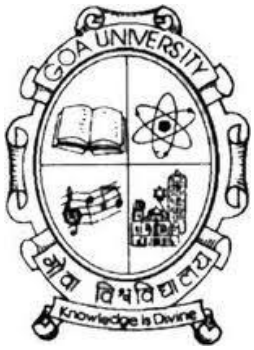
"Midgut proteome analysis of female *Anopheles stephensi* Liston, a vector of malaria"

A thesis submitted to Goa University



For the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

ZOOLOGY

By

Ajeet Kumar Mohanty

Guide
Dr. Ashwani Kumar

and

Co-Guide
Dr. Sandeep Garg

Goa University
Taleigao Plateau, Goa-403206

March 2016

STATEMENT

As required under the University ordinance, I hereby state that the present thesis for Ph.D. degree entitled "**Midgut proteome analysis of female *Anopheles stephensi* Liston, a vector of malaria**" is my original contribution and that the thesis and any part of it has not been previously submitted for the award of any diploma/degree of any University or Institute. To the best of my knowledge, the present study is the first comprehensive work of its kind from this area.

The literature pertaining to the problem investigated has been duly cited. Facilities availed from other sources are duly acknowledged.

Ajeet Kumar Mohanty
Ph. D student
Department of Zoology
Goa University, Goa

CERTIFICATE

This is to certify that the thesis entitled "**Midgut proteome analysis of female *Anopheles stephensi* Liston, a vector of malaria**" submitted by **Mr. Ajeet Kumar Mohanty** for the award of the degree of **Doctor of Philosophy** in **Zoology** is based on his original studies carried out by him under our supervision. The thesis or any part thereof has not been previously submitted for any other degree or diploma in any University or Institution.

Place: Goa University

Date:

Dr. Ashwani Kumar (Guide)
Scientist 'F'& Officer-in-Charge
National Institute of Malaria Research (ICMR)
Directorate of Health Services Building,
Campal, Panaji- 403 001, Goa

Dr. Sandeep Garg (Co-Guide)
Associate Professor
Department of Microbiology
Goa University, Taleigao Plateau,
Goa-403206

CERTIFICATE FROM THE CANDIDATE

As suggested by the External Examiners, appropriate corrections are incorporated into the thesis on relevant pages.

Ajeet Kumar Mohanty
Ph. D student
Department of Zoology
Goa University, Goa

ACKNOWLEDGEMENTS

I feel immense pleasure by taking privilege to record my deepest sense of gratitude and indebtness to my esteemed guide Dr. Ashwani Kumar whose blessings, constant support, intensive guidance, encouragement and timely inspiration was indispensable for the completion of this work. I would also like to express my sincere gratitude to my esteemed co-guide Dr. Sandeep Garg for his scientific and philosophical guidance. His blessings, constant motivation, critical reviewing and positive attitude have always inspired me throughout the endeavour.

I am immensely grateful to the members of the Faculty Research Committee, Dr.N. Ramaiah, V.C's nominee (NIO) for providing his valuable time, cooperation, critical assessment and timely approvals. I am also thankful to Prof. Saroj Bhosle (ex-Dean, Faculty of life sciences and environment and Prof. M. K. Janarthanam (Dean, Faculty of life sciences and environment) Goa University for their support, critical suggestions, appreciations and constant motivation.

I wish to express my gratitude to Prof. S. K. Shyama (Head, Department of Zoology) for the support, paperwork and cooperation throughout the course of this study. I also like to thank the faculty members of Zoology Prof. R. Roy, Prof. I. K. Pai for the cooperation. I am also very thankful to the non-teaching staff of the Department of Zoology and administrative staffs, Goa University for helping me through the administrative procedures.

I am sincerely thankful to Director, National Institute of Malaria Research (ICMR) New Delhi for necessary permission, precious support and providing research facilities.

My sincere thanks to my enthusiastic collaborator Dr. Akhilesh Pandey and Dr. T. S. K Prasad from for their scientific interactions and constant support at many stages of the study, right from facilitating collaboration to critical inputs in the manuscripts. I also thank Dr. Harsha Gowda for his help and support for carrying out research experiments and analysis.

I would like to make a special mention of my dear friends from Institute of Bioinformatics (IOB), Bangalore Manish, Gourav, Manoj, Shreelaxmi and Manjula for the stimulating discussions and constant help. I also extend my thanks and appreciation to each and every member of IOB. This project would have not been possible without the help of so many people in so many ways.

I would like to deeply acknowledge the tremendous support provided by my colleague Dr. Hemanth Kumar, Senior Research Scientist and my staff members of NIMR Goa, Mrs. Maria Flavia DeSouza, Mrs. Sushma L. Parkar, Mrs. Smita P. Naik, Mr. Udesh T. Konvilkar, Mr. Arun N. Phadte, Mr. Dharma V. Chodankar, Mr. Floriano Coutinho, Mr. Mhalu G. Gawas, Mr. Domnic M. Vaz, Mr. Pratap B. Jalmi, Mr. Sahab Singh, Mr. Rajkapoor Maurya and Mrs. Alda Lobo for their help and support during this period. A special thanks to Mrs. Sushma L. Parkar and Mr. Rajkapoor Maurya for their hard work which they had put in maintaining

the mosquito colony and providing the mosquitoes, whenever I needed. I also extend my thanks to all the staff members of MESA-NIH team for their help and support.

I acknowledge my friends and colleagues Pramod Nayak, Santosh Kumar Mohanty, Vasant Salgaonkar, Premdeep Matlane, Dr. Rakhi Dhawan, Dr. Trelit de Sousa, Dr. Teja, Dr. Nandini Korgaonkar, Dr. Deeparani Prabhu, Minisha and Joleen for their constant help and support.

Last but not the least I wish to express my sentiments of gratitude towards my grandparents especially my grandfather Late Sh. Narayan Chandra Mohanty and my grandmother Mrs. Kusum Mohanty and my family members whose unconditional love, affection, blessings and support have driven me in completing this research work. My sister Bijaya Lakshmi Routray, Samita Swain, my brother Manoj and Sujit and my in-laws are thanked for their care and best wishes. My wife Abhipsa Mohanty needs a special mention here for her encouragement, support and love during my thesis writing.

Ajeet Kumar Mohanty

DEDICATED WITH LOVE TO MY

GRAND PARENTS

AND

TEACHERS

List of Abbreviations

Cm	Centimeter	PCR	Polymerase chain reaction
%	Percentage	PBS	Phosphate buffer saline
°C	Degree centigrade	PSM	Peptide spectrum match
µl	Microlitre	RNA	Ribonucleic acid
µM	Micromolar	RPLC	Reversed phase liquid chromatography
AIDS	Acquired Immuno Deficiency syndrome	PI	Isoelectric point
NVBDCP	National vector borne disease control programme	rpm	Revolutions per minute
BLAST	Basic Local Alignment Search Tool	SCX	Strong cation exchange
Bp	Base pairs	SDS	Sodium dodecyl sulfate (sodium lauryl sulphate)
BSA	Bovine serum albumin	sp.	Species
cDNA	Complementary DNA	WHO	World Health Organization
CDS	Coding DNA sequence	RBC	Red blood cell
CID	Collision induced dissociation	RT-PCR	Real time-PCR
DDT	Dichloro Diphenyl Trichloroethane	PAGE	Poly-acrylamide Gel electrophoresis
DNA	Deoxyribonucleic acid	P^H	Potential of hydrogen
DTT	Dithiothreitol	µg	Microgram
EDTA	Ethylene diamine tetra acetic acid disodium salt	FDR	False discovery rate
EST	Expressed sequence tag	nm	Nanometer
Da	Dalton	TFA	Trifluoro acetic acid
LTQ-Orbitrap	Linear trap quadropole-Orbitrap	MS	Mass spectrometry
GSSPs	Genome Search Specific Peptides	M	Molar
hr	Hour	ppm	Parts per million
HCD	High energy collision induced dissociation	iTRAQ	Isobaric tags for relative and absolute quantitation
IPG	Immobilized pH gradient	m/z	Mass-to-charge ration
HPLC	High performance liquid chromatography	BCA assay	Bicinchoninic acid assay
IAA	Iodoacetamide	PTMs	Post-Translational Modification
IP	Intra-peritoneally	C18	Octadecyl carbon chain
Kb	Kilo base pair	ACN	Acetonitrile
KCl	Potassium Chloride	mRNA	Messenger RNA
kDa	Kilo Dalton	UTR	Untranslated region

List of tables

Chapter I

Table 1.1 Burden of malaria morbidity and mortality in the countries of South East Asia Region in 2013

Table 1.2 Epidemiological indicators for malaria in India (2001-2014)

Table 1.3 List of the *Anopheles* vector species and species complexes found in India

Table 1.4 Current status of genome projects of various mosquito vectors of human disease

Table 1.5 Summary of proteomic studies carried out in different mosquito species

Chapter II

Table 2.1 Partial list of most abundant proteins identified in this study based on normalized spectral abundant factor (NSAF) values

Table 2.2 List of immunity related proteins identified in *Anopheles stephensi* midgut

Table 2.3 List of mosquito proteins involved in parasite interaction identified in *An. stephensi* midgut (Agonistic molecules)

Table 2.4 List of mosquito proteins involved in parasite interaction identified in *An. stephensi* midgut (Antagonistic molecules)

Table 2.5 A partial list of proteins over-expressed in female midgut identified in this study

Table 2.6 List of immunity related proteins identified and their expression level in female and male midgut tissue

Table 2.7 List of proteins involved in vector parasite interaction and their relative expression level between female and male midgut tissue

Table 2.8 A partial list of female specific proteins identified and their expression level in female midgut tissue

Table 2.9 List of male specific proteins identified and their expression level in male midgut tissue

Chapter III

Table 3.1 A partial lists of proteins identified to be over-expressed by ≥ 1.5 in blood fed midgut tissue as compared to sugar fed mosquitoes

Table 3.2 A partial list of proteins identified to be down-regulated by ≤ 0.66 fold in blood fed midgut tissue as compared to sugar fed mosquitoes

Table 3.3 A partial lists of proteins identified to be over-expressed in *P. berghei* infected midgut tissue as compared to blood fed mosquitoes

Table 3.4 A partial lists of proteins identified to be down-regulated in *P. berghei* infected midgut tissue as compared to blood fed mosquitoes

Table 3.5 A partial list of the proteins specifically over-expressed between *Plasmodium berghei* infected and sugar fed midgut tissue

Table 3.6 A partial list of the proteins down-regulated between *Plasmodium berghei* infected and sugar fed midgut tissue

Table 3.7 List of proteins involved in vector parasite interaction and their relative fold change between sugar fed, blood fed and *P. berghei* infected midgut tissue

Table 3.8 List of immunity related proteins identified to be differentially expressed between sugar, blood and infected blood fed midgut tissue

Chapter IV

Table 4.1 The details about all the novel genes annotated/ identified in the proteogenomic analysis of *An. stephensi*

List of Figures

Chapter I

Figure 1.1	Countries with ongoing transmission of malaria 2013
Figure 1.2	Burden and proportion of confirmed malaria cases among the countries in South East Asia in 2013
Figure 1.3	Showing distribution of *P. falciparum* malaria in India
Figure 1.4	Burden and proportion of confirmed malaria cases in different states of India 2014
Figure 1.5	State-wise distribution of malaria cases and deaths reported in India
Figure 1.6	Malaria situation in India from 2001 to 2014
Figure 1.7	Distribution of malaria vectors in India
Figure 1.8	Female *Anopheles stephensi* Liston, showing main distinguishing characters
Figure 1.9	*Anopheles* (Cellia) *stephensi* Liston 1901
Figure 1.10	Life Cycle of the malaria parasite in mosquito vector and human host
Figure 1.11	Workflow for protein identification by using bottom-up MS approach

Chapter II

Figure 2.1	Dissected midgut from a female *Anopheles stephensi*
Figure 2.2	SDS PAGE profile of proteins obtained from the midgut of male and female *Anopheles stephensi*
Figure 2.3	Typical elution profile of tryptic digest of female *Anopheles stephensi* midgut protein in bRPLC chromatography
Figure 2.4	Outline of the quantitative proteomic strategy
Figure 2.5	Typical elution profile of tryptic digest of iTRAQ labeled male and female midgut protein in SCX chromatography
Figure 2.6	Workflow explaining the steps involved in the proteomic analysis of midgut tissue
Figure 2.7	The distribution of mass error in parts per million for the entire peptides identified in this study
Figure 2.8	A column graph depicting the number of proteins identified with peptide spectra matches (PSMs)
Figure 2.9	Peptides and proteins identified in different fractionation methods
Figure 2.10 A & B	MS/MS spectra of peptides for the most abundant proteins in *An. stephensi* identified in this study
Figure 2.11 A & B	Representative MS/MS spectra of proteases identified in this study
Figure 2.12 A & B	Representative MS/MS spectra of immune response proteins identified in this study
Figure 2.13 A & B	Representative MS/MS spectra of proteins involved in insecticide

resistance mechanism identified in this study

Figure 2.14 A & B Representative MS/MS spectra of agonistic proteins those that aid *Plasmodium* development in mosquito identified in this study

Figure 2.15 A & B Representative MS/MS spectra of Antagonistic proteins those that prevent *Plasmodium* development in mosquito identified in this study

Figure 2.16 Gene ontology based classification of proteins identified in midgut tissue of sugar female fed *An. stephensi*

Figure 2.17 Differential expression of proteins between male and female midgut

Figure 2.18 A & B MS/MS spectra of peptides for the proteins over-expressed in female midgut

Figure 2.19 A & B MS/MS spectra of peptides for the proteins over-expressed in male midgut

Figure 2.20 Gene ontology based classification of over-expressed proteins identified in female midgut

Figure 2.21 Gene ontology based classification of over-expressed proteins identified from male midgut

Chapter III

Figure 3.1 SDS-PAGE profile of proteins obtained from midgut tissue of sugar fed, blood and *Plasmodium berghei* infected female *An. stephensi*

Figure 3.2 SCX chromatogram of iTRAQ labeled peptides obtained from midguts of sugar, blood fed and *P. berghei* infected female *Anopheles stephensi*

Figure 3.3 Workflow illustrating the steps involved in the quantitative proteomic strategy

Figure 3.4 Heat map representation of all the proteins identified in the midgut of *An. stephensi* in response to sugar, blood and *P. berghei* infected blood feeding

Figure 3.5 Representative MS/MS spectra of over-expressed proteins identified in blood fed midgut as compared to sugar fed midgut tissue

Figure 3.6 Representative MS/MS spectra of down-regulated proteins identified in blood fed midgut as compared to sugar fed midgut tissue

Figure 3.7 Representative MS/MS spectra of over-expressed proteins identified in *P. berghei* infected midgut tissue as compared to blood fed midgut

Figure 3.8 Representative MS/MS spectra of down- regulated proteins identified in *P. berghei* infected midgut tissue as compare to blood fed midgut tissue

Figure 3.9 Classification of differentially expressed proteins between sugar, blood and *P. berghei* infected midgut tissue based on gene ontology annotation

Chapter IV

Figure 4.1	Workflow of proteogenomic analysis
Figure 4.2	Identification of novel genes using peptides identified in intergenic regions
Figure 4.3	Identification of novel gene using peptides identified in intergenic regions
Figure 4.4	An example of joining of genes ASTEI10980 and ASTEI10981 based on transcript, intergenic peptides and ortholog gene in SDA 500 strain
Figure 4.5 A & B	Correction in the gene structure using peptide mapping to an intron of an annotated gene
Figure 4.6 A & B	Identification of peptides mapping to exon-intron boundary of annotated genes ASTEI09047 and ASTEI02109
Figure 4.7 A & B	N-terminal extension of a protein using peptides mapping upstream to an annotated start site
Figure 4.8	Identification of peptides translated in a different frame than the annotated protein sequence
Figure 4.9 A & B	Identification of an N-terminally acetylated peptide indicated an alternate start site for ASTEI08087 gene and ASTEI08114
Figure 4.10	Gel image of RT-PCR validation of the novel gene models identified against first genome assembly AsteI of *An. stephensi* Indian strain

Table of Contents

S. No.	Contents	Page No.
Chapter I	Introduction and Literature Survey	**1-46**
Chapter II	Proteomic analysis of midgut of female *Anopheles stephensi* using Mass Spectrometry	**47-117**
Chapter III	Differential expression of proteins in the midgut of *Plasmodium berghei* infected and uninfected female *Anopheles stephensi* Liston	**118-151**
Chapter IV	Application of mass spectrometry-derived proteomic data from midgut in genome annotation of *Anopheles stephensi*	**152-168**
	Summary and Future Prospects	**169-174**
	Appendix	**175-265**
	Bibliography	**266-293**
	Publications	**294-296**

Chapter-I

Introduction and Literature Survey

1.1 Introduction

1.1.1 Global burden of malaria

Vector-borne diseases such as malaria, filariasis, Japanese encephalitis, chikungunya and dengue hemorrhagic fever cause considerable mortality, morbidity and disability globally. Among all the vector borne diseases, malaria is one of the most debilitating diseases in the sub-tropical and tropical regions of the world. Malaria accounts for 85% of global infectious disease burden. This imposes huge socio-economic burden on humanity including six other diseases i.e. pneumonia, tuberculosis, diarrhea, HIV/AIDS, measles, and hepatitis B (Murray and Lopez, 1996, 1997). Globally an estimated 3.3 billion population lived in areas at risk of malaria transmission in 97 countries and territories (WHO report, 2014). About 1.2 billion population is living in the high risk regions. In these high-risk areas, more than one malaria case occurs per 1000 population (Figure 1.1).

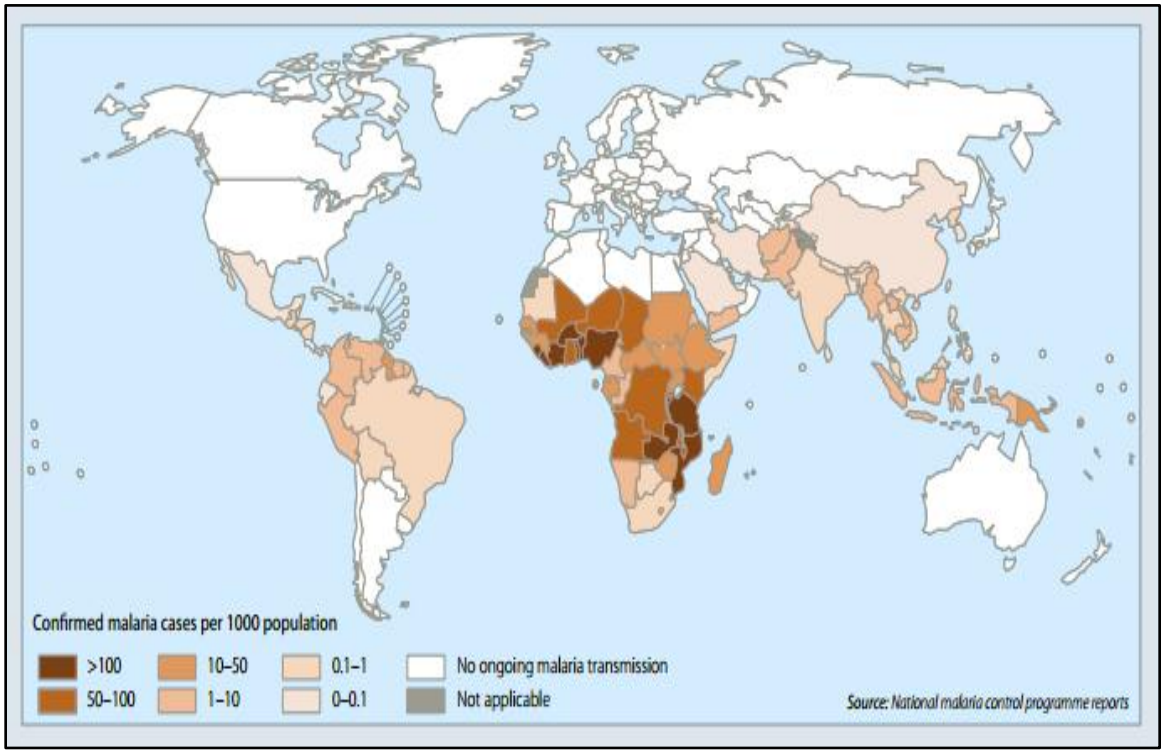


Figure 1.1: Countries with ongoing transmission of malaria, 2013 **(**Source**:** WHO report 2014)

About 82% of malaria cases worldwide and more than 90% of the malaria deaths occur in Africa, mostly among children under the age of five and pregnant women, this is because majority of infections are caused by *Plasmodium falciparum*, the most dangerous among the five human malaria parasites highly prevalent in the African continent. African region hence contributes highest (82%) of all malaria cases, followed by the South-East Asia region (12%) Eastern Mediterranean region (5%), Western pacific (0.5%) and America (0.35%). According to an estimate by the World Health Organization, annually ~198 million people suffer from malaria (range 124–283 million) which resulted in an estimated deaths of about 5,84,000 (range 3,67,000–7,55,000) individuals worldwide in the year 2013. Recent WHO report indicates that there is a significant decrease in global mortality rate by 47% between 2000 and 2013 and in Africa region by 54% (WHO report, 2014).

1.2 Malaria burden in South East Asia region

The South East Asian Region of WHO is comprised of 11 countries and in 10 countries malaria is a major public health challenge as it causes considerable mortality and morbidity. According to WHO World Malaria Report 2014, South East Asian (SEA) region bears the second largest burden of malaria contributing (12%), after African region (82%). About 1.4 billion people live in 10 malaria endemic countries of South-East Asian region, with 352 million at high risk, among whom most of the 1.6 million confirmed malaria cases are reported. India alone contributes about 55% burden, followed by Myanmar and Indonesia with each 21% in the year 2013 (WHO report 2014). The remaining 7 countries together contribute only 3% of malaria cases in the region (Figure 1.2).

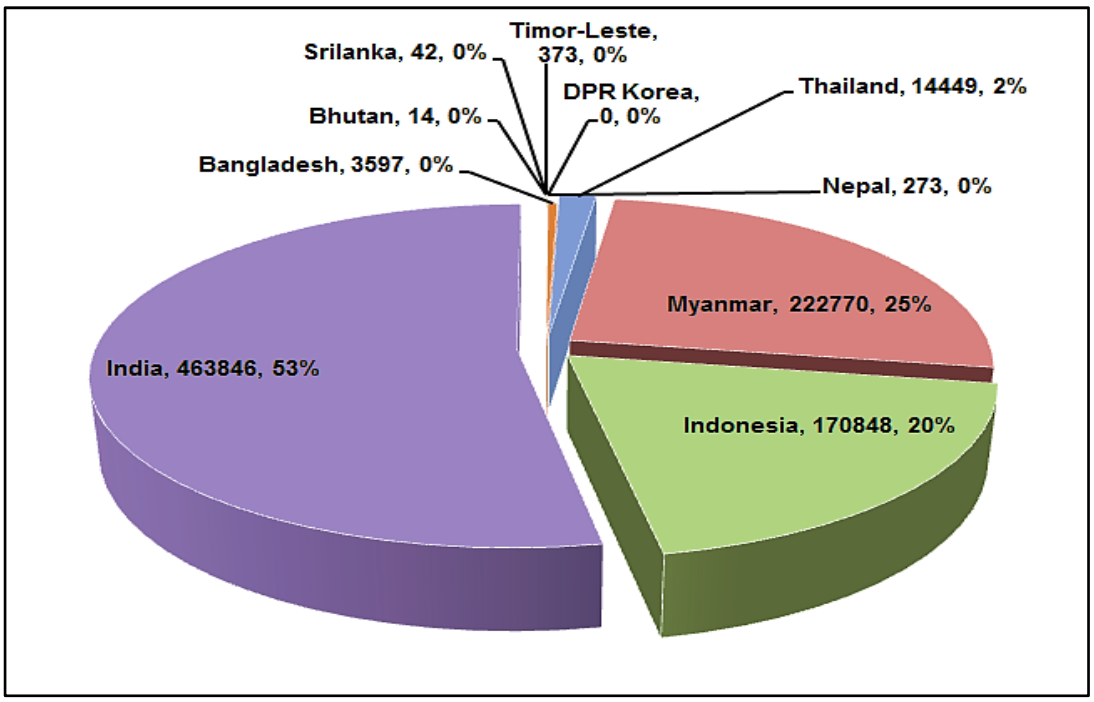


Figure 1.2: Burden and proportion of confirmed malaria cases among the countries in South East Asia in 2013. **(**Source: WHO Malaria report, 2014).

Out of 776 deaths reported in the SEA region, India contributed the highest about 56.7% followed by Myanmar with 30.4% Indonesia with 5.7% Thailand with 4.7%, Bangladesh with 1.9% and Timor-Leste 0.8% deaths. There was no malaria attributed mortality recorded in Bhutan, Nepal, Srilanka and Democratic People's Republic of Korea (Table 1.1). The good news is that Maldives in South-East Asia region is declared 'free of malaria'. Bhutan and Democratic People's Republic of Korea are in the 'pre-elimination phase'. Srilanka is in malaria elimination phase. India, Bangladesh, Democratic Republic of Timor-Leste, Indonesia, Myanmar, Nepal and Thailand are in the 'control phase'. The numbers of confirmed reported malaria cases in this region have decreased from 2.9 to 1.6 million between 2000 and 2013 (WHO report, 2014).

Table 1.1: Burden of malaria morbidity and mortality in the countries of South East Asia region in 2013 (Source: WHO report, 2014).

S. No.	Country	Confirmed Cases	Pf cases	Pv cases	Deaths
1	Bhutan	45	14	31	0
2	Srilanka	95	42	52	0
3	Timor-Leste	1,025	373	512	3
4	Nepal	1,974	273	1,659	0
5	DPR Korea	14,407	0	14,407	0
6	Bangladesh	3,864	3,597	262	15
7	Thailand	33,302	14,449	15,573	37
8	Myanmar	3,33,871	2,22,770	98,860	236
9	Indonesia	3,43,527	1,70,848	1,50,985	45
10	India	8,81,730	4,63,846	4,17,884	440
	Total	**16,13,840**	**8,76,212**	**7,00,225**	**776**

1.3 Malaria situation in India

In India malaria has been a problem since centuries. Historically it contributes to considerable annual morbidity and mortality and this poses a challenge to public health care system. This disease was also described in the ancient Indian medical literature like the *Charaka Samhita* and *Atharva Veda* (Gupta *et al*., 2014). In 1947, when India became independent, 75 million malaria cases were estimated in a population of 330 million and estimated mortality due to the disease was 0.8 million per annum (Lal *et al*., 2000). To combat this threat, the National Malaria Control Programme was launched by the Govt. of India in April 1953. The programme successfully controlled malaria and as a result of that the number of malaria cases significantly declined to just 100,000 in 1964 (Kumar *et al*., 2007). Although malaria was nearly eradicated in the mid 1960's by the National Malaria Eradication Programme (NMEP) as also no deaths were recorded in the country but by the year 1976 malaria resurged with vengeance to touch ~6.45 million cases and since then cases have gradually declined to about 0.88 million reported cases in 2014.

In India, the epidemiology of malaria is different from rest of the world because of geo-ecological diversity, multi-ethnicity, a number of vector species with wide distribution and three *Plasmodium* species with a great deal of heterogeneity in their distribution. There are several factors responsible for the malaria transmission which includes both

climatic and non-climatic factors. Temperature, rainfall and relative humidity are the climatic variables that influence vector breeding potential, development, longevity and feeding propensity as also parasites development (sporogony) in vector (Sharma, 1993; Gubler *et al*., 2001; Koenraadt *et al*., 2004; Kumar *et al*., 2007; Kar *et al*., 2014). Non-climatic factors include mosquito vectors, parasites, human host factors, population movement, urbanization, preventive and control measures undertaken. The endemicity of malaria in different parts of the country is determined based on contribution of these factors and this forms different ecotypes, i.e. rural plains, plain and hilly forest, urban, industrial and coastal (Kumar *et al*., 2012; Sharma, 2012). According to World Malaria Report, 22% of India's population live in high transmission risk areas (> 1 case per 1000 population), 67% live in low transmission areas (0–1 cases per 1000 population) and 11% live in malaria-free areas (0 cases). The country reports annually in the range of 0.7-1.6 million cases and a 400-1000 deaths.

In the year 2014, 1.10 million confirmed cases were reported from India and 535 deaths (NVBDCP). The incidence of *Plasmodium falciparum* accounted for 65% infection, of the total cases recorded in country. In the recent years, the incidence of cases due to *Plasmodium falciparum* exceeds that of *P. vivax* which used to be the predominant species marking a new phenomenon. However the spatial distribution of *P. falciparum* and *P. vivax* widely varies in different parts of country (Figure 1.3).

In India, malaria problem is of major concern in the 15 States and Union Territories. About 95% of the population of India is at risk of contracting malaria. 80% of malaria cases reported in the country are confined to areas consisting of 20% of the population majority of whom are ethnic tribes living in difficult, hilly and poorly accessible areas. About 92% of malaria cases and 97% of deaths due to malaria are reported from North Eastern states, Odisha, Madhya Pradesh, Chhattisgarh, Jharkhand, Andhra Pradesh, Maharashtra, Rajasthan, Gujarat, West Bengal and Karnataka. The state of Odisha with 3.6% of the country's population, contributes highest i.e. 37% of the total malaria cases recorded in the country followed by Chhattisgarh (12%), Madhya Pradesh (9%), Jharkhand (9%), Maharashtra (5%) as per (NVBDCP) as shown in Figure 1.4.

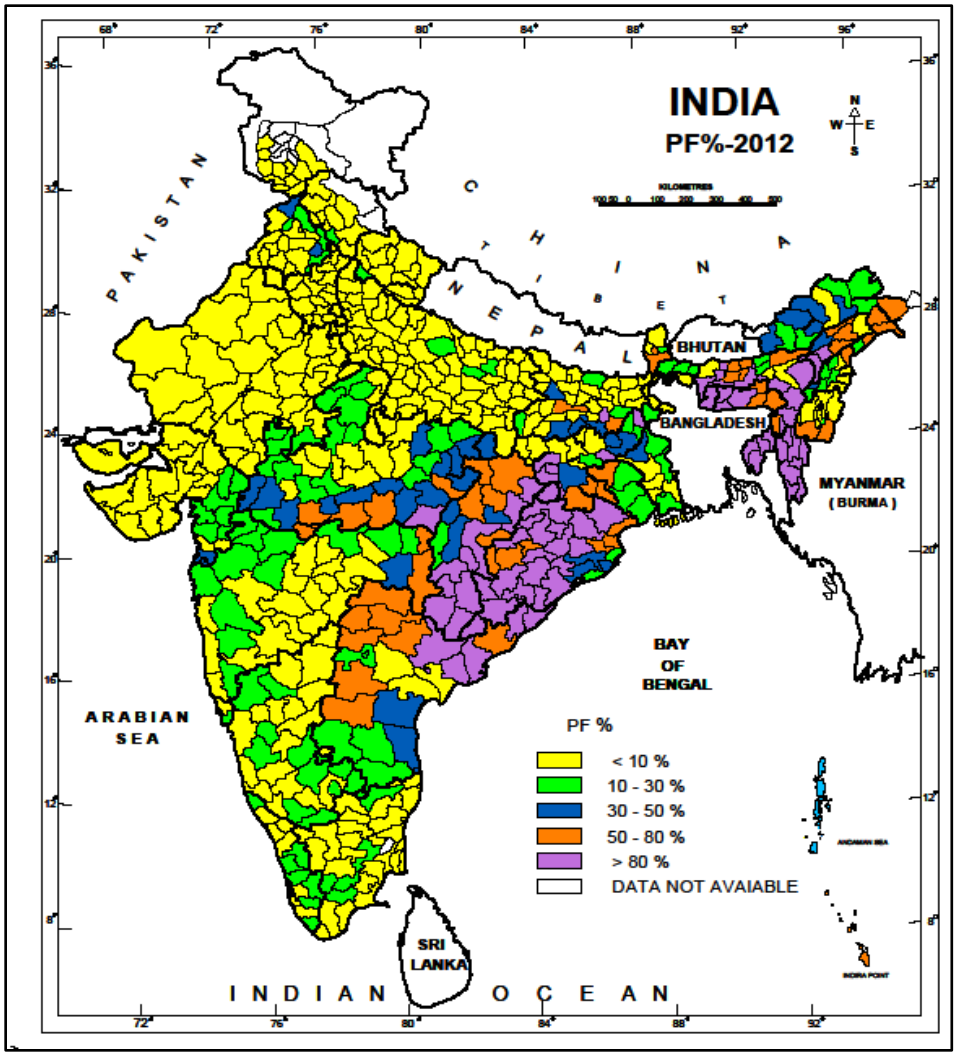


Figure 1.3: Showing distribution of *P. falciparum* malaria in India (Source: Kumar *et al*., 2014).

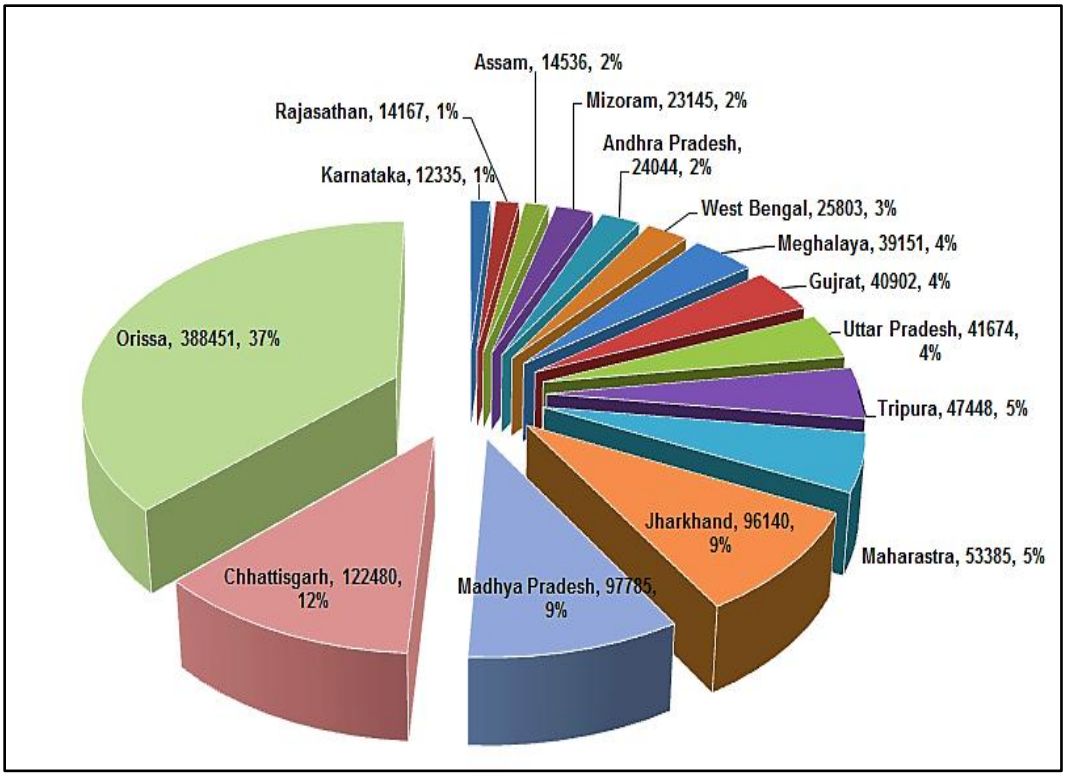


Figure 1.4: Burden and proportion of confirmed malaria cases in different states of India 2014 (Source: NVBDCP).

The western part of India are epidemic prone particularly in the states of Rajasthan, Gujarat, Haryana apart from some regions of Karnataka in the south and Brahmaputra valley in the east (NVBDCP). However, out of 535 deaths reported in 2014, Tripura recorded highest number of 96 deaths (17.9%) due to epidemic outbreak followed by Meghalaya 78 deaths (14.5%), Odisha 73 deaths (13.6%), Maharashtra 63 deaths (11.7%), West Bengal 62 deaths (11.5%) and Chhattisgarh 53 deaths (99.9%) (Figure1.5). It has been seen that malarial mortality is maximum in the economically productive age groups of 15 to 44 years in the country (Kumar *et al*., 2007). In India the biggest burden of malaria is borne by the poor, most backward and remote parts of the country.

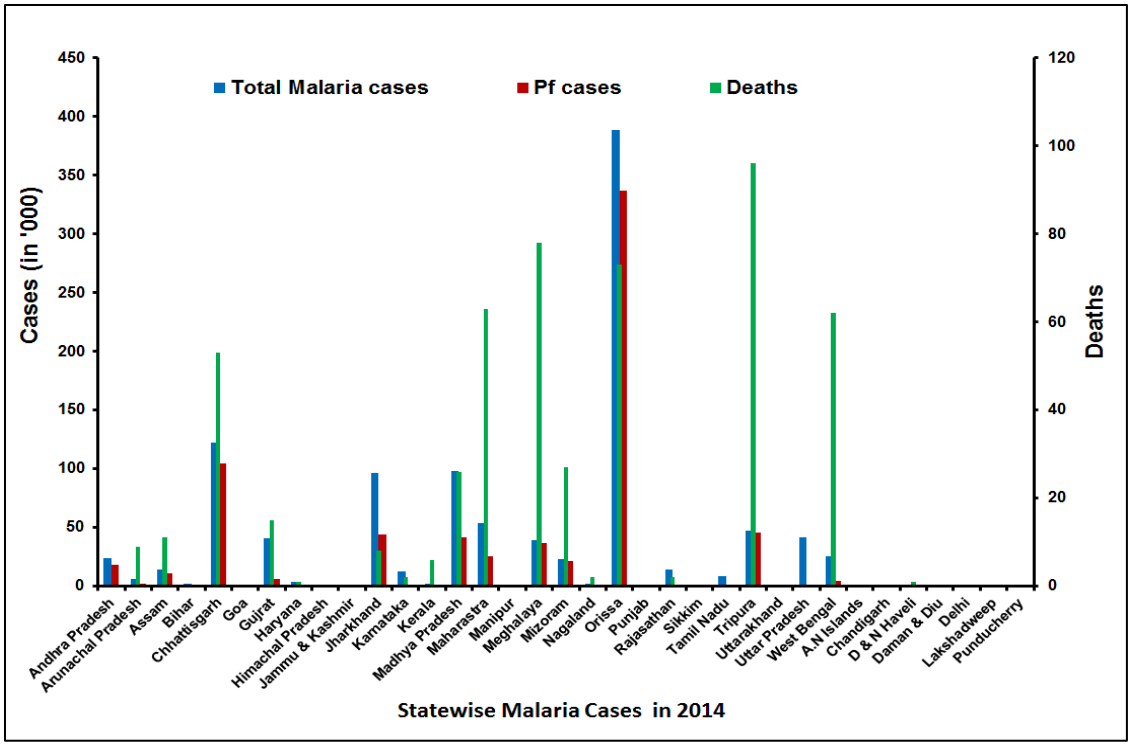


Figure 1.5: State-wise distribution of malaria cases and deaths reported in India **(**Source: NVBDCP).

