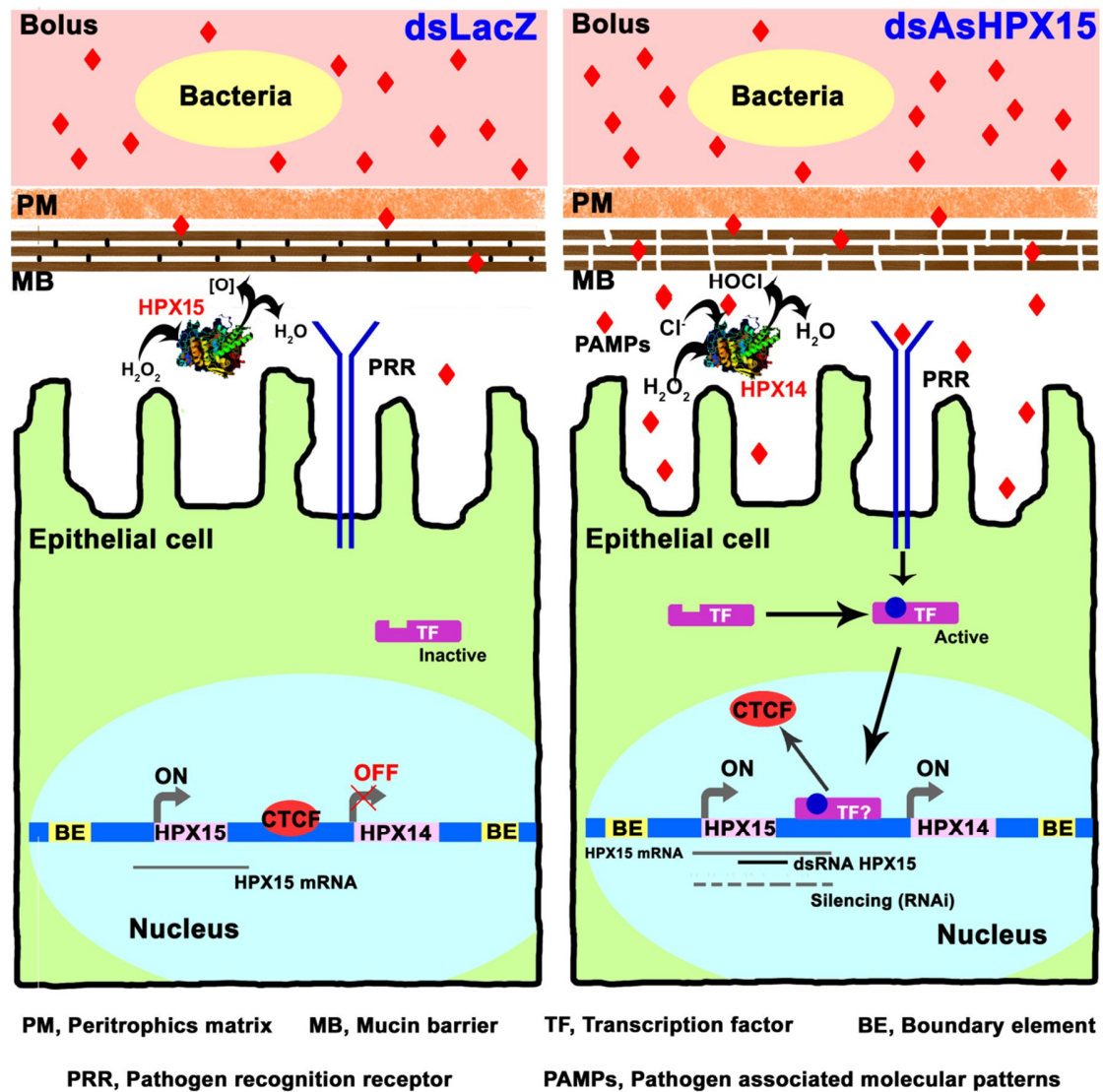
Figure 5.17: Hypothetical model depicts the duplicated gene cluster organization and function in *Anopheles*. The duplicated genes are flanked by the boundary elements (BE). In dsLacZ RNA injected midguts and exogenous bacteria supplemented blood feeding, HPX15 shows the peroxidase activity and catalyzes the cross-linking of mucin layer that blocks the direct interaction of proliferating lumen bacteria or bacterial elicitors with midgut epithelium. The HPX14 is under the control of CTCF and remained in the off condition with the basal expression. In dsAsHPX15 RNA injected midguts and exogenous bacteria supplemented blood feeding, HPX15 silencing disrupts the mucin barrier that allows direct interaction of lumen bacteria or their elicitors with midgut epithelium. In this condition, some molecules get attached to the pathogen recognition receptor (PRR) and activate the unknown transcription factor (TF) which goes and displace the CTCF from the promoter region of HPX14. This, in turn, causes the induction of mRNA levels of HPX14 gene. The AsHPX14 might perform halogenation cycle to produce HOCl that, in turn, kills bacteria.

In our previous report, we showed upregulation of STAT pathway in AsHPX15 silenced exogenous bacteria fed midguts (Kajla et al., 2016c). Transcription factor analysis revealed the presence of putative STAT binding site in the regulatory region of AsHPX14. So, the two data might suggest the involvement of STAT pathway in the induction of AsHPX14 as immune gene against bacterial challenge. This hypothesis is supported by the other studies which showed that JAK/STAT pathway mount immune responses upon bacterial challenge (Stokes et al., 2015).

5.4 Discussion

Previous studies showed that heme peroxidase HXP15 of *A. gambiae* and *A. stephensi* is highly induced after blood feeding in the midgut and carried out the formation of the gut barrier that creates low immunity zone for the proliferation of commensal bacteria (Kumar et al., 2010, Kajla et al., 2016b). In the current study, we analyzed the computational and functional genomics of lineage-specific tandemly duplicated genes of heme-peroxidase multi-gene family from two major malaria vectors *A. gambiae* and *A. stephensi.* The aim of this study was to explore their divergent expression and their *cis*-regulatory motifs (AgHPX15, AgHPX14, AsHPX15 and AsHPX14). Our analysis showed that this duplicated gene cluster is present in all the genome 19 worldwide distributed anophelines species. Additional investigation is carried out to compare and evaluate protein divergence among these duplicated genes by Ka/Ks (non-synonymous to synonymous substitution) analysis. These two duplicated genes are considered functional as the ratio of Ka/Ks values are much lower than 0.3 and seem to be under purifying selection. Neofunctionalization has been normally accepted as the terminal fate of duplicated genes under purifying selection.

To further evaluate their functional divergence after duplication, we also analyzed the expression of these genes in the Indian malaria vector *A. stephensi*. Finally, to decipher their promoter region, transcription factors, insulators, associated chromatin boundaries and other related elements, we performed the computational prophecy using different tools. In our study, we have found that HPX15 of *A. gambiae* and *A. stephensi* contains binding sites for TFs that induce the transcription of genes in response to blood feeding, in immunity and development. HPX15 of *A. gambiae* and *A. stephensi* is induced in blood fed midgut only and involved in the cross-linking of mucin layer (Kumar et al., 2010, Kajla et al., 2016b). So, the presence of motifs that could be recognized by ecdysone, GATA and grainy head showed positive co-relation with the blood-induction, tissue-specific expression and

involvement in cross-linking of mucin layer as reported earlier for other genes (Jinwal et al., 2006; Senger et al., 2006; Giannoni et al., 2001; Swevers et al., 2003; Narasimha et al., 2008). Although all these TF binding sites are present in the promoter region of HPX14 in both *A. gambiae* and *A. stephensi* but due to the presence of putative CTCF binding sites, activity of all these TFs are repressed. Our study suggested that differential expression of AsHPX14 gene is under the regulation of CTCF.

When AsHPX15 is active in the gut, there is no induction of AsHPX14 upon bacterial challenge but in AsHPX15 silenced bacteria supplemented blood fed midguts induced expression of AsHPX14 is observed. This might have happened because transcription factors binding sites present in the promoter region of AsHPX14 gene remain suppressed because of CTCF binding. However, in AsHPX15 silenced bacteria fed midguts, some signal causes the displacement of CTCF and activation of AsHPX14 by binding of transcription factors. In our previous report, we have shown that silencing of AsHPX15 in the exogenous bacterial fed midgut suppressed the growth of exogenous bacteria. This is achieved through the induction of NOS gene via the activation of STAT pathway (Kajla et al., 2016c). The presence of STAT binding site in the promoter region of AsHPX14 might suggest the involvement of STAT pathway in the induction of this gene in AsHPX15 silenced bacteria supplemented blood fed midguts. The overall study can be concluded as during evolution mosquito immunity generates some sort of machinery, which activates when general immune pathways are not enough to fight pathogen development. As shown in our previous report, when exogenous bacteria are present in the gut and barrier formation takes place, induction of known anti-bacterial classical immune molecules such as Toll, GNBP and gambicin is observed. In AsHPX15 silenced bacteria supplemented blood fed midguts, disruption of gut barrier causes induction of immunity. In these silenced midguts, the classical immune molecules such as GNBP, Toll etc. are not induced but induced expression of NOS is observed (Kajla et al., 2016c). Further investigation might confirm the exact signaling between AsHPX15, AsHPX14 and STAT pathway upon bacterial challenge.

5.5 Conclusion

It is clear from the above study that this duplicated gene cluster is a functional domain composed of genes that are acting in the developmental processes, have tissue-specific expression and possess innate immune responses. We have also shown that these

domains are unique to *Anopheles* which might indicate its evolutionary importance. Our observations on this heme-peroxidase cluster also showed that detailed analysis of individual genes from this cluster might be a promising way to identify novel functional units of various biological processes probably in mosquito immune defense mechanisms against blood-borne antigens and can be targeted to develop transmission blocking strategies.

Conclusion of thesis

The work presented in this thesis focused on the role of heme peroxidases in maintaining bacterial homeostasis and regulating *Plasmodium* development in the midgut of *A. stephensi*. A sincere effort was done to explore the interaction of two heme peroxidases, AsHPX2 and AsDuox in immunity against bacteria and *Plasmodium* at the molecular level in *A. stephensi* midgut.

Molecular characterization of AsHPX2 revealed that it is a single exonic gene that encodes for a globular secreted protein of 692 amino acids. The presence of integrin binding motif, LDV revealed that it is a peroxinectin like protein. The expression of AsHPX2 is reduced in blood fed midgut. This gene is induced in the presence of exogenous bacteria in the midgut. The silencing of AsHPX2 gene in the sugar fed midgut and bacteria supplemented blood fed midguts significantly increased the bacterial growth in silenced midgut in comparison to the control midguts. This data suggests that AsHPX2 is one of the molecules that maintain midgut microbial homeostasis in the midgut. Our study also showed that silencing of AsHPX2 increased the *Plasmodium* development and hence, acts as an anti-plasmodial molecule in the midguts. This result revealed that the anti-plasmodial role of HPX2 is conserved in *A. gambiae* and *A. stephensi*. Thus, this mosquito-specific gene can be targeted to manipulate mosquito vectorial capacity.