Malaria which was once predominantly a rural phenomenon > 90-95% cases were being reported in rural region of India has since 1990s emerged as a major communicable disease in urban areas primarily because of the rapid, haphazard and unplanned expansion of cities, besides the sheer pace at which some of the towns are expanding in India. The accelerated construction activity in the cities, on one hand has led to the migration of workers from malaria endemic and hyper-endemic areas of the country, from where parasites are imported with these migrant workers to the urban centres and on the other hand, it has provided firm foothold to the urban vector (Kumar, 1997). Because of extensive water stagnations, there is perpetuation of breeding of the urban malaria vector, *Anopheles stephensi.* Hence this dual phenomenon of vulnerability to infection in high risk population and high vector receptivity has resulted in many small and major outbreaks in the last three decades. It is now estimated that urban areas contribute to about 10-15% of the total malaria cases reported in India (Kumar, 1997; Sharma, 1999).

Kumar *et al*., (2007) carried out estimation of malaria burden, which indicated that there was 68% to 98% gap between India's reported malaria cases and the actual incidence of

malaria, which is primarily due to reporting factors. Gupta *et al*., (2009) reported that many malaria cases treated at private facilities are not included in the government's official statistics. Several efforts have been made to quantify the exact burden of malaria using different approaches leading to highly variable results (Strüchler, 1989; Snow *et al*., 2005; WHO report, 1994, 1999, 2010; Kumar *et al*., 2011). An estimate given by Hay and colleagues suggests almost twice as many cases of malaria in 2007 than reported in the World Malaria Report 2008. Dhingra *et al*., (2010) estimated 2,05,000 (CI=12,500-2,77,000) annual deaths attributed to malaria in India in 2002, a large proportion of which were in adults, compared with the WHO's estimate of 15,000 deaths. Murray *et al*., (2012) estimated the malaria mortality in India in 2010 at 46,970 (CI= 14,757-94,945). However the exact rate of mortality due to malaria in India is still not known and a subject of intense debate.

In India, malaria cases reported to NVBDCP have consistently declined from 2.08 to 1.1 million during 2001 to 2014. Similarly *Plasmodium falciparum* cases have declined from 1.0 to 0.72 million cases during the same period (NVBDCP). The numbers of deaths reported have also declined to less than 2,000 deaths during this period. The high mortality was recorded in North Eastern States in 2006 which was related to severe malaria epidemic due to population movement (Figure 1.6).

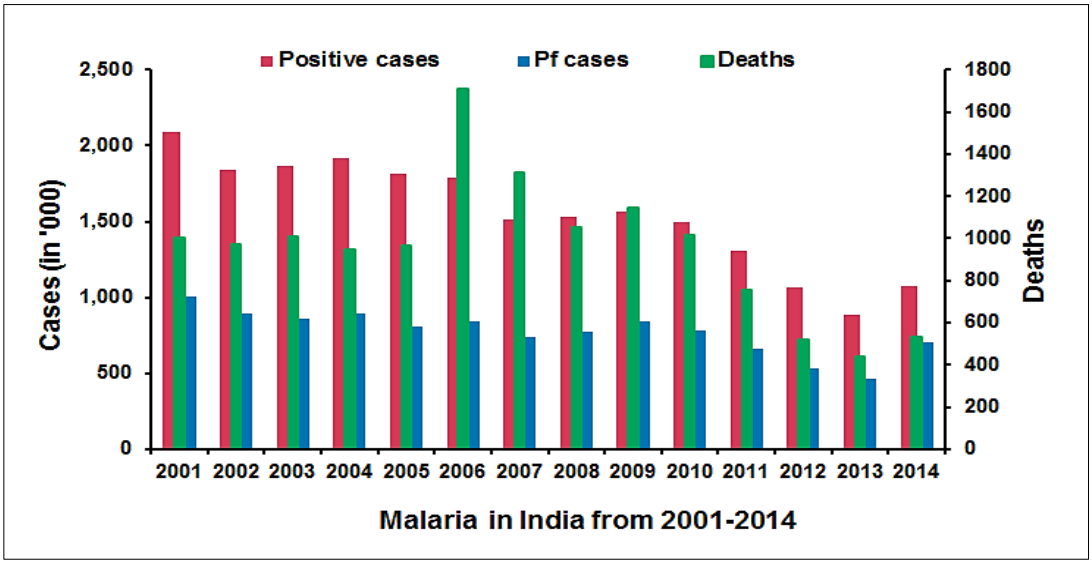


Figure 1.6: Malaria situation in India from 2001 to 2014 (Source: NVBDCP).

The country's Slide Positive Rates (SPR) have declined from 2.31 to 0.91 from 2001 to 2014. Similarly Slide Falciparum Rates (SFR) has declined from 1.11 to 0.60 during these years. This indicates that there is downward trend of overall malaria endemicity in the country during all these years (Table1.2). To combat the high rate of prevalence of the disease, a number of intervention strategies have been devised and implemented for controlling the transmission of malaria. These include early detection and prompt treatment of malaria in human hosts through the introduction of rapid detection tests (RDT) in addition to the ongoing method of blood smear examination, modifications in the drug policy by way of introduction of Artemisinin Combination Therapy (ACT) to treat *P. falciparum* malaria, prevention of mosquito bites using mosquito repellents, long lasting insecticide treated nets (LLINs), selective indoor residual spraying (IRS) to control mosquitoes, destroying the breeding habitants in water bodies using insecticides and anti-larval operations to control mosquitoes. Sustained and long term implementation of preventive and control measures are necessary in order to achieve a significant and sustainable malaria control in a specific region. The recent trends in malaria show that areas where effective malaria control programs have been adopted have experienced a rapid decline in the rate of transmission and clinical burden of malaria. If the control measures are stopped or being interrupted malaria can revert back and can cause severe infection to more number of people. Therefore, better understanding of the vectorial capacity of the anopheline mosquitoes transmitting malaria may enable identification of novel methods to reduce disease burden.

Table 1.2: Epidemiological indicators for malaria in India (2001-14) (Source: NVBDCP).

Year	Population in thousand	Blood Smear Examined	Positive cases	Pf Cases	ABER	API	SPR	SFR	Deaths
2001	984579	9,03,89,019	20,85,484	10,05,236	9.18	2.12	2.31	1.11	1,005
2002	1013942	9,16,17,725	18,41,229	8,97,446	9.04	1.82	2.01	0.98	973
2003	1027157	9,91,36,143	18,69,403	8,57,101	9.65	1.82	1.89	0.86	1,006
2004	1040939	9,71,11,526	19,15,363	8,90,152	9.33	1.84	1.97	0.92	949
2005	1082882	10,41,43,806	18,16,569	8,05,077	9.62	1.68	1.74	0.77	963
2006	1072713	10,67,25,851	17,85,129	8,40,360	9.95	1.66	1.67	0.79	1,707
2007	1087582	9,49,28,090	15,08,927	7,41,076	8.73	1.39	1.59	0.78	1,311
2008	1119624	9,73,16,158	15,26,210	7,75,523	8.69	1.36	1.57	0.80	1,055
2009	1150113	10,33,96,076	15,63,574	8,39,877	8.99	1.36	1.51	0.81	1,144
2010	1167360	10,60,40,223	14,95,817	7,79,549	9.21	1.37	1.41	0.74	1,018
2011	1194901	10,93,13,294	1,31,06,56	6,65,004	9.12	1.10	1.20	0.61	754
2012	1211509	10,90,44,798	10,66,981	5,33,535	9.00	0.88	0.98	0.49	519
2013	1263438	11,31,09,094	8,81,730	4,63,846	8.95	0.70	0.78	0.41	440
2014	1270272	11,73,80,897	10,70,513	7,03,587	9.24	0.84	0.91	0.60	535

ABER: Annual blood examination rate **API: Annual parasite index** **SPR: Slide positive rate** **SFR: Slide falciparum rate**

1.4 Malaria vectors and their distribution in India

Malaria is transmitted among humans by female mosquitoes of the genus *Anopheles.* Transmission of malaria is primarily dependent on three epidemiological factors i.e. agent (*Plasmodium*), two hosts (female *Anopheles* mosquitoes and humans), and the environment. While the malaria parasite is the true agent of infection, the female *Anopheles* mosquito is the transmitter of parasite to the human host. Fossil records and zoogeographic evidence suggest that mosquitoes had evolved during Jurassic period, approximately 210 million years ago (Edwards, 1932). Worldwide, 3,523 mosquito species under 111 genera have been recorded (Reinert *et al*., 2009; Harbach, 2012). Mosquitoes belong to the dipteran family Culicidae, which is classified into three subfamilies, Anophelinae, Culicinae and Toxorhynchitinae. The important mosquito species, which transmit diseases belong to the subfamilies Anophelinae and Culicinae. The members of subfamily: Toxorhynchitinae do not feed on vertebrate blood and therefore, do not transmit any disease. All human malaria parasite species are transmitted through anophelines (Genus *Anopheles*) only, but out of the approximately 450 *Anopheles* species known, ~60 species transmit malaria in nature worldwide (Manguin *et al*., 2008).

In India malaria endemicity is due to diverse ecology and multiple vector species distributed throughout the country (Sharma, 1998). In all 58 species of anophelines are found in India, among which 6 are primary vectors viz., *Anopheles culicifacies* Giles, *Anopheles stephensi* Liston, *Anopheles fluviatilis* James, *Anopheles baimai* Sallum & Peyton, *Anopheles sundaicus* (Rodenwaldt) and *Anopheles minimus* Theobald and 4 are secondary vectors, *Anopheles philippinensis* Ludlow, *Anopheles jeyporiensis* James and *Anopheles varuna* Iyengar and *Anopheles annularis* Van der Wulp (Dash *et al*., 2008; Dev and Sharma, 2013). The regional distribution of major malaria vectors is shown in Figure 1.7. *Anopheles culicifacies* is the principal vector in the rural areas and widely distributed in the plains & riverine areas of India and contributes to almost 60-65% of malaria cases (Sharma, 1998). *Anopheles stephensi* is the urban and industrial malaria vector and contributes 10-12% of total malaria cases. *Anopheles minimus*, *Anopheles fluviatilis* and *Anopheles baimaii* (dirus) are the vectors of foothills, forest and forest

fringe areas like NE states, parts of Odisha, Chhattisgarh, Jharkhand and Madhya Pradesh. *Anopheles fluviatilis* contributes to 15% of malaria cases. Except *An. stephensi, An. jeyporiensis and An. varuna* all these mosquito species have been characterized as species complexes (Table 1.3) with varied number of morphologically indistinguishable sibling species which also vary in their role in malaria transmission (Subbarao,1998).

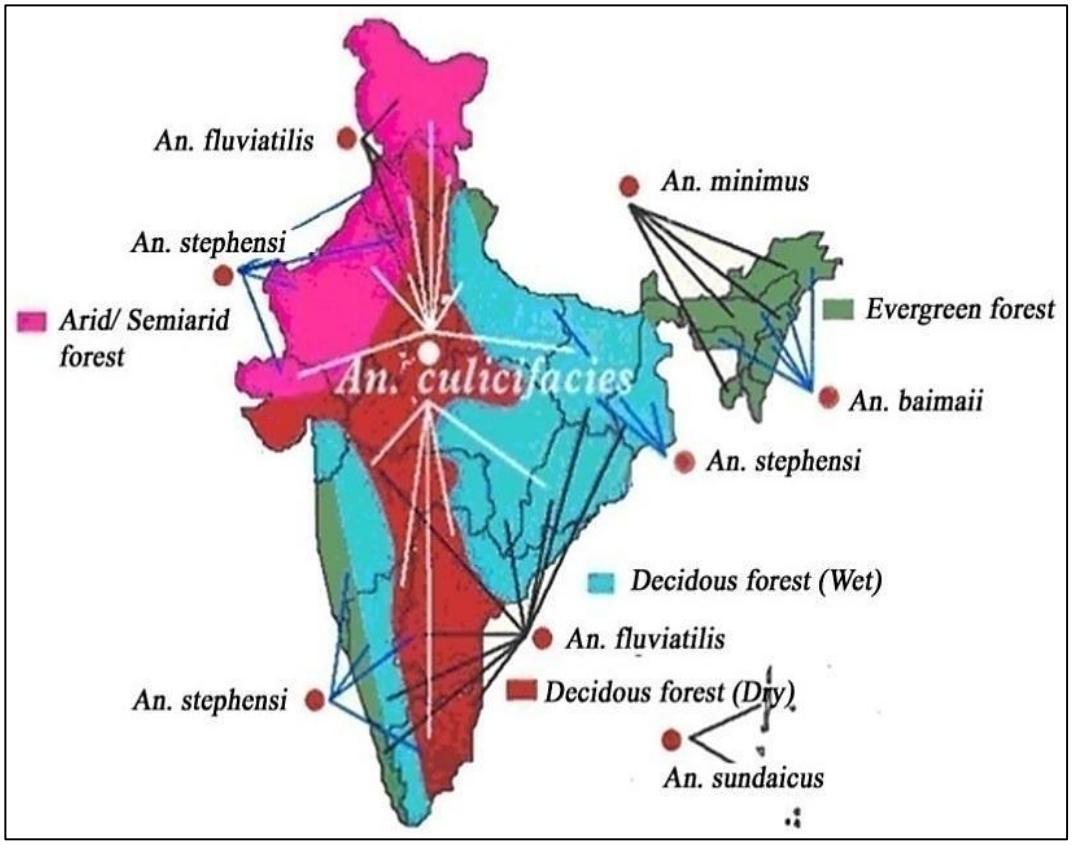


Figure 1.7: Distribution of malaria vectors in India. (Source: Dev and Sharma, 2013).

Table 1.3 List of the *Anopheles* vector species and species complexes found in India.

S. No	*Anopheles* Vectors	Number of sibling species recorded	Number of sibling species found in India	Members present in India	Ecological distribution
1	*An. stephensi*	-	-	-	Urban/peri-urban
2	*An. culicifacies*	5	5	A,B,C,D,E	Rural/peri-urban
3	*An. fluviatilis*	4	4	S,T,U,V	Foot hills/Forest
4	*An. minimus*	3	1	A	North eastern states
5	*An. dirus*	7	2	D, E	North eastern states and Karnataka
6	*An. sundaicus*	3	1	A new cytotype D	Andaman-Nicobar islands
7	*An. annularis*	2	1	A B	Rural (Maharashtra Odisha, Jharkhand)
8	*An. philippinensis-nivipes*	3	2	*An. philippinensis, An. nivipes*A	West Bengal and North eastern states
9	*Anopheles jeyporiensis*	-	-	-	North eastern states
10	*Anopheles varuna*	-	-	-	North eastern states

1.5 *Anopheles stephensi* Liston urban malaria vector in India

Anopheles stephensi Liston (Diptera: Culicidae) belongs to Neocellia series of subgenus Cellia Theobald (Harbach, 2004). *Anopheles stephensi* belongs to the sub-family anophelinae, which is characterized by a rounded scutellum and lacking of respiratory siphon in the larvae. Whereas adult palpi are characterized by apical and sub-apical bands, which are separated by a dark band and are speckled (Liston, 1901). Thorax is beset with broad scales and hind legs are not conspicuously marked in white (Figure 1.8). *Anopheles stephensi* has three ecological variants such as 'type',' mysorensis' and 'intermediate'. These forms can be differentiated on the basis of difference in number of ridges on egg floats and on the basis of spiracular indices. All these three forms are found in India and Iran. (Rao *et al*., 1938; Subbarao *et al*., 1987; Kumar *et al*., 2006; Vatandoost, 2006).

1.5.1 Classification of *Anopheles stephensi*

Kingdom : Animalia

Phylum : Arthropoda

Class : Insecta

Order : Diptera

Sub-order : Nematocera

Family : Culicidae

Sub-family : Anophelinae

Genus : *Anopheles*

Species : *stephensi* Liston

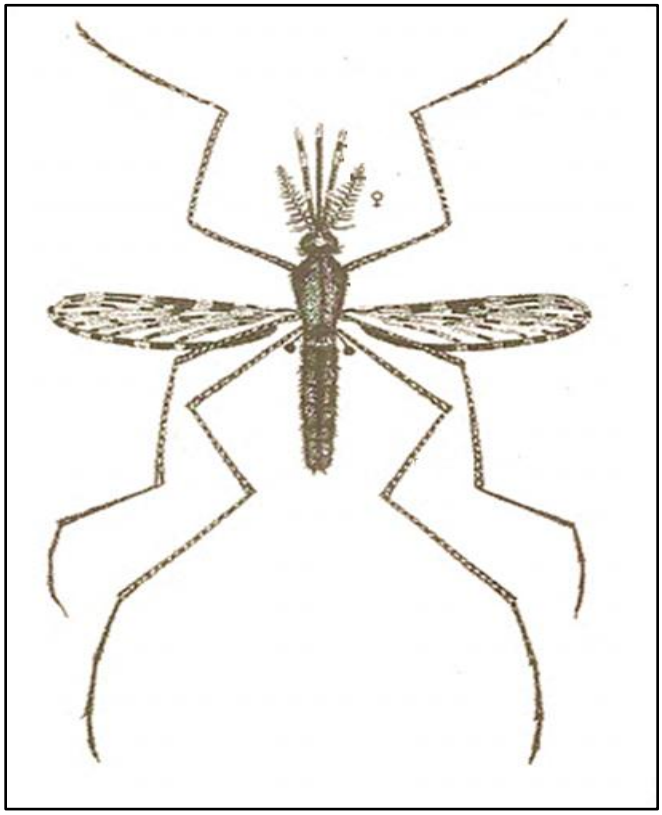


Figure 1.8: Female *Anopheles stephensi* Liston, showing main distinguishing characters (Source: Nagpal and Sharma, 1995).

1.5.2 Geographical distribution

Anopheles stephensi Liston is a predominant species and found across Indian subcontinent (Gillies, 1987; Sinka *et al*., 2011). It is a sub-tropical species and distributed throughout the Middle East and South Asia region viz. Afghanistan, Bahrain, Bangladesh, China, Egypt, India, Iran, Iraq, Myanmar (Burma), Nepal, Oman, Pakistan, Saudi Arabia, Thailand (Krishnan, 1961; Manouchehri *et al*., 1976; Pervez and Shah, 1989; Vatandoost *et al*., 2006). In India, *An. stephensi* is one of the major mosquito vectors that transmit malaria in urban areas. Its presence is reported in almost all parts of India, including Andaman and Nicobar Islands, Andhra Pradesh, Bihar, Delhi, Gujarat, Goa, Himachal Pradesh, Madhya Pradesh, Odisha, Punjab, Rajasthan and Uttar Pradesh (Singh, 2002) but rarely found at high altitudes in the Himalayas (Rao, 1984).

1.5.3 Breeding habitat

Anopheles stephensi prefers to breed in various artificial containers and curing water in construction sites and other industrial locations. In rural areas, larvae mainly thrive in fresh-water pools, stream margins and beds, catch basins, seepage canals, wells, domestic water-storage containers, stone pots and earthenware containers (Chakraborty *et al*., 1998; Kumar *et al*., 2006; Vatandoost *et al*., 2006). In rural areas, *An. stephensi* is prevalent round the year in canal-irrigated, non-canal-irrigated and riverine villages, but generally in low densities. In urban areas the larval breeding is found in overhead water tanks, domestic wells, room coolers, cisterns and roof gutters, buckets, barrels, ornamental tanks, underground cement tanks, evaporator coolers and curing water in construction sites, which are all man made habitats (Batra and Reuben, 1979; Batra *et al*., 2006) however, the greater numbers of larvae are typically found outdoors as compared with indoors (Biswas *et al*., 1992).

1.5.4 Resting and feeding preferences

Anopheles stephensi prefers humid places for resting and a surface temperature of around 25°C (Figure 1.9). This species rests primarily in poorly constructed human and animal shelters and on a variety of odd places (Chatterjee *et al*., 1993; Sumodan *et al*., 2004). *An. stephensi* is considered to be an endophilic and endophagic species (Reisen, 1986; Zaim *et al*., 1986). During the warmer summer months, it bites outdoors due to greater outdoor activity of humans and domestic animals. Only "type form" of *An. stephensi* is responsible for malaria outbreaks in urban areas where ongoing construction projects and associated tropical aggregation of labor from malaria endemic areas. This vector is a thermophilic species with longer flight range that maintains a high degree of contact with human population (Kalra, 1991). Peak biting time is between 2200 and 2400 hrs but varies from area to area and season to season (Kumar *et al*., 1995; Nagpal *et al*., 1995; Korgaonkar *et al*., 2012). Blood-meal analysis of females *An. stephensi* collected in urban areas indicated an increased tendency to feed on humans rather than cattle (Sweet and Rao, 1937). In rural areas, "mysorensis form" exists as zoophilic species and rests outdoors in cattle sheds.

Figure 1.9: *Anopheles* (Cellia) *stephensi* Liston, 1901 (Source: Wikipedia.org).

1.5.5 Vectorial capacity

An. stephensi "type form" is efficient in malaria transmission. The mysorensis form is found in rural areas and has a poor vectorial capacity due to its highly zoophilic behavior (Subbarao *et al*., 1987). The 'intermediate form' is typically found in peri-urban and rural areas but very limited information is available about its vectorial status. In India, *An. stephensi* was first incriminated as a vector of malaria in Mumbai in the year 1911, Madras and Gujarat in 1938, and Barooch in 1967. The Anthropophilic index (AI) of *An. stephensi* in Delhi varies from 0.45% (near the riverine zone) to 1.40% (non-riverine zone). In Kheda (city), the anthropological index (AI) was found to be 1.03%. However, higher AI values of 4.9% and 8.6% were recorded in the cities of Surat and Ahmedabad, respectively (Pervez and Shaw, 1989; Singh, 2002). Subbarao, (1998) indicated that the mysorensis form is considered an important vector in Iran. In southern Iran sporozoite rates have been reported ranging from 0.5 to 47% (Motabar *et al*., 1975; Eshghy *et al*., 1976; Manouchehri *et al*., 1976; Basseri *et al*., 2005).

1.5.6 Insecticide resistance

Anopheles stephensi has been found to be resistant to multiple insecticides such as DDT, dieldrin and malathion reported in Iran, Iraq, Saudi Arabia and Indian subcontinent (Busvine and Pal, 1969; Zahar, 1974; Kumari *et al*., 1998; Abai *et al*., 2008; Tikar *et al*., 2011). In India insecticide resistance has been reported from Delhi, Goa, Haryana, Rajasthan and Karnataka (Sharma *et al*., 1986; Thavaselvam *et al*., 1993; Subbarao *et al*., 1994; Bansal and Singh, 1996; Tiwari *et al*., 2010). Therefore indoor residual spraying is not used for *An. stephensi* control in urban areas.

1.6 Malaria parasite and its distribution

Malaria is a mosquito borne disease, which is caused by a protozoan parasite called *Plasmodium*, with following taxonomical classification.

Phylum: Protozoa

Subphylum: Apicomplexa

Class: Sporozoa

Subclass: Coccidia

Order: Coccidiida

Suborder: Haemosporina

Family: Plasmodiidae

Genus: *Plasmodium*

Subgenera: *Plasmodium*, Laverania

Malaria parasite was discovered by the Charles Louis Alphonse Laveran, who was a French army surgeon in 1880 in the blood of a malaria patient. Later, William MacCallum in 1897 identified the sexual stages in the blood of bird infected with malaria like parasite haematozoan, *Haemoproteus columbae*. In 1890, Raimondo Filetti and Giovanni Batista Grassi first introduced the names *Plasmodium malariae* and *Plasmodium vivax* for two of the malaria parasites that affect humans. Finally, Ronald Ross in 1897, elucidated the entire transmission cycle in culicine mosquitoes and birds infected with *Plasmodium relictum*. William H. Welch, in 1897 named the malignant tertian malaria parasite, *Plasmodium falciparum*. The fourth human malaria parasite, *Plasmodium ovale* was described by John William Watson Stephens in 1922. Recently, *Plasmodium knowlesi*, which is typically a monkey parasite, has been documented as fifth species which currently has been also found to infect humans in many countries of South East Asia region (Daneshvar *et al*., 2009). Between 1898-1900 several scientists such as Giovanni Battista Grassi, Giuseppe Bastianelli, Amico Bignami, Camillo Golgi, Angelo Celli, and Ettore Marchiafava demonstrated that malaria parasites were transmitted by anopheleine mosquitoes. Henry Shortt and Cyril Garnham in 1948 discovered that malaria parasites first developed in the liver cells before entering the blood stream and Wojciech Krotoski demonstrated the final stage in the life cycle and the

presence of dormant stages in the liver in 1982 (Francis, 2010). Worldwide, more than 200 *Plasmodium* species have been identified, which causes infection in reptiles, birds and mammals (Rich and Ayala, 2006). Out of which, only five species were known to cause malaria in humans, namely *Plasmodium vivax*, *P. falciparum*, *P. malariae*, *P. ovale* and *P. knowlesi*.

1.6.1 Distribution of *Plasmodium* species

Nearly 85% of malaria cases reported in Africa are caused by *Plasmodium falciparum* and the rest are due to other three species. *Plasmodium vivax* is a geographically widespread species, occurring in Asia, Central and South America, the Middle East, where 70-90% of malaria incidence is due to this species and the rest due to *Plasmodium falciparum* (Carter and Mendis, 2002; Rich and Ayala, 2006). *Plasmodium malariae* is found in Africa, some parts of India, Western Pacific and South America. *Plasmodium ovale* infection is restricted to tropical Africa, Philippines and New Guinea (Carter and Mendis, 2002). *Plasmodium knowlesi* has been reported from the South East Asian countries such as Thailand, Vietnam, Malaysia, Myanmar, Philippines (Singh *et al*., 2004; Cox-Singh *et al*., 2008; Chaturong *et al*., 2009; Daneshvar *et al*., 2009; Eede *et al*., 2009). In India, both *Plasmodium falciparum* and *Plasmodium vivax* are prevalent species and unevenly distributed throughout the country (Das *et al*., 2012) (Figure 1.3). *Plasmodium falciparum* accounts for about 50-55% of cases and *Plasmodium vivax* causing 40-45% of the infections recorded in the year 2014 (NVBDCP). The sporadic cases of *Plasmodium malariae* have been reported from Odisha, West Bengal, Madhya Pradesh, Karnataka, Tamil Nadu, Kerela and Assam (Roy *et al*., 1977, 1979; Panda *et al*., 1990; Sharma, 1993; Das *et al*., 1996; Dev, 2000; Mohapatra *et al*., 2008; Bharti *et al*., 2013). However, *Plasmodium ovale* is extremely rare with only a few cases reported from hilly and forested areas (Jambulingam *et al*., 1989; Mishra *et al*., 1999; Prakash *et al*., 2003; Marathe *et al*., 2006; Singh *et al*., 2013).

1.6.2 Life cycle of *Plasmodium* in mosquito vector and human host

Mosquitoes act as a vector and transmit *Plasmodium* parasite to humans. Malaria parasite exhibits a complex life cycle which involves two hosts, a vertebrate host (human) and an insect vector (mosquito). The sexual phase of development occurs inside the mosquito host (primary) whereas asexual phase in the humans (secondary host). All the species of *Plasmodium* follow similar life cycle pattern with only minor species specific variations.

1.6.3 Development of *Plasmodium* in *Anopheles* vector

The sexual phase of malaria parasite is completed inside the mosquito host, a process known as sporogony which is a highly complex process that involves transformation of malarial parasites into six different forms. The life cycle of the *Plasmodium* in vector begins when a female mosquito sucks blood of an infected human host having malaria parasite circulating in the blood. Along with the blood meal thus both asexual stages and gametocytes of the parasite enter inside the mosquito gut. The sexual stages i.e. male and female gametocytes then undergo rapid developmental events leading to formation of extracellular gametes. Each male gametocyte (microgametocyte) gives rise to eight male gametes (microgametes), which are like thin threads called flagella. Each female gametocyte forms a single female gamete known as macrogamete after shedding three polar bodies. Within hours, male and female gametes fertilize to form a zygote, which then transforms into an elongated, motile ookinete. The latter then penetrates the gut epithelial cells and develops between inner basal lamina of epithelial cell and outer cuticle in the form of an oocyst. The nucleus of the oocyst undergoes mitotic divisions resulting in the formation of thousands of sporozoites. The mature oocysts then rupture and release sporozoites in the body cavity i.e. haemocoel of the mosquito. The sporozoites swim in the fluid (haemolymph) filled haemocoel and finally invade the salivary glands. The developmental cycle of malaria parasite inside the mosquito may take 8 to 10 days in case of *P. vivax* infection, 9-10 days in *P. falciparum* infection, 14-16 days in *P. malariae* infection and 12 to 14 days in *P. ovale* infection. Transmission occurs when this infected mosquito bites a healthy human host which marks the beginning of a new cycle in the human host.

1.6.4 Development of *Plasmodium* in human host

The development of malaria parasite inside the human host begins with the bite of an infected mosquito that injects saliva containing sporozoites into blood. This phase of life cycle is known as asexual reproduction phase which is divided into two different phases: pre-erythrocytic **(**Exo-erythrocytic**)** phase and the erythrocytic phase**.** During the pre-erythrocytic phase, the sporozoites carried through the blood circulation to the liver and invade hepatocytes and start dividing leading to the formation of schizonts. Each schizont produces thousands of small parasites called merozoites which are released into the blood stream. Each of the merozoites or rarely a few will invade the red blood cells (RBCs) and start the erythrocytic stage of development. In case of *Plasmodium vivax* as many as 10,000 merozoites are produced from a single hepatocyte, whereas in case of *Plasmodium falciparum* some 30,000, in case of *Plasmodium malariae* and *Plasmodium ovale* 15,000 merozoites may be produced. Depending on the parasite species this complete development period varies between 5 and 16 days. The first stage after invasion of RBC is a signet ring stage trophozoite. Most of the parasites develop into trophozoites. The amoeboid trophozoites mature into schizonts, which again rupture releasing merozoites. Each mature schizont gives birth to new generation of merozoites which then infect the new RBCs. The erythrocytic cycle is completed in 24 hours in case of *P. knowlesi*, 48 hrs in case of *P. vivax* and *P. falciparum*, 50 hrs in *P. ovale* and 72hrs in *P. malariae*. Some of the merozoites differentiate into sexual stages i.e male and female gametocytes. This is an infective form of parasite when mature. When a female *Anopheles* mosquito bites a malaria infected person, along with the blood meal, it ingested these gametocytes which marks the beginning of sexual phase of life cycle in a mosquito host as explained earlier (Figure 1.10).

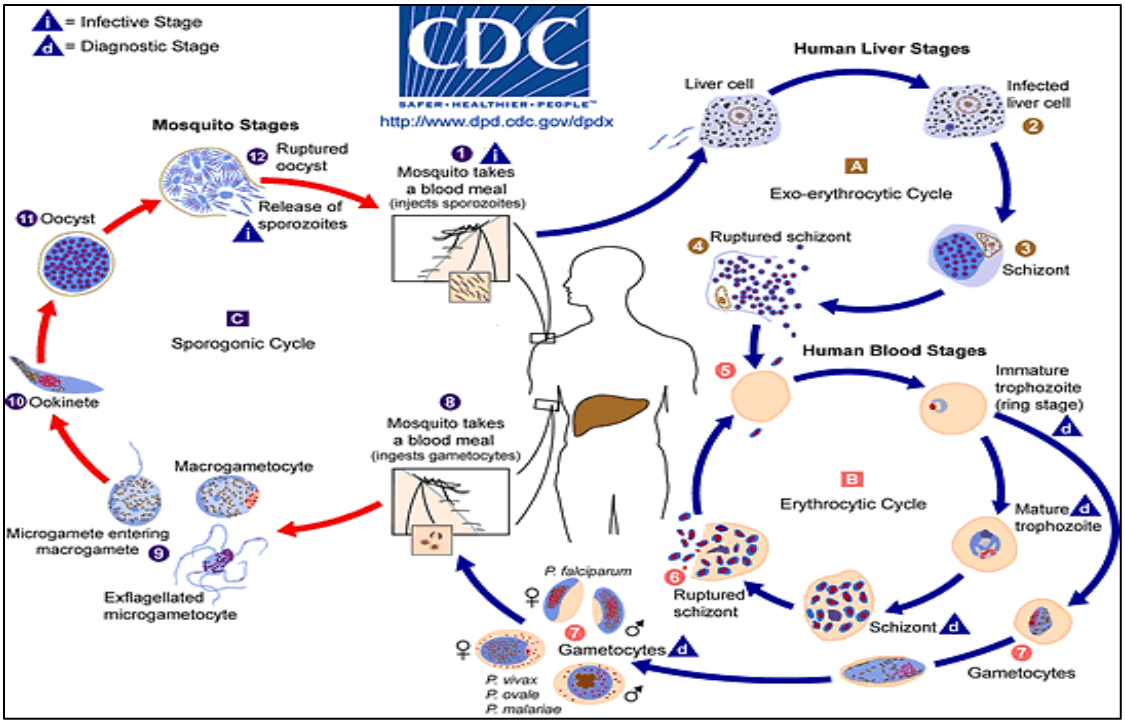


Figure 1.10: Life cycle of the malaria parasite in mosquito vector and human host. (Source: http://www.cdc.gov/malaria/history/index.htm).

1.7 Vector control strategies and approaches

Mosquitoes carry and transmit disease causing protozoan parasites and viruses among humans. Ever since the discovery that mosquitoes transmit malaria parasite between humans by Ronald Ross in 1897, the vector control has becomes an important component of malaria control strategy. Several strategies have been designed and used to reduce mosquito population in the last century. There are several newer methods being discovered and investigated for their application in the field settings. The strategies which are being used or explored are broadly covered under environmental management, chemical, biological, genetic (using Sterile Insect Technique) and molecular approaches.

1.7.1 Environment management

This strategy involves environmental modifications, manipulation and changes in human behavior and habitations. The methods used under this control strategy include source reduction i.e. removal or permanent destruction of mosquito breeding sites, installation of house screens, proper covering of stored water, filling of the breeding places and channelization of breeding source by effective drainage. These approaches have been

successfully utilized in mosquito control in some areas. There are some regions where these strategies are impractical.

1.7.2 Biological control

The control of mosquito larval stages through the use of fish e.g. *Gambusia affinis* (mosquito fish), *Lebistis reticulatus* (guppy), microbial agents such as fungi (e.g. *Lagenidium giganteum)* and, bacterial agent (e.g. *Bacillus thuringiensis israelensis* and *Bacillus sphaericus*), mermithid nematodes (e.g. *Romanomermis culicivorax*) is possible. *Bacillus thuringiensis israelensis* and *Bacillus sphaericus* have been used as an alternative to chemical larvicides as a cost effective strategy in urban and peri-urban areas (Kumar *et al*., 1995, 1996; Caldas de Castro *et al*., 2004). The advantages of biological control over chemical control are safety to humans and non-target organisms, low cost of production in some cases, and the lower risk of resistance development (Yap, 1985). However, biological control methods are not suited to all types of breeding habitats, and their effectiveness varies across various types of breeding sites (Fillinger *et al*., 2003).

1.7.3 Chemical control

Paris green (copper acetoarsenite) and petroleum by products were initially relied upon for Mosquito control. In light of their high toxicity and pollution of water sources, their usage in the field has been discontinued. With the discovery of dichloro-diphenyl-trichloroethane (DDT), the focus of malaria control strategies shifted to adult mosquito population control. DDT spraying has been widely demonstrated as a valuable tool for the control of malaria transmission (Pant, 1988). However, after continuous usage, mosquitoes developed resistance to DDT. Therefore, the focus from DDT has been shifted to other chemicals like OP compounds such as Malathion, carbamates and Synthetic pyrethroids like Deltamethrin, Lambda-Cyahalothrin, Cyfluithrin, etc. In addition, chemical larvicides, mineral oils, insect-growth regulators formulations are available for vector control in larval stages. Among all these larvicides, insect-growth regulators are generally safer for non-targets organisms and may become more useful and

popular in the future malaria control strategy (Graf, 1993). They have been introduced in India by the National Vector Borne Diseases Control Programme.

1.7.4 Molecular entomology approaches

Advances in the field of genetic engineering provide an unprecedented opportunity to conceive and create mosquito strains which can either suppress or replace the natural vector populations. There has been much interest in the potential of genetic technology for the vector control. Genetic control strategies fall into two categories.

1.7.4.1 Population suppression

In this approach mosquito vectors are genetically engineered to carry a lethal gene ensuring that no viable offspring will produce when they mate with wild individuals. This technique is known as Sterile Insect Technique (SIT) which is a species-specific and environmentally friendly method for insect population control. In this technique sterile males are produced using radiation mediated sterilization that renders their wild female mates incapable of having viable progeny. This approach has succeeded in controlling agricultural pests and mosquito vectors (Wilke *et al*., 2009). The Sterile Insect Technique is the most non-disruptive pest control methods. Unlike some other biologically-based methods, it is species specific, neither it releases any exotic agents into new environment nor it introduces any new genetic material into existing populations. This approach has the least impact on ecosystems and does not involve the spread of modified genes through the wild population.

1.7.4.2 Population replacement

In this approach, such mosquitoes are developed by genetic modifications that are incapable to transmit the disease and become refractory to the *Plasmodium* infection. This approach is still many years from application in field settings.

In order to develop/devise new genetic methods for vector control, knowledge about the genome, of a particular mosquito species is essential. Therefore, understanding the vector biology at genomic, proteomic and transcriptomic level will provide several leads for

developing novel approaches of vector and pathogen interactions. In recent years, genomics and post-genomics researches have produced lots of significant data that has been employed for the development of new tools to control vector population or to block pathogen transmission by them.

1.8 Mosquito genomics

Malaria is a complex disease transmitted by the female *Anopheles* mosquito. Malaria transmission is the outcome of a complex interaction between human host, malaria parasite, mosquito vector and environment, where mosquitoes play a central role in disease transmission. Vector control remains the most effective method to prevent the malaria transmission, however lack of information on genetic makeup of mosquitoes is a major obstacle in the development of modern and effective vector management tools. The genome of mosquito holds information on their genetic architecture as well as phenotype-specific proteomics and transcriptomics. Recent advances in genome sequencing technologies provide ideal tools for rapid and reliable genomic exploration of mosquitoes. Better understanding of mosquito genome will lead to better ways to control vector, and also to target parasite development inside the vector.

1.8.1 Advances in vector genomics

Recent advances in vector genomics offer new promise for the control of disease transmission via mosquito vectors. The availability of genome sequences of mosquito vectors has now made it possible to identify genes that are necessary for parasite development in the mosquito. Genome information will accelerate the development of genome wide screens for gene expression in vector tissues that interact with the parasites/pathogens. Understanding how the parasites and pathogens develop in the mosquito would enable formulation of new strategies to intervene in their transmission cycles. A major milestone in vector genomics was the completion of genome sequence of *Anopheles gambiae* (Holt *et al*., 2002). Following this, the genome of the yellow fever mosquito *Aedes aegypti* (Nene *et al*., 2007), *Culex quinquefasciatus* (Arensburger *et al*., 2010), *Anopheles darlingi* (Marinotti *et al*., 2013) and *Anopheles sinensis* (Zhou *et al*., 2014) were sequenced and published. Recently, a consortium of Institutes led by the

Broad Institute has carried out the sequencing of genomes and transcriptomes of *Anopheles* species, including 9 major and 5 minor vectors and 2 non-vectors of human malaria. Whole genome sequences of 16 *Anopheles* species and transcriptome data of 16 *Anopheles* species have been made publicly available through VectorBase, SRA and the Broad Institute website (Neafsey *et al*., 2013). The information about number of genes, transcripts and protein coding genes identified are provided in Table 1.4.