Our study on Dual oxidase of *A. stephensi* revealed that it has both heme peroxidase and NADPH oxidase domain. This gene is highly conserved among world-wide distributed anophelines. This gene is induced in blood fed midguts and has a weak negative correlation with the growth of endogenous bacteria in blood fed midguts. Silencing of AsDuox in sugar fed and blood fed midguts increased the bacterial growth. The increased endogenous bacteria in the AsDuox silenced blood fed midguts induced the expression of anti-plasmodial molecules such as NOS, TEP1, HPX2 and NOX5. This showed that AsDuox silencing modulates the midgut immunity and this can be further explored to block *Plasmodium* development in the midguts. Expression analysis revealed that this gene is induced in bacteria supplemented blood fed midguts and has a strong negative correlation with the growth of bacteria in these midguts. Silencing of AsDuox and infection with exogenous bacteria did not have any effect on the rate of survival of control and silenced mosquito. Analysis of 16S rRNA in these midguts revealed the reduced bacterial growth. This showed that mosquito midgut immunity is highly efficient to handle the bacterial load.

This data showed that Duox is one of the major molecules of midgut immunity. AsDuox gene is induced in response to the *Plasmodium* infection and silencing of this gene reduced the number of *Plasmodium* oocysts. Thus, we conclude that AsDuox gene is an important molecule of innate immunity against blood-borne antigens in midgut. This gene maintains bacterial homeostasis and regulate *Plasmodium* development in the midgut. Hence, the manipulation of this gene will open up new frontiers in blocking the *Plasmodium* development.

Besides these, gene duplication event in the heme peroxidase multi-gene family was also analyzed and we found that HPX15 which is previously reported as a vaccine candidate has a duplicated paralog named HPX14 in both *A. gambiae* and *A. stephensi* have. They are tandemly duplicated paralog present in head to tail orientation in the genome and organized into an independent domain of gene expression. Expression analysis of this duplicated domain revealed that HPX15 and HPX14 have differential pattern of gene expression. The mRNA levels of AsHPX15 are very high in comparison to the AsHPX14 in blood fed midguts and exogenous bacteria supplemented blood fed midguts. To elucidate the cause of basal expression of AsHPX14, we analyzed the transcription factor binding motifs in the regulatory region of this duplicated paralog. Surprisingly, we found the presence of an insulator protein CTCF, and its core motif is present only in the promoter region of AsHPX14. Hence, we assumed that CTCF regulates the expression of duplicated paralog, AsHPX14. Further studies revealed that AsHPX14 is induced in AsHPX15 silenced bacteria supplemented blood fed midguts, however, expression of AsHPX14 remained unaffected in AsHPX15 silenced blood fed midguts. This data suggested that AsHPX14 may have a role in immunity and not in physiology. Our analysis revealed that there is no redundancy in the function of AsHPX14 and AsHPX15 and hence, strengthen the potentiality of HPX15 as a vaccine candidate.

Thus, overall we conclude that in *A. stephensi* heme peroxidases HPX2, Duox and HPX15 are involved in maintaining gut bacterial homeostasis and also regulate *Plasmodium* development. The study suggested that gut physiology can be manipulated by targeting these heme peroxidases to arrest *Plasmodium* development. Hence, further these heme peroxidases can be explored to develop transmission blocking strategies to control malaria.

Future Prospects

A further research will include the functional elucidation of following future perspectives:

❖ It would be interesting to know the exact mechanism of HPX2 and Duox in maintaining the midgut bacterial homeostasis. It is a subject of great interest to know the function of HPX2/Duox in other major Indian malaria vectors to propose these proteins as vaccine candidates.

❖ In future, analysis of bacterial communities will explore the type of bacteria that are proliferating in HPX2/Duox silenced sugar fed or blood fed *A. stephensi* midgut. Further, these bacteria can be used in designing the promising paratransgenesis techniques to block *Plasmodium* development.

❖ It would be of great interest to know the molecular mechanisms involved in the induction of HPX14 in HPX15 silenced bacteria supplemented blood fed midguts. Further investigation may confirm the exact transcriptional regulation of HPX15 and HPX14 genes in midgut. Future studies include expression analysis of these genes in other worldwide distributed anophelines and their role in the development of *Plasmodium*.

❖ Deciphering the molecular mechanism associated with the CTCF in regulating the expression of HPX14 can be the future refinement of this thesis.

Summary of thesis

Chapter 1:   Introduction

Malaria is one of the serious diseases that have an enormous socioeconomic impact. Globally, near about 3.2 billion people (almost half of the world's population) are at the risk of malaria (WHO, 2016). The causing organism of malaria is a protozoan of the genus *Plasmodium,* which is transmitted to the vertebrate host by *Anopheles* mosquito. *Plasmodium* requires two hosts to complete its life cycle; the vertebrate host to complete the asexual life cycle and mosquito to complete the sexual life cycle. *Anopheles* female encounters *Plasmodium* during a blood meal from the malaria infected host. The gametocyte stages of *Plasmodium* fertilize in the mosquito midgut to form zygotes that transform into the motile ookinetes around 18-24 h of ingestion. These motile ookinetes traverse the midgut epithelium and reach the basal lamina where they are transformed into the oocysts. After maturation, the oocyst releases thousands of sporozoites into the hemocoel and they finally enter the salivary gland. Now the mosquito is infected and it injects these sporozoites into the vertebrate host during subsequent blood feeding (Sinden, 2002).

Numerous anti-malarial drugs are used to treat malaria and anti-malaria vaccine development is also underway. Additionally, the malaria eradication program also emphasized to introduce synthetic chemicals (called insecticides) to control the mosquito population in a natural environment. As mosquito development has 4 stages (eggs, 1st to 4th instar larvae, pupae, adult male and female) and among them, larvae remain localized to particular water bodies and thus, are considered to be the best targets for insecticides. Although, the above-said methods have been successful in many areas of the world, however, there are numerous roadblocks that prevent these classical approaches from eradicating malaria completely (Hemingway et al., 2016). The major issues in the elimination of malaria include the development of insecticide-resistant mosquitoes and drug-resistant parasites. This situation demands to develop new prevention methods of controlling malaria. One such strategy is to block *Plasmodium* development inside the mosquito host and thus, called transmission blocking strategies or transmission blocking vaccines.

Plasmodium completes its development in different body compartments of the mosquito host and midgut is the foremost compartment that encounters the parasites.

Inside the midgut, *Plasmodium* undergoes a sizeable bottleneck situation during its development (Sinden, 2002). However, numerous mosquito midgut-specific molecules also determine the successful development of *Plasmodium*. These insect molecules are called agonists and considered potent vaccine candidates for blocking *Plasmodium* development and subsequent transmission among humans (Kajla et al., 2015; Niu et al., 2017; Venkat Rao et al., 2017).

Interestingly, the mosquito midgut microbiota also plays an important role in the regulation of *Plasmodium* development. In particular, midgut bacteria either directly kill *Plasmodium* by the production of toxins or indirectly through induction of mosquito anti-plasmodial immunity (Cirimotich et al., 2011; Dong et al., 2009). Therefore, the mosquito microbiota has received great attention and new concepts of microbiota-mediated transmission blocking are currently under investigation. Recent studies revealed that heme peroxidases modulate midgut immunity in African mosquito, *Anopheles gambiae* (Kumar et al., 2010; Oliveira et al., 2012). However, the knowledge of gut immunity and immune active molecules is least understood in the major Indian malaria vector *A. stephensi*. Thus, functional studies of these heme peroxidases in *A. stephensi* may provide an opportunity to target these molecules as vaccine candidates for widespread control of malaria.

It is noteworthy to mention that heme peroxidases in eukaryotes belong to a multigene family and gene duplication is a very common event in this process. In *A. gambiae* 18 heme peroxidases are reported (Kajla et al., 2015; Neafsey et al., 2015). Thus, to strengthen the potentiality of mosquito heme peroxidases as vaccine candidates, their comparative study and functional analysis will be also very helpful in Indian mosquitoes.

Chapter 2:   Materials and Methods

➢ *A. stephensi* mosquitoes were reared in insectory at 28°C, 80% relative humidity (RH) and 12h light:dark cycle as described before (Kajla et al., 2015).

➢ Heme peroxidase (HPX) gene sequences were retrieved from the partially annotated genome of *A. stephensi* using *A. gambiae* HPX gene sequences as references. Molecular characterization of HPX gene was carried using PCR based approaches.

➢ The obtained HPX gene sequences were analyzed using various computational tools to determine their genetic structure, domain organization and transcription factors in the regulatory regions, post-translational modifications and three-dimensional structure of proteins.

- Phylogeny of *A. stephensi* heme peroxidases was analyzed using the neighbor-joining (NJ) method implemented in MEGA 5.2.

- cDNA was prepared from mosquito samples and expression of heme peroxidase and various immune genes was analyzed through Real Time PCR using ribosomal protein subunit S7 mRNA as an internal loading control for normalization. Relative mRNA levels were calculated against respective controls using $^{\Delta\Delta}$Ct method.

- The expression pattern of peroxidase genes was analyzed in different development stages such as eggs, 1st to 4th instar larvae, pupae, adult males and females.

- In addition, the role and expression kinetics of these peroxidases were also analyzed in bacteria challenged or *Plasmodium* infected *A. stephensi* female mosquitoes.

- To explore the role of heme peroxidases (HPX) in immune responses against bacteria and *Plasmodium,* gene silencing approach was adopted. For this, dsRNA of LacZ (control) or HPX (test or silenced) gene was prepared and injected into the thorax of female mosquitoes. These control and silenced mosquitoes were maintained separately on 10% sugar solution (sugar fed, SF) for four days and then further used in various experiments. To understand the effect of gene silencing on bacterial growth, gene silenced sugar-, blood- or bacteria supplemented blood-fed midguts were dissected and expression of 16S bacterial rRNA was analyzed. Similarly, to analyze the effect of gene silencing on *Plasmodium* development, gene silenced mosquitoes were infected with *P. berghei* and after seven days the midguts were dissected and oocysts (one of the development stage of *Plasmodium*) numbers were counted.

- The gene sequences of peroxidases were compared using various computational tools (JASPAR and MatInspector) to analyze their gene structure, promoter region and the presence of boundary elements. Gene duplication event was analyzed in the heme peroxidase multi-gene family of *Anopheles* using BLAST and other methods as described before (Audemard et al., 2012; Li et al., 2001).