The genome *of Anopheles gambiae* was sequenced using a whole genome shotgun approach (Holt *et al*., 2002). Genome of *Anopheles gambiae* is larger than that of the fruit fly *Drosophila melanogaster*. The difference in genome size between *Anopheles gambiae* and *Drosophila melanogaster* is largely due to intergenic DNA. The first physical mapping of *An. gambiae* genome was conducted by in situ hybridization of about 2,000 bacterial artificial chromosome (BAC) clones on ovarian nurse cell polytene chromosomes. The first genome assembly of *Anopheles gambiae* contained 67 scaffolds equivalent to 227 mega base-pairs (Mbp) were assigned to chromosomes. However, approximately 18% of the assembled genome of *Anopheles gambiae* was represented in scaffolds that did not have a chromosomal location assigned. About 50 mega base pairs (Mbp) in the assembly were assigned with arbitrary order and orientation to an unmapped chromosome (Mongin *et al*., 2004). According to Cot analysis, 33% (about 86 mega base pairs) of the *Anopheles gambiae* genome corresponds to repetitive elements (Besansky and Powell, 1992).

The sequencing of *Aedes aegypti* and *Culex pipiens*, the mosquito vector for the Arbovirus, filarial worm and West Nile virus, respectively, shed further light on mosquito biology and mosquito species-specific gene functions. Interestingly the genome size of *Aedes aegypti* is ~5 times bigger than that of *Anopheles gambiae*. Around 50% of the *Aedes aegypti* genome consists of transposable elements, which eventually increase the average gene length relative to *Anopheles gambiae*. The *Culex quinquefasciatus* genome contains 18,883 protein-coding genes which is 22% larger than that of *Aedes aegypti* and 52% larger than that of *Anopheles gambiae.* It has multiple gene-family expansions, including olfactory and gustatory receptors, salivary gland genes, and genes associated with xenobiotic detoxification.

Table 1.4: Current status of genome projects of various mosquito vectors of human disease.

S. No.	Species	Strain	Genome Size (Mbs)	Total number of genes	Transcripts (bp)	Protein Coding	Scaffold count	Status
1	*Anopheles gambiae*	PEST	273	13,624	15,478	12,843	8	Completed
2	*Culex quinquefasciatus*	Johannesburg	579	19,363	19,427	18,968	3,172	Completed
3	*Aedes aegypti*	Liverpool	1,311	17,479	18,838	15,797	4,757	Completed
4	*Anopheles albimanus*	STECLA	170	12,335	12,418	11,911	204	Community annotation
5	*Anopheles arabiensis*	Dongola	246	13,678	13,736	13,162	1,214	Community annotation
6	*Anopheles atroparvus*	EBRO	224	14,230	14,230	13,776	1,371	Community annotation
7	*Anopheles christyi*	ACHKN1017	172	11,156	1,233	10,738	30,369	Community annotation
8	*Anopheles culicifacies A*	A-37	202	14,882	14,882	14,335	16,162	Community annotation
9	*Anopheles darlingi*	Coari	136	10,948	10,948	10,457	2,221	Community annotation
10	*Anopheles dirus A*	WRAIR2	216	13,239	13,350	12,781	1,266	Community annotation
11	*Anopheles epiroticus*	Epiroticus2	223	12,602	12,654	12,078	2,673	Community annotation
12	*Anopheles farauti*	FAR1	180	13,703	13,703	13,217	550	Community annotation

Table 1.4 contd: Current status of genome projects of various mosquito vectors of human disease.

S. No.	Species	Strain	Genome Size (Mbs)	Total number of genes	Transcripts (bp)	Protein Coding	Scaffold count	Status
13	*Anopheles funestus*	FUMOZ	225	13,756	13,897	13,344	1,392	Community annotation
14	*Anopheles maculatus B*	maculatus3	141	15,046	1,546	14,835	47,797	Community annotation
15	*Anopheles melas*	CM1001059	227	16,666	16,666	16,149	20,281	Community annotation
16	*Anopheles merus*	MAF	251	14,415	14,415	13,886	2,753	Community annotation
17	*Anopheles minimus A*	MINIMUS1	201	13,128	13,221	12,653	678	Community annotation
18	*Anopheles quadriannulatus A*	SANGWE	283	13,857	13,923	13,349	2,823	Community annotation
19	*Anopheles sinensis*	SINENSIS	241	15,280	15,280	14,791	11,270	Community annotation
20	*Anopheles stephensi*	Indian	221	12,350	12,350	11,789	23,371	Community annotation
21	*Anopheles stephensi*	SDA-500	225	13,650	13,788	13,113	1,100	Community annotation
22	*Anopheles coluzzii*	Mali-NIH	224	14,703	-	14,703	10,521	Community annotation

Source: https://www.**vectorbase**.org (accessed on 30th April 2015)

1.8.2 Genome analysis of *Anopheles stephensi*

The genome sequence of *Anopheles stephensi* 'type' form of the Indian strain was published in 2014 (Jiang *et al*., 2014). The type form of *An. stephensi* is responsible for the most of the urban malaria transmission across the Indian subcontinent and the Middle East. The whole genome sequencing was carried out using a combination of, 454 GS FLX, PacBio and Illumina sequencing and verified by bacterial artificial chromosome BAC clones and expressed sequence tags (ESTs) (Jiang *et al*., 2014). The assembled genome size of the *Anopheles stephensi* is 221Mb, which is smaller than the genome size of *An. gambiae* (278 Mb), *Aedes aegypti* (1311Mb), *Culex quinquefasciatus* (579Mb) but considerably larger than the genome size of fruit fly *D. melanogaster* (180Mb). Physical mapping assigned 62% of the *An. stephensi* genome onto chromosomes, which is in complete agreement with the chromosome based genome assembly. A total of 12,350 genes (209,483,518 base pairs) comprising of 11,789 protein coding genes, 434 tRNA genes, 127mi RNA genes and 53 fragments of rRNA genes were annotated using a combination of homology and de novo prediction. The average transcript length was 3,666 bp and the average number of exons per transcript was 4.18. Around 89% of predicted genes have orthologs present in *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster*. Comparative analysis of gene families shows that most of the gene families are similar between *Anopheles stephensi* and *Anopheles gambiae*. Very few gene families showed difference in numbers between the two species. A high level of orthology was observed between *Anopheles stephensi* and *Anopheles gambiae*. The repetitive DNA constitutes the 7.1% of the total assembled genome which is relatively lower as compare to *An. gambiae* (16%). Thus the variation in the size of the genomes is due to the different amounts of repetitive DNA in these two species. The genomic and transcriptomic data generated in this study provides new insights into vector biology and vector-parasite interactions, and also provides a platform for fundamental and translational research for scientific community.

1.8.3 The International *Anopheles* genome project

1.8.3.1 Genome analysis of 16 major *Anopheles* vectors of malaria parasites

The whole genome sequencing and analysis of the 16 different *Anopheles* vectors of malaria was a joint effort of Broad Institute, VectorBase and members of the *Anopheles* Genomes Cluster Coordinating Committee (AGCC). The project proposal was funded in the year 2008, by the National Human Genome Research Institute (NHGRI) and the National Institute of Allergy and Infectious Diseases (NIAID) of the U.S. under National Institutes of Health. The broad goal of this proposal was to undertake comparative analysis of the genomes of major malaria vectors with respect to their vectorial capacity and genetic relatedness. The 16 mosquito species selected included *An. arabiensis*, *An. quadriannulatus* A, *An. merus*, *An. melas*, *An. christyi*, *An. epiroticus*, *An. stephensi* (SDA 500), *An. maculatus***,** *An. funestus*, *An. minimus*, *An. culicifacies* A, *An. farauti*, *An. dirus* s.s., *An. atroparvus*, *An. albimanus*, *An. sinensis*. These 16 *Anopheles* species were selected based on geographical locations, ecological conditions and variable vectorial capacities for the transmission of malaria. The four earlier sequenced species i.e. *Anopheles stephensi*, *Anopheles gambiae* PEST strain and its two incipient species *An. gambiae* M and S forms are also included in these16 mosquito species. Recently, the genome sequence of these 16 *Anopheles* vectors has been published. Neafsey *et al*., (2015) reported in each mosquito species between 10,738 to 16,149 protein coding genes (Table 1.4). Comparative analysis showed that gene fusion and fission played a substantial role in the evolution of mosquitoes. Following the submission of genome sequence to VectorBase database, Fontaine *et al*., (2015) carried out an in-depth analysis of the new genomic data, to find out the species relationships and introgression (inter-specific gene flow) within the *An. gambiae* complex. They found that the ancestor of *An. gambiae* and *An. coluzzii* were separated from other species about two million years ago and are distantly related to *An. arabiensis*. They have also found out that pervasive autosomal introgression occurred between these vectors and the direction of genetic transfer occurred from *An. arabensis* in to the ancestor of *An. gambiae* and *An. coluzzii*. The availability of multiple *Anopheles* vector genome sequence has completely changed the landscape of fundamental research on mosquitoes. This has opened up new avenues to find mechanisms to interfere with the spread of malaria transmission and reducing

malaria disease burden. Availability of genome sequences of *An. gambiae* and two human malaria parasites *Plasmodium falciparum* (Gardner *et al*., 2002) and *Plasmodium vivax* (Carlton *et al*., 2008) has already enabled several large-scale multi-OMICs and molecular biology investigations to discover a large number of molecules involved in vector-*Plasmodium* interactions. Several such potentially important molecules have been identified in vector mosquitoes, particularly in *Anopheles gambiae,* using differential expression analysis (Vlachou *et al*., 2005; Dana *et al*., 2006; Mikolajczak *et al*., 2008; Gonzalez-Lazaro *et al*., 2009; Serrano-Pinto *et al*., 2010; Vijay *et al*., 2011). Using gene silencing approaches, over 94 mosquito genes have been validated for their role in the promotion or inhibition of growth of *Plasmodium* parasites in the mosquito (Sreenivasamurthy *et al*., 2013). The details about such molecules mediate the vector-pathogen-host interactions are provide in Appendix Table 1 A.

1.9 VectorBase: A resource for integration of genome information of disease vectors

Whole genome sequences of increasing number of vector species has resulted in availability of huge amount of data from multi-OMICs platforms. VectorBase (http://www.vectorbase.org), a NIH-NIAID funded resource centre for invertebrate vectors of human pathogens, is established to store and manage information on sequences, gene sets, and controlled vocabulary for features and traits of vectors. An important feature of VectorBase is the Ensembl Genome Browser, developed at the European Bioinformatics Institute, which displays genome-related annotations mapped onto the genome sequences. The data stored in VectorBase is available for download as flat files, either in fasta format or as database dumps. VectorBase has also developed a Community Annotation Pipeline that provides the format and tools for manual annotations to be submitted directly by the research community.

The next section is an introduction to the proteomics using mass spectrometry and proteogenomic analysis followed by a literature review on the proteomics study on *An. stephensi*. The next section also includes a brief history on the use of proteomics data for genome annotation.

1.10 Proteomics using tandem mass spectrometry

Proteomics is the large scale study of proteins expressed in a living system, particularly their structure and function. Wilkins *et al*., (1996) first coined the term proteome to describe the set of proteins encoded by the genome. In the past few years, there have been remarkable advances in proteomic technologies. Mass spectrometry has become one of the most powerful proteomic tools. There are other techniques being used for proteome characterization such as two -dimensional electrophoresis (2DE), two-hybrid analysis and protein micro arrays. However these techniques failed to achieve the depth of informative proteome analysis as compared to mass spectrometry. Currently two major types of approaches are being used to identify proteins using mass spectrometry. First approach is top-down proteomics i.e. analysis of intact proteins and second one is bottom-up proteomics i.e. analysis of peptide mixtures from digested proteins. These two proteomics approaches can identify protein expression, protein-protein interactions and post-translational modifications.

In top-down approach, intact proteins from complex biological samples are separated using traditional techniques such as 2-D gel electrophoresis or liquid chromatography followed by differential expression analysis using spectrum analysis or gel imaging platforms. The mass measurement of intact proteins occurs inside the mass spectrometry followed by their direct fragmentation without prior digestion with proteases. Using this approach, direct detection of the native molecular mass of intact protein is possible. This provides a complete characterization of protein isoforms and any post translational modifications, which is not possible in bottom-up approach. This technique is very useful for protein biomarker discovery using body fluids (plasma, serum, cerebrospinal fluid, urine) from humans and animals.

In bottom-up mass spectrometry approach, complex protein mixtures are first enzymatically digested with proteases, resulting peptide samples are separated by single and multi-dimensional liquid chromatography (LC) and then analyzed by MS and MS-MS to sequence the peptides. Finally peptides are identified from MS/MS spectra using automated database search programs. This process is known as liquid chromatography-

tandem MS or LC-MS/MS. The workflow for protein identification by using bottom-up Mass spectrometry approach is shown in Figure 1.11. Bottom-up proteomic approach has become the method of choice for protein identification, quantification, post-translational modification and genome annotation. The most popular commercially and freely available peptide search engines are Mascot (Perkins *et al*., 1999) and SEQUEST (Eng *et al*., 1994), whereas X!Tandem is an example of an open source program (Craig and Beavis, 2004). The acquired MS/MS spectrum is compared against the theoretical fragmentation patterns of peptides constructed in a database. A match score based on user specific search parameters is generated. The highest scoring peptide is selected for further analysis. The peptides are grouped based on their hit to the corresponding protein. Protein inference is made by distributing the peptides among all their corresponding proteins and generating a minimal list of proteins observed in the sample study (Nesvizhskii, 2007). The advantage of bottom-up approach is the ability to achieve high-resolution separations. This approach also requires a very few number of samples for protein identification and relative expression as compare to top-down approach. This approach is widely used and accepted because hundreds of proteins can be identified with in a small numbers of samples. Although, this method has been successfully used for protein profiling and identification, but it has also one disadvantage that it fails to identify smaller proteins and peptides which have fewer proteolytic cleavage sites.

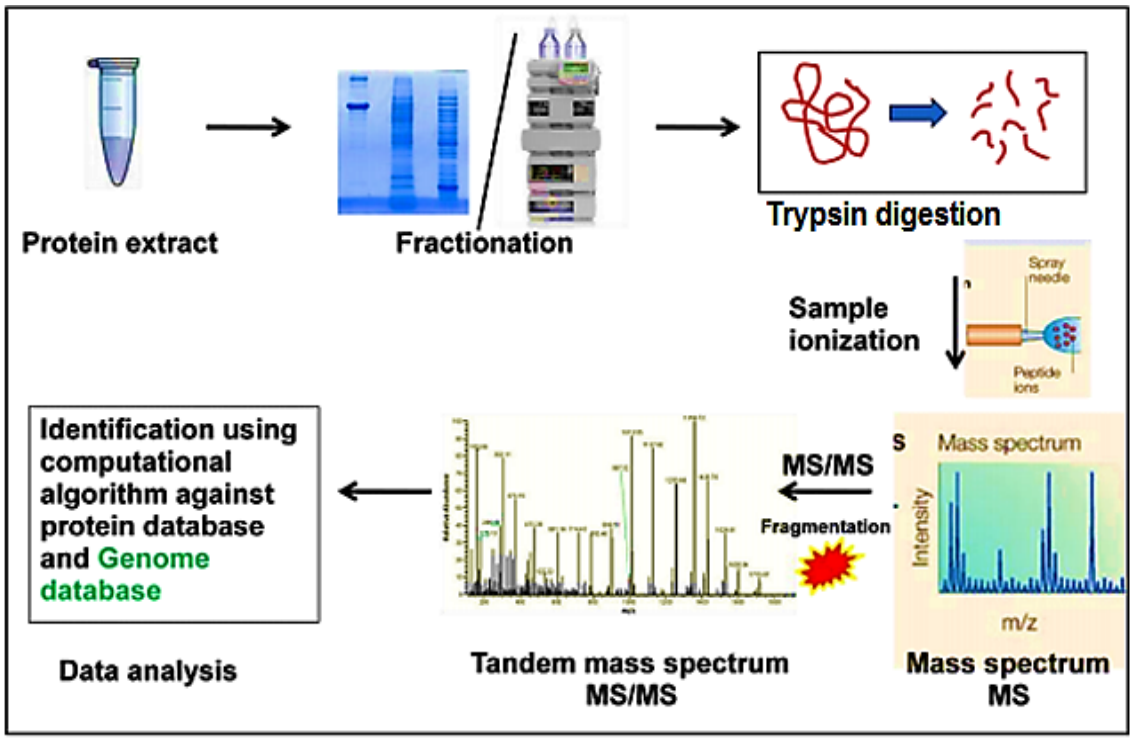


Figure 1.11: Workflow for protein identification by using bottom-up MS approach.

1.10.1 Proteomics of mosquito vectors

The proteomic studies of mosquito vectors are now a rapidly evolving field of research. Efforts are being made to analyze and catalogue the proteins expressed in different mosquito species, tissues, cell types and different physiological stages. The proteome analysis of mosquitoes has revealed several processes viz., changes in the hemolymph, midgut, peritrophic membrane, egg chorion/shell salivary and ovarian proteins after blood feeding, understanding vector-parasite interactions, the resistance/susceptibility response of mosquito to different parasites. Mosquito proteomic studies have given insights into the molecular mechanisms of embryogenesis, embryonic development, diapause, their regulation mechanism. This has also helped in tracing out the evolutionary history of proteins by phylogenetics and toxicology studies (Sehrawat and Gakhar, 2014). However, mosquito proteomics are still in the developing stage with research, primarily from India focusing on a variety of fields.

1.10.2 Progress in mosquito proteomics

During the last decade, various approaches have been followed to describe protein content in different organs of mosquito vectors. Midgut of mosquito is the chief locus of digestion of blood meal, absorption, secretion of digestive enzymes, egg development and parasite development and transmission. The protein profile of mosquito midgut from male, sugar and human blood fed female *Anopheles gambiae* was carried out using 2-D gel electrophoresis followed by mass spectrometry (Prevot *et al*., 2003). In this study, 375 proteins were identified from the male and female midgut of *An. gambiae*. Ten proteins were identified as specifically produced in male midgut, which were absent in sugar-fed or blood-fed females. Similarly 23 proteins were identified as female-specific proteins, which were not present in male midgut. This study provided information about the sex specific and blood meal induced protein expression in the midgut of *An. gambiae.* Pinto *et al*., (2010) carried out proteomic analysis of *Anopheles albimanus* midgut infected with *Plasmodium berghei* and identified 21 spot proteins that were differentially expressed in the midgut of mosquito during the immune challenge. These proteins have important implications for understanding the process of blood meal digestion and understanding parasite-vector interactions.

Hemolymph of mosquito plays a very vital role in transporting nutrients and hormones, eliminates metabolic wastes and protecting against harmful microorganisms. Puskewitz and Shi, (2005) utilized the proteomic approach to investigate the hemolymph proteome of adult *An. gambiae* mosquito and identified 280 proteins spots, out of which 28 spots, representing 26 individual proteins. Most of these proteins have known or predicted functions in immunity, iron transport like ferritin or lipid biology. In this study one novel protein belong to ML family (involved in lipid recognition) has been identified, which is absent in other insect species. Kalume *et al*., (2005) analyzed the salivary gland proteome of *An. gambiae,* and identified 69 unique proteins, 57 of which were novel. The identified proteins were found to be involved in carbohydrate, protein and nucleic acid metabolism, transport and energy pathways. Choumet *et al*., 2007 conducted proteomic study on the salivary glands and saliva of *An. gambiae* by using multiple techniques of 1-DE, 2-DE and LC-MS/MS based profiling. A comparative analysis using iTRAQ labeled proteins

from the salivary gland of mosquito infected with *P. berghei* and uninfected mosquitoes shows that five secreted proteins expression level were altered in parasite infected mosquito (Choumet *et al*., 2007).

Lefevre *et al*., (2007) compared the head proteome of *An. gambiae* mosquito infected with *P. berghei* with that of uninfected mosquitoes and identified 12 protein spots. These proteins were functionally classified as belonging to metabolic, synaptic, signalling, molecular chaperone and cytoskeletal groups. This study shows that there is an altered energy metabolism in sporozoite-infected mosquito's head**.** Two proteins identified in the study have been demonstrated to be involved in behavioral modifications in other host-parasite systems.

Mass spectrometry based study by Amenya *et al*., (2010) identified 44 proteins in the eggshells of the female *Anopheles gambiae* mosquitoes. Among 44 proteins, a group of seven putative chorion proteins and 2 vitelline membrane proteins were identified. Functional characterization of these proteins could lead to a greater knowledge of the mosquito's reproductive strategies. They have also identified seven odorant binding proteins associated with the mosquito eggshell. Dinglasan *et al*., (2009) characterized protein components of the peritrophic matrix that forms around the blood bolus in adult females. A total of 209 proteins were identified, of which 123 identified were with predicted signal peptides, 17 proteins were identified with transmembrane domain and 9 were new chitin-binding peritrophic matrix proteins.

No global proteomic work on mosquito vectors have yet been reported from any laboratory in India other than proteogenomic analysis of *An. gambiae* (Chaerkady *et al*., 2011). The majority of the work in the field has been carried out on *Anopheles gambiae*, however very limited studies have been carried out in other mosquito species including vectors in India (Butler *et al*., 2009; Saboia-Vahia *et al*., 2012; Vijay *et al*., 2014). The summary about the proteomic profiling carried out in different mosquito species is provided in Table 1.5.

Table 1.5: Summary of proteomic studies carried out in different mosquito species.

S. No.	Mosquito species studied	Organ/tissue details	Technique used	Number of proteins identified	Reference
1	***Anopheles gambiae***	Midgut	2-D Gel electrophoresis and Mass spectrometry	375	Prevot *et al*., 2003
		Hemolymph	Two dimensional SDS–PAGE and micro-sequencing	280	Puskewitz and Shi, 2005
		Salivary gland	LC-MS/MS and 1-DE, 2-DE and LC-MS/MS	69 &127	Kalume *et al*., 2005 & Choumet *et al*., 2007
		Salivary gland, Malpighian tubule, Head, Midgut, Male accessory glands, Testis, Viscera, Ovaries, Larvae and Pupae	LTQ-Orbitrap XL ETD mass spectrometer	5963	Chaekady *et al* ., 2011
		Head	2-D DIGE coupled with MS	12	Lefevre *et al*., 2007
		Egg shell	Mass spectrometry	44	Amenya *et al*., 2010
		Peritrophic membrane	1-DE followed by LC-MS	209	Dinglasan *et al*., 2009
2	***Anopheles stephensi***	Salivary gland	SDS- PAGE followed by Edman degradation of polyvinylidene difluoride (PVDF)	13	Valenzuela *et al*., 2003

		Midgut	2-D Gel electrophoresis	29	Prevot *et al*., 1998
		Salivary gland	1DE LC-MS/MS, 2D gel profile and MALDI-TOF	159 & 31	Vijay *et al*., 2014 & 2015
3	***Anopheles dirus***	Salivary gland	SDS-PAGE	7 spots	Jariyapan *et al*., 2007
4	***Anopheles barbirostris species A2***	Salivary gland	2DE and Mass spectrometry	30	Jariyapan *et al.*, 2012
5	***Anopheles albimanus***	Midgut	2DE and Mass spectrometry	19	Cázares-Raga *et al*., 2014
6	***Anopheles campestris***	Salivary gland	2-DE and NanoLC-MS analysis	20	Sor-suwan *et al*., 2013
7	***Aedes aegypti***	Brush border membrane vesicle (BBMV)	Two-dimensional gel electrophoresis (isoelectric focusing and SDS-PAGE) and a shotgun two-dimensional liquid chromatographic (LS/LS)	119	Butler *et al*., 2009
		Eggshells	SDS-PAGE/mass spectrometry	130	Marinotti *et al*., 2014
		Egg shell chorion	2-D Gel electrophoresis and MALDI-TOF MS	38	Li, 2003
		Fat body	2DE profiles	500 protein spots	Shih and Fallon, 2001
8	***Aedes albopictus***	Midgut	2DE & (LC-MS/MS)	56	Saboia-Vahia *et al*., 2012
9	***Culex quinquefasciatus***	Salivary gland	2DE and Mass spectrometry	67	Djegbe *et al*., 2011

1.10.3 Proteomic studies on *Anopheles stephensi*

As mentioned elsewhere, *Anopheles stephensi* is a major malaria transmitting vector in Asia, whose genome has been sequenced recently (Jiang *et al*., 2014). To understand the interaction between vector and pathogen, it is important to know the proteins expressed in mosquito host produced in response to pathogen invasion. There has been very limited information on *An. stephensi* proteome so far, which limits understanding of the mechanism of *An. stephensi-Plasmodium* interaction, vector immunity, parasite development and survival.

Prévot *et al*., (1998) carried out first study on midgut proteins of blood and sugar fed *Anopheles stephensi* lines of different susceptibility to *Plasmodium falciparum* infection using two-dimensional gel analysis approach. They identified 29 proteins which were exclusively present in *P. falciparum* susceptible lines as compared to the less susceptible lines. The midgut protein profile of sugar fed mosquitoes was found similar in both the lines of *Anopheles stephensi.* Valenzuela *et al*., (2003) identified twelve proteins in *Anopheles stephensi* salivary gland by performing SDS-PAGE and Edman degradation and reported 33 full length novel cDNA sequences. However, none of these studies have used advanced proteomic technologies including multiple fractionation strategies and high-resolution mass spectrometry. Recently Vijay *et al*., 2014 carried out proteomic analysis of salivary glands from female *An. stephensi* using mass spectrometry. They identified a catalogue of 159 proteins, of which 36 were known and 123 were novel proteins. Most of the proteins identified were involved in regulation of blood coagulation cascade, signal transduction, mosquito immunity related and some were found to be involved in energy pathways. Vijay *et al*., (2015) have carried out a study using quantitative proteomic approach to identify differentially expressed proteins in DDT/deltamethrin resistant strains of *Anopheles stephensi.* They identified 31 differentially expressed putative modulated proteins which have an impact on insecticide resistance and xenobiotic detoxification in *An. stephensi*.

Fontaine *et al*., 2012 compared the salivary proteome of six major *Anopheles* species and demonstrated that salivary proteins from closely related species exhibit molecular

diversity although they have common pharmacological activities. However, so far very limited number of studies has been carried out on the proteomic profiling and functional characterization of proteins of *Anopheles stephensi*. Therefore, more in-depth proteomic studies are needed to be carried out on this important vector in order to understand the molecular mechanisms involved in mosquito-pathogen interactions, disease transmission, refractory status of naturally occurring and transgenic mosquitoes, and development of insecticide resistance.

1.11 Proteogenomics and genome annotation of mosquito vectors

The collective process of identifying genes and assigning function to each gene is commonly referred to as genome annotation. Conventional approaches used in genome annotation, i.e. identifying and characterizing genes, are mainly dependent on computational algorithms. These methods are far from perfect leaving most existing genome annotations riddled with errors such as missing genes, erroneous exon-intron boundaries, and incorrect assignment of protein start sites. An integrated analysis of transcriptome and proteome can provide a near complete annotation of genome. Proteogenomics commonly refer to studies that use proteomic information, often derived from mass spectrometry, to improve gene annotation (Renuse *et al*., 2011). Proteogenomic analysis not only validates computationally derived gene models at proteomic level, but also allows the identification of novel protein-coding genomic regions, which are missed by gene prediction algorithms. Therefore, genome annotation using mass spectrometry derived proteomic data is complementary to other gene prediction methods (Mann and Pandey, 2001). Several studies have demonstrated the use of mass spectrometry based proteomic approaches to correct or validate gene annotations including *Homo sapiens* (Desiere *et al*., 2005; Molina *et al*., 2005; Fermin *et al*., 2006; Menon *et al*., 2009; Kim *et al*., 2014), *Anopheles gambiae* (Kalume *et al*., 2005; Chaerkady *et al*., 2011), *Mycobacterium tuberculosis* (Kelkar *et al*., 2011), *Candida glabrata* (Prasad *et al*., 2012), *Schmidtea mediterranea* (Bocchinfuso *et al*., 2012), *Toxoplasma gondii* (Xia *et al*., 2008), *Plasmodium falciparum* (Lasonder *et al*., 2002), *Yersinia pestis* (Payne *et al*., 2010), *Deinococcus deserti* (Baudet *et al*., 2010),

Leishmania donovani (Nirujogi *et al*., 2013), *Pristionchus pacificus* (Borchert *et al*., 2010) and other organisms.

The preliminary annotation of the *Anopheles gambiae* genome was based on automated gene prediction programs. The genome annotation of *Anopheles gambiae* was largely based on the *Drosophila melanogaster* data. Mass spectrometry analysis of salivary gland proteins of *Anopheles gambiae* was carried out in 2005 (Kalume *et al*., 2005). This proteogenomic annotation approach led to the validation of 23 known and 50 predicted transcripts at the protein level. A novel gene was identified based on peptide matches to the genomic region where no gene was predicted. Further, an extensive proteomic analysis of *Anopheles gambiae* was carried out in year 2011. In this study, 2,682 genome specific unique peptides were identified, which validated more than 6,000 gene annotations including revision of 80 novel gene models and 500 translational start sites (Chaerkady *et al*., 2011). Recently, Dwivedi *et al*., (2014) have carried out proteogenomic analysis of brain tissue of *Anopheles gambiae,* which led to the identification of 15 novel genes and 71 gene refinements. However, very limited studies have been carried out on other mosquito vectors using proteogenomic annotation approach. Such type of proteogenomic analysis, if extended to the other mosquito vectors such as *Aedes aegypti*, *Culex quinquefasciatus*, *Anopheles stephensi*, *Anopheles culicifacies*, *Anopheles sinensis*, *Anopheles albimanus*, *Anopheles darlingi*, etc. would substantially improve the quality of these annotated genomes.

The availability of genome sequencing data of different arthropod vectors, coupled with knowledge of expressed genes via transcriptomic and proteomic data has greatly enhanced our understanding of arthropod vectors. Proteogenomic approaches provide platform for validating current genome annotation strategies. We anticipate that improvements in transcriptomic and proteomic technologies in future will provide greater depth for such analyses in terms of sequence coverage, identification of less abundant molecules, post-translational modifications, which will accelerate biomedical research in vector-pathogen interactions, development of transmission blocking vaccines and prevention of disease transmission.

Aims and objectives of the research work

Anopheles stephensi is one of the major vectors of malaria in India and its complete genetic makeup is still unknown. Geographical variation of this vector species in terms of its forms, vectorial capacity and behavior has also been reported. Therefore, molecular level discoveries are needed to efficiently understand biology of this vector and to combat the diseases it carries. Development of *Plasmodium* inside the vector constitutes a major population bottleneck in the parasite life cycle and can be the target for transmission intervention (Angrisano *et al*., 2012; Baton and Ranford-Cartwright, 2005). Midgut tissue plays an important role in *Plasmodium* development and its proteins are responsible for various biological processes including blood meal digestion, egg development and vector pathogen interactions. A very few proteins are being reported in *Anopheles stephensi* that are involved in development of *Plasmodia* and their invasion in the host cells. Midgut being the first site for interaction between mosquito host and the *Plasmodium* is very important organ system of the female host. Unavailability of proteomic data of midgut tissue and its differential expression in response to the activity of *Plasmodium* gametocytes, their transformation into gametes, their fusion and formation of ookinete in the gut lumen and progress of sporogony on outer gut surface, limits our knowledge about sequence of proteomic events that might be occurring due to these events. It is assumed that proteins which if differentially expressed in response to blood meal and *Plasmodium* infection might be playing significant role in *Plasmodium* sexual reproduction process in the vector and further in malaria transmission. Therefore, it is necessary to study midgut proteome and its response to the blood meal and parasite infection. This will provide experimental evidence for computationally predicted genes and also identify set of novel proteins which are differentially expressed in response to blood meal and parasite infection. It is thought that the identification of the proteins involved in parasite development in mosquito midgut could form basis for transmission blocking strategies by blocking the parasite development. In addition to this, information on differential expression of proteins between male and female midgut tissue will provide experimental evidence for gender specific differential expression of proteins between two sexes. Further, this and similar information can be utilized for future biomedical

investigations with regard to host-pathogen interactions, vectorial capacity, and development of traits such as insecticide resistance at molecular level.

The genome sequence of *Anopheles stephensi* (Indian strain) has been recently published and now available in VectorBase; however, it is not extensively annotated (Jiang *et al*., 2014). The biological importance of completely sequenced genome can be recognized only by the precise identification of protein coding genes. This annotation process which is partially achieved through computational approaches immediately after sequencing a genome. Once a genome is sequenced, annotation of the genome becomes a daunting and cumbersome task. There is an increasing demand on different gene prediction algorithms to generate gene models and predicting exons.

Mass spectrometry is emerging as a powerful tool for genome annotation. 'Large-scale proteomics' powered by high resolution mass spectrometry now enables the unbiased and accurate system-wide characterization of thousands of proteins and their post-translational modifications giving direct evidence for the protein coding potential of the genome sequence. Mass spectrometry derived peptides not only validate the existence of several predicted genes but also help in identifying many novel genes or gene structures. Previous studies have demonstrated the use of mass spectrometry based proteomic approaches to correct or validate gene annotations in *An. gambiae* (Kalume *et al*., 2005; Chaerkady *et al*., 2011; Dwivedi *et al*., 2014). However, no proteogenomic studies have been carried out on *Anopheles stephensi* in order to annotate its protein-coding genes ever since its genome sequence has been released. Therefore, it is important to validate computationally predicted protein coding genes using experimental methods in order to refine the existing gene annotation.

It is expected that a well annotated genome and proteome of *Anopheles stephensi* will provide a better platform for investigating proteins and genes involved in the transmission of parasite, development of resistance to insecticides, functioning of the olfactory system as basis for attraction to human host and the ability to choose humans as

the source of blood meal. This would eventually help in defining and devising novel ways to control the transmission of malaria via mosquitoes.

In view of the above, the following objectives were proposed for study:

1. **A proteomic analysis of midgut of freshly emerged female *Anopheles stephensi* fed with sugar solution.**

2. **To study/identify differentially expressed proteins in the midgut of uninfected and infected female *Anopheles stephensi*.**

3. **Application of mass spectrometry-derived data of midgut in genome annotation.**

Chapter-II

Proteomic analysis of midgut of female *Anopheles stephensi* using Mass Spectrometry

2.1 Introduction

Malaria is caused by Protozoan parasite known as *Plasmodium,* which is transmitted by the infective bite of an *Anopheles* mosquito. Mosquitoes play a central role in the development and *transmission* of malaria parasites. *Anopheles stephensi* Liston is one of the major vectors of malaria in urban areas of India responsible for the transmission of 12-15% of total cases (Sumodan *et al*., 2004; Adak *et al*., 2005; Kumar *et al*., 2006). The species is not a sibling species complex but has three well recognized variants, *Anopheles stephensi* sensu stricto, (Type form), which is an urban form and *Anopheles stephensi mysorensis,* a rural form besides a third variant, *Anopheles stephensi* intermediate form (Subbarao *et al*., 1987). In India, *Anopheles stephensi* type and intermediate forms are potent malaria vectors in urban and peri-urban areas respectively. Vector control is central to the disease control. Preventive measures utilized to control malaria have been hampered due to the development of multiple insecticide resistance in mosquitoes and drug resistance in parasites. New strategies therefore need to be developed in order to achieve effective and sustained control of malaria. Such efforts require detailed molecular studies to understand vector biology and behavioral responses of mosquitoes.

Vector genomics has already laid foundation for understanding vector population structure and function at genetic level. The genome sequence of *Anopheles stephensi* (Indian strain) and SDA-500 strain from Pakistan has been recently published (Jiang *et al*., 2014; Neasefy *et al*., 2015). The genome assemblies of these two strains have been made available through public repository (https://www.vectorbase.org/). Mosquito midgut is the first point of contact between malaria parasites and vector host as the ingested blood meal from a *Plasmodium* infected person is obtained. Midgut presents the first physical barrier for the parasite during its life cycle in vector. The transmission of malaria to a new vertebrate host depends on the successful development of the malaria parasite in the mosquito midgut. In order to understand interactions between vector and pathogen, it is important and interesting to have knowledge on protein composition of the mosquito midgut. Characterization of the mosquito midgut proteins will identify proteins involved in the invasion and survival of the parasites inside the host vector.

Mass spectrometry-based proteomic analysis has emerged as a powerful tool, not only to perform large-scale analysis of proteins present in the complex tissues but also to characterize and understand protein-protein interactions. Many mosquito proteins are known to be involved in the parasite development and invasion of the host cells. Several potentially important proteins have been identified in vector mosquitoes, particularly in *Anopheles gambiae* (Vlachou *et al*., 2005; Dana *et al*., 2006; Mikolajczak *et al*., 2008; Gonzalez-Lazaro *et al*., 2009; Serrano-Pinto *et al*., 2010; Sreenivasamurthy *et al*., 2013). Proteomics studies on *An. gambiae* have provided numerous insights into the development, innate immunity, vectorial capacity of this important malaria vector and its susceptibility to different control strategies (Prevot *et al*., 2003; Psskewitz and Shi, 2005; Dinglasan *et al*., 2009; Chaerkady *et al*., 2011). However, a very little information is available on the protein composition and their possible functions in *An. stephensi* midgut, which limits understanding on attributes associated with vectorial capacity and susceptibility to various vector control methods. The first study on midgut proteins of blood and sugar fed *Anopheles stephensi* lines having different susceptibility to *Plasmodium falciparum* infection were carried out using two-dimensional gel analysis approach (Prevot *et al*., 1998). Around 29 proteins were identified exclusively from *P. falciparum* susceptible lines as compare to the less susceptible lines. This study has identified a limited number of proteins from the midgut tissue. Therefore, a more in-depth proteomic analysis was needed using advanced techniques.

A concerted effort in profiling midgut proteome of this important vector mosquito would provide a molecular platform to understand fundamental processes involved in the vector-parasite interaction and thus vectorial capacity. In the present study, proteomic analysis of midgut of sugar fed female *Anopheles stephensi* (Type form) was carried out using multiple fractionation strategies, high resolution mass spectrometry and multiple search algorithms in order to maximize proteome coverage and identification. This proteomic analysis led to the identification of thousands of proteins found to be involved in different biological and molecular functions, immunity and vector parasite interaction.

Female mosquitoes encounter certain challenges, which are never confronted by male primarily being blood digestion and possible *Plasmodium* infection when blood meal is obtained from an infected person in the wild. Studies have been undertaken to characterize sex-biased gene expression in *An. gambiae* (Hahn and Lanzaro, 2005), *Anopheles anthropophagus* (Geng *et al*., 2009), but such a study has not been carried out in *An. stephensi* thus far. There are a number of biological differences between male and female mosquitoes, such as much longer life span and much larger body size of the females compared to the males, which feed exclusively on flower nectar and sweet juices of plant tissues, while the females feed both on sugars for energy and blood meal for the development of their eggs.

To understand molecular mechanisms underlying, comparative proteomic approach was employed to investigate the differences in the midgut proteomes of male and female *Anopheles stephensi*. In this study, comparative proteomic analysis of male and female midgut proteins was carried out using iTRAQ based quantitative proteomic approach to identify the differentially expressed proteins between the two sexes. The relative quantitative values of proteins expressed in female midgut were compared to male and vice versa. This analysis led to the identification of many proteins which are more abundant in midgut of female tissue as compared to that of male and vice versa. The functional role of these differentially expressed proteins was also studied in detail. The information generated in this study may be useful in devising malaria transmission blocking strategies.

2.2 Materials and Methods

2.2.1 Establishment of *Anopheles stephensi* colony

Anopheles stephensi larvae were collected from various man made breeding habitats such as curing waters, overhead tanks, wells, etc. in Goa. These larvae were transferred in the labeled plastic containers along with water from respective habitats and transported carefully to the insectary of National Institute of Malaria Research, Field Station at Goa for further development of larvae. For this, the larvae were transferred to the plastic trays containing tap water and reared under laboratory conditions at temperature of 27±2°C,

relative humidity of 70±5% and a photoperiod of 12h of day and night cycle. A pinch of food (mixture of fish food and Cerelac powder™) was given to the growing larvae once in a day till the development of pupal stage. Pupae were collected in plastic bowls containing tap water and then kept inside a closed net cage for emergence of adults. The adult mosquitoes emerged from these pupae were given feed of 10% glucose soaked in a cotton pad. After two days, mosquitoes were allowed to feed on blood using a glass feeder in which blood was maintained at 37^{o}C with the help of water bath. Bowls with water was kept inside the cage for oviposition. In this fashion, a continuous cyclic colony of the *An. stephensi* was maintained in the insectary.

2.2.2 Dissection of midgut

Midguts were dissected from the 1,000 adult male and female *An. stephensi.* The dissection was carried out in 0.65% normal saline under dissecting microscope and immediately transferred to the PBS buffer and stored at -80˚C until used. Figure 2.1 shows dissected midgut from a female *Anopheles stephensi.*

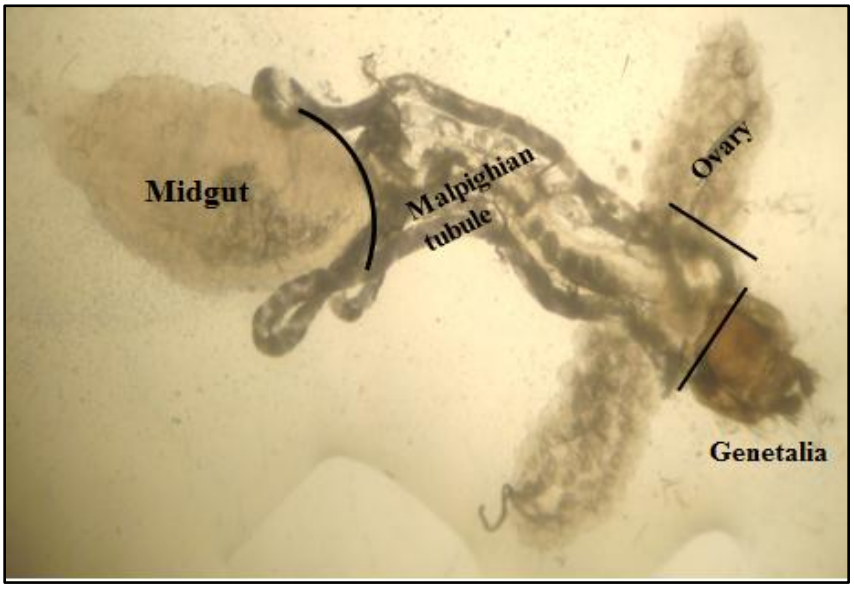


Figure 2.1: Dissected midgut from a female *Anopheles stephensi*.

2.2.3 Protein extraction and quantification

The midguts collected from 1,000 male and female mosquitoes were homogenized in 200 µl of 0.5% SDS using ultrasonication. The extracted suspension was centrifuged at 14,000 rpm for 10 min at 4˚C. Supernatant was collected without disturbing the pellet. The protein concentration was determined using BioRad BCA assay kit**.** The total amount of extracted proteins was calculated and divided into three approximately equal parts for performing three different methods of fractionation. Both protein and peptide level fractionations were attempted to increase the coverage of the protein and peptide identifications. From the extracted protein about 50µg of midgut protein and 5µg of BSA were taken to run on SDS-PAGE. Profile of proteins obtained from midgut of male and female *An. stephensi* can be seen in (Figure 2.2).

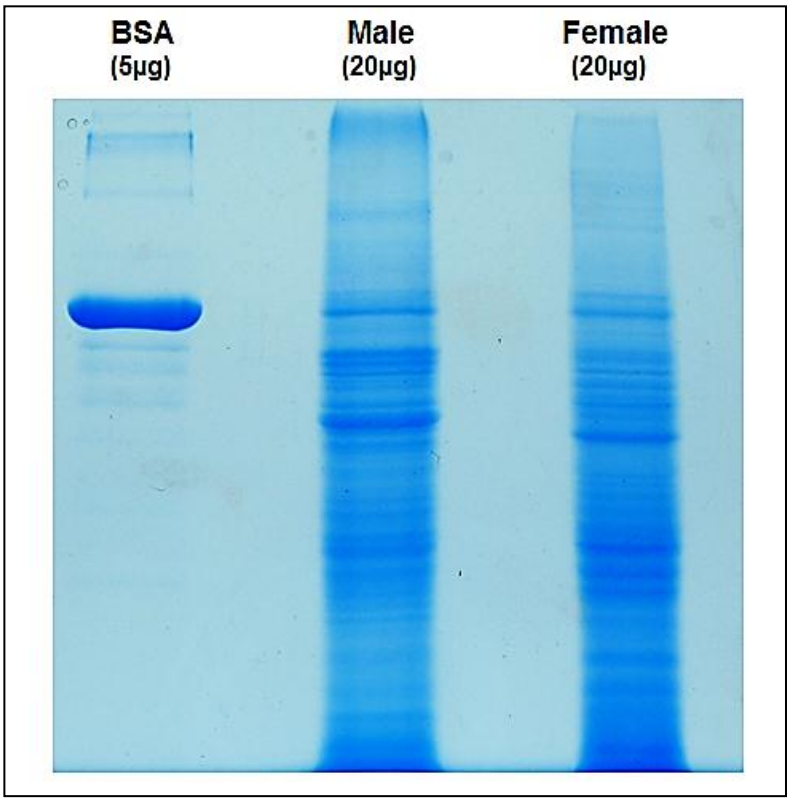


Figure 2.2: SDS PAGE profile of proteins obtained from the midgut of male and female *Anopheles stephensi*. Lane 1: Bovine serum albumin (BSA), Lane 2: Protein from male midgut tissue, Lane 3: Protein from female midgut.

2.2.4 Protein fractionation and digestion

2.2.4.1 In-gel digestion

About 200µg of protein from female midgut sample was resolved on 10% SDS-PAGE gel. The gel was stained using Colloidal Coomassie 33 stain (Invitrogen, Carlsbad, CA). Excess stain was removed by giving multiple washings with 10% methanol. The protein lanes were cut into 22 gel pieces depending on the protein profile of sample and then in-gel trypsin digestion was carried out as described in detail below:

1. The destained gel was first rinsed with water and then carefully placed onto a clean and sterilize glass plate. A clean scalpel blade was used to cut the visible bands.
2. Protein bands were excised into small equal sized pieces and transferred into micro-centrifuge tubes containing destaining solution (40mM ammonium bicarbonate, pH- 8, 40% acetonitrile).
3. Tubes were placed on rocker and gently agitated to aid in destaining. The solution was discarded and this procedure was repeated until the gel pieces were completely destained.
4. Then 500ul of 100% acetonitrile was added and incubated for 10-15 min until the gel pieces were completely dehydrated. The remaining acetonitrile was removed as much as possible using gel loading tips.
5. Sufficient amount of freshly prepared reducing agent (5mm DTT in 40mM ABC; prepare shortly before use) was added to completely cover the gel pieces which were then incubate at 60^{o}C for 30 min on a shaker.
6. The tubes were allowed to cool down to room temperature and the reducing agent was discarded using micropipette.
7. The freshly prepared Alkylation solution was added (20mM Iodoacetamide (IAA) in 40mM ABC; prepare shortly before use). Then tubes were incubated for 15 min at room temperature in the dark and quick shake every 5 minutes.
8. Alkylation solution was removed after short spin and gel pieces were dehydrated by adding 100% acetonitrile (volume of alkylation solution should be sufficient to cover gel pieces).

9. Acetonitrile was completely removed after the gel pieces were completely dehydrated and then tubes were kept on ice.
10. Sequencing grade trypsin (Promega) was suspended in chilled 40mM ammonium bicarbonate (2 ml 40 mM ABC add to one vial of trypsin contains 20 ug trypsin). And then trypsin solution was added to the tubes containing gel pieces.
11. Tubes were left on ice for 45mins to1hr and once the gel pieces were completely rehydrated (more trypsin solution was added if necessary)
12. Once the gel pieces were completely rehydrated, excess trypsin was remove and replaced with 40mM ammonium bicarbonate just enough to cover the gel pieces. The tubes were incubated at 37^{o}C overnight. After the process of overnight in-gel digestion, the digested peptides were extracted from the gel pieces by the following process.
13. Peptide extraction**:** Tubes were cooled to room temperature and 100 µl of 5% formic acid was added. Tubes were spun and aqueous phase was collected in different set of tubes labeled according to fraction number.
14. 100 µl of extraction buffer (5 % formic acid; 40% acetonitrile) was added to each tube and the tubes were kept on shaker for 30 min. Aqueous phase was collected in the tubes containing supernatant from the earlier step. This second step was repeated once or twice.
15. Final extraction was carried out by adding 100% ACN. After gel pieces completely dehydrated, supernatant was collected in the new tubes. Dehydrated gel pieces can be discarded. Peptide extracts were dried using speedvac and stored for further analysis at -80°C. Dried peptides appeared as translucent yellow to cream color film at the bottom of the tubes.

2.2.4.2 In-solution digestion

400 µg of protein from female midgut tissues were subjected to in-solution trypsin digestion. In-solution digestion was carried out as follows.

1. The proteins were extracted using 8M urea solution. The extracted proteins were then quantified according to Lowry's method (using Bio-Rad DC Protein assay kit).

2. About 400 µg of proteins were taken in a 1.5 ml microcentrifuge tube. Reducing agent dithiothreitol solution (50 mM) was added to the protein solution so that final concentration of dithiothreitol will be 5 mM and the tube was kept at 60°C for 40 minutes after gentle mixing.
3. The tube was brought to room temperature and alkylating agent iodoacetamide solution (100 mM) was added to the solution so that the final concentration is 10mM. The tube was kept in dark for 15 minutes.
4. After reduction and alkylation, protein solution was diluted with 40mM ammonium bicarbonate in order to reduce the final concentration of urea to 2M from 8M. This is done as trypsin action will be hindered in the presence of 8 M urea.
5. Sequencing grade modified trypsin (Promega, USA) was added at the ratio of 1:20. And reaction was allowed to take place for up to 12-16 hours at 37°C.
6. Formic acid was added to the final concentration of 5% to stop the reaction. An aliquot equivalent to 20 µg of trypsin digested protein was taken out to run on SDS-PAGE to check for digestion efficiency (complete digestion is indicated by absence of bands after staining the gel). The tryptic peptides obtained was desalted and stored at -20°C for future use.
7. These tryptic digested peptide mixtures were divided into two equal amounts for further separation by using off-gel fractionator and reverse-phase liquid chromatography.

2.2.5 Cleaning/desalting of peptide preparation

Peptides obtained after in-solution digestion were cleaned using Sep-Pak columns packed with reverse phase C_{18} material (Waters) as follows.

1. Tryptic digested peptide solution was acidified to reach the pH range of 2-3 using 10% trifluoroacetic acid (TFA). The acidified solution was centrifuged at 1000 rpm to remove precipitate or particulate matter which can block the column.
2. In a 15 ml syringe C_{18} Sep-Pak column was fitted. Column was pre-wetted using 100% acetonitrile. Solvent A (0.1% TFA) is passed through the column to remove acetonitrile.

3. Tryptic digest solution was loaded on top of the column and mobile phase is allowed to trickle down under gravity.
4. Column was washed with 12ml of solvent A thrice, applied as 1 + 5 + 6 ml.
5. Peptides were eluted with Solvent B (0.1% TFA, 40% acetonitrile) twice applied as 3 + 3 ml.
6. Eluted peptides were vacuum dried and stored at -80°C until mass spectrometry analysis.

2.2.6 Off-gel fractionation of trypsin digested proteins

The trypsin digested peptides were fractionated pI-based Off-gel electrophoresis using an Off-gel fractionator (Agilent 3100) as per manufacturer's instructions. Peptides were separated using IPG strip (pH 3-10) by focusing for 50 kVh with maximum current of 50 µA and maximum voltage set to 4000 V. After fractionation, a total of 12 fractions were collected and acidified using 1% TFA and stored at -80 °C until LC-MS/MS analysis.

2.2.7 Reverse-phase liquid chromatography

The digested peptides were fractionated by basic pH reverse-phase liquid chromatography method using solvent B (10 mMtriethyl-ammonium bicarbonate, pH 8.5 in 95% Acetonitrile) with a gradient of 5% to 60% and 1 ml flow rate per minute for over 60 minutes. The profile of chromatograph is show in Figure 2.3. A total of 96 fractions were collected using automatic fraction collector, which were further concatenated to 24 fractions, vacuum dried and stored in -80 ºC freezer until further LC-MS/MS analysis.

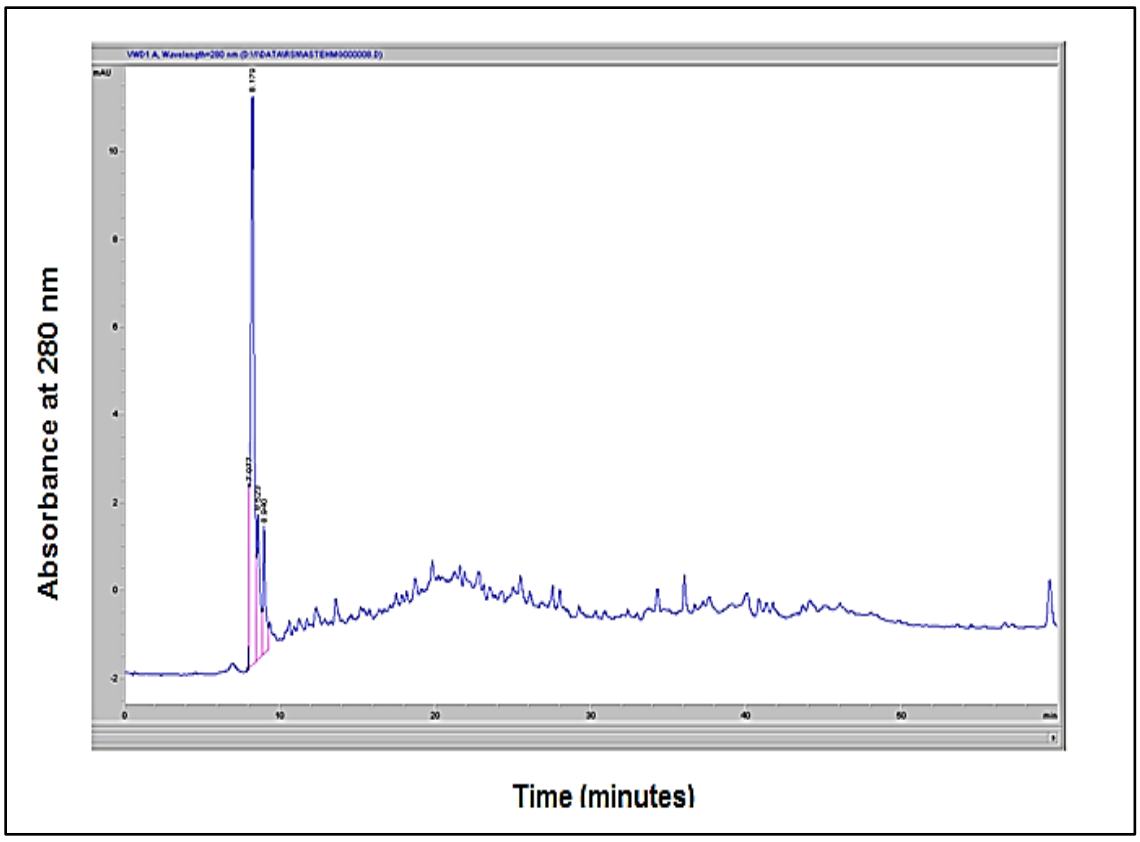


Figure 2.3: Typical elution profile of tryptic digest of female *Anopheles stephensi* midgut protein in bRPLC chromatography.

2.2.8 Cleaning of peptides obtained from Off-gel and bRPLC fractions using C_{18} stage tip

Individual bRPLC and Off-gel fractions were cleaned/desalted using C_{18} stage tips.

Solvent A/washing solution was 0.1% TFA and solvent B/eluting agent was 0.1% TFA, 60% acetonitrile.

2.2.8.1 Cleaning procedure followed was as follows

1. C18 stage tips were prepared by packing two ~2mm diameter discs punched out from 3M Empore C18 discs in 10µl micropipette tips. For punching out the discs pipetting needle (18X4") was used and syringe attached at the back of the needle in used to release the punched discs into the 10µl tip. Gel loader tip or silica capillary was used to slide the material and fix it at the end of the tip.

2. Fractions were acidified by adding 10% TFA, so that final concentration of TFA in the sample will be 1%.
3. C18 material in stage tips was pre-wetted by 100% acetonitrile followed by washing with solvent A.
4. Acidified peptide sample was loaded on the C18 material using gel loader tip and pushed through the material using 10 ml syringe attached at the back of the stage tip.
5. Two washes of solvent A, each of 50 µl are given in the similar way.
6. Peptides were eluted by adding 30 to 50 µl solvent B and passing it through the C18 material using syringe and slower speed. The stage tips can be incubated at room temperature for 10 minutes after pushing little bit of solvent B through the material to ensure efficient elution of the peptides.
7. Eluted peptides were vacuum dried and stored at -80 0 C until they were analyzed on mass Spectrometry.

2.2.9 Quantitative proteomic analysis of male and female midgut

Midgut tissues from both male and female mosquito samples were suspended in 0.5% SDS and sonicated. Suspension was centrifuged at 10,000 rpm for 10 min at 4°C to remove cell debris. Supernatant was collected without disturbing the pellet. The extracted proteins were quantified according to Lowry method using Bio-Rad DC Protein assay kit. Equal amount of protein obtained from each respective midgut tissue was loaded on the SDS-PAGE for normalization. The SDS-PAGE profile of proteins from male and female midgut tissue is shown in Figure 2.2.

2.2.10 iTRAQ (isobaric tag for relative and absolute quantitation) labeling

Equal amount of protein samples from each male and female midgut tissue was taken for iTRAQ labeling. Briefly 100µg of protein samples from each male and female midguts of *An. stephensi* were treated with 2µl of reducing solution (tris(2-carboxyethyl) phosphine) and incubated at 60°C for 1h and alkylated with 1µl of methyl methane thio-sulfonate (MMTS)) at room temperature for 10 min. Proteins were then digested using trypsin (Promega) (1:20) overnight at 37°C. Peptide samples were labeled using iTRAQ reagents

(Applied Biosystems Catalog No. 4352135) containing reporters "iTRAQ 116" and "iTRAQ 117" for male and female, respectively. iTRAQ labeling of peptides was carried out according to manufacturer's protocol. Labeling was carried out for 2 h at room temperature. After labeling, the peptides from the two samples were pooled in a single tube and then fractionated using strong cation exchange chromatography. The workflow illustrating the steps involved in iTRAQ labeling of midgut proteins from male and female *Anopheles stephensi* (Figure 2.4).

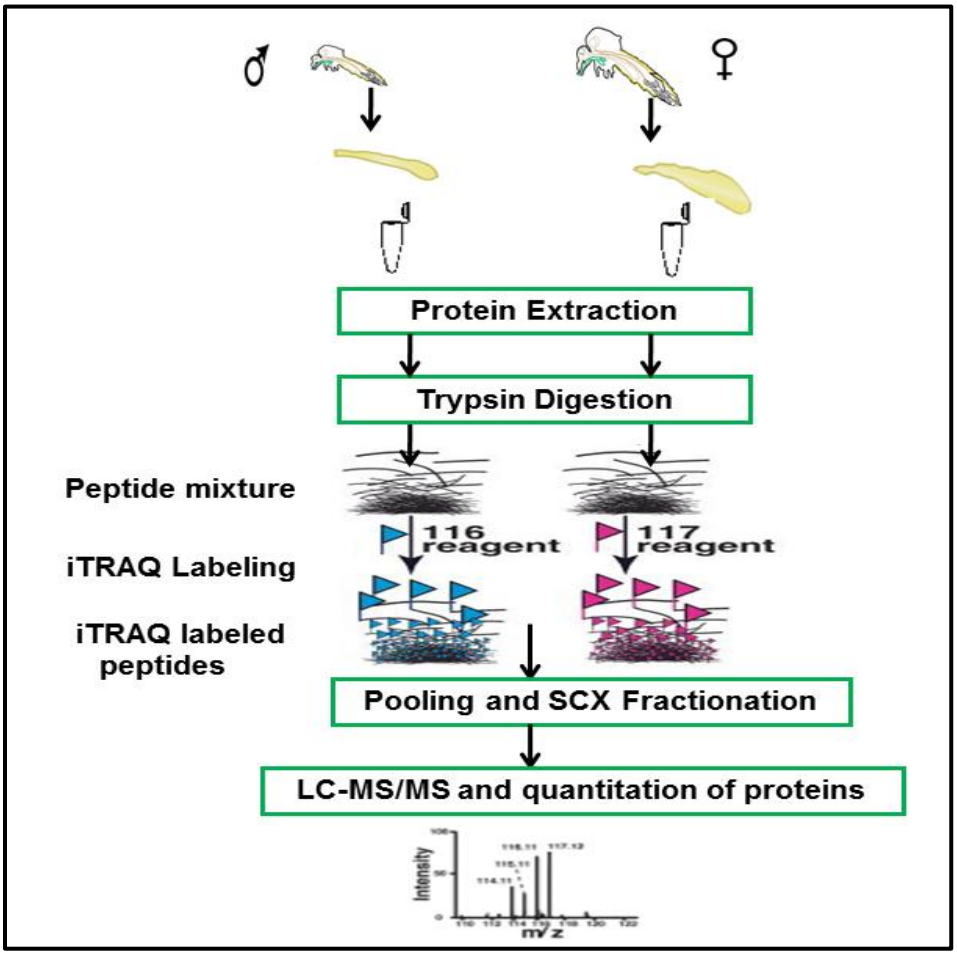


Figure 2.4: Outline of the quantitative proteomic strategy. Midgut tissues from male and female *An. stephensi* were used to extract proteins and digested with trypsin. Peptides obtained from the male and female midgut tissue were labeled with iTRAQ reagents followed by LC-MS/MS analysis.

2.2.11 Strong cation exchange chromatography (SCX)

The fractionation of iTRAQ labeled peptides was carried out using strong cation exchange chromatography (SCX) having polysulfoethyl A column (PolyLC, Columbia, MD) (100 × 2.1 mm, 5 µm particles with 300 Å pores) using a nLC Packing HPLC system connected to a probot fraction collector. Fractions were collected at a flow rate of 0.2 ml/min using 80-min gradient of KCl from 0 to 350 mM concentration in 10 mM potassium phosphate buffer and 25% acetonitrile (pH 2.85). The fractions were dried and reconstituted in 10 µl of 2% tri-fluoroacetic acid before mass spectrometric analysis.The profile of chromatograph is show in Figure 2.5.

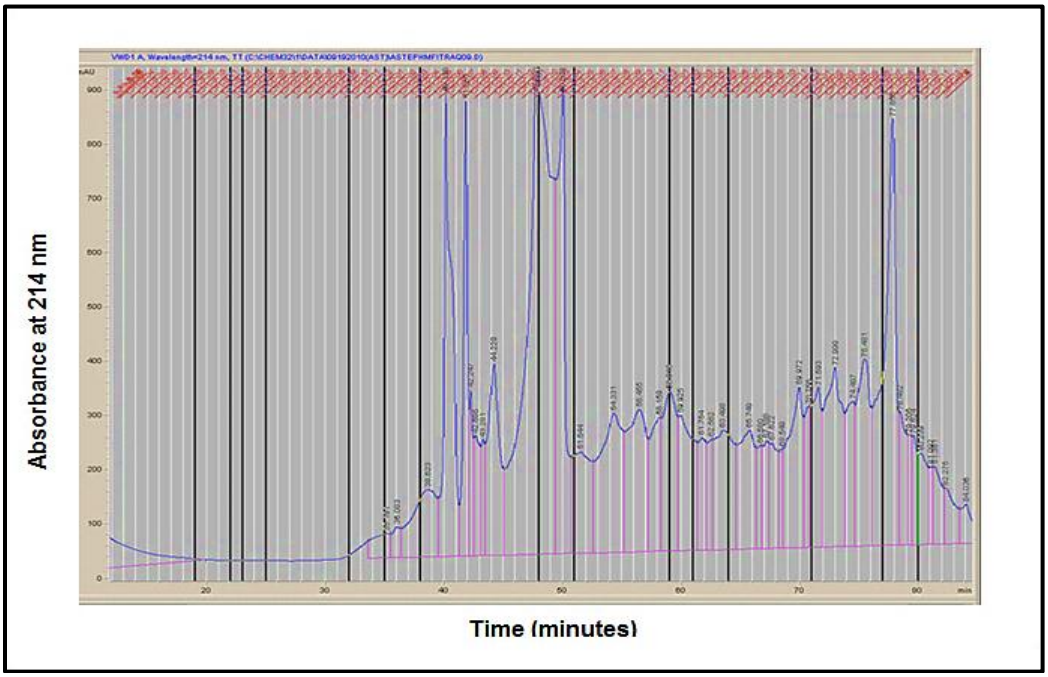


Figure 2.5: Typical elution profile of tryptic digest of iTRAQ labeled male and female midgut protein in SCX chromatography.

2.2.12 Mass spectrometry analysis

In this study, we performed a total of 80 LC-MS/MS analyses. 24 bRPLC fractions from midgut, were performed on LTQ-Orbitrap Elite (Thermo Scientific, USA) mass spectrometer interfaced with Easy- nanoLC II nano flow liquid chromatography system (Thermo Scientific), while the remaining 56 fractions (including in-gel, offgel and iTRAQ male and female midgut fractions) were analyzed on LTQ-Orbitrap Velos mass spectrometer interfaced with Proxeon Easy nLC system (Thermo Scientific, Bremen,

Germany). The peptides from each fraction were reconstituted in 0.1% formic-acid and loaded on to a pre-column (75u x 2cm) packed with magic C18 AQ (Michrom Bioresources, USA) 5µ particle and 100Å pore size at flow rate of 5µl per minute. Peptides were resolved at 250nl/min flow rate using a linear gradient of 10 to 35% solvent B (0.1% formic acid in 95% Acetonitrile) over 75 minutes on an analytical column, of 75 µ x 60 cm, 5 µ particle and 100 Å pore size for Elite and 75 µ x 15 cm, 3 µ particle and 100 Å pore size for Velos was packed using nitrogen pressure cell at 2,500 psi. To reduce the back pressure 60 cm analytical column was operated in a heated insulator at 60 ºC temperature using butterfly column heater (Phoenix S&T, Inc. PA, USA) and was fitted on flex ion source that was operated at 2.5 kv voltage (Only for Elite). The analysis on mass spectrometry was carried out in a data dependent manner with a full scans in the range of m/z 350 to 2000. Both MS and MS/MS were acquired and measured using Orbitrap mass analyzer. Full MS scans were measured at a resolution of 120,000 for Elite and 30,000 for Velos at m/z 400. For iTRAQ lebelled peptides, both MS and MS/MS scans were acquired in Orbitrap mass analyzer after fragmentation by higher-energy collisional dissociation (HCD) mode at 45% normalized collision energy. Fifteen to twenty most intense precursors were selected for MS/MS and were fragmented using higher-energy collisional dissociation (HCD) method and detected at a mass resolution of 30,000 for Elite and 15,000 for Velos at m/z 400. Automatic gain control for full MS was set to 1 million ions and for MS/MS was set to 0.05 million ions with a maximum ion injection time of 100 and 200 ms respectively. Dynamic exclusion was set to 30 sec and singly charged ions were rejected. Internal calibration was carried out using lock mass option (m/z 445.1200025) from ambient air (Olsen *et al.*, 2005). Figure 2.6 illustrates the steps involved in the proteomic analysis of midgut tissue using mass spectrometry.

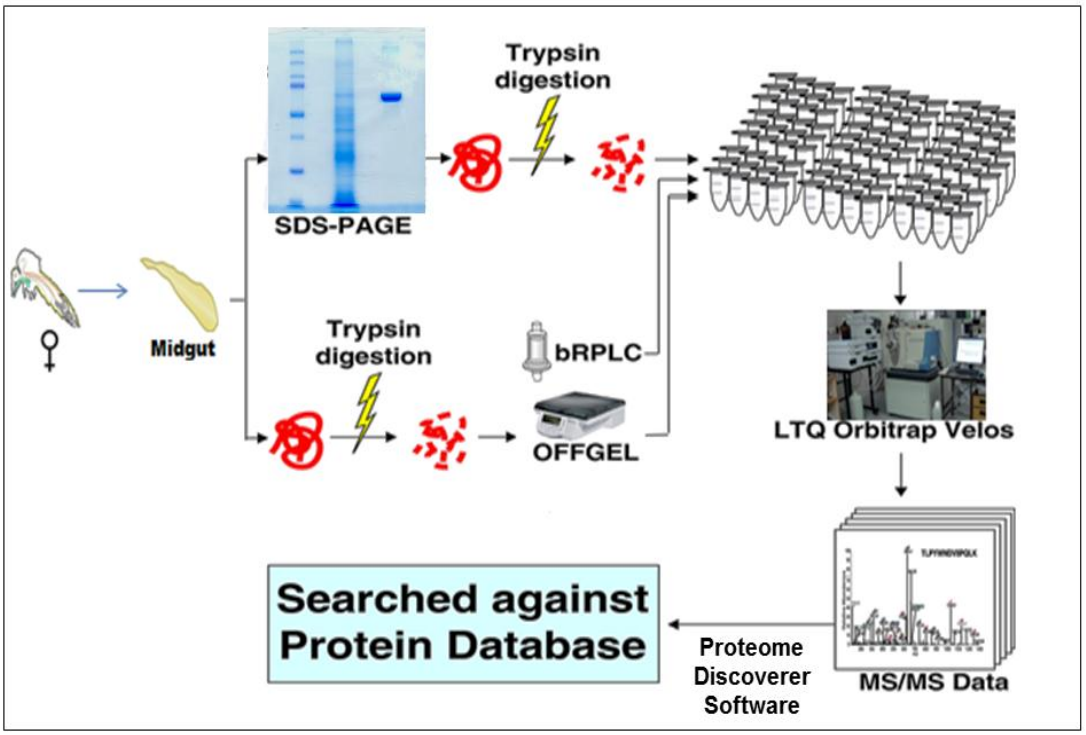


Figure 2.6: Workflow explaining the steps involved in the proteomic analysis of midgut tissue.

2.2.13 Data analysis

The data obtained was processed using Proteome Discoverer software (version 1.4, Thermo Fisher Scientific, Bremen, Germany) workflow and searched using Sequest and Mascot search algorithm against Vector Base protein database Astel2.0. The search parameters included trypsin as the proteolytic enzyme allowing up to two missed cleavages for protein database searches; methionine oxidation was set as a dynamic modification while carbamido-methylation at cysteine was set as static modification.

In addition, for iTRAQ data, iTRAQ labeling at peptide N-terminus and lysine side chain was given as static modification. Peptide mass error tolerance and fragment mass error tolerance were set to 20 ppm and 0.1 Da, respectively. The protein and peptide data were extracted with search result parameters as peptide rank one and peptide confidence set as high. For the entire data set, false discovery rate (FDR) was calculated by enabling the peptide sequence analysis using a decoy database and a cut-off of 1% was used for identifications.

2.2.14 Determination of relative abundance of proteins using label free quantification method

Relative abundance of proteins was carried out using spectral counting based label-free quantitation. Briefly, the normalized spectral abundance factor (NSAF) was calculated for each protein identified in this study. NSAF value for a protein was calculated as described by Paoletti *et al*., (2006). The protein with highest NSAF value was termed as most abundant while the one with lowest NSAF was termed as least abundant. Different bioinformatics tools were used to find out the domain structure, sequence homology and ortholog information. SMART prediction algorithm was used to determine the domain structure of the hypothetical proteins identified. BLASTP was carried out to find out the protein sequence similarity between closely related mosquito species. In-paranoid tool was used to identify the orthologous proteins in closely related species as previously described by Remm *et al*., (2001). The best reciprocal hits of proteins between given two species with more than 70% identity were considered. Further, these identified proteins were analyzed for cellular localization and functional categorization based on biological processes and molecular functions using gene ontology annotations available for *An. stephensi* in VectorBase database.

2.2.15 Relative quantification of proteins identified from male and female midgut

Relative quantitation of proteins was calculated by using reporter ions quantifier node in Proteome Discoverer software. In brief, relative intensities of reporter ions 116 and 117 released during MS/MS fragmentation of peptides were calculated. Relative intensities of the two reporter ions for each peptides identified were used to determine relative quantity of a protein in male and female midgut tissue as described by Gautam *et al*., (2012). The unique peptides identified against each protein were used to calculate relative protein content in the two samples. The over-expressed proteins identified in male and female midgut were further analyzed for cellular localization and functional categorization based on biological processes and molecular functions using gene ontology annotations available for *An. stephensi* in VectorBase database.

2.2.16 Data availability

The mass spectrometry data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD001128 (http://proteomecentral.proteomexchange.org).

2.3 Results and Discussion

2.3.1 Overview of the mass spectrometry data

The goal of the study was to achieve a deep coverage of midgut proteome of *Anopheles stephensi* which is one of the primary vectors of malaria in India and other neighbouring countries in Asia. To this end, proteins from midgut tissue of this vector were isolated and protein profile was observed on SDS-PAGE (Figure 2.2). The protein extract from midgut tissue was either digested directly in-solution with trypsin and then fractionated or first resolved by gel electrophoresis and subsequently digested with trypsin. For in-solution digested peptides, Off-gel and bRPLC was used as the fractionation strategy. In addition, each fraction thus obtained was further resolved on a LC and then analyzed on high resolution mass spectrometer. In total 58 LC-MS/MS runs were performed using high resolution mass spectrometry. In total, 7,16,932 MS/MS spectra were generated from all the runs. The resulting mass spectral data was searched against the predicted protein database of *Anopheles stephensi* Indian strains available on VectorBase (release Astel 2.0 Oct 2013), using the two different search algorithms i.e. Mascot and Sequest. Peptide spectrum matches were filtered with 1% false discovery rate (FDR).

A total of 5,43,028 peptide spectra matches (PSMs) resulted in the identification of approximately 63,525 peptides and 5,434 proteins. The distribution of mass error in parts per million for the entire peptide data set was calculated. Figure 2.7 shows that nearly 95% of the peptides were within ±5 ppm mass error, confirming the high accuracy of peptide data obtained from the mass spectrometer. Orbitrap analyzer was used for both MS and MS/MS acquisition, which has resulted in minimal mass error.

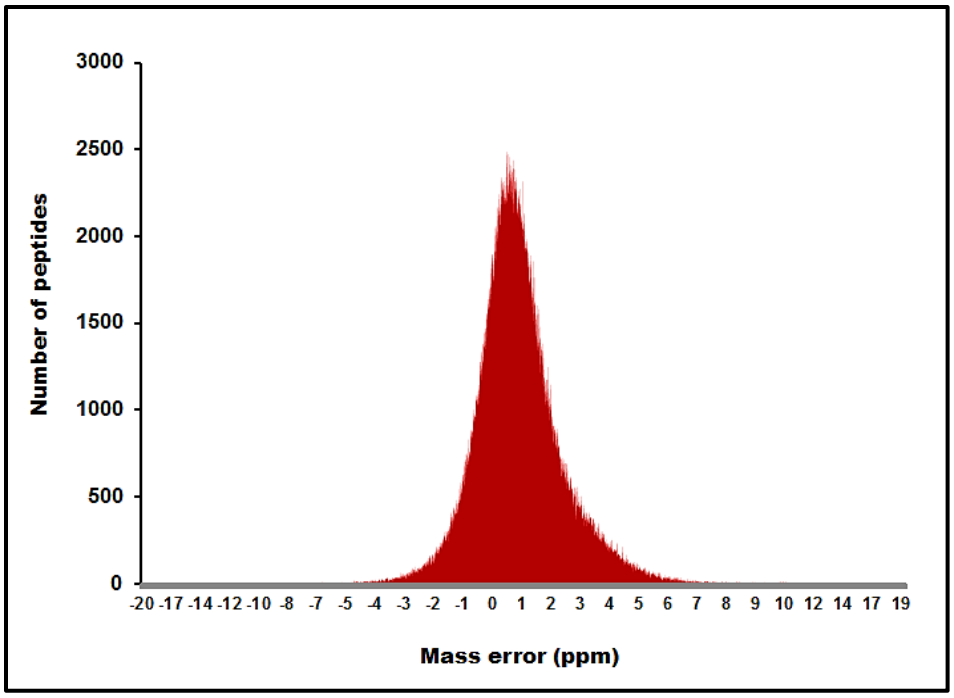


Figure 2.7: The distribution of mass error in parts per million for the entire peptides identified in this study.

A total of 5,434 proteins were identified from midgut tissue comprising ~ 46% of the total proteome of the *An. stephensi*. As shown in Figure 2.8, only 615 proteins i.e. 11.3% of the proteins were identified with single PSM evidence. On the other hand, 88.6 % of the proteins identified had two or more PSMs. A large fraction of the identified proteins had more than 10 PSMs and this increased the confidence level in the peptides identified through mass spectrometer.

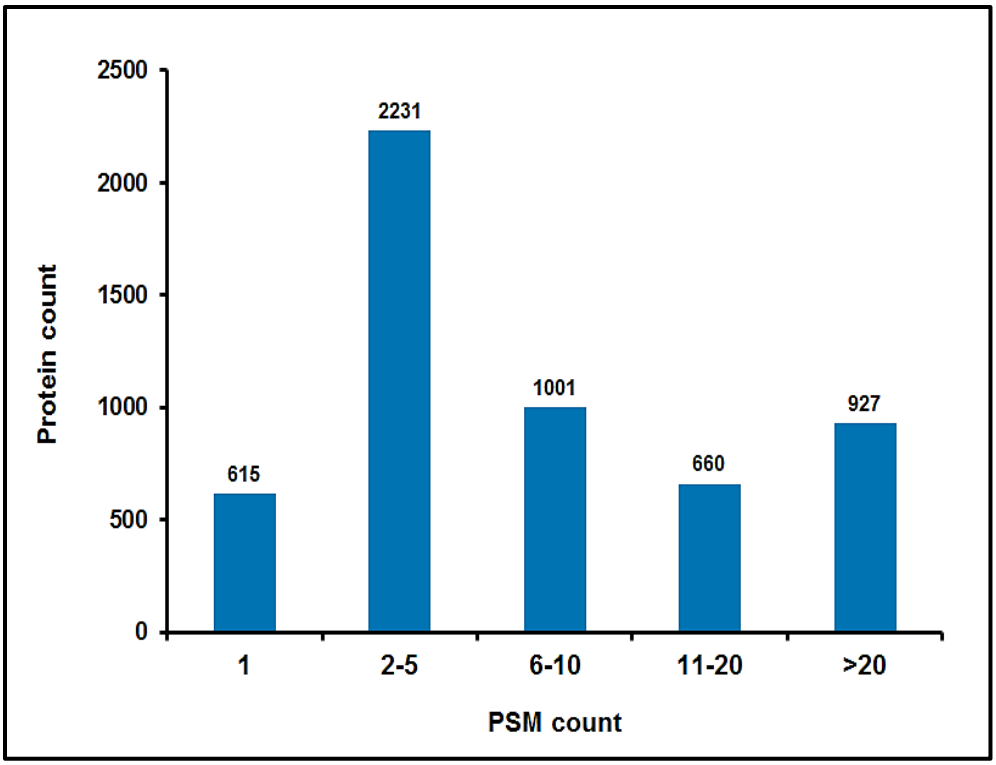


Figure 2.8: A column graph depicting the number of proteins identified with Peptide Spectra Matches (PSMs).