- Statistical significance between test and respective controls was analyzed by Student's t-test or one-way ANOVA post-Tukey's Multiple Comparison Test using GraphPad Prism 5.0 software as before (Motulsky et al., 1999).

Chapter 3:   Molecular cloning and functional characterization of the heme peroxidase HPX2 from *Anopheles stephensi* and its role in immunity against blood-borne antigens

One of the *A. gambiae* heme peroxidase named as AgHPX2 is reported to involve in anti-plasmodial immunity (Oliveira et al., 2012). We characterized AgHPX2 ortholog in major Indian malaria vector *A. stephensi* and studied its role in antiplasmodial immunity. We found that *A. stephensi* HPX2 (AsHPX2) is an intron-less gene that encodes for a 692 amino acid long secreted protein. The conserved domain analysis of AsHPX2 protein revealed the presence of peroxinectin like domain. Peroxidase that contains integrin binding motif is known as peroxinectin. Mosquito peroxinectins are involved in various immune responses such as phagocytosis, encapsulation and they can also perform halogenation cycle to produce hypochlorous acid that in turn, kills bacteria. Phylogenetic analysis showed that AsHPX2 is a mosquito-specific peroxidase. It is expressed in all the developmental stages of *A. stephensi* and exhibit maximum relative expression in adult females. Blood feeding suppressed AsHPX2 gene expression, however, it was induced in exogenous bacteria supplemented blood fed midguts. Silencing of AsHPX2 gene in non-blood fed (also called sugar fed) midguts increased the load of endogenous bacteria against non-silenced controls. This effect was also observed when AsHPX2 silenced mosquitoes were fed on bacteria supplemented blood. These results revealed that AsHPX2 is one of the molecules that actively participate in maintaining bacterial homeostasis in the midgut. Interestingly, the AsHPX2 gene was induced in *P. berghei* infected midguts and its silencing enhanced the number of developing oocysts significantly when compared to the controls. Thus, we concluded that AsHPX2 is an anti-bacterial and -plasmodial gene. These findings corroborate with the previous study in African mosquito (Oliveira et al., 2012). The functional conservation of HPX2 gene in two different anophelines indicates that it can be targeted to arrest *Plasmodium* development inside the mosquito.

Chapter 4:   Functional characterization of *Anopheles stephensi* midgut Dual Oxidase gene and its role in bacterial homeostasis and *Plasmodium* development

The most important mechanism of innate immunity in insect midgut is the production of reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2) and superoxide ($O_2^{\cdot-}$) anions. ROS is mostly generated by NADPH oxidase (NOX) or dual oxidase (Duox) family of proteins (Bedard and Krause, 2007). The Duox gene is least explored in Indian malaria vector *Anopheles stephensi*. Thus, we identified and characterized *A. stephensi* Duox (AsDuox) gene, analyzed its expression and involvement in mosquito immunity against

bacteria and *Plasmodium*. Molecular characterization of Duox in *A. stephensi* genome revealed that it is a 9.45 kb long gene, which is organized into 12 exons. The AsDuox mRNA is 4428 bp long and encodes for a 1475 amino acids long protein. Domain organization analysis showed that AsDuox gene has cytoplasmic N-terminal heme peroxidase domain, a calcium-binding domain, seven transmembrane domains and a non-cytoplasmic C-terminal NADPH oxidase domain. Phylogenetic analysis revealed that AsDuox protein is highly conserved (sequence identity 97-100%) among genus *Anopheles* and is highly diverged from Human Duoxes (<40% similarity). AsDuox gene is expressed in different developmental stages of *A. stephensi* with relative maximum expression in the pupal stage. This gene is induced in midguts in response to the blood feeding. Silencing of this gene increased bacterial growth in both sugar- or blood fed midguts. Interestingly, the silenced midguts did not induce any immune response against increased growth of endogenous bacteria. It revealed that mosquito midgut has a remarkable capacity to regulate its immune signaling pathways to maintain an immune tolerant environment. Silencing of AsDuox reduced *Plasmodium* oocysts number and that was mediated through the activation of TEP1 (Thioester-containing protein 1) gene. In conclusion, these results revealed that AsDuox gene is the important molecule of midgut immunity and can be targeted to arrest *Plasmodium* development in the midgut.

Chapter 5:   CTCF, an insulator protein, regulates the divergent expression patterns of lineage-specific duplicated *Anopheles* peroxidases HPX15 and HPX14

Heme peroxidases belong to a multi-gene family and gene duplication event is a very common phenomenon in the multi-gene family. Analysis of duplication event in *Anopheles stephensi* heme peroxidases revealed that a peroxidase HPX15 has one duplicated copy named HPX14. HPX15 and HPX14 are tandemly duplicated paralogs situated in head to tail orientation on the same DNA strand. This duplicated domain is under the purifying selection and thus, might be under functional constraints. Boundary element analysis revealed that duplicated domain is flanked by the presence of boundary element and hence acts as an independent domain of gene structure. Protein sequence analysis of AsHPX15 and AsHPX14 revealed that both of them are globular secreted proteins. The expression analysis of duplicated genes in various developmental stages of *A. stephensi* revealed a differential expression pattern of these two genes with relative high mRNA levels of AsHPX15 in comparison to AsHPX14. The low level expression of HPX14 gene was found to be due to the core binding motifs of CTCF, an insulator, in its regulatory region. To explore the function of AsHPX14 gene, we silenced AsHPX15 gene and found that the

expression of this gene was induced in silenced midguts only in the presence of exogenous bacteria. However, blood alone without exogenous bacteria did not induce AsHPX14 in AsHPX15 silenced midguts. These results suggested that AsHPX14 has a role in immunity against bacteria but may not have any role in mosquito physiology. Our analysis revealed that there is no redundancy in the function of AsHPX14 and AsHPX15 and hence, strengthen the potentiality of HPX15 as a vaccine candidate.

In this thesis work, we have identified the candidate genes for transmission blocking and in future, these heme peroxidases can be targeted to develop transmission blocking strategies to control malaria.

Bibliography

Adak T, Kaur S, Singh OP. 1999. Comparative susceptibility of different members of the *Anopheles culicifacies* complex to *Plasmodium vivax*. Trans Roy Soc Trop Med Hyg 93:573-577.

Ahanger SH, Srinivasan A, Vasanthi D, Shouche YS, Mishra RK. 2013. Conserved boundary elements from the Hox complex of mosquito, *Anopheles gambiae*. Nucleic Acids Res 41(2):804–816.

Ahmed AM, Baggott SL, Maingon R, Hurd H. 2002. The costs of mounting an immune response are reflected in the reproductive fitness of the mosquito *Anopheles gambiae.* OIKOS 97:371–377.

Akagi K, Ueda H. 2011. Regulatory mechanisms of ecdysone-inducible Blimp-1 encoding a transcriptional repressor that is important for the prepupal development in *Drosophila*. Develop Growth Differ 53:697-703.

Allen RC, Stephens JT Jr. 2011. Myeloperoxidase Selectively Binds and Selectively Kills Microbes. Infect Immun 79:474-485.

Andres AJ, Fletcher JC, Karim FD, Thummel CS. 1993. Molecular analysis of the initiation of insect metamorphosis: a comparative study of *Drosophila* ecdysteroid-regulated transcription. Dev Biol 160(2):388–404.

Anh NT, Nishitani M, Harada S, Yamaguchi M, Kamei K. 2011. Essential role of Duox in stabilization of *Drosophila* wing. J Biol Chem 286(38):33244–33251.

Anvikar AR, Shah N, Dhariwal AC, Sonal GS, Pradhan MM, Ghosh SK, Valecha N. 2016. Epidemiology of *Plasmodium vivax* Malaria in India. Am J Trop Med Hyg 95(6):108–120.

Arama C, Troye-Blomberg M. 2014. The path of malaria vaccine development: challenges and perspectives. J Intern Med 275(5):456-66.

Armistead JS, Morlais I, Mathias DK, Jardim JG, Joy J, Fridman A, Finnefrock AC, Bagchi A, Plebanski M, Scorpio DG, Churcher TS, Borg NA, Sattabongkot J, Dinglasan RR. 2014. Antibodies to a single, conserved epitope in *Anopheles* APN1 inhibit universal transmission of *Plasmodium falciparum* and *Plasmodium vivax* malaria. Infect Immun 82:818–829.

Audemard E, Schiex T, Faraut T. 2012. Detecting long tandem duplications in genomic sequences. BMC Bioinformatics 13:83.

Azambuja P, Garcia ES, Ratcliffe NA. 2005. Gut microbiota and parasite transmission by insect vectors. Trends Parasitol 21:568-572.

Bae YS, Choi MK, Lee WJ. 2010. Dual oxidase in mucosal immunity and host-microbe homeostasis. Trends Immunol 31:278–287.

Bahia AC, Dong Y, Blumberg BJ, Mlambo G, Tripathi A, BenMarzouk-Hidalgo OJ, Chandra R, Dimopoulos G. 2014. Exploring *Anopheles* gut bacteria for *Plasmodium* blocking activity. Environ Microbiol 16:2980-2994.

Bahia AC, Dong Y, Blumberg BJ, Mlambo G, Tripathi A, BenMarzouk-Hidalgo OJ, Chandra R, Dimopoulos. 2014. Exploring *Anopheles* gut bacteria for *Plasmodium* blocking activity. G Environ Microbiol 16(9):2980-94.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. "MEME SUITE: tools for motif discovery and searching". Nucleic Acids Res 37:W202-W208.

Bayer CA, Holley B, Fristrom JW. 1996. A switch in broad-complex zinc-finger isoform expression is regulated posttranscriptionally during the metamorphosis of *Drosophila* imaginal discs. Dev Biol 177(1):1–14.

Beard J. 2006. DDT and human health. Sci Total Environ 355:78–89.

Beier JC. 1998. Malaria parasite development in mosquitoes. Annu Rev Entomol 43:519–43.

Beier MS. Pumpuni CB, Beier JC, Davis JR. 1994. Effects of para-aminobenzoic acid, insulin, and .gentamycin on *Plasmodium falciparum* development in Anopheline mosquitoes (Diptera: Culicidae). J Med Entomol 31:561–565.

Bergmiller T, Ackermann M, Silander OK. 2012. Patterns of Evolutionary Conservation of Essential Genes Correlate with Their Compensability. PLoS Genet 8(6):e1002803.

Billker O, Lindo V, Panico M, Etienne AE, Paxton T, Dell A, Rogers M, Sinden RE, Morris HR. 1998. Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. Nature 392:289-292.

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 115:1-7.