Multiple fractionation strategies used enabled to achieve a broad coverage of peptides. In addition, bRPLC fractions were analyzed on Orbitrap Elite mass spectrometer, which has comparatively higher resolution and scan speed. Distribution of proteins and peptides identified from the various fractionation methods is shown in Figure 2.9 A and B. In brief, 27,672 peptides were identified specifically from bRPLC fractions, 5,393 from In-gel digested bands and 2,589 from Off-gel fractionation. Analyzing bRPLC fractions on Orbitrap Elite instrument resulted in identification of 2,193 proteins, which were not identified in other two fractionation strategies i.e. In-gel & Off-gel. Similarly, 189 unique proteins were identified from in-gel bands and 118 proteins from Off-gel fractions. This shows the advantage of using multiple fractionation strategies and advanced mass spectrometry platform.

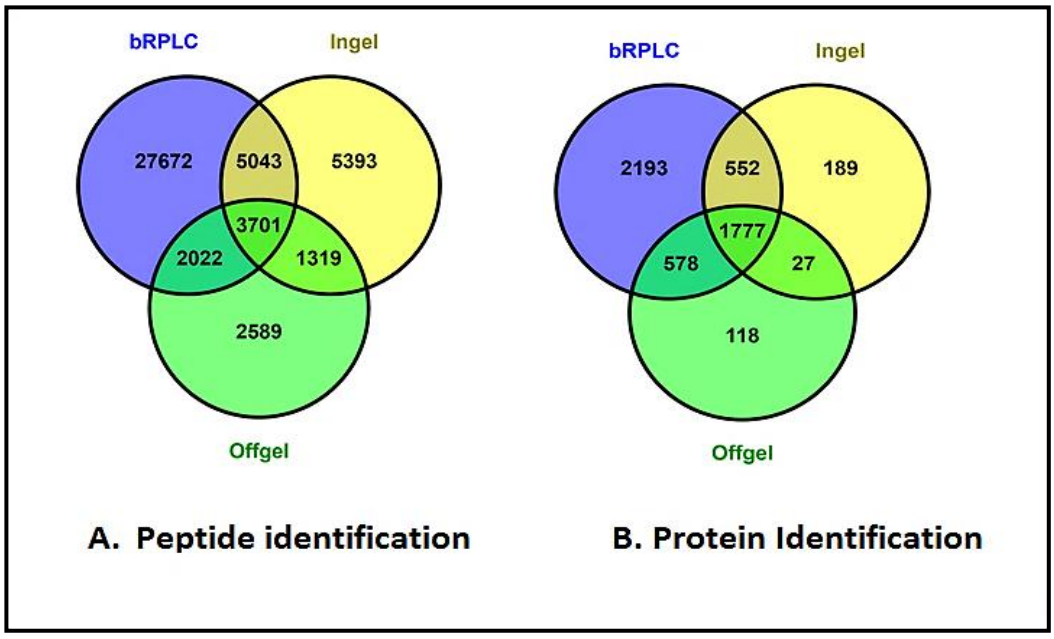


Figure 2.9: Peptides and proteins identified in different fractionation methods. (**A)** Represents the distribution of peptides identified from *An. stephensi* using three different fractionation methods i.e. SDS-PAGE, bRPLC and Off-gel **(B)** Distribution of proteins identified using three fractionation methods.

2.3.2 Relative abundance of the proteins identified and their biological roles

To check the relative abundance of the proteins, a label free based quantitation was carried out. For all the proteins identified through LC-MS/MS analysis, normalized spectral abundant factor (NSAF) values were calculated. The quantitation is based on the spectral count identified for each protein and the variability in the length of the proteins was accounted to normalize the abundance values. The most abundant proteins identified were Actin, F-type H+-transporting ATPase subunit beta, Sodium/potassium-transporting ATPase subunit alpha, Glyceraldehyde 3-Phosphate Dehydrogenase, Histone H4, ATP carrier protein, Galectin (GALE6), Calmodulin, and Sterol carrier protein-2 (Table 2.1). These proteins were identified with high coverage, multiple unique peptides and PSMs. A representative MS/MS spectrum of some of the highly abundant proteins identified in this study is shown in Figure 2.10 A and B.

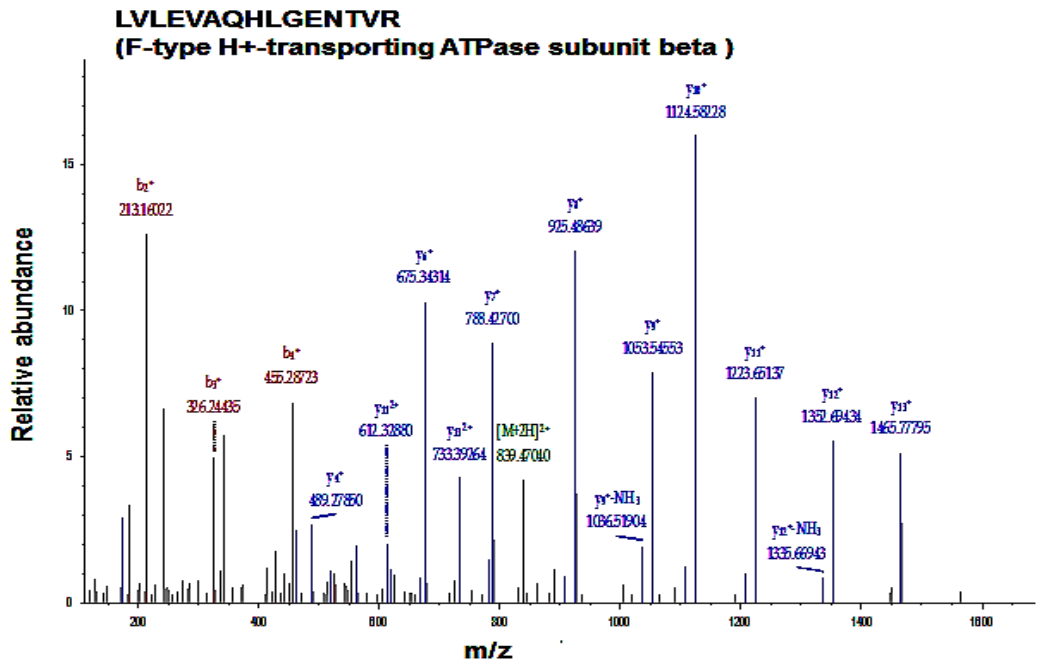
Table 2.1: Partial list of most abundant proteins identified in this study based on normalized spectral abundant factor (NSAF) values.

S. No.	Accession Id	Protein Description	∑ Coverage	∑# Unique Peptides	∑# Peptides	∑# PSMs	#AAs	MW [kDa]	NSAF Values	Domain/ Function
1	ASTEI07742-PA	Actin (similar to *An. gambiae*)	76.86	14	42	10120	376	41.8	2691.49	Actin/actin-like conserved site
2	ASTEI11539-PA	F-type H+-transporting ATPase subunit beta (similar to *An. gambiae*)	88.56	9	36	6227	437	47.0	1424.94	ATP synthase subunit alpha-like domain
3	ASTEI00450-PA	Sodium/potassium-transporting ATPase subunit alpha (similar to *An. gambiae*)	65.57	42	97	14396	1031	114.4	1396.31	Cation-transporting P-type ATPase, C-terminal
4	ASTEI04722-PA	Glyceraldehyde 3-phosphate dehydrogenase (similar to *An. gambiae*)	98.19	40	40	3944	332	35.4	1187.95	Glyceraldehyde 3-phosphate dehydrogenase, NAD(P) binding domain
5	ASTEI04162-PA	Histone H4 (similar to *Cx. quinquefasciatus*)	62.14	15	15	1176	103	11.4	1141.75	TATA box binding protein associated factor (TAF)
6	ASTEI11772-PA	Histone H2B (similar to *An. gambiae*)	100.00	2	13	663	67	7.4	989.55	Histone H2B
7	ASTEI08607-PA	ADP,ATP carrier protein 1 (similar to *An. gambiae*)	79.22	38	47	2951	332	36.3	888.86	Adenine nucleotide translocator 1
8	ASTEI10351-PA	Galectin (GALE6) (similar to *An. gambiae*)	76.22	10	12	1161	143	16.0	811.89	Concanavalin A-like lectin/glucanase domain

Table 2.1 contd: Partial list of most abundant proteins identified in this study based on normalized spectral abundant factor (NSAF) values.

S. No.	Accession Id	Protein description	∑ Coverage	∑# Unique Peptides	∑# Peptides	∑# PSMs	#AAs	MW [kDa]	NSAF Values	Domain/ Function
9	ASTEI09025-PA	Calmodulin (similar to *An. gambiae*)	81.32	12	12	728	91	10.4	800.00	EF-Hand 1, calcium-binding site
10	ASTEI01359-PA	Elongation factor 1-alpha (similar to *An. gambiae*)	84.02	21	51	3357	463	50.4	725.05	Elongation factor, GTP-binding domain
11	ASTEI08379-PA	Hypothetical protein	57.00	25	25	2730	407	45.0	670.76	No conserved domain
12	ASTEI05637-PA	F-type H+-transporting ATPase subunit d (similar to *An. gambiae*)	97.67	23	23	1122	172	19.5	652.33	ATPase, F0 complex, subunit D, mitochondrial
13	ASTEI10732-PA	Fructose-bisphosphatealdolase, class I (similar to *An. gambiae*)	93.96	33	47	2361	364	39.5	648.63	Aldolase-type TIM barrel
14	ASTEI09499-PA	Hypothetical protein	70.32	97	101	8560	1385	150.9	618.05	Insulin-like growth factor binding protein, N-terminal
15	ASTEI01665-PA	Sodium/potassium-transporting ATPase subunit beta (similar to *An. gambiae*)	59.69	25	27	1948	320	36.2	608.75	Sodium/potassium-transporting ATPase subunit beta
16	ASTEI01631-PA	Enolase(similar to *An. gambiae*)	94.92	50	50	2476	433	46.5	571.82	Enolase
17	ASTEI08310-PA	Myosin heavy chain (similar to *An. gambiae*)	80.73	104	203	7502	1318	151.4	569.20	IQ motif, EF-hand binding site
18	ASTEI02421-PA	60S ribosomal protein L18a (partial) (similar to *An. gambiae*)	71.05	23	27	1045	190	22.8	550.00	50S ribosomal protein L18Ae/60S ribosomal protein L20 and L18a
19	ASTEI02293-PA	Dihydropyrimidinase(similar to *An. gambiae*)	32.08	4	4	580	106	12.1	547.17	no conserved domain
20	ASTEI09967-PA	14-3-3 protein epsilon (similar to *An. gambiae*)	82.03	31	37	1376	256	29.2	537.50	14-3-3 domain

(A)



(B)

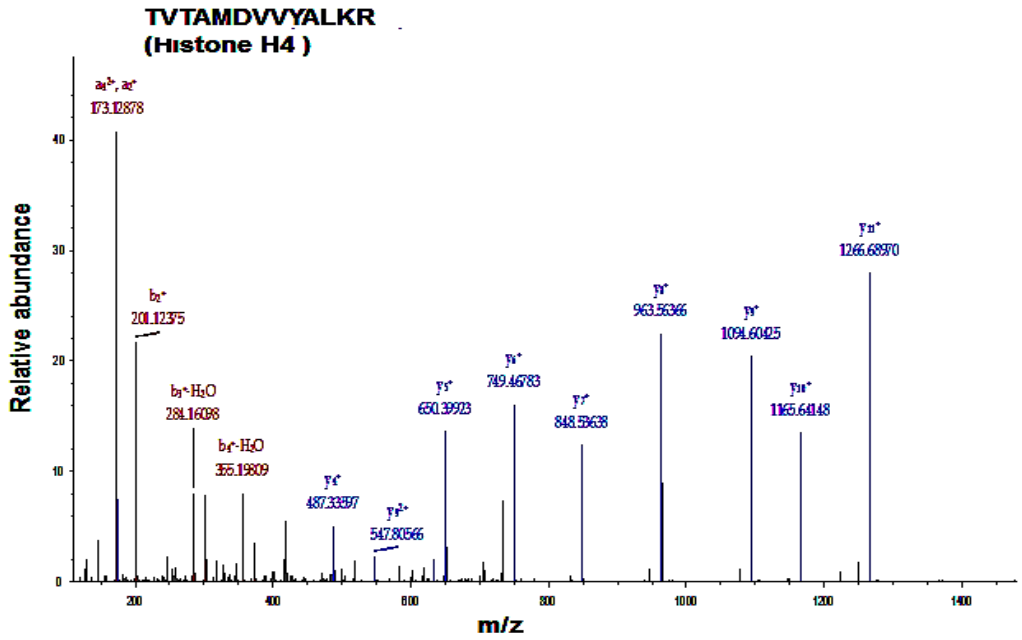
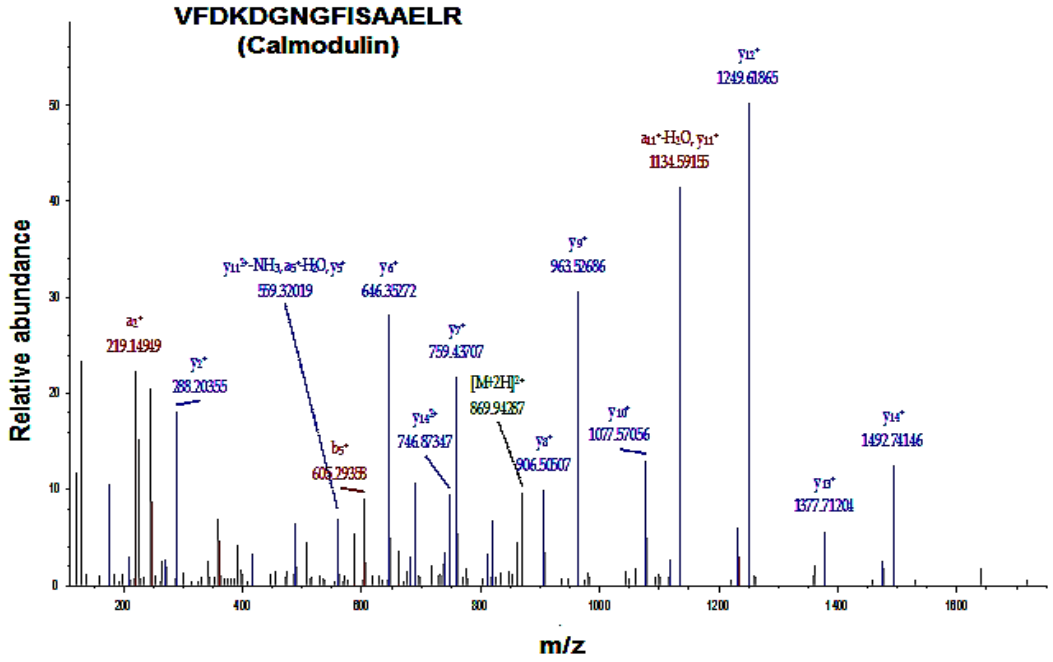


Figure 2.10 A: MS/MS spectra of peptides for the most abundant proteins in *An. stephensi* that were identified in this study. (A) Peptide sequence **LVLEVAQHLGENTVR** identified against F-type H+-transporting ATPase subunit beta; (B) Peptide sequence **TVTAMDVVYALKR** identified against protein Histone H4.

(C)



(D)

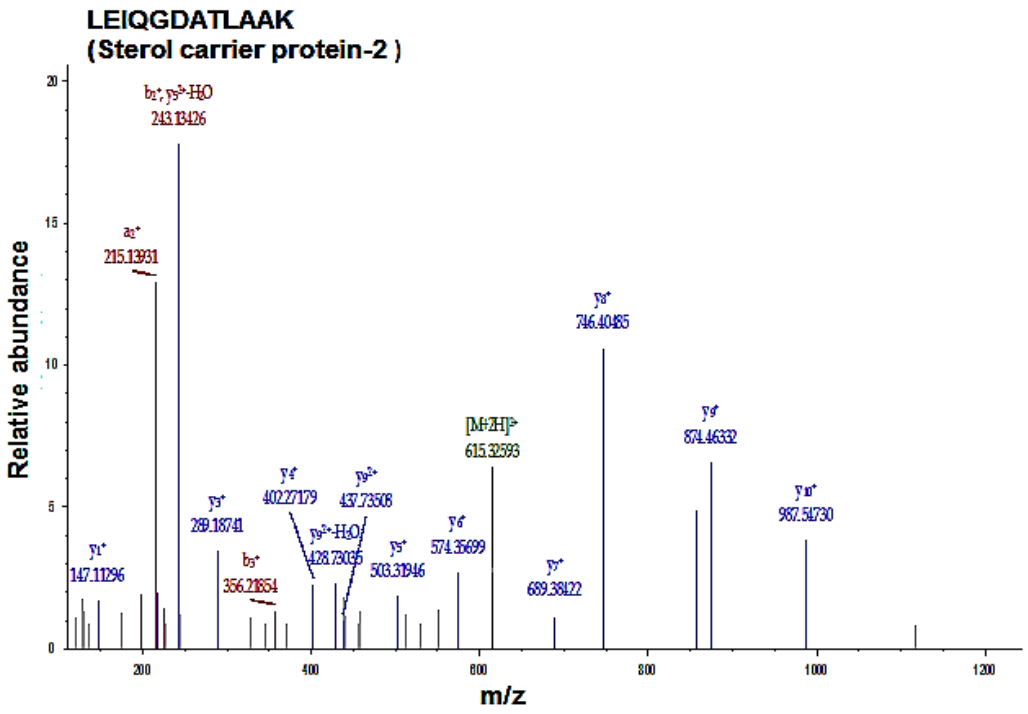


Figure 2.10 B: MS/MS spectra of peptides for the most abundant proteins in *An. stephensi* that were identified in this study. (C) Peptide sequence **VFDKDGNGFISAAELR** identified against Calmodulin; (D) Peptide sequence **LEIQGDATLAAK** identified against Sterol carries protein-2.

Large numbers of proteins in *Anopheles stephensi* are still listed as hypothetical proteins. Their orthologs were identified in other mosquito species (*An. gambiae*, *Culex quinquefasciatus* and *Ae. aegypti*) based on sequence similarity. Interestingly, a hypothetical protein identified in this study was found similar to *Aedes aegypti* sterol carrier protein-2 with 553 PSMs and 72% sequence coverage. This protein is involved in cholesterol uptake in mosquito larvae. Knockdown and gene silencing of sterol carrier protein-2 in *Aedes aegypti* mosquito causes high mortality in developing adults and also reduced egg viability (Blitzer *et al*., 2005; Kim *et al*., 2005). Another protein identified was found similar to Histone H4 of *Culex quinquefasciatus* with 100% similarity. Two proteins were found to be similar to galectin and calmodulin of *An. gambiae* with 91 and 100% sequence similarity respectively. In sandfly, *Phlebotomus papatasi*, galectin acts as a binding receptor for parasite *Leishmania major* for its survival in the sandfly. It is also involved in embryonic development and immunity against pathogens (Kamhawi *et al*., 2004). Calmodulin is a Ca2+ binding protein found in all eukaryotes. It belongs to CaM family which is a major class of calcium sensor proteins involved in intracellular signaling. In *Plasmodium gallinaceum*, it controls morphological differentiation from zygote to ookinete formation (Silva-Neto *et al*., 2002; Moreno and Docampo, 2003). Calmodulin also regulates odorant receptor function in *Drosophila* (Mukunda *et al*., 2014).

In this study, several housekeeping proteins which act as cytoskeletal proteins like actin, chitin binding, cuticular protein, myosin, tubulin, laminin, annexin, tropomysin, dynein light chain, dynactin, connectin and cadherin were identified. These proteins are found to be highly abundant in midgut tissue and suggest that these proteins play a significant housekeeping role in the *Anopheles stephensi* midgut. We have identified three members of ERM family proteins such as ezrin, radixin and moesin. These EMR proteins act as molecular linkers between the actin cytoskeleton and transmembrane receptors. Activation of ERM proteins mediates attractive growth cone guidance through the regulation of actin and adhesion receptors. These proteins also play a role in the Rho and Rac signaling pathways (Bretscher, 1999; Marsick *et al*., 2012). Another important

protein such as actin-depolymerizing factor (ADF)/cofilins as also identified which regulates assembly and disassembly of actin filaments. This protein is expressed in all insects specially *Drosophila melanogaster* and involved in several cellular processes such as motility, cell-polarity during migration and chemotaxis (Blair, 2006; Bamburg, 2010).

In most of the blood sucking insects both endo and exo peptidases are required for complete digestion of blood proteins. Several digestive proteases (like aminopeptidase, carboxypeptidase A & B, trypsin, cathepsin D) and glycosidases (like alpha-amylase and maltase) were identified. These are known digestive proteases found in midgut tissue and involved in blood and sugar digestion in mosquitoes. The glycosidases are involved in sugar digestion, and the proteolytic enzymes are involved in host-specific proteolysis events that occur during the blood-feeding process such as clot prevention and coagulation cascades, and also involved in the digestion of extracellular matrix components. Several other proteases such as catalase, clip-domain serine protease, serine protease inhibitor (serpin) SRPN10, SRPN4, SRPN6, PAS domain containing serine/threonine kinase, serine/threonine-protein kinase PRP4, serine/threonine-protein kinase SRPK1 are also identified. These proteases are related largely to blood digestion and immunity and have been demonstrated in earlier studies in different mosquito species (Koutsos *et al*., 2007; Warr *et al*., 2007; Neira *et al*., 2008). MS/MS spectra of some of the proteolytic enzymes identified in this study are shown in Figure 2.11 A and B.

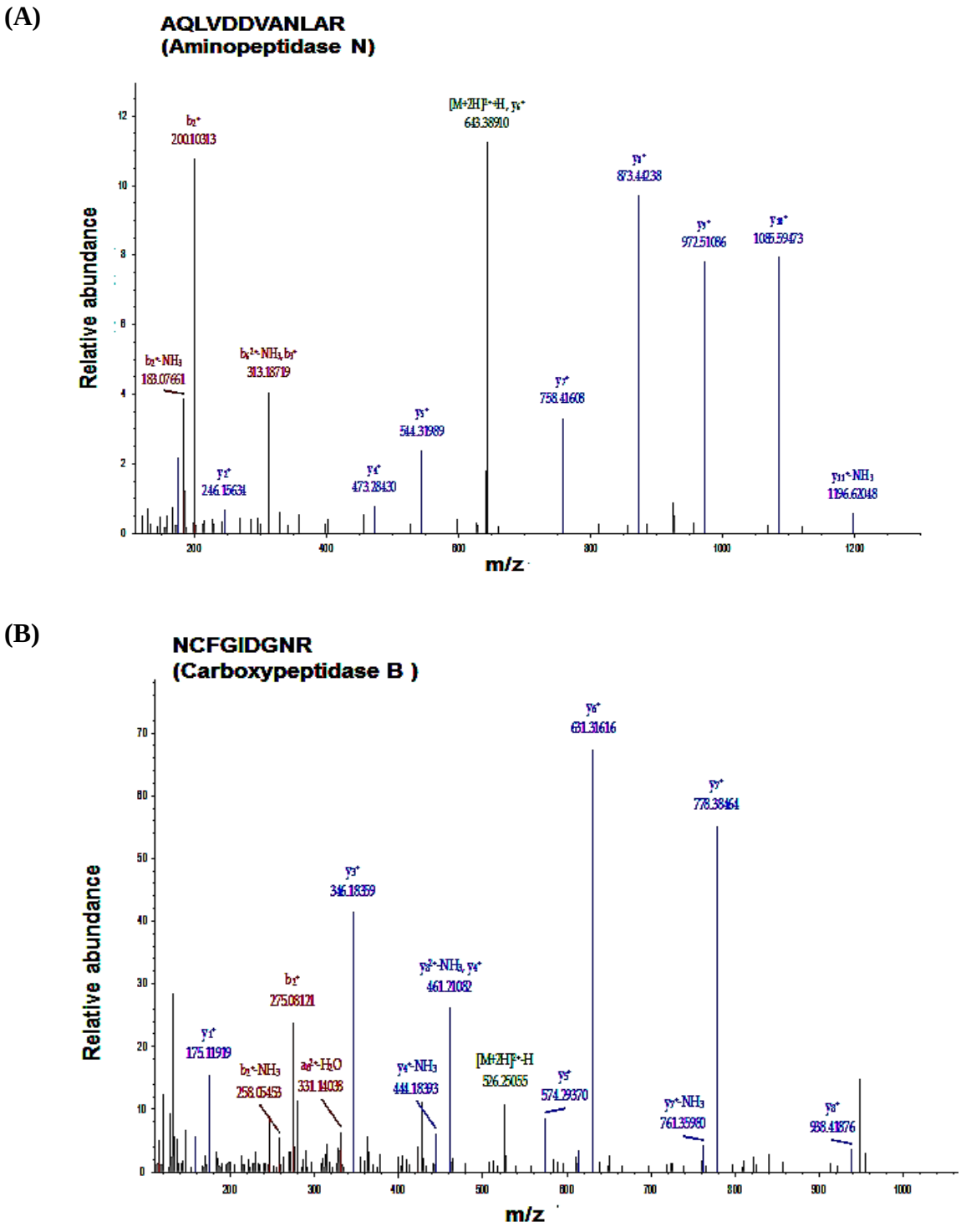


Figure 2.11 A: Representative MS/MS spectra of proteases identified in this study. (A) Peptide sequence **AQLVDDVANLAR** identified against Aminopeptidase N; (B) Peptide sequence **NCFGIDGNR** identified against Carboxypeptidase B.

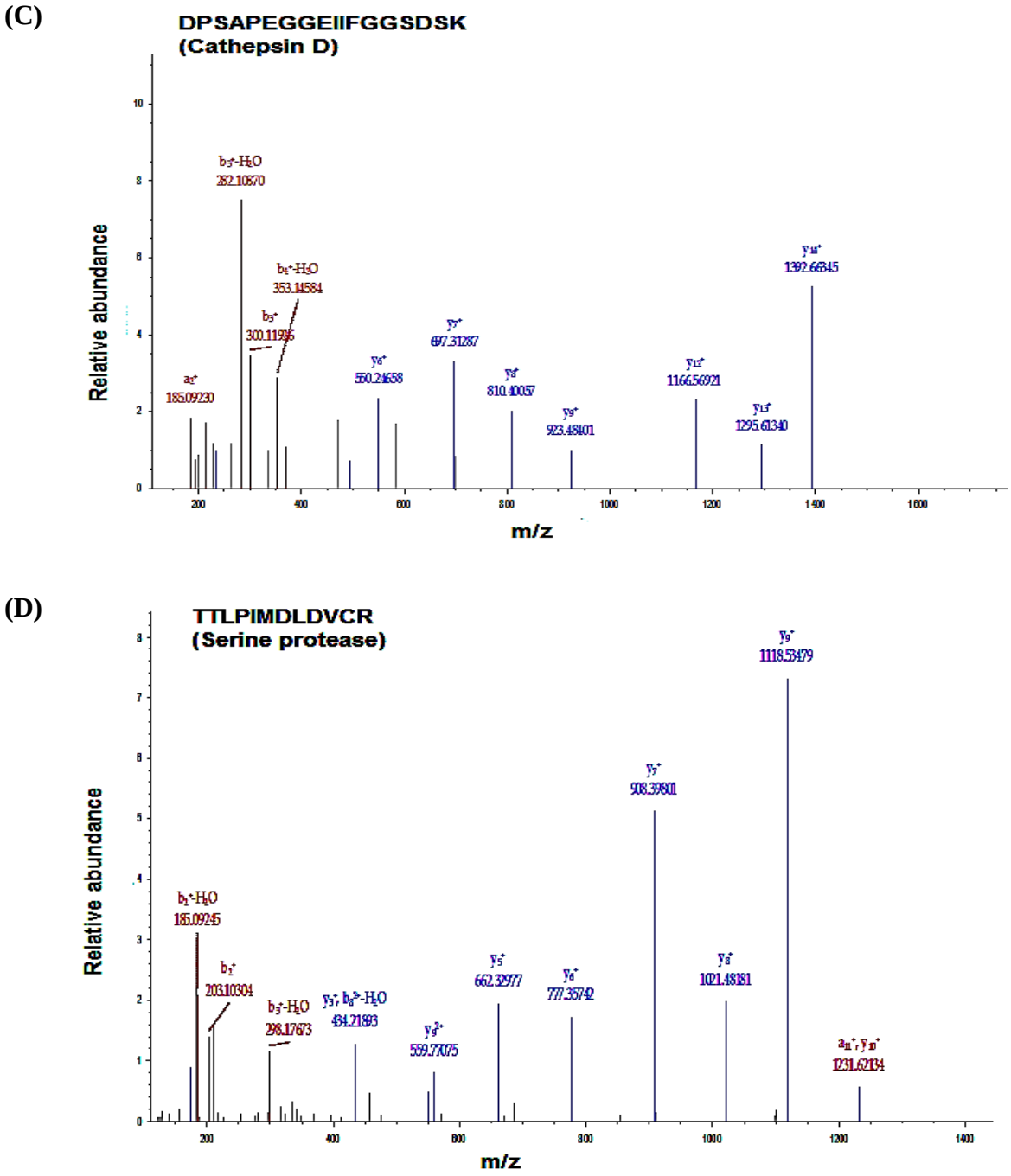


Figure 2.11 B: Representative MS/MS spectra of proteases identified in this study. (C) Peptide sequence **DPSAPEGGEIIFGGSDSK** identified against Cathepsin D; (D) Peptide sequence **TTLPIMDLDVCR** identified against Serine protease.

Jiang *et al.,* (2014) have reported 94 genes which play an important role in immune responses in *Anopheles stephensi*. Of these, 49 proteins have been identified in this study, which included ubiquitin-conjugating enzyme E2 N, GTP-binding protein Rheb precursor, NFkappaB essential modulator, FAS-associated factor 1, Pellino, TOLL pathway signalling NF-kappaB Relish-like transcription factor, Insulin-like peptide 3 precursor, Leucine-rich immune protein (Coil-less), Long leucine-rich immune protein (LRIM1), TOLL pathway signaling, TORC1 growth control complex subunit Kog1 and IMD pathway signaling IKK-beta (ird5) (Table 2.2). These immune response proteins are well studied in *An. gambiae* midgut and *Drosophila melanogaster* (Grewal, 2009; Laplante and Sabatini, 2012; Garver *et al*., 2012). They are thought to be either directly involved in anti-microbial defenses or anti-Plasmodium activity. Till date, very limited knowledge is available about the role of these immune genes in *An. stephensi* midgut. Representative MS/MS spectra of some of the immune responses proteins identified in this study are shown in Figure 2.12 A and B.

Table 2.2: List of immunity related proteins identified in *Anopheles stephensi* midgut.

S. No	Accession ID	Name	Σ Coverage	Σ# Unique Peptides	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]	NSAF values	Protein description as per *An. gambiae*
1	ASTEI092500-PA	Bendless	57.24	8	8	121	152	17.4	79.61	Ubiquitin-conjugating enzyme E2 N
2	ASTEI02599-PA	Uev1A	53.15	9	9	103	143	16.3	72.03	Ubiquitin-conjugating enzyme E2 variant
3	ASTEI09147-PA	Grb2	52.75	10	10	70	182	21.0	38.46	Growth factor receptor-binding protein 2
4	ASTEI01675-PA	4EBP1	54.10	3	4	25	122	13.2	20.49	Eukaryotic translation initiation factor 4E-binding protein 1
5	ASTEI08720-PA	Rheb	27.47	5	5	30	182	20.8	16.48	GTP-binding protein Rheb precursor
6	ASTEI03567-PA	Effettev2	9.62	3	3	32	364	41.8	8.79	Ubiquitin-conjugating enzyme E2D
7	ASTEI06370-PA	kenny	17.77	10	10	34	546	61.2	6.23	NFkappaB essential modulator
8	ASTEI03681-PA	Pellino	14.04	4	5	18	399	44.1	4.51	Pellino
9	ASTEI07221-PA	Casr	14.10	8	8	35	787	87.4	4.45	FAS-associated factor 1
10	ASTEI10386-PA	LRIM26	9.30	4	4	17	398	43.6	4.27	tRNA pseudouridine13 synthase
11	ASTEI04964-PA	Hrsv1	16.54	8	9	28	780	86.7	3.59	hepatocyte growth factor-regulated tyrosine kinase substrate
12	ASTEI03826-PA	REL1	7.04	6	9	46	1350	146.4	3.41	TOLL pathway signalling NF-kappaB Relish-like transcription factor
13	ASTEI02289-PA	Insulin receptor	12.30	5	7	24	740	84.0	3.24	Insulin receptor
14	ASTEI05349-PA	ILP1	12.00	1	1	4	150	16.9	2.67	Insulin-like peptide 3 precursor
15	ASTEI07671-PA	Cactus	8.66	3	3	10	381	41.4	2.62	TOLL pathway signalling.
16	ASTEI05261-PA	CYLDv1	10.22	11	11	33	1291	140.1	2.56	Ubiquitin thioesterase CYLD
17	ASTEI08192-PA	Spt6	11.00	15	15	46	1828	209.2	2.52	Transcription elongation factor SPT6

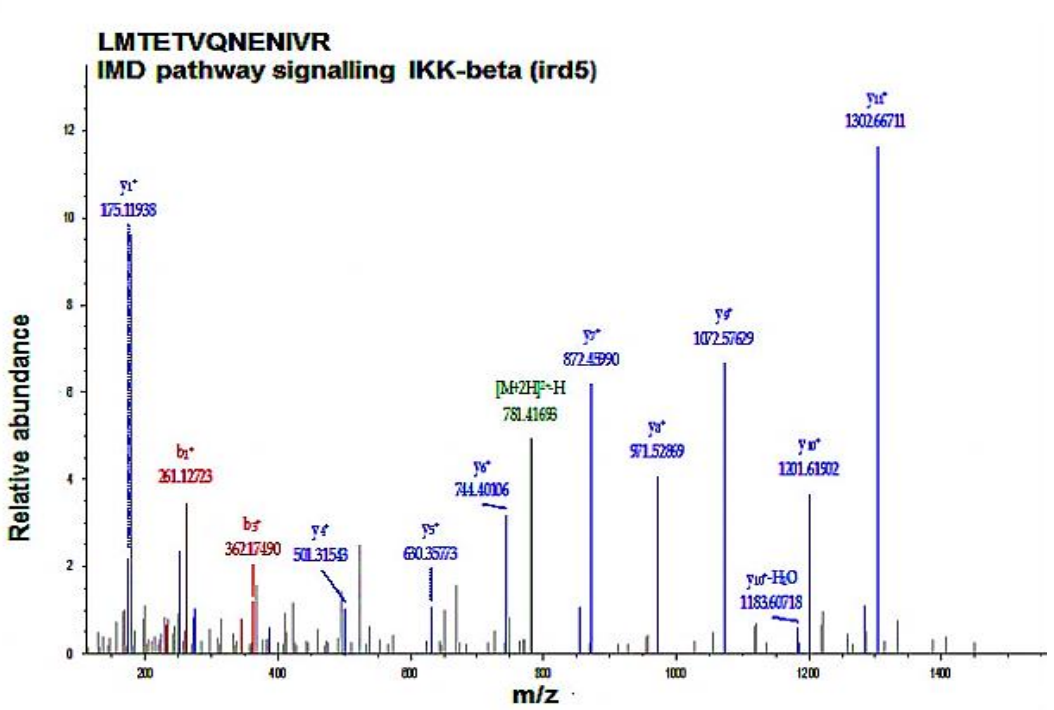
Table 2.2 contd: List of immunity related proteins identified in *Anopheles stephensi* midgut.

S. No	Accession ID	Name	Σ Coverage	Σ# Unique Peptides	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]	NSAF values	Protein description as per *An. gambiae*
18	ASTEI11789-PA	Slmb	12.50	1	1	2	80	9.7	2.5	F-box and WD-40 domain protein 1/11
19	ASTEI07471-PA	Slmb	9.23	5	5	12	509	55.5	2.36	F-box and WD-40 domain protein 1/11
20	ASTEI08241-PA	wispyv2	3.81	3	3	14	814	92.9	1.72	Not found in gambiae
21	ASTEI05393-PA	LRIM18	7.16	3	3	6	405	44.2	1.48	Leucine-rich immune protein (Coil-less)
22	ASTEI05348-PA	ILP3	4.79	1	1	2	167	18.8	1.20	Insulin-like peptide 3 precursor
23	ASTEI01297-PA	p70S6K	7.03	5	5	11	925	104.8	1.19	p70 ribosomal S6 kinase
24	ASTEI05202-PA	ird5	4.84	3	3	9	785	90.6	1.15	IMD pathway signalling IKK-beta (ird5 orthologue)
25	ASTEI09274-PA	LRIM6	8.81	2	2	4	352	40.1	1.14	Leucine-rich immune protein (Short)
26	ASTEI05746-PA	PDK1	4.92	5	5	14	1261	141.7	1.11	3-phosphoinositide dependent protein kinase-1
27	ASTEI03908-PA	Myopic	5.75	7	8	18	1757	192.9	1.02	Protein-tyrosine phosphatase
28	ASTEI03942-PA	Raptor	7.33	7	7	13	1365	151.2	0.95	TORC1 growth control complex subunit Kog1
29	ASTEI07728-PA	Target of Ramycin	4.76	8	8	21	2542	287.2	0.83	Autophagy related gene TOR
30	ASTEI02725-PA	FADD	6.05	1	1	2	248	29.5	0.81	IMD pathway signalling Fas-Associated Death Domain (FADD).
31	ASTEI06299-PA	TRAF6	1.33	1	1	4	527	59.8	0.76	TOLL pathway signalling TNF Receptor-Associated Factor (TRAF).
32	ASTEI02570-PA	LRIM11	4.89	1	1	3	409	46.9	0.73	leucine-rich immune protein (Short)
33	ASTEI05624-PA	LRIM19	5.15	1	1	2	291	33.9	0.69	leucine-rich immune protein (Coil-less)
34	ASTEI08579-PA	Dredd	3.15	2	2	4	603	66.8	0.66	Hypothetical protein

Table 2.2 contd: List of immunity related proteins identified in *Anopheles stephensi* midgut.