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

Boissière A, Tchioffo MT, Bachar D, Abate L, Marie A, Nsango SE, Shahbazkia HR, Awono-Ambene PH, Levashina EA, Christen R, Morlais I. 2012. Midgut Microbiota of the Malaria Mosquito Vector *Anopheles gambiae* and Interactions with *Plasmodium falciparum* Infection. PLoS Pathog 8: e1002742.

Brandt SM, Jaramillo-Gutierrez G, Kumar S, Barillas-Mury C, Schneider DS. 2008. Use of a *Drosophila* Model to Identify Genes Regulating *Plasmodium* Growth in the Mosquito. Genetics 180(3):1671–1678.

Broderick N, Lemaitre B. 2012. Gut-associated microbes of *Drosophila melanogaster*. Gut Microbes 3(4):307-321.

Broderick NA, Buchon N, Lemaitre B. 2014. Microbiota induced changes in *Drosophila melanogaster* host gene expression and gut morphology. MBio 5: e01117–14.

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

Butler D. 2009. Initiative targets malaria eradication. Nature 462:19.

Capo F, Charroux B, Royet J. 2016. Bacteria sensing mechanisms in *Drosophila* gut: Local and systemic consequences. Dev Comp Immunol 64:11-21.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M,Werner T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21:2933–2942.

Cerenius L, Lee BL, Söderhäll K. 2008. The proPO-system: pros and cons for its role in invertebrate immunity. Trends Immunol 29(6):263-271.

Charroux B, Royet J. 2010. *Drosophila* immune response: From systemic antimicrobial peptide production in fat body cells to local defense in the intestinal tract. Fly Austin 4(1):40-47.

Chaurio RA, Pacheco MA, Cornejo OE, Durrego E, Stanley CE Jr, Castillo AI, Herrera S, Escalante AA. 2016. Evolution of the Transmission-Blocking Vaccine Candidates Pvs28 and Pvs25 in *Plasmodium vivax*: Geographic Differentiation and Evidence of Positive Selection. PLoS Negl Trop Dis 10(6):e0004786.

Chen A, Rogan WJ. 2003. Nonmalarial infant deaths and DDT use for malaria control. Emerg Infect Dis 9:960–964.

Chen L, Zhu J, Sun G, Raikhel AS. 2004. The early gene Broad is involved in the ecdysteroid hierarchy governing vitellogenesis of the mosquito *Aedes aegypti*. J Mol Endocrinol 33:743–761.

Chen MS, Liu S, Wang H, Cheng X, El Bouhssini M, Whitworth RJ. 2016. Massive Shift in Gene Expression during Transitions between Developmental Stages of the Gall Midge, *Mayetiola Destructor*. PLoS ONE 11(5):e0155616.

Christophides GK, Zdobnov E, 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.

Cirimotich CM, Dong Y, Clayton AM, Sandiford SL, Souza-Neto JA, Mulenga M, Dimopoulos G. 2011. Natural microbe-mediated refractoriness to *Plasmodium* infection in *Anopheles gambiae*. Science 332:855-858.

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

Clayton AM, Cirimotich CM, Dong Y, Dimopoulos G. 2013. *Caudal* is a negative regulator of the *Anopheles* IMD Pathway that controls resistance to *P. falciparum* infection. Dev Comp Immunol *39*(4):323–332.

Coon KL, Brown MR, Strand MR. 2016. Gut bacteria differentially affect egg production in the anautogenous mosquito *Aedes aegypti* and facultatively autogenous mosquito *Aedes atropalpus* (Diptera: Culicidae). Parasit Vectors 9(1):375.

Cowman AF, Berry D, Baum J. 2012. The cellular and molecular basis for malaria parasite invasion of the human red blood cell. J Cell Biol 198:961–971.

Cox S, Niskar AS, Narayan KM, Marcus M. 2007. Prevalence of self-reported diabetes and exposure to organochlorine pesticides among Mexican Americans: Hispanic Health and Nutrition Examination Survey, 1982–1984. Environ Health Perspect 115:1747–1752.

Cox-Singh J, Singh B. 2008. Knowlesi malaria: newly emergent and of public health importance? Trends Parasitol 24:406-410.

Crotti E, Rizzi A, Chouaia B, Ricci I, Favia G, Alma A, Sacchi L, Bourtzis K, Mandrioli M, Cherif A, Bandi C, Daffonchio D. 2010. Acetic Acid Bacteria, Newly Emerging Symbionts of Insects. Appl Environ Microbiol 76:6963–6970.

Dash AP, Adak T, Raghavendra K, Singh OP. 2007. The biology and control of malaria vectors in India. Curr Sci 92:1571–1578.

Davies MJ. 2011. Myeloperoxidase-derived oxidation: mechanisms of biological damage and its prevention. J Clin Biochem Nutr 48(1):8-19.

Day, J. F. 2016. Mosquito Oviposition Behavior and Vector Control. Insects 7(4):65.

Dhawan R, Gupta K, kajla M, Kumar S, Gakhar SK, Kakani P, Choudhury TP, Gupta L. 2015. Molecular characterization of SOCS gene and its expression analysis on *Plasmodium berghei* infection in *Anopheles culicifacies.* Acta Trop 152:170–175.

Dimopoulos G, Christophides GK, Meister S, Schultz J, White KP, Barillas-Mury C, Kafatos FC. 2002. Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. Proc Natl Acad Sci U S A 99:8814-8819.

Dimopoulos G, Richman A, Müller HM, Kafatos FC. 1997. Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. Proc Natl Acad Sci U S A 94(21):11508-11513.

Dinglasan RR, Alaganan A, Ghosh AK, Saito A, van Kuppevelt TH, Jacobs-Lorena M. 2007. *Plasmodium falciparum* ookinetes require mosquito midgut chondroitin sulfate proteoglycans for cell invasion. Proc Natl Acad Sci U S A 104(40):15882–15887.

Dinglasan RR, Fields I, Shahabuddin M, Azad AF, Sacci Jr JB. 2003. Monoclonal antibody MG96 completely blocks *Plasmodium yoelii* development in *Anopheles stephensi*. Infect Immun 71(12):6995–7001.

Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, Lwin KM, Ariey F, Hanpithakpong W, Lee SJ, Ringwald P, Silamut K, Imwong M, Chotivanich K, Lim P, Herdman T, An SS, Yeung S, Singhasivanon P, Day NP, Lindegardh N, Socheat D, White NJ. 2009. Artemisinin resistance in *Plasmodium falciparum* malaria. N Engl J Med 361:455-467.

Dong B, Liu F, Gao H, Wang B, Xiang J. 2009b. cDNA cloning and gene expression pattern following bacterial challenge of peroxinectin in Chinese shrimp *Fenneropenaeus chinensis.* Mol Biol Rep 36:2333-2339.

Dong Y, Aguilar R, Xi Z, Warr E, Mongin E, Dimopoulos G. 2006. *Anopheles gambiae* immune responses to human and rodent *Plasmodium* parasite species. PLoS Pathog 2:e52.

Dong Y, Das S, Cirimotich C, Souza-Neto JA, McLean KJ, Dimopoulos G. 2011. Engineered *Anopheles* Immunity to *Plasmodium* Infection. PLoS Pathog 7(12): e1002458.

Dong Y, Manfredini F, Dimopoulos G. 2009a. Implication of the mosquito midgut microbiota in the defense against malaria parasites. PLoS Pathog 5:e1000423.

Donkó Á, Péterfi Z, Sum A, Leto T, Geiszt M. 2005. Dual oxidases. Philos Trans R Soc Lond B Biol Sci 360(1464):2301–2308.

Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T. 2005. Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics 21(9):2101-2103.

Edens WA, Sharling L, Cheng G, Shapira R, Kinkade JM, Lee T, Edens HA, Tang X, Sullards C, Flaherty DB, Benian GM, Lambeth JD. 2001. Tyrosine cross-linking of extracellular matrix is catalyzed by Duox, a multidomain oxidase/peroxidase with homology to the phagocyte oxidase subunit gp91phox. J Cell Biol 154(4):879-91.

Edwards YJ, Cottage A. 2003. Bioinformatics methods to predict protein structure and function. A practical approach. Mol Biotechnol 23(2):139-166.

Eriksson P, Talts U. 2000. Neonatal exposure to neurotoxic pesticides increases adult susceptibility: a review of current findings. Neurotoxicology 21:37–47.

Fan C, Chen Y, Long M. 2008. Recurrent Tandem Gene Duplication Gave Rise to Functionally Divergent Genes in *Drosophila*. Mol Biol Evol 25(7):1451–1458.

Farrance CE, Rhee A, Jones RM, Musiychuk K, Shamloul M, Sharma S, Mett V, Chichester JA, Streatfield SJ, Roeffen W, Vegte-Bolmer MV, Sauerwein RW, Tsuboi T, Muratova OV, Wu Y, Yusibov V. 2011. A Plant-Produced Pfs230 Vaccine Candidate Blocks Transmission of *Plasmodium falciparum*. Clin Vaccine Immunol 18:1351–1357.

Farre D, Alba MM. 2010. Heterogeneous patterns of gene-expression diversification in mammalian gene duplicates. Mol Biol Evol 27:325-335.

Flores MV, Crawford KC, Pullin LM, Hall CJ, Crosier KE, Crosier PS. 2010. Dual oxidase in the intestinal epithelium of zebrafish larvae has anti-bacterial properties. Biochem Biophys Res Commun 400:164–168.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

Fraiture M, Baxter RH, Steinert S, Chelliah Y, Frolet C, Quispe-Tintaya W, Hoffmann JA, Blandin SA, Levashina EA. 2009. Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of *Plasmodium*. Cell Host Microbe 5(3):273–284.

Frevert, U. 2004. Sneaking in through the back entrance: the biology of malaria liver stages. Trends Parasitol 20:417-424.

Frolet C, Thoma M, Blandin S, Hoffmann JA, Levashina EA. 2006. Boosting NF-κB-dependent basal immunity of *Anopheles gambiae* aborts development of *Plasmodium berghei*. Immunity 25:677–685.

Gaio Ade O, Gusmão DS, Santos AV, Berbert-Molina MA, Pimenta PF, Lemos FJ. 2011. Contribution of midgut bacteria to blood digestion and egg production in *Aedes aegypti* (diptera: culicidae) (L.). Parasit Vectors 4:105.

Garcia GE, Wirtz RA, Barr JR, Woolfitt A, Rosenberg R. 1998. Xanthurenic acid induces gametogenesis in *Plasmodium*, the malaria parasite. J Biol Chem 273:12003-12005.

Garver LS, Bahia AC, Das S, Souza-Neto JA, Shiao J, Dong Y, Dimopoulos G. 2012. *Anopheles* Imd pathway factors and effectors in infection intensity-dependent anti-*Plasmodium* action. PLoS Pathog 8:e1002737.