S. No	*An. stephensi* VectorBase ID	Name	Σ Coverage	Σ# Unique Peptides	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]	NSAF values	Protein description as per *An. gambiae*
35	ASTEI00302-PA	TUBE	4.71	2	2	4	701	76.9	0.57	TOLL pathway signalling
36	ASTEI03429-PA	nnierv2	2.34	1	1	4	727	78.9	0.55	Hypothetical protein
37	ASTEI05215-PA	SOS	2.44	3	3	8	1846	203.1	0.43	Protein son of sevenless
38	ASTEI02186-PA	PI3kP60	4.95	3	3	5	1171	130.2	0.43	Phosphoinositide-3-kinase, regulatory subunit
39	ASTEI09290-PA	LRIM1	1.95	1	1	2	514	58.0	0.39	Long leucine-rich immune protein (LRIM1)
40	ASTEI01382-PA	LRIM8A	2.31	1	1	2	520	58.6	0.38	Leucine-rich immune protein (Short)
41	ASTEI05629-PA	TSC	1.06	1	2	8	2265	246.7	0.35	Tuberous sclerosis 2
42	ASTEI05979-PA	Myd88	3.04	1	1	1	460	52.1	0.22	TOLL pathway signalling.
43	ASTEI03109-PA	ptip	0.56	1	1	4	1955	216.8	0.20	PAX-interacting protein 1
44	ASTEI08789-PA	POSH	1.10	1	1	2	998	110.7	0.20	Conserved oligomeric Golgi complex subunit 8
45	ASTEI04717-PA	PTEN	1.94	1	1	2	1032	111.2	0.19	phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase and dual-specificity
46	ASTEI08642-PA	Relish	0.70	1	1	2	1140	126.0	0.18	IMD pathway signalling NF-kappaB Relish-like transcription factor
47	ASTEI01430-PA	Scrawnyv1	0.66	1	1	2	1216	126.1	0.16	Ubiquitin carboxyl-terminal hydrolase 36/42
48	ASTEI10301-PA	Skd	0.39	1	1	2	2845	295.2	0.07	Mediator of RNA polymerase II transcription subunit 13
49	ASTEI03518-PA	TOLL1A	0.32	1	1	1	2187	248.0	0.05	TOLL-like receptor TOLL1A

(A)



(B)

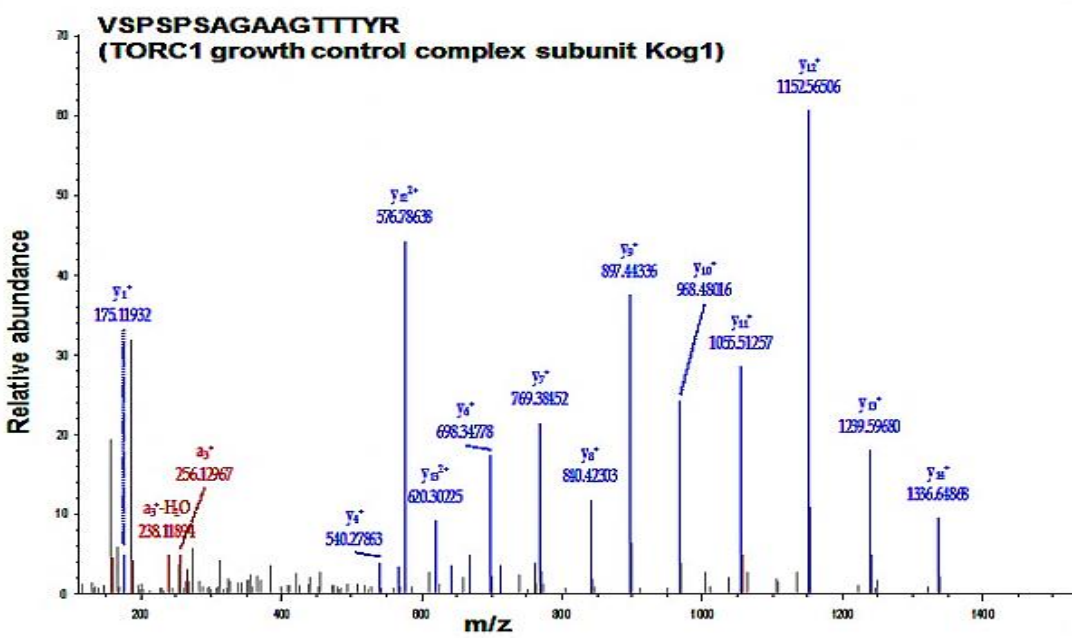
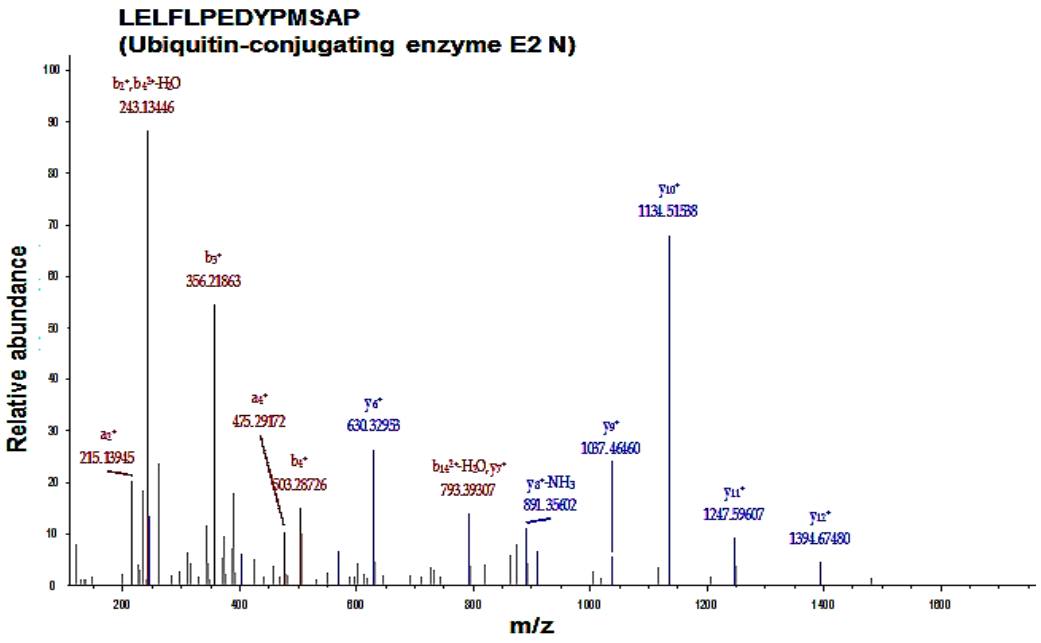


Figure 2.12 A: Representative MS/MS spectra of immune response proteins identified in this study. (A) Peptide sequence LMTETVQNENIVR identified against IMD pathway signaling IKK-beta (ird5); (B) Peptide sequence VSPSPSAGAAGTTTYR identified against TORC1 growth control complex subunit Kog1.

(C)



(D)

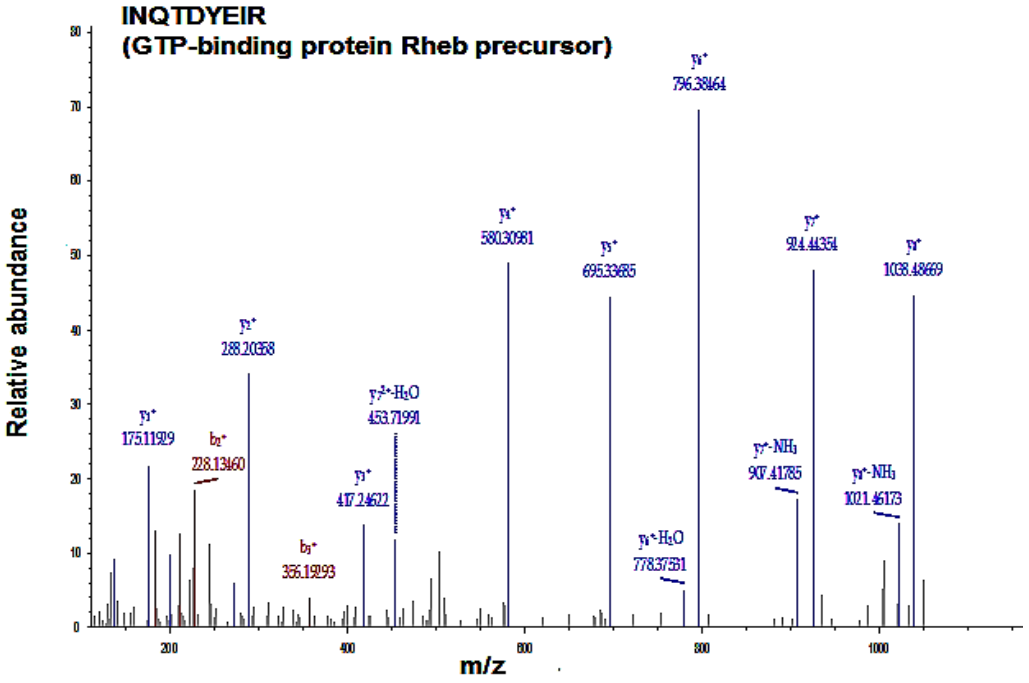
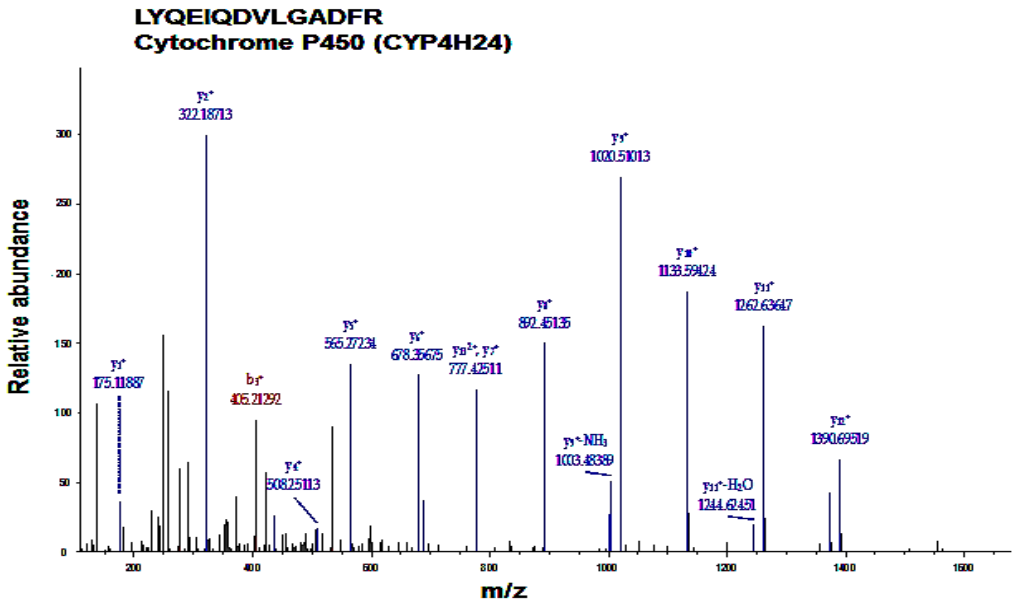


Figure 2.12 B: Representative MS/MS spectra of immune response proteins identified in this study. (C) Peptide sequence LELFLPEDYPMSAPK identified against Ubiquitin-conjugating enzyme E2 N; (D) Peptide sequence INQTDYEIR identified against GTP-binding protein Rheb precursor.

Several proteins identified in this study were found to be involved in antioxidant, oxidoreductive and stress-related processes. The proteins like glutathione S-transferase, cytochrome P450, alpha esterase, carboxylesterase, acetylcholinesterase, carboxypeptidase B, heme peroxidase, Cytochrome c oxidase subunit and mitochondrial inner membrane protein. These are proteins known to be associated with insecticide resistance which helps insects to metabolize insecticides at a higher rate. Some of these proteins have been identified in *An. gambiae*, *Aedes aegypti* and several other mosquito species (Biron *et al*., 2005; David *et al*., 2005). Representative MS/MS spectra of proteins associated with insecticide resistance mechanisms are shown in Figure 2.13 A and B.

Several heat-shock proteins (HSPs) were also identified in midgut. These heat shock proteins are involved in cellular defense against external stress from various sources (Lund *et al*., 2001). There are several proteins identified like ferritin light and heavy chain subunits, catalase, enolase, arginine/creatine kinases, aquaporin found to be involved in various metabolic processes in mosquitoes. There are many hypothetical proteins which are also identified with unknown function. Further functional studies could provide insights into the functional significance of such proteins in *Anopheles stephensi* midgut.

(A)



(B)

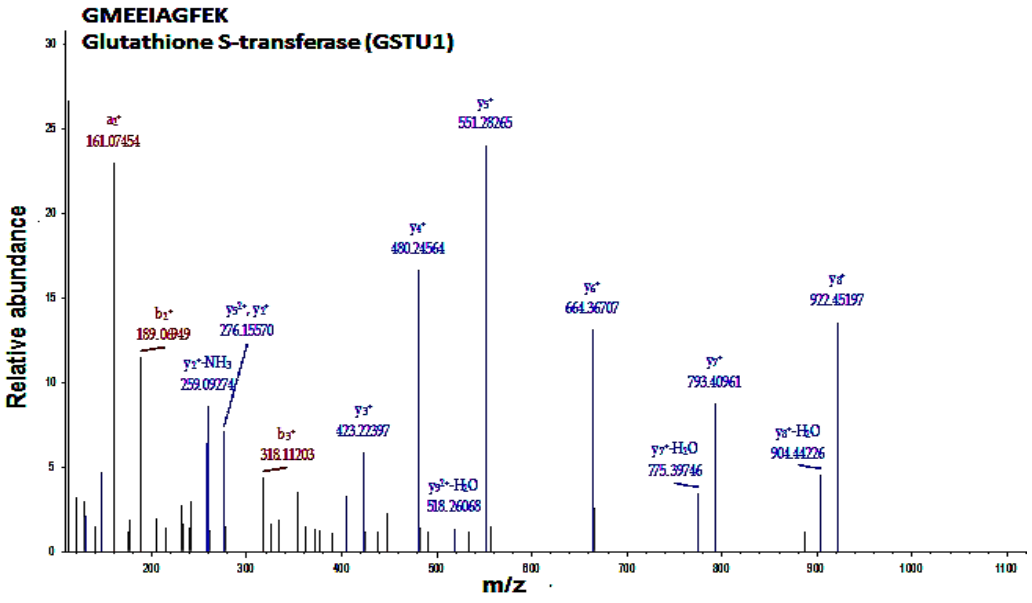
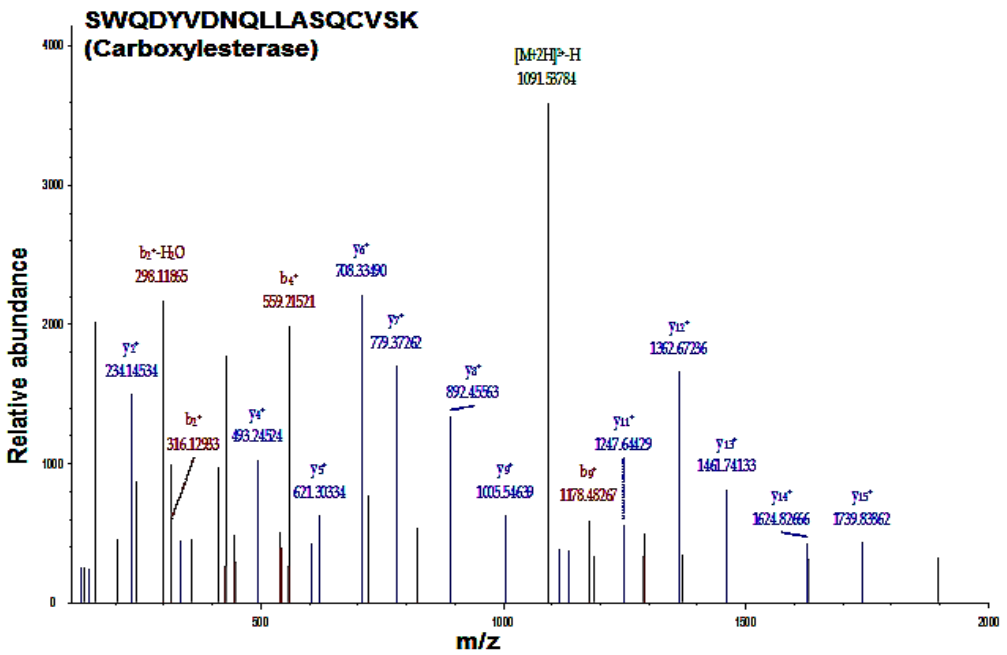


Figure 2.13 A: Representative MS/MS spectra of proteins involved in insecticide resistance mechanism identified in this study. (A) Peptide sequence **LYQEIQDVLGADFR** identified against Cytochrome P450 (CYP4H24); (B) Peptide sequence **GMEEIAGFEK** identified against Glutathione S-transferase (GSTU1).

(C)



(D)

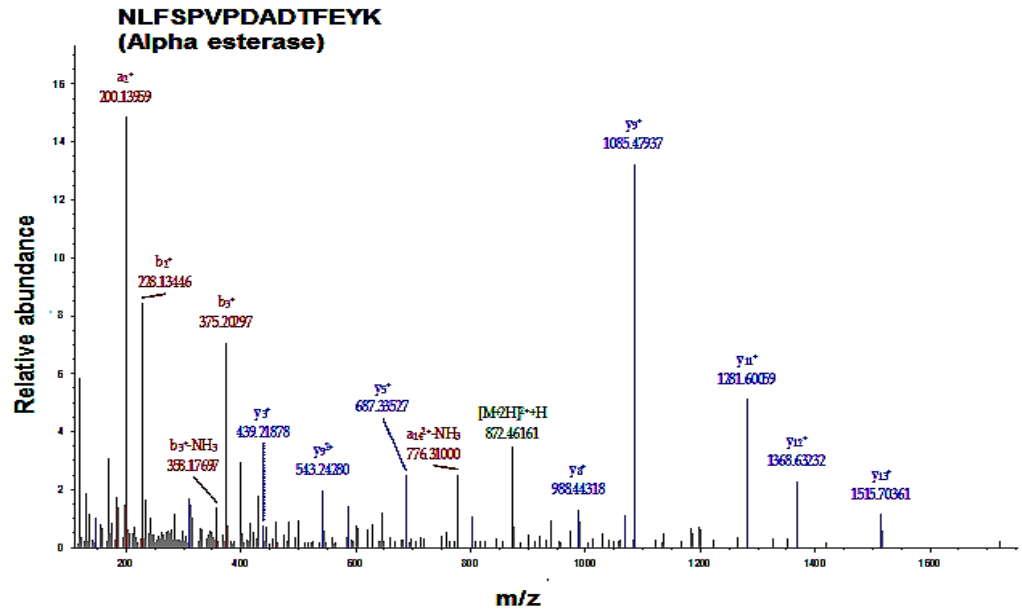


Figure 2.13 B: Representative MS/MS spectra of proteins involved in insecticide resistance mechanism identified in this study. (C) Peptide sequence **SWQDYVDNQLLASQCVSK** identified against Carboxylesterase; (D) Peptide sequence **NLFSPVPDADTFEYK** identified against Alpha esterase.

2.3.3 Proteins involved in parasite development/interaction

Several proteins have been identified in *Anopheles gambiae* which mediate the interaction of the *Plasmodium* parasites with mosquito tissues and also regulate their maturation inside host vectors (Sreenivasamurthy *et al*., 2013). In this study, 43 proteins were identified based on sequence similarity with *An. gambiae* proteins that are known to be involved in parasite development in mosquito. Of these, 18 proteins were found to be agonistic in nature that supports *Plasmodium* development in mosquito. Another 25 proteins were found to be antagonistic in nature that inhibits *Plasmodium* development in mosquito.

Among the eighteen proteins belonging to the agonistic group was adenine nucleotide translocator, Glutathione-S-transferase theta-1, Retinoid and fatty-acid binding glycoprotein (lipophorin or ApoII/I), Lysozyme c-1, Croquemort ortholog, Peptide-O-xylosyl transferase 1, Caspar, Cactus, Gelsolin etc. The list of agonistic proteins identified is provided in Table 2.3. These Agonistic molecules are associated with the *Plasmodium* development in mosquito *Anopheles gambiae* and their role has been confirmed using knock-down studies (Vlachou *et al*., 2005; Cheon *et al*., 2006; Dinglasan *et al*., 2007; Mendes *et al*., 2008; Garver*et al*., 2009; Jaramillo-Gutierrez *et al*., 2009; González-Lázaro *et al*., 2009; Rono *et al*., 2010; Kalja *et al*., 2011; Oliveira *et al*., 2011; Zou *et al*., 2011; Garver *et al*., 2012; Lapcharoen *et al*., 2012). Representative MS/MS spectra of some of the agonistic proteins identified in this study are shown in Figure 2.14 A and B. Twenty five proteins belonging to the antagonistic group includes Wiskott-Aldrich syndrome protein, Gram-negative bacteria binding protein A2, *Anopheles Plasmodium*-responsive Leucine-rich repeat protein 1B, Thioester-containing protein 1, Relish 1 etc. The list of antagonistic proteins identified is provided in Table 2.4. These antagonistic molecules prevent *Plasmodium* development in *Anopheles gambiae* and have been associated with increase in oocyst/sporozoite count upon knock-down (Meister *et al*., 2005; Vlachou *et al*., 2005; Warr et *al*., 2008; Dong *et al*., 2009; Garver *et al*., 2012; Molina-Cruz *et al*., 2012). MS/MS spectra of some of the antagonistic proteins identified in this study are shown in Figure 2.15 A and B.

Table 2.3: List of mosquito proteins involved in parasite interaction identified in *An. stephensi* midgut. **Agonistic molecules**: those that support *Plasmodium* development in mosquito midgut.

S. No	*An. stephensi* VectorBase ID	Protein name	*An. gambiae* ID	Protein description as per *An. gambiae*
1	ASTEI00150-PA	RFABG	AGAP001826	Retinoid and fatty-acid binding glycol protein (lipophorin or ApoII/I)
2	ASTEI00999-PA	GSTT1	AGAP000761	Glutathione-S-transferase theta-1
3	ASTEI02883-PA	PRS1	AGAP006102	*Plasmodium* responsive salivary 1
4	ASTEI01309-PA	LYSC-1	AGAP007347	Lysozyme c-1
5	ASTEI03999-PA	DUOX	AGAP009978	Dual oxidase
6	ASTEI08607-PA	ANT	AGAP006782	Adenine nucleotide translocator
7	ASTEI10836-PA	SCRBQ2	AGAP010133	Croquemort ortholog
8	ASTEI04167-PA	CLIPA8	AGAP010731	CLIP-domain serine protease subfamily A 8
9	ASTEI04270-PA	Gelsolin	AGAP011369	Gelsolin
10	ASTEI04472-PA	Saglin	AGAP000610	Saglin
11	ASTEI04537-PA	OXT1	AGAP005811	Peptide-O-xylosyltransferase 1
12	ASTEI04993-PA	ESP	AGAP010240	Epithelial serine protease
13	ASTEI056831-PA	CLIPA8	AGAP001648	CLIP-domain serine protease subfamily B17
14	ASTEI07221-PA	Caspar	AGAP006473	Caspar
15	ASTEI07671-PA	CACT	AGAP007938	Cactus
16	ASTEI08424-PA	IMPer	AGAP013327	Immunomodulatory peroxidase
17	ASTEI09198-PA	CLIPA5	AGAP011787	CLIP-domain serine protease subfamily A 5
18	ASTEI08922-PA	CLIPB3	AGAP003249	CLIP-domain serine protease subfamily B 3

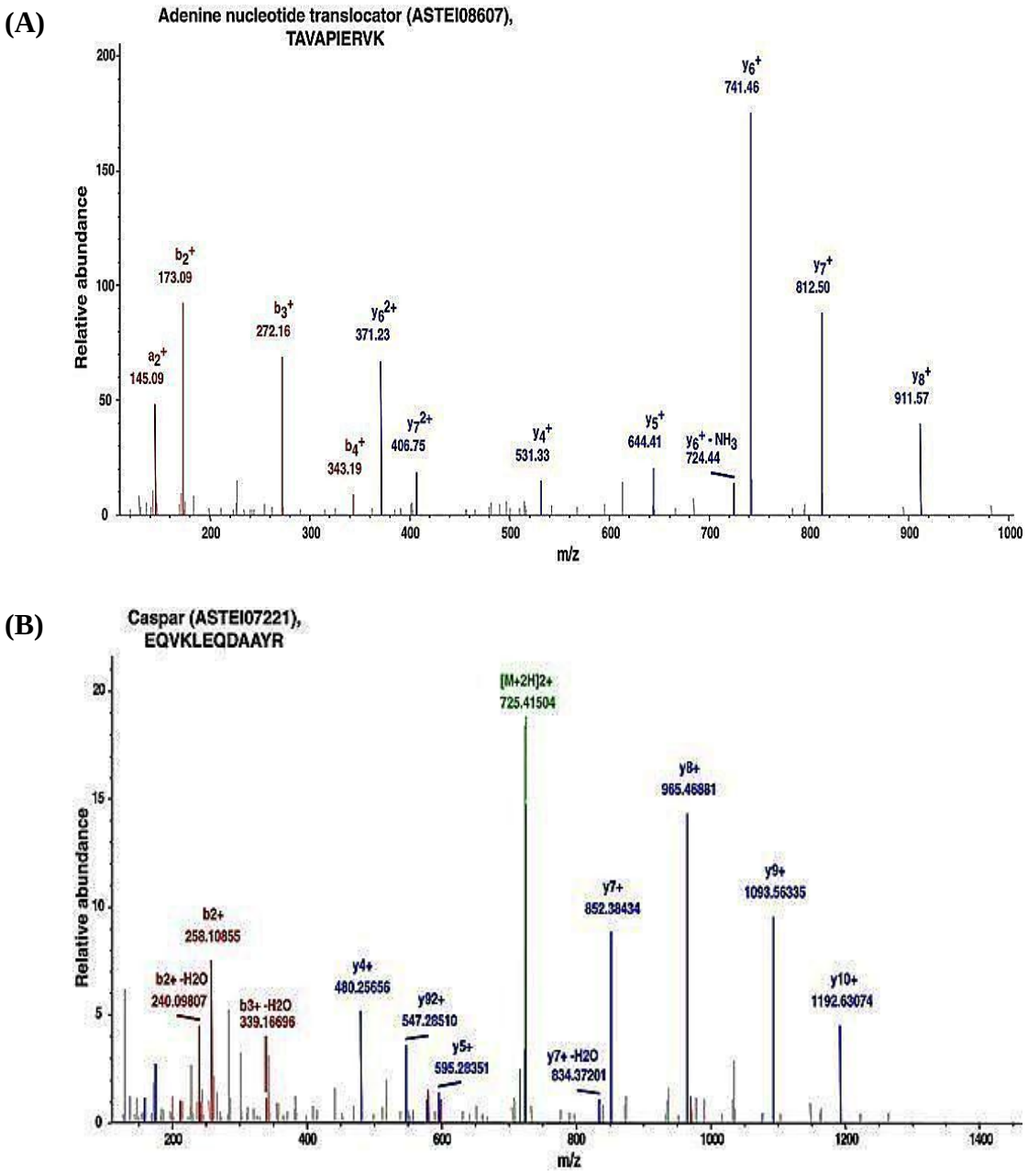
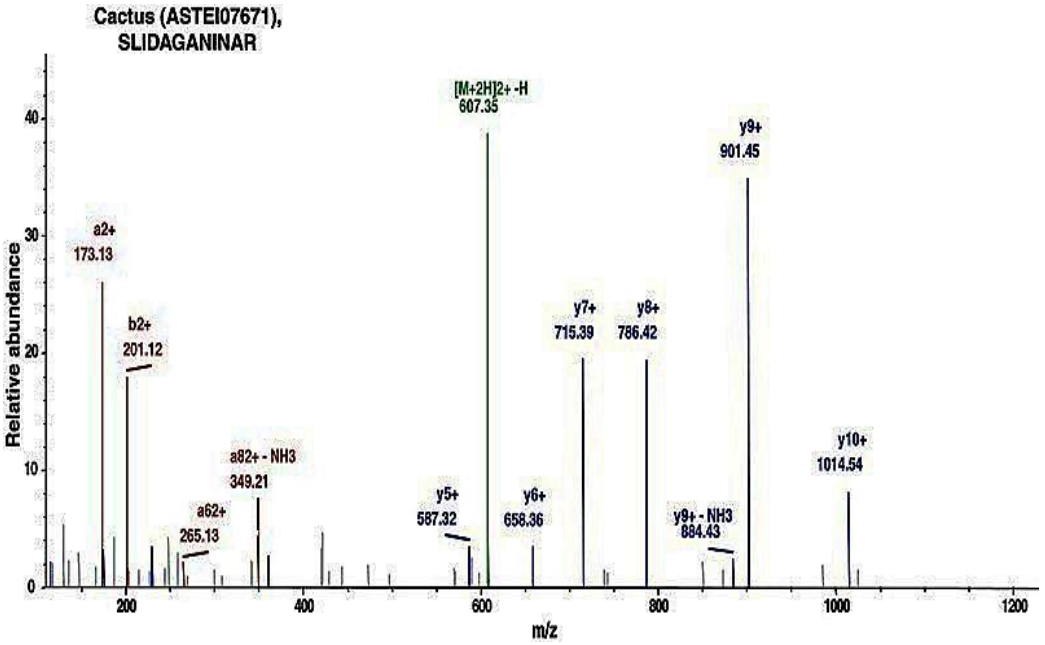


Figure 2.14 A: Representative MS/MS spectra of agonistic proteins those that aid *Plasmodium* development in mosquito identified in this study. (A) Peptide sequence **TAVAPIERVK** identified againstAdenine nucleotide translocator; (B) Peptide sequence **EQVKLEQDAAYR** identified against Caspar.

(C)



(D)

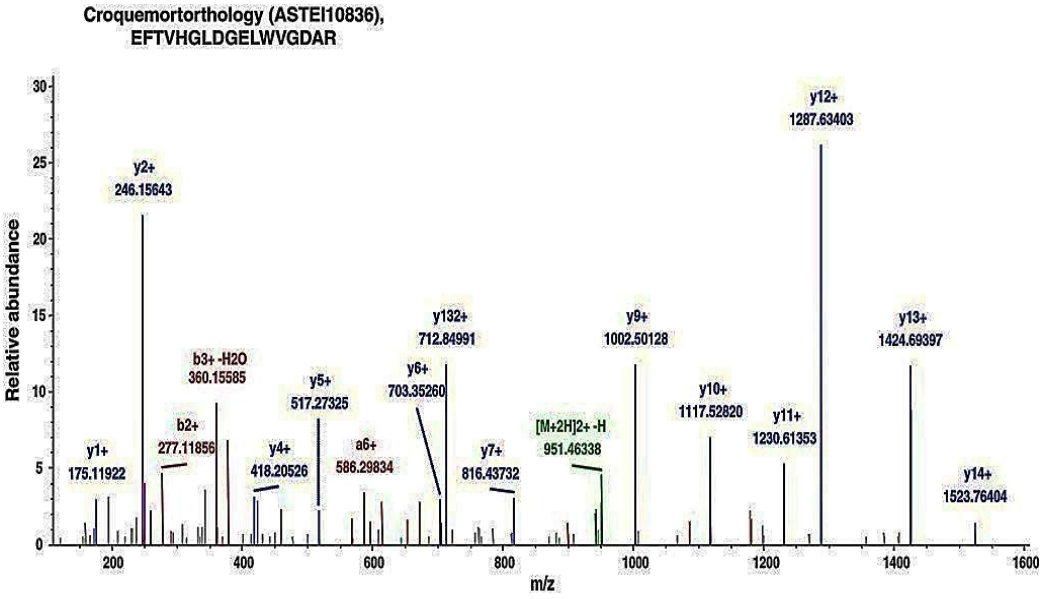


Figure 2.14 B: Representative MS/MS spectra of agonistic proteins those that aid *Plasmodium* development in mosquito identified in this study. (C) Peptide sequence **SLIDAGANINAR** identified against Cactus; (D) Peptide sequence **EFTVHGLDGELWVGDAR** identified against Croquemort ortholog.

Table 2.4: List of mosquito proteins involved in parasite interaction identified in *An. stephensi* midgut. **Antagonistic molecules**: those that inhibit *Plasmodium* development in mosquito midgut.

S. No	*An. stephensi* VectorBase ID	Protein name	*An. gambiae* ID	Protein description as per *An. gambiae*
1	ASTEI02301-PA	GNBPA2	AGAP012409	Gram-negative bacteria binding protein A2
2	ASTEI02361-PA	MDL1	AGAP012352	MD2-like receptor 1
3	ASTEI01142-PA	WASP	AGAP001081	Wiskott-Aldrich syndrome protein
4	ASTEI01898-PA	ARCP41	AGAP008908	Actin related 2/3 complex 41 KDa subunit P41
5	ASTEI02571-PA	APL1B	AGAP007035	Anopheles Plasmodium-responsive Leucine-rich repeat protein 1B
6	ASTEI02725-PA	FADD	AGAP007173	Fas-Associated Death Domain
7	ASTEI03111-PA	ARCP21	AGAP001712	Actin related 2/3 complex 21 KDa subunit P21
8	ASTEI04066-PA	APOD	AGAP002593	Apo lipoprotein
9	ASTEI05785-PA	MC1	AGAP001297	Mitochondrial carrier 1
10	ASTEI04897-PA	GNBPB1	AGAP004455	Gram-negative bacteria binding protein B1
11	ASTEI06240-PA	STAT	AGAP000099	Signal Transducers and Activators of Transcription
12	ASTEI06370-PA	IKK-gamma	AGAP005933	Inhibitor of kappa B kinase gamma
13	ASTEI06809-PA	FBN6	AGAP011231	Fibrinogen domain immunolectin 6
14	ASTEI08335-PA	GPRFZ2	AGAP010442	Frizzled-2
15	ASTEI08432-PA	TEP1	AGAP010815	Thioester-containing protein 1
16	ASTEI08579-PA	CASPL1	AGAP011693	Caspase long class
17	ASTEI08642-PA	REL2	AGAP006747	Relish-2
18	ASTEI05239-PA	SRPN6	AGAP009212	Serine protease inhibitor 6 (also known as serpin 6)
19	ASTEI09290-PA	LRIM1	AGAP006348	Leucine-Rich Immune Molecule 1
20	ASTEI09780-PA	LL3	AGAP009053	LITAF-like 3
21	ASTEI10265-PA	IRSP1	AGAP006421	Infection responsive secreted peptide 1
22	ASTEI10301-PA	Skd	AGAP006436	Skuld (also known as Med13 or TRAP240)
23	ASTEI10972-PA	GNBPB3	AGAP002799	Gram-negative bacteria binding protein B3
24	ASTEI00224-PA	ApoLp-III	AGAP013365	Apolipophorin-III
25	ASTEI03826-PA	REL1	AGAP009515-PA	Relish 1

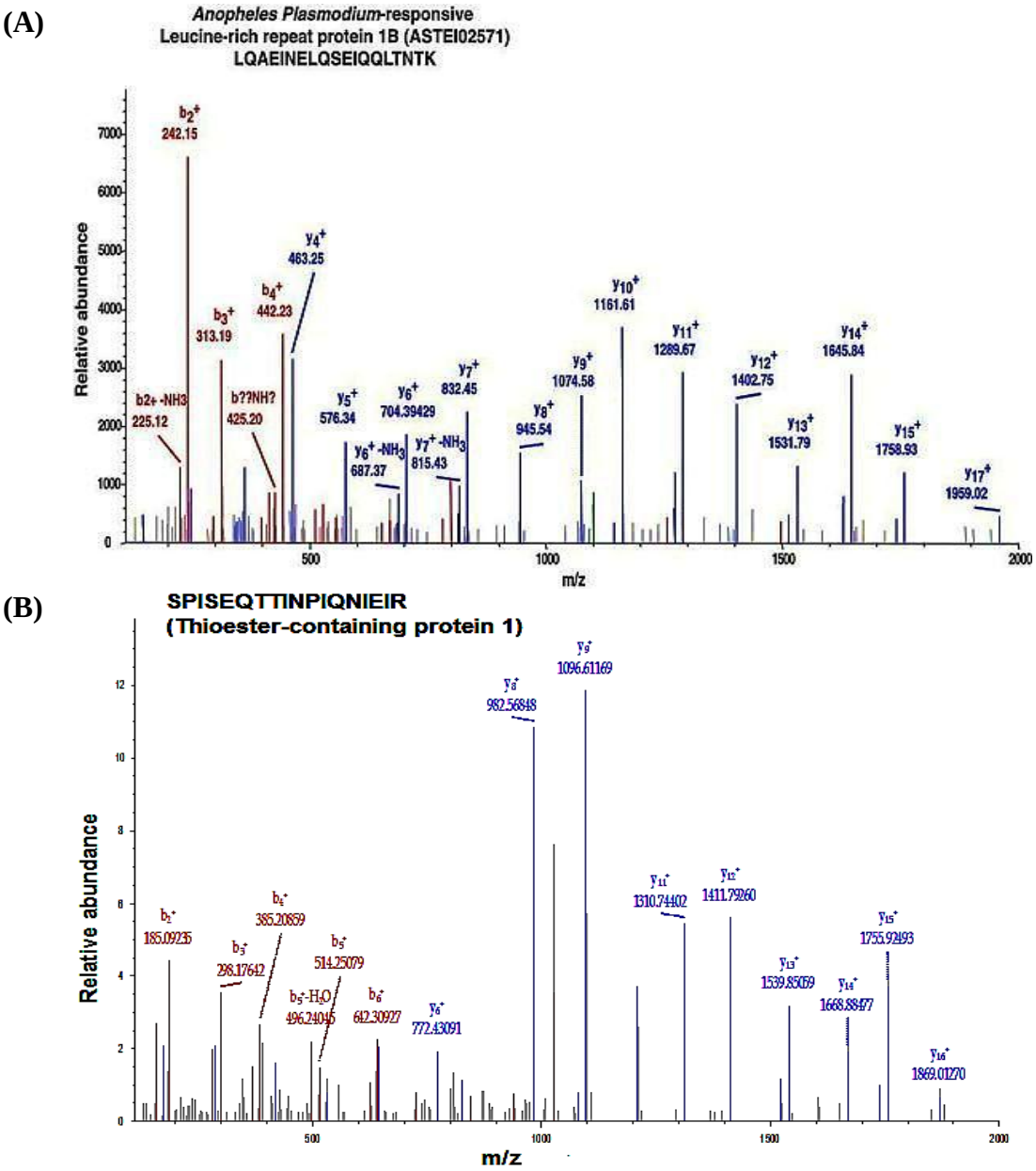
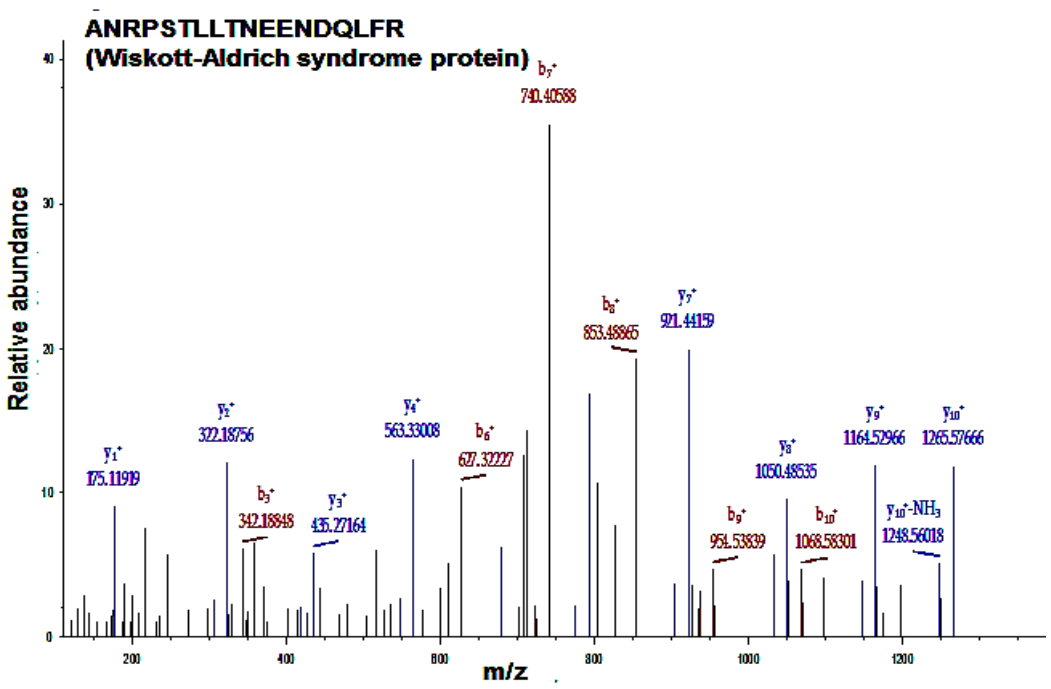


Figure 2.15 A: Representative MS/MS spectra of Antagonistic proteins those that prevent *Plasmodium* development in mosquito identified in this study. (A) Peptide sequence **LQAEINELQSEIQQLTNTK** identified against *Anopheles Plasmodium*-responsive leucine-rich repeat protein-1B; (B) Peptide sequence **SPISEQTTINPIQNIEIR** identified against Thioester containing protein-1.

(C)



(D)

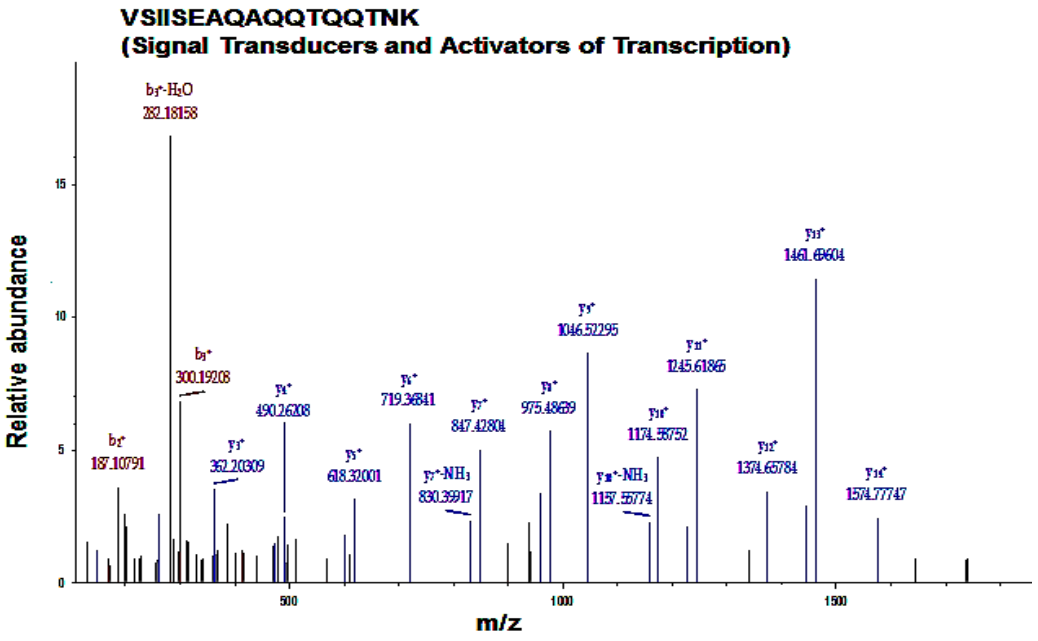
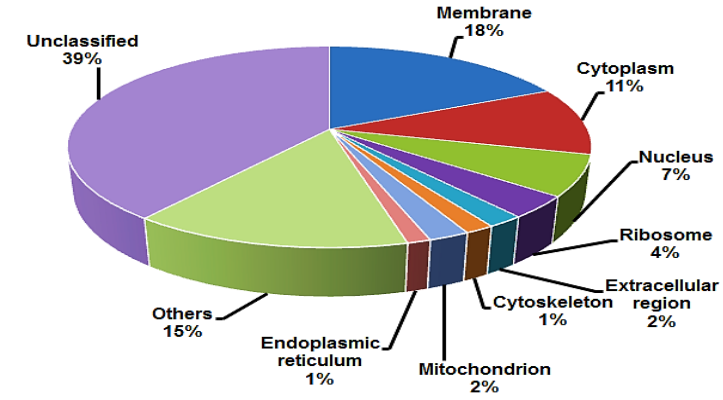


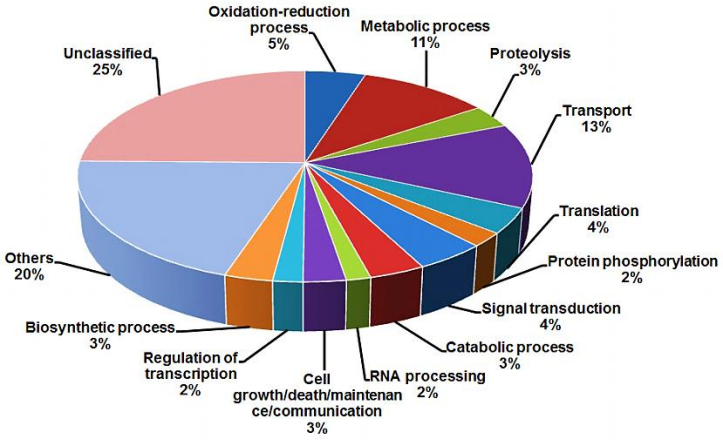
Figure 2.15 B: Representative MS/MS spectra of Antagonistic proteins those that prevent *Plasmodium* development in mosquito identified in this study. (C) Peptide sequence **ANRPSTLLTNEENDQLFR** identified against Wiskott- aldrich syndrome protein; (D) Peptide sequence **VSIISEAQAQQTQQTNK** identified against Signal transducers and activators of transcription.

2.3.4 Classification of proteins based on gene ontology annotation

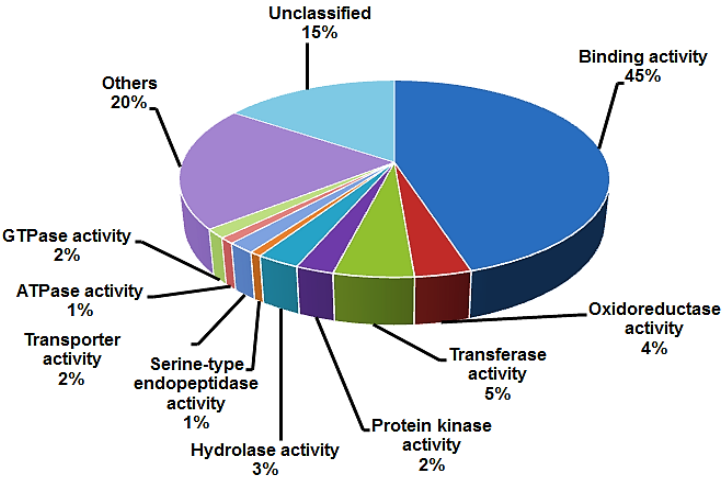
The identified proteins were functionally categorized based on their subcellular localization; biological processes and molecular function using gene ontology (GO) based annotation. The information for all the identified proteins was fetched from the VectorBase database. As shown in Figure 2.16 A, the majority of the proteins identified in this study were localized in the membrane (18%) followed by cytoplasm (11%), nucleus (7%), ribosome (4%) and mitochondrion (2%). A large majority (39%) of proteins are still unclassified in terms of their cellular localization. Biological process-based categorization showed that a majority of them played a role in transport (13%), metabolic process (11%), oxidation-reduction process (5%), signal transduction (4%) and translation (4%) (Figure 2.16 B). However, 25% of proteins are still unclassified in terms of their biological process. Classification based on molecular function showed that the majority of the identified proteins were involved in binding activity such as protein binding, ATP/GTP binding, nucleotide binding and calcium ion binding (45%), transferase activity (5%), oxido-reductase activity (4%), hydrolase activity (3%) and transporter activity (2%). About 15% of identified proteins are still unclassified in terms of their molecular function (Figure 2.16 C).



(A) Subcellular localization



(B) Biological process



(C) Molecular function

Figure 2.16: Gene ontology based classification of proteins identified in midgut tissue of sugar fed female *An. stephensi*. (A) Subcellular localization, (B) Biological process, (C) Molecular function.

2.3.5 Comparative proteomic analysis of male and female midgut

A quantitative proteomic analysis to compare male and female midgut proteome was carried out using iTRAQbased quantitative proteomic approach, to study differentially expressed proteins between these two sexes. Midgut proteins extracted from male and female mosquitoes were digested using trypsin enzyme followed by labeling with two different iTRAQ tags and SCX fractionation. Quantitative proteomic analysis of these iTRAQ labeled peptides was carried out by employing high resolution fourier transform mass spectrometry (LTQ-OrbitrapVelos). From this, a total of 1,30,453 tandem mass spectra were acquired, which were searched using two search algorithms Sequest and Mascot against VectorBase protein database of *Anopheles stephensi* Indian strain (database version Astel2.0). A total of 1,06,774 peptide-spectrum matches (PSMs) led to identification of 21,311 peptides and 3,316 proteins. Of these 3,316 proteins, identification of 2,434 proteins was supported by 2 or more unique peptides, while 881 protein identifications were based on single peptide evidence with multiple PSMs.

2.3.6 Identification of differentially expressed proteins between male and female midgut

In all, 758 proteins out of 3,316 total proteins were found to be differentially expressed in female midgut with respect to male counterpart. Of these differentially expressed proteins, 489 proteins were identified by multiple peptides and 269 protein identifications were based on single peptide evidence but with multiple PSMs. Relative quantification of peptides was done on the basis of relative intensity of reporter ions 116 and 117 for male and female midgut respectively. Protein ratios were calculated as the median of all the peptide ratios corresponding to respective proteins. Fold change of ≥2 was considered as over expressed while ≤0.5 was considered as under expressed (Puttamallesh *et al*., 2013). Among the differentially expressed proteins, 575 proteins were found over expressed (≥2 fold) in female midgut as compared to male midgut, while the remaining 183 proteins were under expressed (≤0.5 fold) as shown in Figure 2.17.

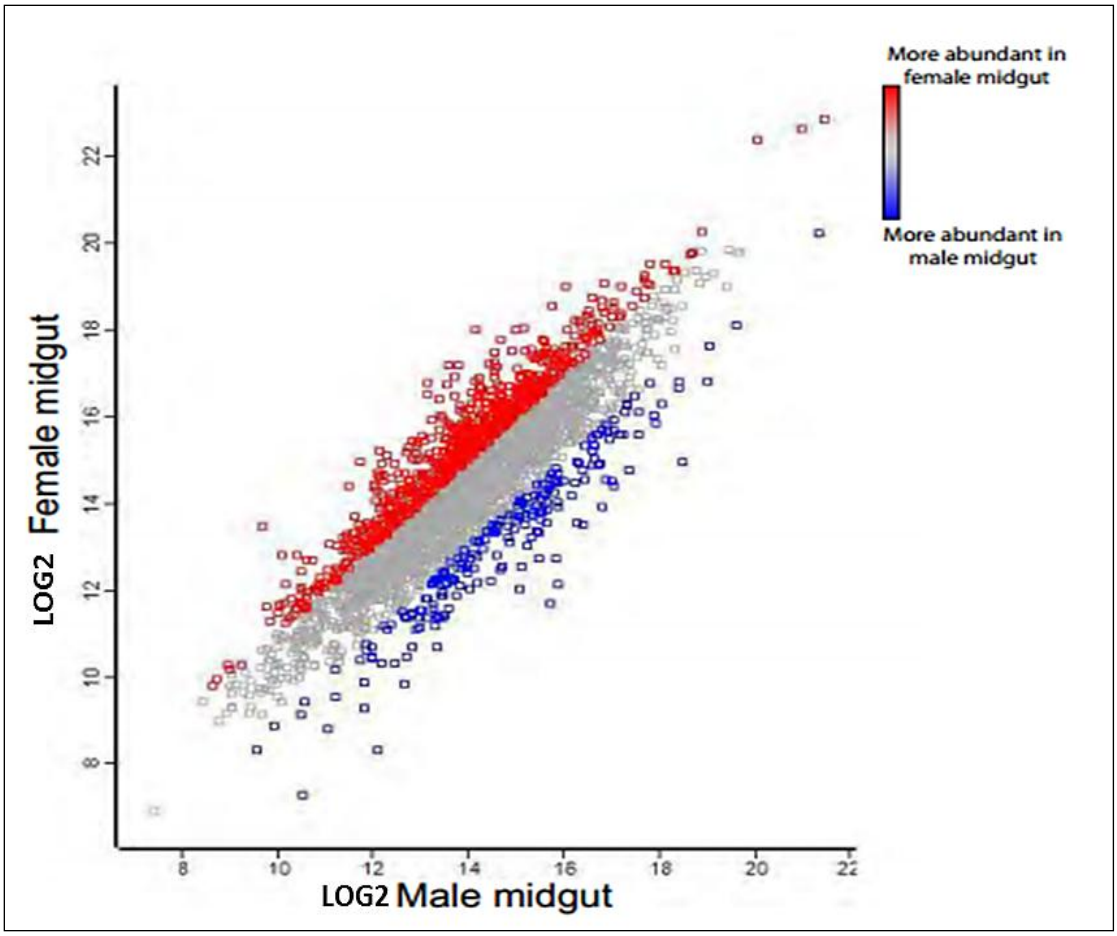


Figure 2.17: Differential expression of proteins between male and female midgut. Differentially expressed proteins identified are plotted against the LOG2 values of their fold change in female midgut when compared to that of male midgut. Red colored dots represent the over-expressed proteins (≥ 2-fold), whereas blue dots represent the under-expressed proteins (≤ 0.5-fold) in female midgut with respect to male midgut.

2.3.7 Proteins over expressed in female midgut

In female midgut, a total of 575 proteins were identified as over-expressed by >2fold as compared to male midgut. The over-expressed proteins includes F-box and leucine-rich repeat protein 6, calcineurin beta subunit, bromodomain-containing protein 8, Sodium-independent sulfate anion transporter, serine protease, nuclear transport factor 2, elongation factor 1 alpha-like protein. Partial list of proteins over-expressed in female midgut is provided in Table 2.5 and a complete list of over-expressed proteins can be seen in Appendix Table 2 A.

The most abundant protein identified was F-box and Leucine-rich Repeat Protein 6. This protein was found to be 44.1 fold over-expressed in female midgut as compared to male. This protein belongs to a member of a family of proteins that are characterized by an F-box motif and involved in phosphorylation-dependent ubiquitination. Another six proteins were found to be involved in the glycolysis pathway including enolase, fructose-bis-phosphate aldolase, triose-phosphate isomerase, phosphoglycerate kinase, 6-phosphofructo kinase 1and Phospho-glycerate mutase. These enzymes are over-expressed in female midgut as compared to the male mosquito, which suggests that glucose metabolism is higher in female mosquito than in the male mosquito. Several proteases like aminopeptidase N, carboxypeptidase B and serine proteases were found over expressed in female midgut by more than 2 fold than male midgut. Proteases are involved in proteolytic events that occur during the blood-feeding process in female mosquitoes. These proteases are also used by the *Plasmodium* parasites to activate its own enzymes during invasion of host cells for its survival (Shahabuddin *et al*., 1993). Over-expression of multiple ribosomal subunits and various translation initiation factors suggested enhanced protein synthesis in female midgut. Proteins involved in insecticide resistance like cytochrome p450 and glutathione S-transferase zeta class were found to be over-expressed in females. Over-expression of alkaline phosphatase was found in female midgut, this is known to serve as binding sites for *Bacillus thuringiensis* in insects (Thammasittirong *et al*., 2011). Heat shock proteins, which play a role in protein degradation, protein folding and signal transduction was found over expressed in female midgut.

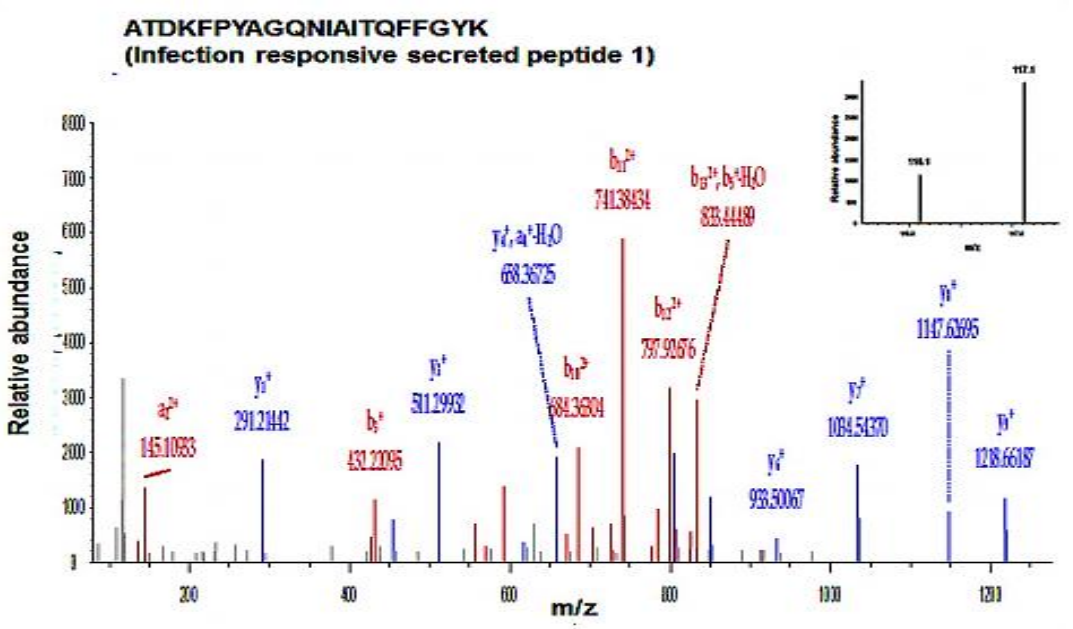
In this study, 18 immunity related proteins were expressed differentially between female and male midgut reported in *An. stephensi* by Jiang *et al*., (2014). Of these, protein FADD was found 2.6 fold over expressed in female midgut, whereas, leucine-rich immune protein (Coil-less) was found 0.4 fold times under-expressed in female midgut (Table 2.6). Previous studies have demonstrated the role of these genes in mosquito defense against parasite and viruses (Garver *et al*., 2009; Dong *et al*., 2011). Representative MS/MS spectra of some of the proteins over-expressed in female midgut are shown in Figure 2.18 A and B.

Table 2.5: A partial list of proteins over-expressed in female midgut identified in this study.

S. No	*An. stephensi* VectorBase ID	Protein description as per *An. gambiae*	Relative fold change female/male	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]
1	ASTEI10274-PA	F-box and leucine-rich repeat protein 6	44.163	1	4	917	99.2
2	ASTEI00645-PA	Calcineurin beta subunit	13.695	1	2	193	21.7
3	ASTEI00772-PA	Bromodomain-containing protein 8	11.125	2	4	1449	160.7
4	ASTEI10988-PA	Piggybac-derived 2	10.243	1	6	359	41.3
5	ASTEI09499-PA	Hypothetical protein	9.948	6	54	1385	150.9
6	ASTEI03820-PA	Hypothetical protein	9.809	2	4	1723	197.7
7	ASTEI06278-PA	Dnaj-like protein subfamily C	8.575	1	2	551	62.1
8	ASTEI07509-PA	Hypothetical protein	7.638	1	2	240	26.4
9	ASTEI03729-PA	Hypothetical protein	7.346	4	27	636	72.4
10	ASTEI02407-PA	Sodium-independent sulfate anion transporter	7.269	7	30	655	71.6
11	ASTEI11298-PA	Hypothetical protein	7.207	2	5	130	14.9
12	ASTEI08015-PA	Hypothetical protein	7.206	1	21	146	15.5
13	ASTEI025761-PA	Nuclear transport factor 2	6.677	4	16	130	14.6
14	ASTEI05236-PA	DNA-directed RNA polymerase II subunit RPB11	6.619	1	1	117	13.5
15	ASTEI08033-PA	Serine protease	6.599	4	49	260	28.8
16	ASTEI04077-PA	Elongation factor 1 alpha-like protein	6.364	3	8	877	96.3

Table 2.5 contd: Partial list of proteins over-expressed in female midgut identified in this study.							
S. No	***An. stephensi* VectorBase ID**	**Protein description as per *An. gambiae***	**Relative fold change female/male**	**Σ# Peptides**	**Σ# PSMs**	**# AAs**	**MW [kDa]**
17	ASTEI06639-PA	Hypothetical protein	6.058	1	2	415	46.2
18	ASTEI08034-PA	Serine protease	6.023	3	58	268	29.1
19	ASTEI09575-PA	Hypothetical protein	6.004	1	2	582	67.3
20	ASTEI11463-PA	Hypothetical protein	6.003	2	3	108	12.3
21	ASTEI05455-PA	Hypothetical protein	5.985	4	12	229	22.7
22	ASTEI03153-PA	Hypothetical protein	5.625	1	2	244	27.5
23	ASTEI01570-PA	Anaphase-promoting complex subunit 3	5.463	1	2	1197	132.4
24	ASTEI11423-PA	Hypothetical protein	5.389	1	4	181	20.7
25	ASTEI09278-PA	Ccdc109a	5.364	2	5	312	35.7
26	ASTEI02774-PA	Pyridoxine kinase	5.364	14	124	903	96.6
27	ASTEI07669-PA	Hypothetical protein	5.358	2	6	99	11.2
28	ASTEI03928-PA	Hypothetical protein	5.356	12	137	914	102.2
29	ASTEI02547-PA	Hypothetical protein	5.296	1	1	3093	347.7
30	ASTEI10087-PA	Hypothetical protein	5.260	1	4	64	7.6
31	ASTEI08929-PA	N-myc downstream regulated	5.207	10	87	330	36.3
32	ASTEI10850-PA	Mitochondrial ribosomal protein L40	5.149	1	4	196	22.8

(A)



(B)

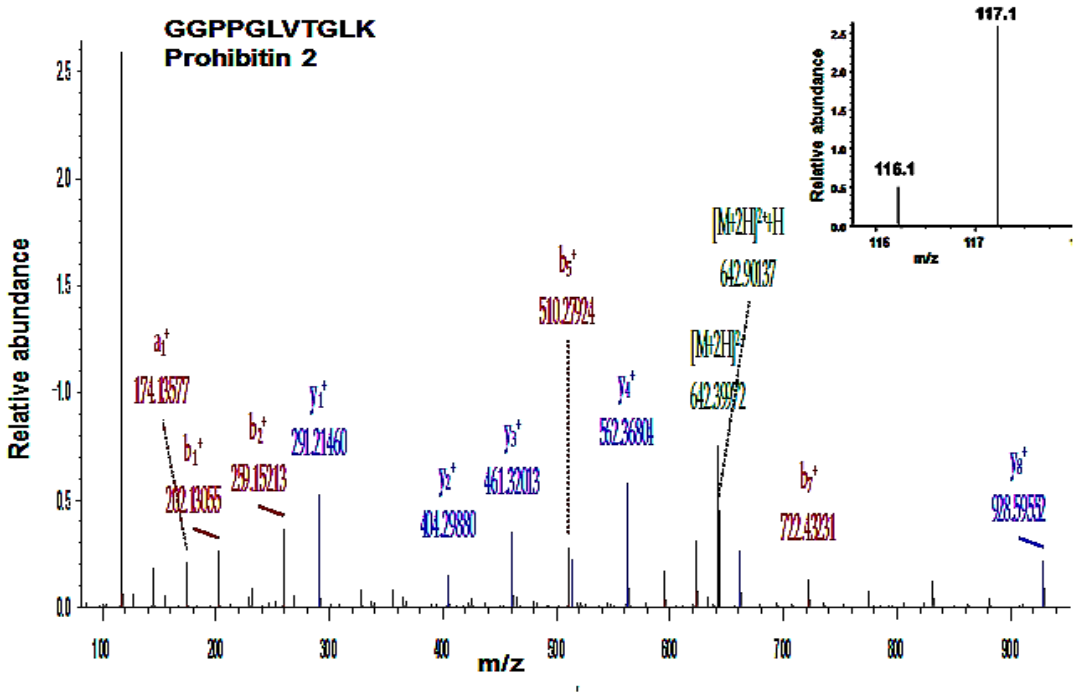
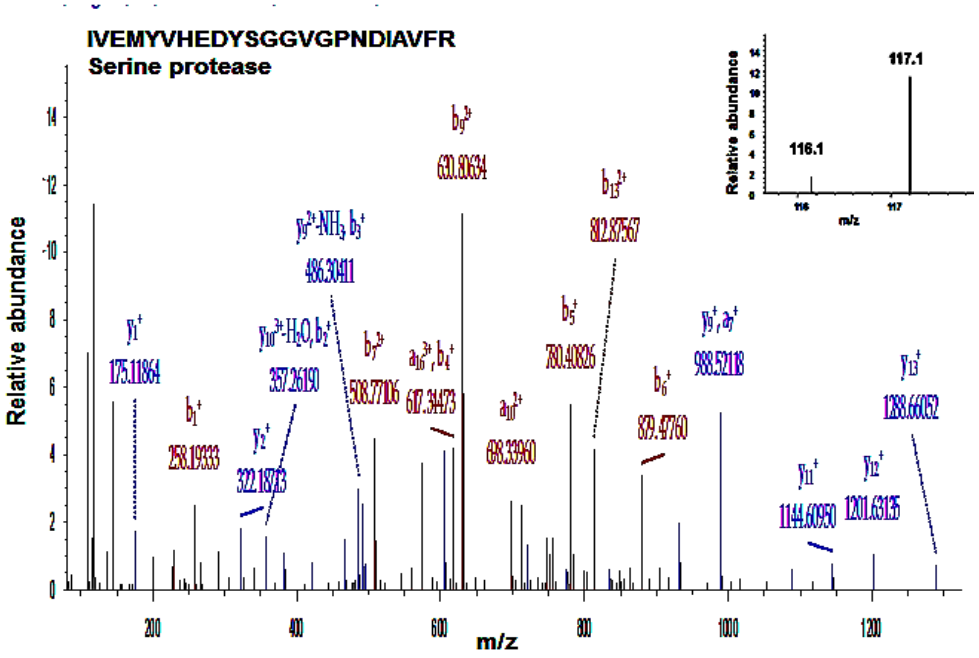


Figure 2.18 A: MS/MS spectra of peptides for the proteins over-expressed in female midgut (A) Peptide sequence **ATDKFPYAGQNIAITQFFGYK** identified against Infection responsive secreted peptide 1, (B) Peptide sequence **GGPPGLVTGLK** identified against Prohibitin 2.

(C)



(D)

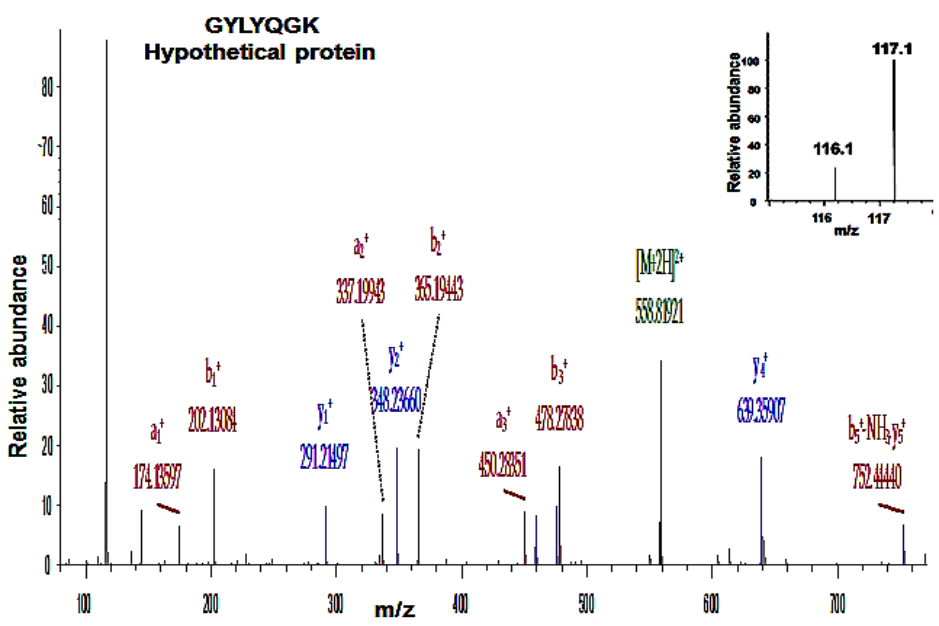


Figure 2.18 B: MS/MS spectra of peptides for the proteins over-expressed in female midgut (C) Peptide sequence **IVEMYVHEDYSGGVGPNDIAVFR** identified against Serine protease; (D) Peptide sequence **GYLYQGK** identified against hypothetical protein.

Table 2.6: List of immunity related proteins identified and their expression level in female and male midgut tissue.

S. No.	*An. stephensi* VectorBase ID	Relative expression level male/female	Relative fold change female/male	Name	Protein description as per *An. gambiae*	Molecular function	Biological process
1	ASTEI02725-PA	0.377	2.652	FADD	IMD pathway signaling Fas-Associated Death Domain (FADD).	Protein binding	Signal transduction,
2	ASTEI092504-PA	0.543	1.840	PGRPLC common	Peptidoglycan recognition protein (Long)	Hydrolase activity	Peptidoglycan catabolic process
3	ASTEI05746-PA	0.571	1.750	PDK1	Hypothetical protein	Protein kinase activity/transferase activity	Cellular protein modification process
4	ASTEI092500-PA	0.605	1.654	Bendless	Ubiquitin-conjugating enzyme E2 N	Unknown	Unknown
5	ASTEI08192-PA	0.608	1.643	Spt6	Transcription elongation factor SPT6	DNA binding	Cellular nitrogen compound metabolic process
6	ASTEI08720-PA	0.629	1.589	Rheb	GTP-binding protein Rheb precursor	Ion binding	Signal transduction/transport
7	ASTEI01297-PA	0.719	1.390	p70S6K	p70 ribosomal S6 kinase	Kinase activity	Cellular protein modification process
8	ASTEI09147-PA	0.805	1.242	Grb2	Growth factor receptor-binding protein 2	Protein binding	Unknown
9	ASTEI03826-PA	0.826	1.211	REL1	TOLL pathway signaling NF-kappaB Relish-like transcription factor	DNA binding transcription factor activity	Regulation of DNA transcription,
10	ASTEI03942-PA	0.902	1.108	Raptor	TORC1 growth control complex subunit Kog1	Protein binding	TOR signaling
11	ASTEI09290-PA	0.933	1.072	LRIM1	Leucine-rich immune protein (Long)	Protein binding	Unknown
12	ASTEI04964-PA	1.006	0.994	Hrsv1	Hepatocyte growth factor-regulated tyrosine kinase substrate	Metal ion binding	Protein transport
13	ASTEI02325-PA	1.048	0.954	TOLL8	Toll	Protein binding	Signal transduction
14	ASTEI06299-PA	1.088	0.919	TRAF6	TOLL pathway signaling TNF Receptor-Associated Factor (TRAF)	Protein binding	Unknown
15	ASTEI07471-PA	1.122	0.891	Slmb	F-box and WD-40 domain protein 1/11	Protein binding	Unknown
16	ASTEI02567-PA	1.133	0.883	LRIM4	Leucine-rich immune protein (Long)	Protein binding	Unknown
17	ASTEI10386-PA	1.274	0.785	LRIM26	Leucine-rich immune protein (Coil-less)	Protein binding	Unknown
18	ASTEI05393-PA	2.357	0.424	LRIM18	Leucine-rich immune protein (Coil-less)	Protein binding	Unknown

Proteins involved in vector parasite interaction were also identified. 31 proteins were identified which are known to play a role in vector-*Plasmodium* interaction (Table 2.7).
Of these 12 proteins are agonistic in nature while 19 are antagonistic in nature. Two proteins like Fas-Associated Death Domain (FADD) and Infection responsive secreted peptide 1(IRSP1) were found to be over-expressed by 2.6 and 2.2 times in female midgut respectively. These two proteins are known to suppress *Plasmodium* infection in mosquito midgut which may be involved in mosquito immunity in response to infection (Dong *et al*., 2006; Garver *et al*., 2012).

Interestingly, three proteins such as Gram-negative bacteria binding protein B4, Beta thymosin family and Cell division cycle 42, are known to inhibit *Plasmodium* infection were found to be under-expressed in female. In contrast, another protein Caspar known to support oocyst development in midgut was also found under-expressed. siRNA based gene silencing of Gram negative bacteria binding protein B4 in *An. gambiae* have shown an enhanced rate of oocyst infection upon its gene silencing (Warr *et al*., 2008). Similarly, Beta thymosin family (Ciboulot) is known to prevent *Plasmodium* development in mosquito (oocyst number increases upon knock-down) and was found to be 0.46 times under-expressed in female midgut (Vlachou *et al*., 2005). FAS-associated factor 1(Caspar) is known to support *Plasmodium* development in mosquito midgut, gene silencing of Caspar gene results in decrease in oocyst formation in mosquito *An. gambiae*. This protein was found 0.37 times under-expressed in female midgut (Garver *et al*., 2009, 2012).

Table 2.7: List of proteins involved in vector parasite interaction and their relative expression level between female and male midgut tissue.

(A). Antagonistic molecules: Inhibit the oocyst formation in mosquito midgut.				
S. No.	***An. stephensi* VectorBase ID**	**Relative fold change female/male**	**Protein name**	**Protein description as per *An. gambiae***
1	ASTEI02725-PA	2.652	FADD	Fas-Associated Death Domain
2	ASTEI10265-PA	2.227	IRSP1	Infection responsive secreted peptide 1
3	ASTEI04705-PA	1.887	Prohibit	Prohibitin 2
4	ASTEI01142-PA	1.510	WASP	Wiskott-Aldrich syndrome protein
5	ASTEI06240-PA	1.464	STAT	Signal Transducers and Activators of Transcription
6	ASTEI03826-PA	1.211	REL1	Relish 1
7	ASTEI07645-PA	1.207	Hsc-3	heat shock 70kDa protein 5
8	ASTEI01898-PA	1.193	ARC P41	Actin related 2/3 complex 41 KDa subunit P41
9	ASTEI07389-PA	1.184	JNK	Jun N-terminal Kinase
10	ASTEI03111-PA	1.099	ARC P21	Actin related 2/3 complex 21 KDa subunit P21
11	ASTEI09290-PA	1.072	LRIM1	Leucine-Rich Immune Molecule 1
12	ASTEI03480-PA	0.925	GSTT2	Glutathione-S-transferase theta-2
13	ASTEI05239-PA	0.790	SRPN6	Serine protease inhibitor 6 (also known as serpin 6)
14	ASTEI00224-PA	0.725	ApoLp-III	Apolipophorin-III
15	ASTEI08432-PA	0.680	TEP1	Thioester-containing protein 1
16	ASTEI00731-PA	0.589	Cdc42	Cell division cycle 42
17	ASTEI06691-PA	0.473		aquaporin
18	ASTEI01099-PA	0.460	Ciboulot	Beta thymosin family
19	ASTEI10972-PA	0.338	GNBPB4	Gram-negative bacteria binding protein B4

Table 2.7 contd: List of proteins involved in vector parasite interaction and their expression level between female and male midgut tissue.

(B) Agonistic molecules: support the oocyst formation in mosquito midgut.				
S. No.	***An. stephensi* VectorBase ID**	**Relative fold change female/male**	**Protein name**	**Protein description as per *An. gambiae***
1	ASTEI08607-PA	1.997	ANT	Adenine nucleotide translocator
2	ASTEI092504-PA	1.840	PGRPLC	PGN Recognition Protein LC
3	ASTEI07966-PA	1.462	SDR1	Short-chain dehydrogenases/reductases
4	ASTEI04504-PA	1.164	CPR	Cytochrome P450 reductase
5	ASTEI01530-PA	1.116	LANB2	Laminin gamma 1
6	ASTEI10836-PA	1.100	SCRBQ2	Class B Scavenger Receptor (CD36 domain). Croquemort ortholog
7	ASTEI02221-PA	0.966	ArgK	Arginine kinase
8	ASTEI04270-PA	0.937	Gelsolin	Gelsolin
9	ASTEI00999-PA	0.881	GSTT1	Glutathione-S-transferase theta-1
10	ASTEI07737-PA	0.846	CP	F-actin capping protein
11	ASTEI00150-PA	0.716	RFABG	Retinoid and fatty-acid binding glycoprotein, also known as lipophorin or ApoII/l
12	ASTEI07221-PA	0.374	Caspar	Caspar

2.3.8 Proteins over expressed in male midgut

In the male midgut, 183 proteins were found to be over-expressed (>2fold) than the female midgut. The over-expressed proteins included gustatory receptor, translocon-associated protein subunit gamma, sodium-coupled monocarboxylate transporter 1, cuticular protein RR-1 family (CPR24), innexin inx2, ankyrin, tubulin polymerization-promoting protein, 3-Glucan Binding Protein (BGBP), Fas-associated factor 1. Partial list of proteins over-expressed in male midgut are provided in Table 2.8 and a complete list of proteins can be seen in Appendix Table 2 B. An insect gustatory receptor protein belongs to gustatory and odorant receptors were found to be 19.4 times over-expressed in male as compared to female. In insects this protein plays a role in detection of sugar feeding sources, blood-feeding hosts, and oviposition sites (Freeman *et al*., 2014). Maltase was found to be over expressed by 2.4 fold than female which is a glycosidase involved in the digestion of sugars as male mosquitoes feeds on plant juices. Another protein, cuticular protein RR-1 family (CPR24) is a member of CPR family found to be 6.1 fold over-expressed in male midgut and plays a role in forming the insect exoskeleton and other cuticular structures. Representative MS/MS spectra of some of the proteins over-expressed in male midgut are shown in Figure 2.19 A and B.