Garver LS, de Almeida Oliveira G, Barillas-Mury C. 2013. The JNK Pathway Is a Key Mediator of *Anopheles gambiae* Antiplasmodial Immunity. PLoS Pathogens 9(9): e1003622.

Garver LS, Dong Y, Dimopoulos G. 2009. Caspar controls resistance to *Plasmodium falciparum* in diverse anopheline species. PLoS Pathog 5:e1000335.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server; In: John M. Walker, editor. The Proteomics Protocols Handbook, Humana Press pp. 571-607.

Ghosh A, Chowdhury N, Chandra G. 2012. Plant extracts as potential mosquito larvicides. Indian J Med Res 135(5):581–598.

Giannoni F, Müller HM, Vizioli J, Catteruccia F, Kafatos FC, Crisanti AJ. 2001. Nuclear factors bind to a conserved DNA element that modulates transcription of *Anopheles gambiae* trypsin genes. Biol Chem 276(1):700-7.

Gillies MT, De Meillon B. 1968. The Anophelinae of Africa south of the Sahara (Ethiopian Zoogeographical Region). 2nd edn. South African Institute for Medical Research, Johannesburg. Publications of the South African Institute for Medical Research no. 54.

Gonzalez-Lazaro M, Dinglasan RR, Hernandez-Hernandez Fde L, Rodriguez MH, Laclaustra M, Jacobs-Lorena M, Flores-Romo L. 2009. *Anopheles gambiae* Croquemort SCRBQ2, expression profile in the mosquito and its potential interaction with the malaria parasite *Plasmodium berghei*. Insect Biochem Mol Biol 39(5-6):395–402.

Gray CE, Coates CJ. 2005. Cloning and characterization of cDNAs encoding putative CTCFs in the mosquitoes, *Aedes aegypti* and *Anopheles gambiae*. BMC Mol Biol 6:16.

Gu Z, Nicolae D, Lu HH, Li WH. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet 18:609–613.

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.

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. Genome Biol 8(2):R24.

Ha EM, Lee KA, Park SH, Kim SH, Nam HJ, Lee HY, Kang D, Lee WJ. 2009a. Regulation of DUOX by the Gaq-phospholipase Cb-Ca2+ pathway in *Drosophila* gut immunity. Dev Cell 16:386–397

Ha EM, Lee KA, Seo YY, Kim SH, Lim JH, Oh BH, Kim J, Lee WJ. 2009b. Coordination of multiple dual oxidase-regulatory pathways in responses to commensal and infectious microbes in *Drosophila* gut. Nat Immunol 10:949–957.

Ha EM, Oh CT, Bae YS, Lee WJ. 2005a. A direct role for dual oxidase in *Drosophila* gut immunity. Science 310:847–850.

Ha EM, Oh CT, Ryu JH, Bae YS, Kang SW. 2005b. An antioxidant system required for host protection against gut infection in *Drosophila*. Dev Cell 8:125–132.

Hagedorn HH, O'Connor JD, Fuchs MS, Sage B, Schlaeger DA andBohm MK. 1975. The ovary as a source of α-ecdysone in an adult mosquito. Proc Natl Acad Sci U S A 72:3255–3259.

Han ZS, Ip YT. 1999. Interaction and specificity of Rel-related proteins inregulating *Drosophila* immunity gene expression. J Biol Chem 274:21355–21361.

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

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

Hillyer JF, Barreau C, Vernick KD. 2007. Efficiency of salivary gland invasion by malaria sporozoites is controlled by rapid sporozoite destruction in the mosquito haemocoel. Int J Parasitol 37:673-681.

Hillyer JF. 2010. Mosquito immunity. Adv Exp Med Biol 708:218-38.

Hodgetts RB, Clark WC, O'Keefe SL, Schouls M, Crossgrove K, Guild GM, von Kalm L. 1995. Hormonal induction of dopa decarboxylase in the epidermis of Drosophila is mediated by the Broad-Complex. Development 121:3913–3922

Hu X, Yang R, Zhang X, Chen L, Xiang X, Gong C, Wu X. 2013. Molecular Cloning and Functional Characterization of the Dual Oxidase (BmDuox) Gene from the Silkworm *Bombyx mori*. PLoS ONE 8(8): e70118.

Huerta-Cepas J, Dopazo H, Dopazo J, Gabaldón T. 2007. The human phylome. Genome Biol 8:R109.

Hughes AL. 2012. Evolution of the heme peroxisases of Culicidae (Diptera). Psyche 146387. Doi:10.1155/2012/146387.

Huminiecki L, Wolfe KH. 2004. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. Genome Res 14(10A):1870-9.

Hurd TR, Liang FX, Lehmann R. 2015. Curly Encodes Dual Oxidase, Which Acts with Heme Peroxidase Curly Su to Shape the Adult *Drosophila* Wing. PLoS Genetics 11(11):e1005625.

Hurles M. 2004. Gene Duplication: The Genomic Trade in Spare Parts. PLoS Biology 2(7):0900.

Inada M, Kihara K, Kono T, Sudhakaran R, Mekata T, Sakai M, Yoshida T, Itami T. 2013. Deciphering of the Dual oxidase (Nox family) gene from kuruma shrimp, *Marsupenaeus japonicus*: Full-length cDNA cloning and characterization. Fish Shellfish Immunol 34:471e485.

Ishii K, Laemmli UK. 2003. Structural and dynamic functions pore complex via a conserved interaction with CAN/Nup159p. establish chromatin domains. Mol Cell 11:237–248.

Jacobs-Lorena M, Oo MM. 1996. The peritrophic matrix of insects. In: Beaty BJ, Marquardt WC, editors.The Biology of Disease Vectors. Niwot, CO:University Press of Colorado. p 318–332.

Jaramillo-Gutierrez G, Rodrigues J, Ndikuyeze G, Povelones M, Molina-Cruz A, Barillas-Mury C. 2009. Mosquito immune responses and compatibility between *Plasmodium* parasites and anopheline mosquitoes. BMC Microbiol 9:154-164.

Jiang SY, González JM, Ramachandran S. 2013. Comparative Genomic and Transcriptomic Analysis of Tandemly and Segmentally Duplicated Genes in Rice. PLoS One 8(5):e63551.

Jinwal UK, Zakharkin SO, Litvinova OV, Jain S, Benes H. 2006. Sex-, stage- and tissue-specific regulation by a mosquito hexamerin promoter. Insect Mol Biol 15(3):301–311.

Jiravanichpaisal P, Lee BL, Söderhäll K. 2006. Cell-mediated immunity in arthropods: hematopoiesis, coagulation, melanization and opsonization. Immunobiology 211(4):213-36.

Johanson, Michael. 2013. An Investigation of Insulator Proteins in Mosquito Genomes. Master's thesis, Texas A & M University. Available electronically from http : / /hdl .handle.net /1969 .1 /151362.

Johansson MW. 1999. Cell adhesion molecules in invertebrate immunity. Dev Comp Immunol 23(4-5):303-315.

Kajla M, Bhattacharya K, Gupta K, Banerjee U, Kakani P, Gupta L and Kumar S. 2016a. Identification of temperature Induced Larvicidal Efficacy of *Agave angustifolia* against *Aedes*, *Culex* and *Anopheles* Larvae. Front Public Health 3:286.

Kajla M, Choudhury TP, Kakani P, Gupta K, Dhawan R, Gupta L and Kumar S. 2016c. Silencing of *Anopheles stephensi* Heme Peroxidase HPX15 Activates Diverse Immune Pathways to Regulate the Growth of Midgut Bacteria. Front Microbiol 7:1351.

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

Kajla M, Gupta K, Kakani P, Dhawan R, Choudhury TP, Gupta L, Gakhar SK, Kumar S. 2015a. Identification of an *Anopheles* Lineage-Specific Unique Heme Peroxidase HPX15: A plausible candidate for arresting malaria parasite development. J Phylogen Evolution Biol 3:160.

Kajla M, Kakani P, Choudhury TP, Kumar V, Gupta K, Dhawan R, Gupta L, Kumar S. 2017. *Anopheles stephensi* Heme Peroxidase HPX15 Suppresses Midgut Immunity to Support *Plasmodium* Development. Front Immunol 8:249.

Kajla M, Kakani P, Gupta K, Choudhury TP, Gupta L, Kumar S. 2016b. Characterization and expression analysis of gene encoding hemeperoxidase HPX15 in major Indian malaria vector *Anopheles stephensi* (Diptera: Culicidae). Acta Tropica 158:107–116.

Kakani P, Suman S, Gupta L, Kumar S. 2016. Ambivalent outcomes of cell apoptosis: A barrier or blessing in malaria progression. Front Microbiol 7:302.

Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. Nucleic Acids Res 35:W429-32.

Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. 2012. Template-based protein structure modeling using the RaptorX web server. Nature Protocols 7:1511–1522.

Kaushik JS, Gomber S, Dewan P. 2012. Clinical and Epidemiological Profiles of Severe Malaria in Children from Delhi, India. J Health Popul Nutr 30(1):113-116.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modelling, prediction and analysis. Nature protocols 10(6):845-858.

Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128(6):1231-1245.

Klebanoff SJ. 1968. Myeloperoxidase-halide-hydrogen peroxide antibacterial system. J Bacteriol 95:2131–2138.

Klein EY. 2013. Antimalarial drug resistance: a review of the biology and strategies to delay emergence and spread. Int J Antimicrob Agents 41:311– 317.

Kotsyfakis M, Ehret-Sabatier L, Siden-Kiamos I, Mendoza J, Sinden RE, Louis C. 2005. *Plasmodium berghei* ookinetes bind to *Anopheles gambiae* and *Drosophila melanogaster* annexins. Mol Microbiol 57(1):171–9.

Kovendan K, Murugan K, Shanthakumar SP, Vincent S, Hwang JS. 2012. Larvicidal activity of *Morinda citrifolia* L. (Noni) (Family: Rubiaceae) leaf extract against *Anopheles stephensi*, *Culex quinquefasciatus* and *Aedes aegypti*. *Parasitol Res* 111:1481-1490.

Kuehn A, Pradel G. 2010. The coming-out of malaria gametocytes. J Biomed Biotechnol, 2010:976827.

Kumar S, Christophides GK, Cantera R, Charles B, Han YS, Meister S, Dimopoulos G, Kafatos FC, Barillas-Mury C. 2003. The role of reactive oxygen species on *Plasmodium* melanotic encapsulation in *Anopheles gambiae*. Proc Natl Acad Sci U. S. A. 100:14139–14144.

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 279:53475–53482.

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.

Lefevre P, Witham J, Lacroix CE, Cockerill PN, Bonifer C. 2008. The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Mol Cell 32: 129–139

Lehane MJ, Aksoy S, Levashina E. 2004. Immune responses and parasite transmission in blood-feeding insects. Trends Parasitol 20:433–439.

Lehane MJ. Peritrophic matrix structure and function. Annu Rev Entomol 1997;42: 525–550

Lemaitre B, Kromer-Metzger E, Michaut L, Nicolas E, Meister M, Georgel P, Reichhart JM, Hoffmann JA. 1995. A recessive mutation, immune deficiency (imd), defines two distinct control pathways in the *Drosophila* host defense. Proc Natl Acad Sci U. S. A. 92:9465–9469.

Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. 1996. The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. Cell 86:973-983.

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

Letunic I, Doerks T, Bork P. 2014. SMART: recent updates, new developments and status in 2015 Nucleic Acids Res doi:10.1093/nar/gku949.

Leulier F, Royet J. 2009. Maintaining immune homeostasis in fly gut. Nat Immunol 10(9):936-938.

Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124:837–848.

Li WH, Gu Z, Wang H, Nekrutenko A. 2001. Evolutionary analyses of the human genome. Nature 409:847–849.

Li, WH. 1997. Molecular Evolution, Sinauer Associates, Sunderland, Massachusetts

Lindh JM, Terenius O, Faye I. 2005. 16S rRNA gene-based identification of midgut bac-teria from field-caught *Anopheles gambiae* sensu lato and *A. funestus* mosquitoes reveals new species related to known insect symbionts. Appl Environ Microbiol 71:7217-7223.

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.

Logue K, Keven JB, Cannon MV, Reimer L, Siba P, Walker ED, Zimmerman PA, Serre D. 2016. Unbiased Characterization of *Anopheles* Mosquito Blood Meals by Targeted High-Throughput Sequencing. PLoS Negl Trop Dis 10(3):e0004512.

Lowenberger C, Charlet M, Vizioli J, Kamal S, Richman A, Christensen BM, Bulet P. 1999. Antimicrobial activity spectrum, cDNA cloning and mRNA expression of a newly isolated member of the cecropin family from the mosquito vector *Aedes aegypti*. J Biol Chem 274(29):20092-20097.

Lowenberger C, Bulet P, Charlet M, Hetru C, Hodgeman B, Christensen BM, Hoffmann JA. 1995. Insect immunity: isolation of three novel inducible antibacterial defensins from the vector mosquito, *Aedes aegypti*. Insect Biochem Mol Biol 25(7):867-873.

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 U S A 95:5700–5705.

Luna C, Hoa NT, Lin H, Zhang L, Nguyen HL, Kanzok SM, Zheng L. 2006. Expression of immune responsive genes in cell lines from two different Anopheline species. Insect Mol Biol 15:721-729.

Ma Z, Li M, Roy S, Liu KJ, Romine ML, Lane DC, Patel SK, Cai HN. 2016. Chromatin boundary elements organize genomic architecture and developmental gene regulation in *Drosophila Hox* clusters. World J Biol Chem 7(3):223-230.

Makalowski W, Boguski MS. 1998. Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. Proc Natl Acad Sci U S A 95:9407–9412.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43:D222-226.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J. 2013. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res 42:D142–D147.

Mathias DK, Jardim JG, Parish LA, Armistead JS, Trinh HV, Kumpitak C, Sattabongkot J, Dinglasan RR. 2014. Differential roles of an Anopheline midgut GPI-anchored protein in mediating *Plasmodium falciparum* and *Plasmodium vivax* ookinete invasion. Infect Genet Evol 28:635–47.

Mehlotra RK, Lorry K, Kastens W, Miller SM, Alpers MP, Bockarie M, Kazura JW, Zimmerman PA. 2000. Random distribution of mixed species malaria infections in Papua New Guinea. Am J Trop Med Hyg 62:225-231.

Meister S, Kanzok SM, Zheng XL, Luna C, Li TR, Hoa NT, Clayton JR, White KP, Kafatos FC, Christophides GK, Zheng L. 2005. Immune signaling pathways regulating bacterial and malaria parasite infection of the mosquito *Anopheles gambiae*. Proc Natl Acad Sci U S A

102:11420-11425.

Michel T, Reichhart JM, Hoffmann JA, Royet, J. 2001. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature 414:756-759.

Mier P, Andrade-Navarro MA, Pérez-Pulido AJ. 2015. orthoFind Facilitates the Discovery of Homologous and Orthologous Proteins. PLoS ONE 10(12):e0143906.

Minard G, Mavingui P, Moro CV. 2013. Diversity and function of bacterial microbiota in the mosquito holobiont. Parasit Vectors 6:146.

Molina-Cruz A, Canepa GE, Kamath N, Pavlovic NV, Mu J, Ramphul UN, Ramirez JL, Barillas-Mury C. 2015. *Plasmodium* evasion of mosquito immunity and global malaria ransmission: The lock-and-key theory. Proc Natl Acad Sci U S A 112(49):15178-83.

Molina-Cruz A, DeJong RJ, Charles B, Gupta L, Kumar S, Jaramillo-Gutierrez G, Barillas-Mury C. 2008. Reactive oxygen species modulate *Anopheles gambiae* immunity against bacteria and *Plasmodium*. J Biol Chem 283:3217–3223.

Molina-Cruz A, DeJong RJ, Ortega C, Haile A, Abban E, Rodrigues J, Jaramillo-Gutierrez G, Barillas-Mury C. 2012. Some strains of *Plasmodium falciparum*, a human malaria parasite, evade the complement-like system of *Anopheles gambiae* mosquitoes. Proc Natl Acad Sci U S A 109: E1975-62.

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

Mrema EJ, Rubino FM, Brambilla G, Moretto A, Tsatsakis AM, Colosio C. 2013. Persistent organochlorinated pesticides and mechanisms of their toxicity. Toxicol 307:74-88.

Müller HM, Catteruccia F, Vizioli J, della Torre A, Crisanti A. 1995. Constitutive and Blood Meal-Induced Trypsin Genes in *Anopheles gambiae.* Exp Parasitol 81:371–385.

Muyskens JB, Guillemin K. 2008. Bugs inside Bugs: what the fruit fly can teach us about immune and microbial balance in the gut. Cell Host Microbe 3(3):117-118.

Nagpal BN, Sharma VP. 1995. Indian Anophelines. New Delhi:Oxford & IBH Publishing Co. Pvt. LTD., p. 416.

Nakahashi H, Kwon KR, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, Qian J, Dubois W, Welsh S, Phair RD, Pugh BF, Lobanenkov V, Hager GL, Casellas R. 2013. A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. Cell Reports 3:1678–1689.

Narasimha M, Uv A, Krejci A, Brown NH, Bray SJ. 2008. Grainy head promotes expression of septate junction proteins and influences epithelial morphogenesis. J Cell Sci 121:747–752.

National Vector Borne Disease Control Program, (2010- 2014).Malaria situation in India. Available from: http://www. nvbdcp.gov.in/Doc/malaria-situation-March14. pdf, accessed on March 11, 2014.

Nauseef WM, Petrides PE. 1999. Peroxidases and human disease: a meeting of minds. The peroxidase multigene family of enzymes: biochemical basis and clinical applications. Fraueninsel, Germany, 27 September-2 October 1998. Mol Med Today 5(2):58-60.

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

Nekrutenko A, Makova KD, Li W-H. 2002. The K_A/K_S Ratio Test for Assessing the Protein-Coding Potential of Genomic Regions: An Empirical and Simulation Study. Genome Res 12(1):198-202.

Nijhout HF, Riddiford LM, Mirth C, Shingleton AW, Suzuki Y, Callier V. 2014. The Developmental Control of Size in Insects. Wiley Interdiscip Rev Dev Biol 3(1):113-134.

Niu G, Franca C, Zhang G, Roobsoong W, Nguitragool W, Wang X, Prachumsri J, Butler NS, Li J. 2017. Fibrinogen domain of FREP1 is a broad spectrum malaria transmission-blocking vaccine antigen. J Biol Chem 292(28):11960-11969.

Noden BH, Vaughan JA, Pumpuni CB, Beier JC. 2011. Mosquito ingestion of antibodies against mosquito midgut microbiota improves conversion of ookinetes to oocysts for *Plasmodium falciparum*, but not *P. yoelii*. Parasitol Int 60:440-6.

Noedl H, Se Y, Sriwichai S, Schaecher K, Teja-Isavadharm P, Smith B, Rutvisuttinunt W, Bethell D, Surasri S, Fukuda MM, Socheat D, Chan Thap L. 2010. Artemisinin resistance in Cambodia: a clinical trial designed to address an emerging problem in Southeast Asia. Clin Infect Dis 51:e82-89.

Oliveira GA, Lieberman J, Barillas-Mury C. 2012. Epithelial nitration by a peroxidase/NOX5 system mediates mosquito antiplasmodial immunity. Science 335(6070):856-9.

Oliveira JH, Gonçalves RL, Lara FA, Dias FA, Gandara AC, Menna-Barreto RF, Edwards MC, Laurindo FR, Silva-Neto MA, Sorgine MH, Oliveira PL. 2011. Blood Meal-Derived Heme Decreases ROS Levels in the Midgut of *Aedes aegypti* and Allows Proliferation of Intestinal Microbiota. PLoS Pathog 7(3): e1001320.

Osta MA, Christophides GK, Vlachou D, Kafatos FC. 2004. Innate immunity in the malaria vector *Anopheles gambiae*: comparative and functional genomics. J Exp Biol 207:2551-2563.

Pace CN, Scholtz JM. 1998. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 75:422-427.

Page AP, Johnstone IL. The cuticle. 2007 May 3. In: WormBook: The Online Review of C. elegans Biology [Internet]. Pasadena (CA): WormBook; 2005-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK19745.

Pankova MV, Brykov VA, Pankova VV, Atopkin DM. 2013. Fish Growth Hormone Genes: Divergence of Intron Sequence in Charrs of *Salvelinus* Genus. Genetika 49(6):743-750.

Passardi F, Bakalovic N, Teixeira FK, Margis-Pinheiro M, Penel C, Dunand C. 2007. Prokaryotic origins of the non-animal peroxidase superfamily and organelle-mediated transmission to eukaryotes. Genomics 89:567-579.

Patil V. 2012. Complicated falciparum Malaria in western Maharashtra. Trop Parasitol 2(1):49-54.

Pegueroles C, Laurie S, Albà MM. 2013. Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. Mol Biol Evol 30(8):1830-1842.

Persson T, Andersson P, Bodelsson M, Laurell M, Malm J, Egesten A. 2001. Bactericidal activity of human eosinophilic granulocytes against *Escherichia coli*. Infect Immun 69:3591–3596.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786.

Phillips JE, Corces VG. 2009. CTCF: Master Weaver of the Genome. Cell 137(7):1194-1211

Phillips RS. 2001. Current status of malaria and potential for control. Clin Microbiol Rev 14:208-26.

Phyo AP, Nkhoma S, Stepniewska K, Ashley EA, Nair S, McGready R, ler Moo C, Al-Saai S, Dondorp AM, Lwin KM, Singhasivanon P, Day NP, White NJ, Anderson TJ, Nosten F. 2012. Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. Lancet 379:1960-1966.

Povelones M, Waterhouse RM, Kafatos FC, Christophides GK. 2009. Leucine-rich repeat protein complex activates mosquito complement in defense against *Plasmodium* parasites. Science 324:258–261.

Pradel G. 2007. Proteins of the malaria parasite sexual stages: expression, function and potential for transmission blocking strategies. Parasitology 134:1911-1929.

Pumpuni CB, Beier MS, Nataro JP, Guers LD, Davis JR. 1993. *Plasmodium falciparum*: inhibition of sporogonic development in *Anopheles stephensi* by Gram-negative bacteria. Exp Parasitol 77:195–199.

Pumpuni CB, DeMaio J, Kent M, Davis JR, Beier JC. 1996. Bacterial population dynamics in three anopheline species: the impact on *Plasmodium* sporogonic development. Am J Trop Med Hyg 54:214–218.

Ramirez JL, Dimopoulos G. 2010. The Toll immune signaling pathway controls conserved anti-dengue defenses across diverse *Aedes aegypti* strains and against multiple dengue virus serotypes. Dev Comp Immunol 34:625-629.

Ramirez JL, Garver LS, Dimopoulos G. 2009. Challenges and Approaches for Mosquito Targeted Malaria Control. Curr Mol Med 9(2):116-130.

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 U S A 112(5):1273–1280.

Ramsey PG, Martin T, Chi E, Klebanoff SJ. 1982. Arming of mononuclear phagocytes by eosinophil peroxidase bound to *Staphylococcus aureus*. J Immunol 128: 415–420.

Raz A, Dinparast Djadid N, Zakeri S. 2013. Molecular Characterization of the Carboxypeptidase B1 of *Anopheles stephensi* and Its Evaluation as a Target for Transmission-Blocking Vaccines. Infection and Immunity 81(6):2206–2216.

Reese MG. 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comput Chem 26:51-56.

Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, Pedone PV. 2007. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at

imprinted loci. J Biol Chem 282(46):33336-33345.

Riehle MM, Xu J, Lazzaro BP, Rottschaefer SM, Coulibaly B, Sacko M, Niare O, Morlais I, Traore SF, Vernick KD. 2008. *Anopheles gambiae* APL1 is a family of variable LRR proteins required for Rel1-mediated protection from the malaria parasite, *Plasmodium berghei*. PLoS One 3:e3672.

Roth C, Rastogi S, Arvestad L, Dittmar K, Light S, Ekman D, Liberles DA. 2007. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. J Exp Zool Mol Dev Evol 308:58–73.

Royet J. 2011. Epithelial homeostasis and the underlying molecular mechanisms in the gut of the insect model *Drosophila melanogaster*. Cell Mol Life Sci 68(22):3651-3660.

RTS,S Clinical Trials Partnership. 2015. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. Lancet 386(9988):31-45.

Ruoslahti E. 1996. RGD AND OTHER RECOGNITION SEQUENCES FOR INTEGRINS. Annu Rev Cell Dev Biol 12:697-715.

Ryu JH, Kim SH, Lee HY, Bai JY, Nam YD, Bae JW, Lee DG, Shin SC, Ha EM, Lee WJ. 2008. Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. Science 319:777–82.

Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007. Dynamic evolution of the innate immune system in *Drosophila*. Nat Genet 39:1461–1468.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

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 21:4147.

Schoborg TA, Labrador M. 2010. The phylogenetic distribution of non-CTCF insulator proteins is limited to insects and reveals that BEAF-32 is *Drosophila* lineage specific. J Mol Evol 70:74–84.

Senger K, Harris K, Levine M. 2006. GATA factors participate in tissue-specific immune responses in *Drosophila* larvae. Proc Natl Acad Sci U S A 103(43):15957–15962.

Seufi AM, Hafez EE, Galal FH. 2011. Identification, phylogenetic analysis and expression profile of an anionic insect defensin gene, with antibacterial activity, from bacterial-challenged cotton leafworm, *Spodoptera littoralis.* BMC Mol Biol 12:47.

Sharma VP. 1999. Current scenario of malaria in India. Parassitologia 41:349–353.

Shin SC, Kim SH, You H, Kim B, Kim AC, Lee KA, Yoon JH, Ryu JH, Lee WJ. 2011. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. Science 334:670–674.

Sinden RE. 2002. Molecular interactions between *Plasmodium* and its insect vectors. Cell Microbiol 4:713–724.

Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, Mbogo CM, Hemingway J, Patil AP, Temperley WH Gething PW, Kabaria CW, Burkot TR, Harbach RE, Hay SI. 2012. A global map of dominant malaria vectors. Parasit Vectors 5:69.

Smith Gueye C, Newby G, Gosling RD, Whittaker MA, Chandramohan D, Slutsker L, Tanner M. 2016. Strategies and approaches to vector control in nine malaria-eliminating countries: a cross-case study analysis. Malar J 15:2.

Smith RC, Barillas-Mury C, Jacobs-Lorena M. 2015. Hemocyte differentiation mediates the mosquito late-phase immune response against *Plasmodium* in *Anopheles gambiae*. Proc Natl Acad Sci U S A 112:3412-3420.

Smith RC, Eappen AG, Radtke AJ, Jacobs-Lorena M. 2012. Regulation of anti-*Plasmodium* immunity by a LITAF-like transcription factor in the malaria vector *Anopheles gambiae*. PLoS Pathog 8: e1002965.

Smith RC, Rodríguez JV, Jacobs-Lorena M. 2014. The *Plasmodium* bottleneck: malaria parasite losses in the mosquito vector. Mem Inst Oswaldo Cruz 109:644-661.

Srinivasan A, Mishra RK. 2012. Chromatin domain boundary element search tool for *Drosophila*. Nucleic Acids Res 40(10):4385-4395.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24(5):637-644.

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res 35: 506-511;

Stokes BA, Yadav S, Shokal U, Smith LC, Eleftherianos I. 2015. Bacterial and fungal pattern recognition receptors in homologous innate signaling pathways of insects and mammals. Front Microbiol 6:19.

Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F. 2011. *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. Cell Metabol 14: 403–414.

Sumimoto H. 2008. Structure, regulation and evolution of Nox-family NADPH oxidases that produce reactive oxygen species. FEBS J 275: 3249–3277.

Sundaramoorthy M, Terner J, Poulos TL. 1995. The crystal structure of chloroperoxidase: a heme peroxidase--cytochrome P450 functional hybrid. Structure 3(12):1367-77.

Swami KK, Srivastava M. 2012. Blood meal preference of some anopheline mosquitoes in command and non-command areas of Rajasthan, India. J Arthropod Borne Dis 6(2):98–103.

Swevers L, Iatrou K. 2003. The ecdysone regulatory cascade and ovarian development in lepidopteran insects: insights from the silkmoth paradigm. Insect Biochem Mol Biol 33:1285–1297.

Ta TH, Hisam S, Lanza M, Jiram AI, Ismail N, Rubio JM. 2014. First case of a naturally acquired human infection with *Plasmodium cynomolgi*. Malar J 13:68.

Tachida H, Kuboyama T. 1998. Evolution of multigene families by gene duplication. A haploid model. Genetics 149(4):2147–2158.

Tajiri R, Ogawa N, Fujiwara H, Kojima T. 2017. Mechanical Control of Whole Body Shape by a Single Cuticular Protein Obstructor-E in *Drosophila melanogaster*. Desplan C, ed. *PLoS Genetics*13(1):e1006548.

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.

Tang T, Li X, Yang X, Yu X, Wang J, Liu F, Huang D. 2014. Transcriptional response of *Musca domestica* larvae to bacterial infection. PLoS One 9(8):e104867.

Tanji T, Hu X, Weber AN, Ip YT. 2007. Toll and IMD Pathways Synergistically Activate an Innate Immune Response in *Drosophila melanogaster*. Mol Cell Biol 27(12):4578–4588.

Telang A, Frame L, Brown MR. 2007. Larval feeding duration affects ecdysteroid levels and nutritional reserves regulating pupal commitment in the yellow fever mosquito *Aedes aegypti* (Diptera: Culicidae). J Exp Biol 210:854-864.

Terra WR. 2001. The origin and functions of the insect peritrophic membrane and peritrophic gel. Arch Insect Biochem Physiol 47:47–61.

The RTS,S Clinical Trials Partnership. 2014. 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.

Thein MC, Winter AD, Stepek G, McCormack G, Stapleton G, Johnstone IL, Page AP. 2009. Combined Extracellular Matrix Cross-linking Activity of the Peroxidase MLT-7 and the Dual Oxidase BLI-3 Is Critical for Post-embryonic Viability in *Caenorhabditis elegans*. J Cell Biol 284(26):17549–17563.

Torres MA, Dangl JL. 2005. Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. Curr Opin Plant Biol 8:397–403.

Tremaroli V, Bäckhed F. 2012. Functional interactions between the gut microbiota and host metabolism. Nature 489: 242–249.

Tzou P, Ohresser S, Ferrandon D, Capovilla M, Reichhart JM, Lemaitre B, Hoffmann JA, Imler JL. 2000. Tissue-specific inducible expression of antimicrobial peptide genes in *Drosophila* surface epithelia. Immunity 13(5):737-48.

Van der Hoeven R, Cruz MR, Chávez V, Garsin DA. 2015. Localization of the Dual Oxidase BLI-3 and Characterization of Its NADPH Oxidase Domain during Infection of Caenorhabditis elegans. Lemos JA, ed. PLoS ONE 10(4):e0124091.

Vega-Rodríguez J, Ghosh AK, Kanzok SM, Dinglasan RR, Wang S, Bongio NJ, Kalume DE, Miura K, Long CA, Pandey A, Jacobs-Lorena M. 2014. Multiple pathways for *Plasmodium ookinete* invasion of the mosquito midgut. Proc Natl Acad Sci U S A 111(4):E492–500.

Venema K. 2010. Role of gut microbiota in the control of energy and carbohydrate metabolism. Curr Opin Clin Nutr Metab Care 13: 432–438.

Venkat Rao V, Kumar SK, Sridevi P, Muley VY, Chaitanya RK. 2017. Cloning, characterization and transmission blocking potential of midgut carboxypeptidase A in *Anopheles stephensi*. Acta Trop 168:21-28.

Vijay S, Rawat M, Adak T, Dixit R, Nanda N, Srivastava H, Sharma JK, Prasad GB, Sharma A. 2011. Parasite killing in malaria non-vector mosquito *Anopheles culicifacies* species B: implication of nitric oxide synthase upregulation. PloS one 6:e18400.

Viljakainen L. 2015. Evolutionary genetics of insect innate immunity. Brief Funct Genomics 14(6):407-412.

Vizioli J, Bulet P, Hoffmann JA, Kafatos FC, Müller HM, Dimopoulos G. 2001. Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae.* Proc Natl Acad Sci U S A 98:12630-12635.

Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J, Spellman P, Boccard F, Lemaitre B. 2005. *Drosophila* host defense after oral infection by an entomopathogenic *Pseudomonas* species. Proc Natl Acad Sci U S A 102(32):11414-11419.

Wagner A. 2000. Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist selectionist debate. Proc Natl Acad Sci U S A 97:6579–6584.

Walsh J B, Stephan W. 2008. Multigene Families: Evolution. In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net [doi: 10.1002/9780470015902.a0001702.pub2.

Wang J, Marowsky NC, Fan C. 2013. Divergent Evolutionary and Expression Patterns between Lineage Specific New Duplicate Genes and Their Parental Paralogs in *Arabidopsis thaliana*. PLOS ONE 8(8): e72362.

Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, Barillas-Mury C. 2007. Evolutionary Dynamics of Immune-Related Genes and Pathways in Disease-Vector Mosquitoes. Science. 316:1738-1743.

Weiss BL, Wang J, Aksoy S. 2011. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. PLoS Biol 9: e1000619.

White NJ. 2004. Antimalarial drug resistance. J Clin Invest 113:1084–1092.

White NJ. 2008. *Plasmodium knowlesi*: the fifth human malaria parasite. Clin Infect Dis 46:172–173.

WHO. 2011: "World malaria report: 2011," in WHO Global Malaria Programme. (Geneva: World Health Organization).

WHO. 2016: "World malaria report: 2016," in WHO Global Malaria Programme. (Geneva: World Health Organization).

William T, Rahman HA, Jelip J, Ibrahim MY, Menon J, Grigg MJ, Yeo TW, Anstey NM, Barber BE.. 2013. Increasing incidence of *Plasmodium knowlesi* malaria following control of *P. falciparum* and *P. vivax* malaria in Sabah, Malaysia. PLoS Negl Trop Dis 7(1):e2026.

Wu DD, Wang GD, Irwin DM, Zhang YP. 2009. A Profound Role for the Expansion of Trypsin-Like Serine Protease Family in the Evolution of Hematophagy in Mosquito. Mol Biol Evol 26(10):2333–2341.

Xiao X, Yang L, Pang X, Zhang R, Zhu Y, Wang P, Gao G, Cheng G. 2017. A Mesh-Duox pathway regulates homeostasis in the insect gut. Nat Microbiol 2:17020.

Xu G, Guo C, Shan H, Kong H. 2012. Divergence of duplicate genes in exon–intron structure. Proc Natl Acad Sci U S A 109:1187–1192.

Xu JR, Zhang Y. 2010. How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics 26(7):889–95.

Yang HT, Yang MC, Sun JJ, Shi XZ, Zhao XF, Wang JX. 2016. Dual oxidases participate in the regulation of intestinal microbiotic homeostasis in the kuruma shrimp *Marsupenaeus japonicus*. Dev Comp Immunol 59:153-63.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: Protein structure and function prediction. Nat Methods 12(1):7-8.

Yao Z, Ailin Wang A, Yushan Li Y, Zhaohui Cai Z, Bruno Lemaitre B and Hongyu Zhang H. 2016. The dual oxidase gene BdDuox regulates the intestinal bacterial community homeostasis of *Bactrocera dorsalis*. The ISME Journal 10:1037–1050.

Zámocký M, Hofbauer S, Schaffner I, Gasselhuber B, Nicolussi A, Soudi M, Pirker KF, Furtmüller PG, Obinger C. 2015. Independent evolution of four heme peroxidase superfamilies. Arch Biochem Biophys 574:108-119.

Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, Copley RR, Christophides GK et al., 2002. Comparative Genome and Proteome Analysis of *Anopheles gambiae* and *Drosophila melanogaster*. Science 298(5591):149-159.

Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci U. S. A. 95: 3708–3713.

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

Zhang LJ, Gallo RL. 2016. Antimicrobial peptides. Curr Biol 26(1):R14-9

Zheng L. 1997. Molecular approaches to mosquito parasite interactions. Arch Insect Biochem Physiol.34:1–18.

Zhu J, Chen L, Raikhel AS. 2007. Distinct roles of Broad isoforms in regulation of the 20-hydroxyecdysone effector gene, Vitellogenin, in the mosquito *Aedes aegypti*. Mol Cell Endocrinol 267:97–105.

The Role of *Anopheles* Heme Peroxidases in Mosquito Immunity against Blood-Borne Antigens and Divergent Expression of anopheline Lineage-Specific Duplicated Genes

THESIS

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

PARIK KAKANI

Under the Supervision of
Prof. Sanjeev Kumar
and
Co-supervision of
Dr. Pankaj Kumar Sharma

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
2018

Conclusion of thesis

The work presented in this thesis focused on the role of heme peroxidases in maintaining bacterial homeostasis and regulating *Plasmodium* development in the midgut of *A. stephensi*. A sincere effort was done to explore the interaction of two heme peroxidases, AsHPX2 and AsDuox in immunity against bacteria and *Plasmodium* at the molecular level in *A. stephensi* midgut.

Molecular characterization of AsHPX2 revealed that it is a single exonic gene that encodes for a globular secreted protein of 692 amino acids. The presence of integrin binding motif, LDV revealed that it is a peroxinectin like protein. The expression of AsHPX2 is reduced in blood fed midgut. This gene is induced in the presence of exogenous bacteria in the midgut. The silencing of AsHPX2 gene in the sugar fed midgut and bacteria supplemented blood fed midguts significantly increased the bacterial growth in silenced midgut in comparison to the control midguts. This data suggests that AsHPX2 is one of the molecules that maintain midgut microbial homeostasis in the midgut. Our study also showed that silencing of AsHPX2 increased the *Plasmodium* development and hence, acts as an anti-plasmodial molecule in the midguts. This result revealed that the anti-plasmodial role of HPX2 is conserved in *A. gambiae* and *A. stephensi*. Thus, this mosquito-specific gene can be targeted to manipulate mosquito vectorial capacity.

Our study on Dual oxidase of *A. stephensi* revealed that it has both heme peroxidase and NADPH oxidase domain. This gene is highly conserved among world-wide distributed anophelines. This gene is induced in blood fed midguts and has a weak negative correlation with the growth of endogenous bacteria in blood fed midguts. Silencing of AsDuox in sugar fed and blood fed midguts increased the bacterial growth. The increased endogenous bacteria in the AsDuox silenced blood fed midguts induced the expression of anti-plasmodial molecules such as NOS, TEP1, HPX2 and NOX5. This showed that AsDuox silencing modulates the midgut immunity and this can be further explored to block *Plasmodium* development in the midguts. Expression analysis revealed that this gene is induced in bacteria supplemented blood fed midguts and has a strong negative correlation with the growth of bacteria in these midguts. Silencing of AsDuox and infection with exogenous bacteria did not have any effect on the rate of survival of control and silenced mosquito. Analysis of 16S rRNA in these midguts revealed the reduced bacterial growth. This showed that mosquito midgut immunity is highly efficient to handle the bacterial load.

This data showed that Duox is one of the major molecules of midgut immunity. AsDuox gene is induced in response to the *Plasmodium* infection and silencing of this gene reduced the number of *Plasmodium* oocysts. Thus, we conclude that AsDuox gene is an important molecule of innate immunity against blood-borne antigens in midgut. This gene maintains bacterial homeostasis and regulate *Plasmodium* development in the midgut. Hence, the manipulation of this gene will open up new frontiers in blocking the *Plasmodium* development.

Besides these, gene duplication event in the heme peroxidase multi-gene family was also analyzed and we found that HPX15 which is previously reported as a vaccine candidate has a duplicated paralog named HPX14 in both *A. gambiae* and *A. stephensi* have. They are tandemly duplicated paralog present in head to tail orientation in the genome and organized into an independent domain of gene expression. Expression analysis of this duplicated domain revealed that HPX15 and HPX14 have differential pattern of gene expression. The mRNA levels of AsHPX15 are very high in comparison to the AsHPX14 in blood fed midguts and exogenous bacteria supplemented blood fed midguts. To elucidate the cause of basal expression of AsHPX14, we analyzed the transcription factor binding motifs in the regulatory region of this duplicated paralog. Surprisingly, we found the presence of an insulator protein CTCF, and its core motif is present only in the promoter region of AsHPX14. Hence, we assumed that CTCF regulates the expression of duplicated paralog, AsHPX14. Further studies revealed that AsHPX14 is induced in AsHPX15 silenced bacteria supplemented blood fed midguts, however, expression of AsHPX14 remained unaffected in AsHPX15 silenced blood fed midguts. This data suggested that AsHPX14 may have a role in immunity and not in physiology. Our analysis revealed that there is no redundancy in the function of AsHPX14 and AsHPX15 and hence, strengthen the potentiality of HPX15 as a vaccine candidate.

Thus, overall we conclude that in *A. stephensi* heme peroxidases HPX2, Duox and HPX15 are involved in maintaining gut bacterial homeostasis and also regulate *Plasmodium* development. The study suggested that gut physiology can be manipulated by targeting these heme peroxidases to arrest *Plasmodium* development. Hence, further these heme peroxidases can be explored to develop transmission blocking strategies to control malaria.

Future Prospects

A further research will include the functional elucidation of following future perspectives:

❖ It would be interesting to know the exact mechanism of HPX2 and Duox in maintaining the midgut bacterial homeostasis. It is a subject of great interest to know the function of HPX2/Duox in other major Indian malaria vectors to propose these proteins as vaccine candidates.

❖ In future, analysis of bacterial communities will explore the type of bacteria that are proliferating in HPX2/Duox silenced sugar fed or blood fed *A. stephensi* midgut. Further, these bacteria can be used in designing the promising paratransgenesis techniques to block *Plasmodium* development.

❖ It would be of great interest to know the molecular mechanisms involved in the induction of HPX14 in HPX15 silenced bacteria supplemented blood fed midguts. Further investigation may confirm the exact transcriptional regulation of HPX15 and HPX14 genes in midgut. Future studies include expression analysis of these genes in other worldwide distributed anophelines and their role in the development of *Plasmodium*.

❖ Deciphering the molecular mechanism associated with the CTCF in regulating the expression of HPX14 can be the future refinement of this thesis.