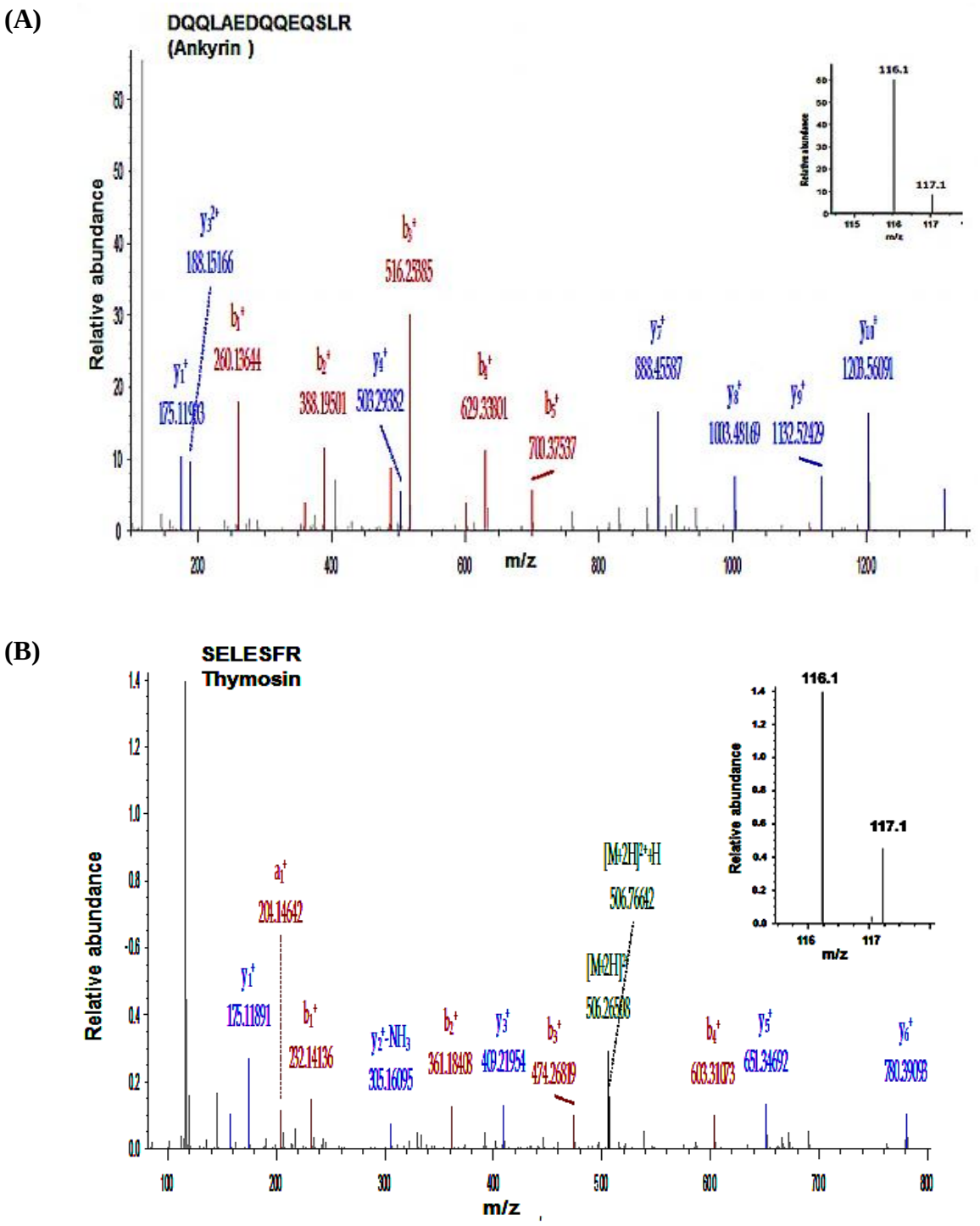
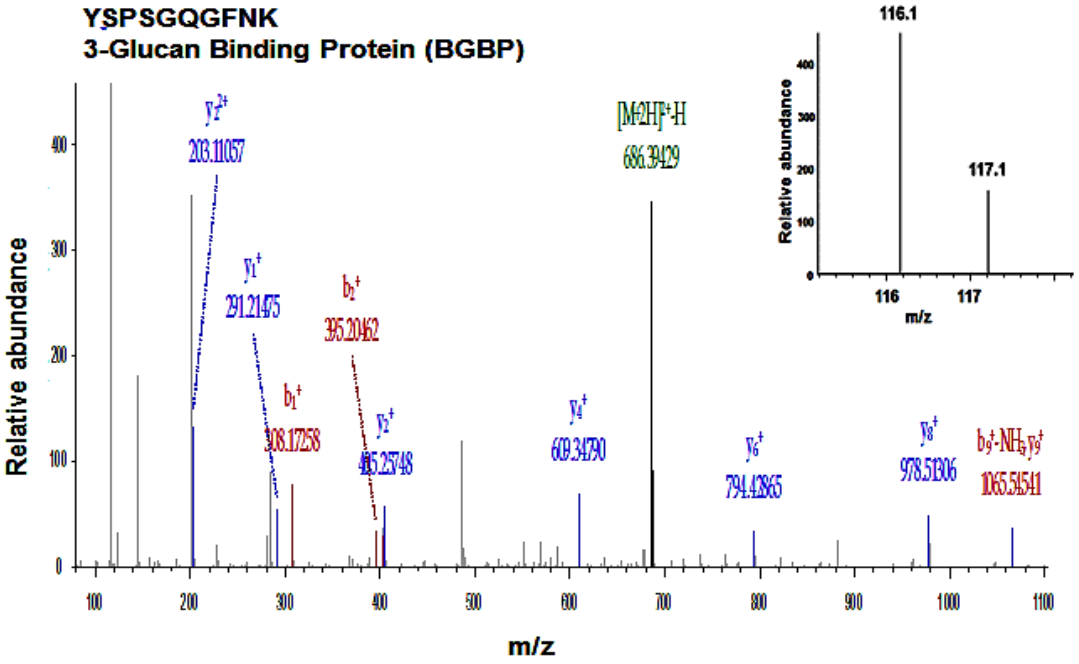


Figure 2.19 A: MS/MS spectra of peptides for the proteins over-expressed in male midgut (A) Peptide sequence **DQQLAEDQQEQSLR** identified against Ankyrin, (B) Peptide sequence **SELESFR** identified against Thymosin.

(C)



(D)

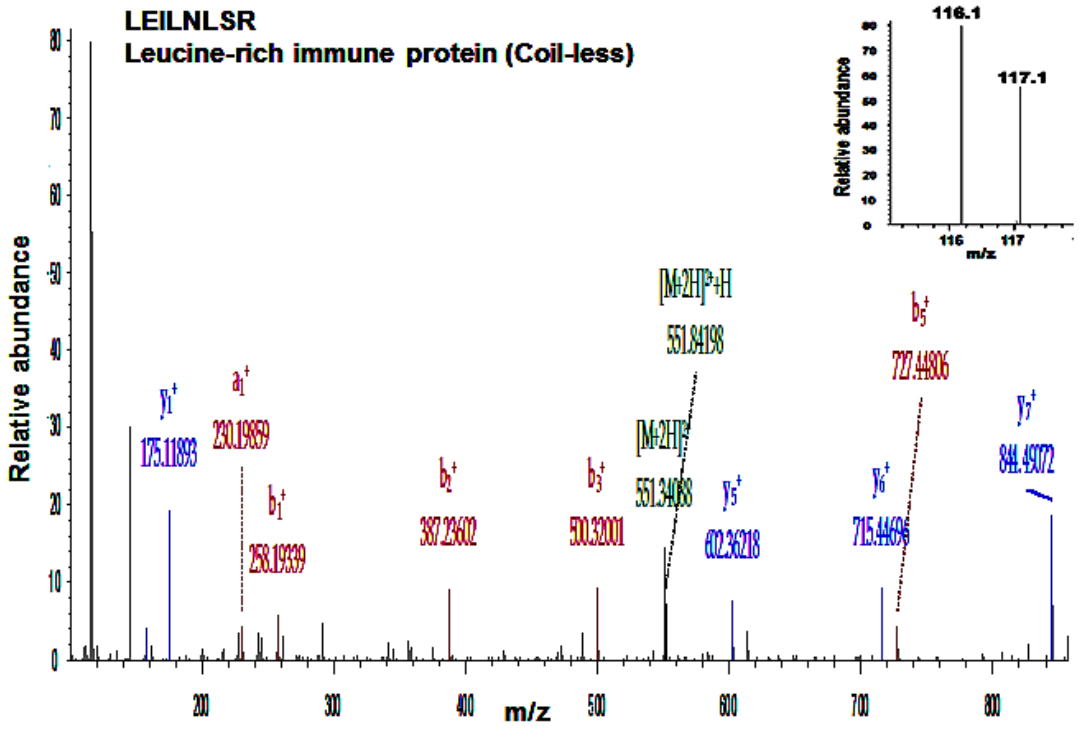


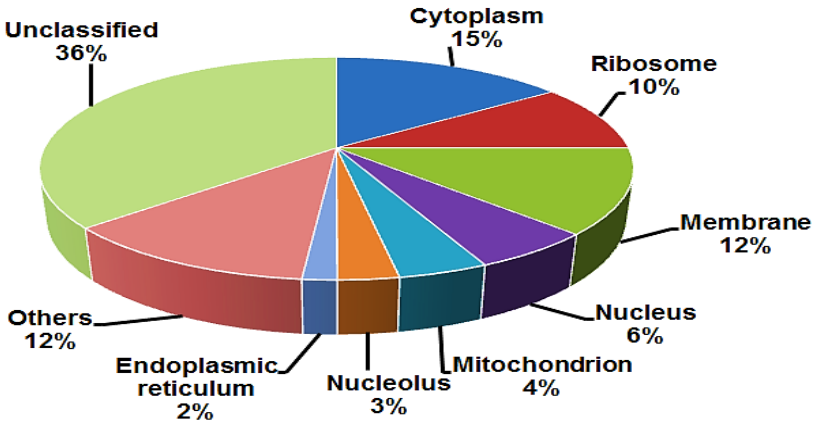
Figure 2.19 B: MS/MS spectra of peptides for the proteins over-expressed in male midgut (C) Peptide sequence **YSPSGQGFNK** identified against 3-Glucan Binding Protein (BGBP); (D) Peptide sequence **LEILNLSR** identified against Leucine-rich immune protein (Coil-less).

2.3.9 Bioinformatic analysis of differentially expressed proteins

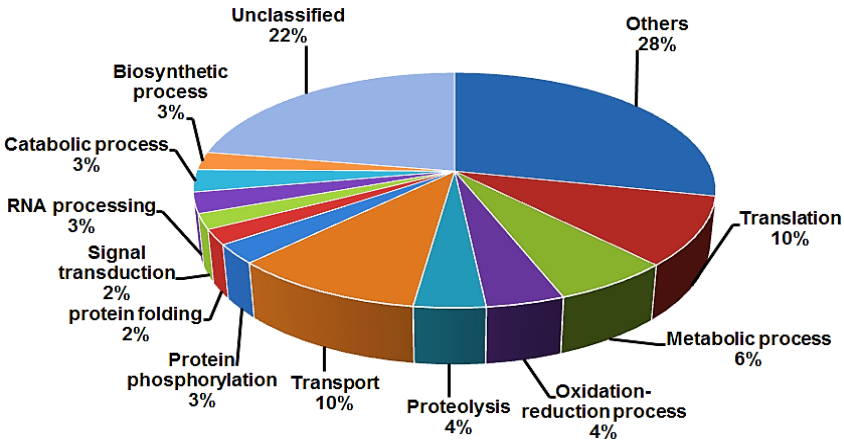
2.3.9.1 Classification and functional annotation of differentially expressed proteins

The information on cellular localization, molecular functions and biological processes for the differentially expressed proteins were retrieved from the Vectorbase database (www.VectorBase.org). These proteins were classified based on Gene Ontology annotations. In order to evaluate which GO term was more pervasive in midgut of female or male *An. stephensi*, the differentially expressed proteins were divided into two groups: 1) Proteins which were over-expressed in female midgut and 2) Proteins which were over-expressed in male midgut tissue.

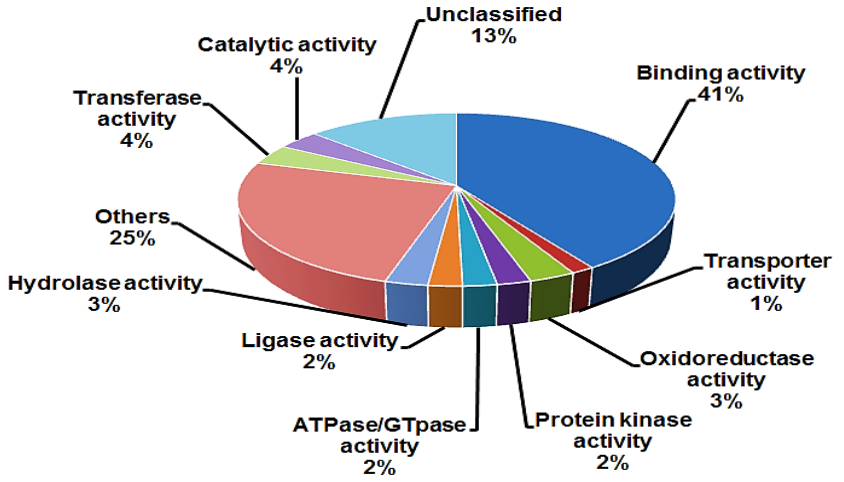
The subcellular localization of proteins over-expressed in female midgut indicated that 15% of the proteins were localized in the cytoplasm. Remaining proteins were predicted to be localized in membrane (12%), ribosome (10%) and nucleus (6%). Of the total, 39% of proteins are still reported as unclassified (Figure 2.20 A). Classification based on biological processes showed that majority of the over-expressed proteins were involved in biological process such as transport (10%), translation (10%), metabolic process (6%), oxidation-reduction process (4%) and RNA processing (3%) (Figure 2.20 B). Classification based on molecular function showed that majority of proteins over-expressed in female midgut were involved in molecular functions such as binding activity (41%), transferase activity (4%), catalytic activity (4%) and hydrolase activity (3%). 13% of identified proteins are still unclassified in terms of their molecular function (Figure 2.20 C).



(A) Subcellular localization



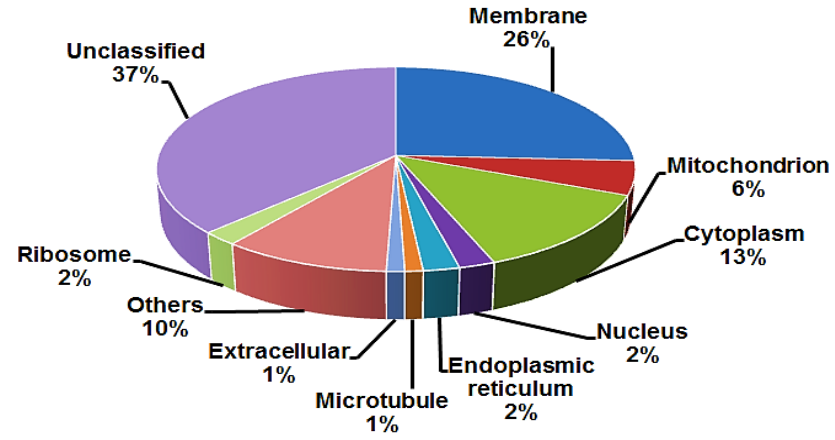
(B) Biological process



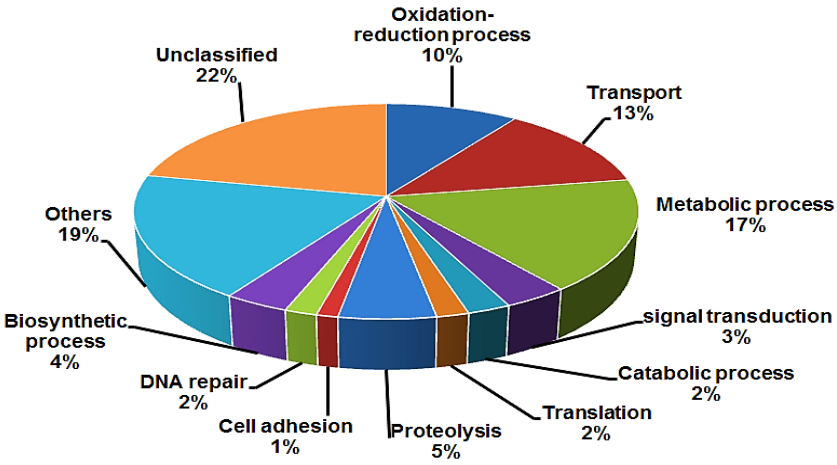
(C) Molecular function

Figure 2.20: Gene ontology based classification of over-expressed proteins identified in female midgut. A) Subcellular localization (B) Biological process (C) Molecular function.

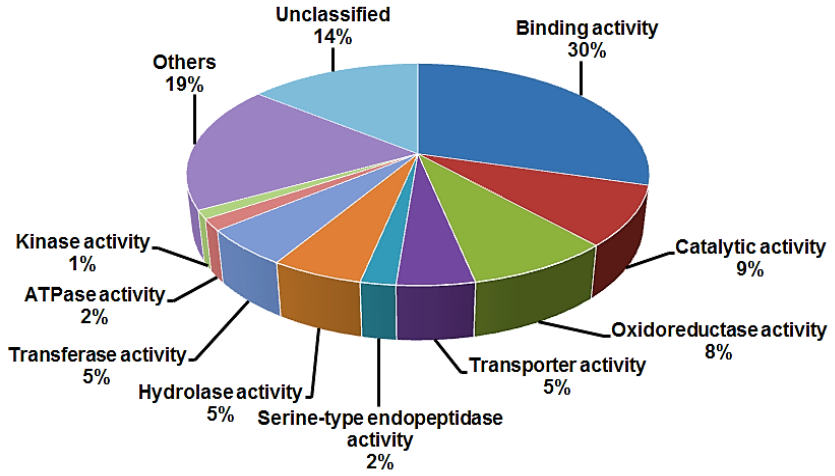
Similarily, the subcellular localization of over-expressed proteins in male midgut indicated that 26% of the proteins were localized to membrane followed by proteins localized to cytoplasm (13%), mitochondria (6%), nucleus (2%) and ribosome (2%) (Figure 2.21 A). Biological process based categorization showed that the majority of proteins were involved in metabolic process (17%), transport (13%), oxidation-reduction process (10%), proteolysis (5%) and biosynthetic process (4%) (Figure 2.21 B). Classification based on molecular function showed that majority of the proteins were predicted to have molecular functions such as protein binding activity (30%), catalytic activity (9%), oxidoreductase activity (8%), transferase activity (5%) and transporter activity (5%). 14% of proteins are still reported as unclassified in terms of molecular function. (Figure 2.21 C).



(A) Subcellular localization



(B) Biological process



(C) Molecular function

Figure 2.21: Gene ontology based classification of over-expressed proteins identified from male midgut. Subcellular localization, biological process and molecular function of differentially expressed proteins. (A) Subcellular localization (B) Biological processes (C) Molecular function.

2.3.10 Sex-biased proteins

Further, the differentially expressed proteins identified were compared with sex-biased genes reported in *An. stephensi* Indian strain and found a strong correlation. Jiang *et al*., (2014) identified 313 and 241 genes with male and female biased expression respectively. Of these, 76 female biased proteins were identified in this study. Among these, 50 proteins were found to be over-expressed >2 fold in female midgut as compare to male, while remaining 25 proteins expression were found to be unchanged between two sexes. The female specific proteins identified includes F-box and leucine-rich repeat protein 6, serine proteases, lipase3, aminopaptidase N, alkaline phosphatase, female reproductive tract protease, Class B Scavenger Receptor (CD36 domain), Venom allergen as shown in Table 2.8. These female specific proteins were mainly involved in proteolysis and metabolic processes relevant to blood meal digestion, vector immunity and some involved in parasite interaction. Similarily, 19 proteins identified in this study were reported to be male biased (Table 2.9). Of these, eight proteins were found to be over-expressed >2fold in male midgut as compared to the female and of the remaining 12 proteins, expression was found to be unchanged between two sexes. The male specific proteins include Regucalcin protein, Actin cytoplasmic, Alpha-crystallin chain B, Trypsin-alpha, Troponin C isoform 4", Leucine amino peptidase-like protein, and two hypothetical proteins. Further investigation of these sex-specific molecules may enhance the current knowledge regarding role of midgut proteins in malaria transmission and vector development.

Table 2.8: A partial list of proteins over-expressed in male midgut tissue

S. No.	Accesssion Id	Protein description	Expression level male/female	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]
1	ASTEI06544-PA	Gustatory receptor	19.453	1	1	791	92.6
2	ASTEI07019-PA	Hypothetical protein	18.422	1	1	175	19.5
3	ASTEI10790-PA	Hypothetical protein	11.219	2	7	184	21.2
4	ASTEI08704-PA	Hypothetical protein	9.645	1	2	735	78.9
5	ASTEI09506-PA	Translocon-associated protein subunit gamma	9.460	4	15	185	21.2
6	ASTEI10959-PA	Sodium-coupled monocarboxylate transporter 1	6.846	1	4	684	74.7
7	ASTEI08483-PA	Hypothetical protein	6.694	1	2	62	6.8
8	ASTEI05983-PA	Potassium voltage-gated channel, subfamily H (eag-related), member 4	6.472	2	4	805	90.7
9	ASTEI06317-PA	Cuticular protein RR-1 family (CPR24)	6.112	2	4	898	98.0
10	ASTEI11712-PA	Innexin inx2	5.829	1	2	116	13.6
11	ASTEI09441-PA	Transcription factor AP-1	5.754	1	2	325	35.2
12	ASTEI04365-PA	Phosphate carrier, mitochondrial	5.238	13	171	348	38.2

Table 2.8 contd: A partial list of proteins over-expressed in male midgut tissue

S. No.	Accesssion Id	Protein description	Expression level male/female	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]
13	ASTEI09681-PA	FK506-binding protein 1	5.102	4	15	108	11.6
14	ASTEI04241-PA	Hypothetical protein	5.024	1	2	153	15.7
15	ASTEI06325-PA	Pyrimidine nucleoside transport protein	5.008	1	2	1205	131.2
16	ASTEI09665-PA	Hypothetical protein	4.811	2	8	204	21.4
17	ASTEI07626-PA	Glutathione S-transferase delta class	4.483	1	6	436	49.6
18	ASTEI05843-PA	Cytochrome P450 (CYP6AF1)	4.472	14	37	493	56.3
19	ASTEI07367-PA	40S ribosomal protein S21	4.223	1	2	506	56.6
20	ASTEI10118-PA	Hypothetical protein	3.980	2	14	394	42.4
21	ASTEI00680-PA	Solute carrier family 23 (nucleobase transporter)	3.973	1	6	524	56.5
22	ASTEI04034-PA	Ankyrin	3.953	2	4	173	19.1
23	ASTEI01741-PA	Fas-associated factor	3.931	2	5	439	50.5
24	ASTEI00353-PA	Dynein heavy chain, axonemal	3.787	1	4	4294	495.2

Table 2.9: List of male specific proteins identified and their expression level in male midgut tissue.

S. No.	*An. stephensi* VectorBase ID	Relative fold change male/female	Protein description as per *An. gambiae*	Molecular function	Biological process
1	ASTEI05843-PA	4.47	Cytochrome P450 (CYP6AF1)	Oxidoreductase activity	Oxidation-reduction process
2	ASTEI01662-PA	3.39	Regucalcin protein	Unknown	Unknown
3	ASTEI03074-PA	2.53	Actin cytoplasmic	Unknown	Unknown
4	ASTEI02736-PA	2.50	Alpha-crystallin chain B	Unknown	Unknown
5	ASTEI07236-PA	2.36	Trypsin-alpha	Serine-type endopeptidase activity	Proteolysis
6	ASTEI09992-PA	2.25	Troponin C isoform 4"	Ion binding	Unknown
7	ASTEI07705-PA	2.16	Leucine aminopeptidase-like protein	Peptidase activity	Proteolysis
8	ASTEI09406-PA	1.24	Hypothetical protein	Serine-type endopeptidase activity	Proteolysis
9	ASTEI08327-PA	1.22	Dynein heavy chain 9, axonemal	Ion binding	Microtubule-based movement
10	ASTEI06940-PA	1.15	Juvenile hormone diol kinase	Ion binding	Unknown
11	ASTEI09889-PA	1.09	Unknown	GTPase activity	Protein polymerization
12	ASTEI07479-PA	0.95	Butyryl-CoA dehydrogenase	Peptidase activity	Proteolysis
13	ASTEI02126-PA	0.90	FMRFamide	Unknown	Neuropeptide signaling pathway
14	ASTEI07186-PA	0.78	Transglutaminase	Transferase activity	Cellular protein modification process
15	ASTEI11276-PA	0.66	Tubulin alpha	GTPase activity	Protein polymerization
16	ASTEI08863-PA	0.65	Dynein, axonemal, heavy chain 5	Ion binding	Microtubule-based movement
17	ASTEI10640-PA	0.63	Hypothetical protein	Unknown	Unknown
18	ASTEI05773-PA	0.46	Hypothetical protein	Unknown	Outer dynein arm assembly
19	ASTEI03866-PA	0.23	Dynein intermediate chain 2, axonemal	Protein binding	Unknown

2.4 Conclusion

The aim of this study was to provide a catalog of the proteins in the midgut of female *An. stephensi.* In this study a high resolution Fourier transform mass spectrometer was used for identifying proteins expressed in the midgut of female *An. stephensi* fed exclusively on sugar solution. This analysis has led to the identification of largest 5,434 proteins in any of the mosquito species so far. This was achieved by the use of multiple fractionation methods and high-accuracy and high resolution mass spectrometer. Mass spectrometry has been considered as an efficient tool for proteomic analysis due to its sensitivity, high mass accuracy and ability to deal with complex sample mixtures. The functional characterisation of hypothetical proteins identified in midgut tissue is a step forward towards understanding the nature of physiological processes and molecular interactions which was lacking in the *An. stephensi*. The cellular components, biological processes and the molecular functions, in which these proteins are involved in, were summarized and discussed. This study demonstrated that the most of the proteins identified are likely to be involved in binding activity, metabolism and transport, and most of these are membrane bound or located intracellularly. Hypothetical proteins identified can be investigated for their role in malaria transmission by instituting gene silencing and expression studies.

In the present study, the difference in protein expression between male and female midgut tissue was also carried out in order to explore the differences in midgut protein expression in relation to sex. For the first time midgut tissue of male and female *An. stephensi* was analyzed using iTRAQ based quantitative proteomic approach. Comparative proteomic analysis of male and female midgut led to the identification of 3,316 proteins and their expression levels between two sexes. The results showed that 758 proteins were quantitatively different between male and female *An. stephensi*. The data generated also provides experimental evidence for several sex specific proteins expressed differentially between male and female mosquitos. Proteomic differences between two sexes also support the broad functional differences between the male and

female midgut. The female midgut specific proteins showed a significantly greater functional diversity than the male midgut-specific proteins, which were mostly involved in housekeeping processes. These findings are on expected lines as the female midgut is site for host blood digestion and parasite development. In comparison to this the male gut has to perform much simpler function of plant sap digestion. Hence overall the functional differences between the proteome of female and male midguts were mainly related to immunity and blood digestion. Prevot *et al*., (2003) also observed similar differences in *An. gambiae* protein expression between unfed female and male midguts by using 2-dimensional gel electrophoresis. The proteomic information generated from this study open up vistas for exploring the candidate genes responsible for the pathogen-transmitting processes in *An. stephensi*. Further functional analysis of these female-specific proteins is necessary for better understanding the differences between male and female *An. stephensi.* These studies are helpful and crucial for designing transmission blocking strategies via selective gene silencing.

Chapter-III

Differential expression of proteins in the midgut of *Plasmodium berghei* infected and uninfected female *Anopheles stephensi* Liston

3.1 Introduction

Midgut of mosquito is the first site of vector-parasite interaction, which plays a central role in the development and subsequent transmission of malaria parasites. Midgut represents one of the most challenging environments for the development and survival of *Plasmodium*. The sporogonic cycle of *Plasmodium* from male and female gametes (derived in turn from male and female forms of gametocytes) to oocyst formation takes place in the lumen and midgut epithelium. It is assumed that these sporogonic developmental steps occur inside the midgut, probably governed by certain mosquito-specific factors. Very little is known about midgut proteins involved in blood meal digestion, vector immunity and parasite-vector interaction. The incidence of malaria could be reduced by controlling transmission of parasite by the mosquitoes. Better understanding of mosquito-*Plasmodium* relationship and genes involved in the transmission of parasites, would eventually help in defining ways to control the transmission of malaria.

During the recent years, efforts have been focused on designing new strategies at vector level, which can be explored to interrupt the transmission of parasites via mosquitoes. One promising approach in this direction is the use of malaria transmission-blocking vaccine that would prevent transition from ookinete to oocyst by targeting mosquito midgut ligands that mediate cell adhesion in the vector. This approach is quite different than the classical transmission blocking vaccines, which target surface molecules on sexual stages of parasite (Carter, 2001). Previous studies showed that mosquito produces xanthurenic acid, which is important for the ex-flagellation of parasite microgametes (Garcia *et al*., 1997, Billker *et al*., 1998). Shahabuddin *et al*., (1994, 1996) found that trypsin produced in the mosquito's digestive tract, probably activates parasite chitinases, which facilitates the passage of the parasite through the peritrophic matrix which surrounds the parasite-containing blood meal in the midgut of mosquito. Studies have also reported that early sporogonic stages of *Plasmodium* modulate the mosquito immune response (Dimopoulos *et al*., 1998, 2001, Tahar *et al*., 2002). These immune molecules

were found to inhibit the development of *Plasmodium* in transgenic mosquitoes (Kokoza *et al*., 2000, Raikhel *et al*., 2002). All these studies were based on analyses of mRNA production, but very few in-depth studies published earlier were concerning proteome analysis of mosquito midgut infected with *Plasmodium berghei*. Using monoclonal antibodies, Lal *et al*., (2001) identified a set of midgut proteins that may be involved in *Plasmodium* development. Ghosh *et al*., (2001) screened a phage display library and selected a peptide that recognized midgut proteins. The production of this peptide in transgenic *Anopheles stephensi* mosquitoes which reduced the development of *Plasmodium berghei* oocysts on the mosquito midgut wall (Ito *et al*., 2002). Dinglasan *et al*., (2007) identified a glycoprotein amino-peptidase N (AgAPN1) in *Anopheles gambiae* midgut luminal surface having role in ookinete invasion. Gonzalez-Lazaro *et al*., (2009) identified a Croquemort SCRBQ2 (scavenger receptor class B croquemort) highly expressed in the *Anopheles gambiae* female midgut that plays an important role in oocyst development. Identification of such proteins involved in parasite development in mosquito midgut would be a most attractive site to restrict the parasite development and their further transmission to humans. Therefore midgut is one of the most attractive sites for antigenic targets, in order to develop vector based transmission-blocking vaccines (TBVs). In the present study an iTRAQ based quantitative proteomic analysis of midgut of female *Anopheles stephensi* after feeding on sugar solution, uninfected blood and *Plasmodium berghei* infected blood was carried out using high resolution Fourier transform mass spectrometry. Several proteins were identified which expressed differentially in the midgut tissue of *Plasmodium berghei* infected mosquito as compared to the blood and sugar fed ones. These proteins would be responsible for blood meal digestion, sporogonic development or inhibition of malaria parasites. This information can be utilized in future for the development of novel targets for control of disease transmission.

3.2 Materials and methods

3.2.1 Source of mosquitoes

Anopheles stephensi female mosquitoes were obtained from the insectary maintained in NIMR Proteogenomic laboratory and used for the experiments. Mosquitoes were divided into three experimental groups and each group contained 500 mosquitoes of the same age.

3.2.2 Mosquito feeding on glucose solution

First experimental group comprised of 500 female *Anopheles stephensi* mosquitoes, which were allowed to feed on glucose solution (10% v/w) soaked in a cotton pad. Midguts were dissected in 0.65% normal saline under dissecting microscope and stored at -80˚C for further use. This group served as control.

3.2.3 Mosquito feeding on uninfected/normal BALB/c mice

In a second group, 500 female mosquitoes were starved for 12 hrs and then allowed to feed on blood of normal uninfected BALB/c mice. Mosquitoes that fed on mice were visually checked for the presence of a blood meal in their abdomen. The adults that did not take a blood meal were removed from the experiment. Thereafter, the fully fed mosquitoes were maintained at 27±1°C, 80 ±10% humidity. Mosquitoes were collected after 48h post feeding and cold anesthetized. Midguts were dissected and stored at -80˚C for further use. Ethical clearance for using BALB/c mice for experiment was obtained from the Animal Ethics Committee of the department of Zoology, Goa University.

3.2.4 Mosquito feeding on BALB/c mice infected with *Plasmodium berghei*

3.2.4.1 Laboratory infection of BALB/c mice with *Plasmodium berghei*

BALB/c mice were injected intra-peritoneally (IP) with 0.5 ml of *Plasmodium berghei* infected blood stock having 10% parasitaemia (*Plasmodium berghei* ANKA strain was obtained from Parasite Bank NIMR, New Delhi). The mice were then maintained under standard conditions (20 ±2°C, 50% relative humidity) and provided with food and water.

Parasitaemia was monitored by making and examining Giemsa-stained blood smears for the presence of blood stage parasites on day 4 and 6 from the date of injection of parasites. Production of gametocytes was checked on day 11 and 12 post injection. Once gametocytes were produced then 12hrs starved female *Anopheles stephensi* mosquitoes were allowed to feed on infected mouse for 1 hr**.** Similarly mosquitoes were visually checked for the presence of a blood meal in their abdomen and unfed mosquitoes were removed from the experiment. The fully fed mosquitoes were maintained in a separate mosquito cage at 27±1°C, 80 ±10% humidity. Mosquitoes were collected after 48h post feeding and dissected in 0.65% normal saline and stored at-80˚C for further use.

3.2.5 Protein extraction from midgut tissues

Midguts obtained from each experimental group were homogenized in 200µl of 1% SDS using ultrasonication. The extracted suspension was centrifuged for 10 min using 14,000 rpm at 4˚C. Supernatant was collected without disturbing the pellet. The extracted proteins were quantified using Bio-Rad DC Protein assay kit. Equal amount of protein was loaded from each condition on SDS-PAGE (Figure 3.1).

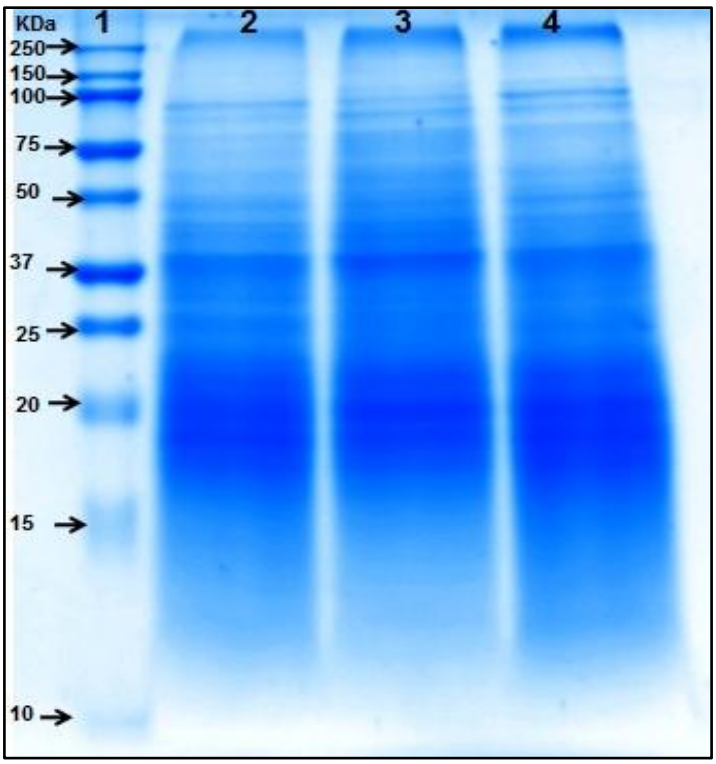


Figure 3.1: SDS-PAGE profile of proteins obtained from midgut tissue of sugar fed, blood and *Plasmodium berghei* infected female *An. stephensi*. Lane1: Protein marker; Lane 2: Protein from sugar fed, Lane 3: Protein from blood fed; Lane 4 Protein from *Plasmodium berghei* infected midgut tissue.

3.2.6 Trypsin digestion, iTRAQ labeling and strong cation exchange fractionation

100µg of protein samples from each condition were separately digested with trypsin and then labeled with iTRAQ reagents for quantitative proteomic analysis. Trypsin digested peptide from each conditions were labeled using iTRAQ reagents containing reporters iTRAQ116, iTRAQ117 and iTRAQ118 for sugar fed, blood fed and *Plasmodium berghei* infected respectively. Labeling of peptides was carried out for 2h at room temperature according to manufacturer's protocol and as previously described by (Puttamallesh *et al*., 2013). After labeling, the peptides from all three samples were pooled and fractionated using strong cation exchange chromatography on Polysulfoethyl A column (PolyLC, Columbia, MD) using an LC Packing HPLC system connected to a Probot fraction collector. SCX fractions were collected at a flow rate of 0.2 ml/min using a 50 min gradient of KCl from 0 to 350 mM concentration in 10 mM potassium phosphate buffer,

25% acetonitrile pH 2.8. SCX profile of iTRAQ labeled peptides is shown in Figure 3.2. The fractions were dried, desalted using C_{18} tips and reconstituted in 10 µl of 2% tri-fluoroacetic acid before mass spectrometric analysis.

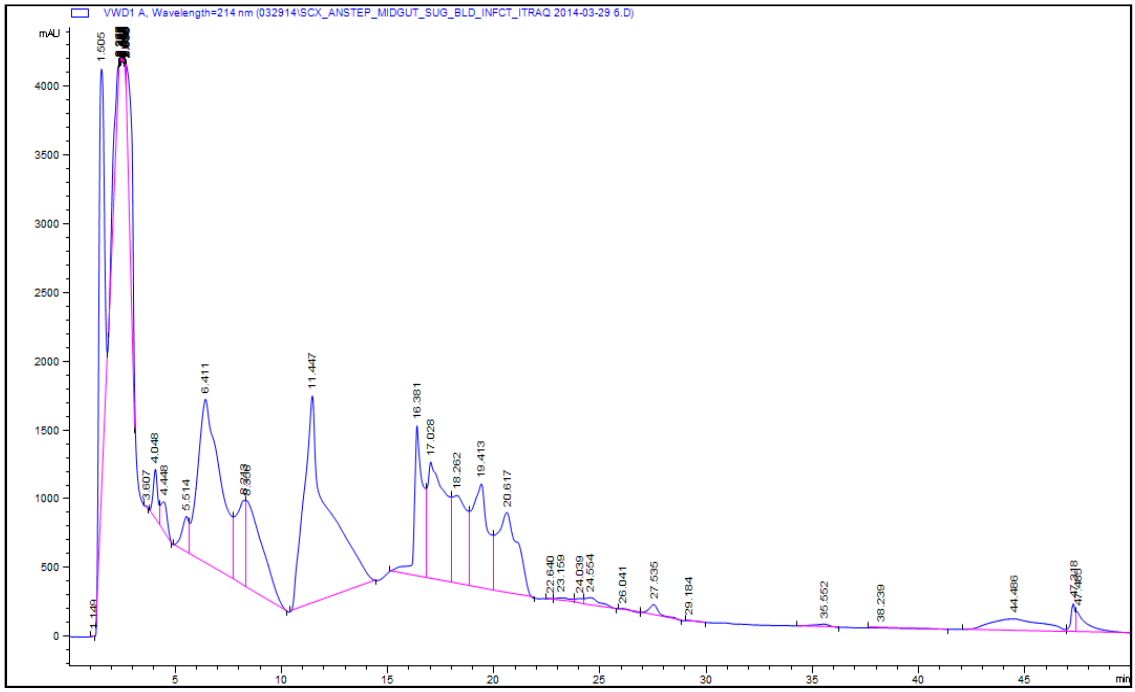


Figure 3.2: SCX chromatogram of iTRAQ labeled peptides obtained from midguts of sugar, blood fed and *P. berghei* infected female *Anopheles stephensi*.

3.2.7 Mass spectrometry analysis

The fractions were analysed on LTQ-Orbitrap Velos mass spectrometer interfaced with Proxeon Easy nLC system (Thermo Scientific) as described in earlier section 2.10.

3.2.8 Data analysis

The data obtained was searched against the combined protein database of *Anopheles stephensi*, *Plasmodium berghei* and Mouse using Proteome Discoverer, version 1.3.0.339 (Thermo Fischer Scientific). The work flow consisted of spectrum selector and reporter ion quantification nodes in addition to Sequest and Mascot search nodes. Search

parameters included trypsin as the enzyme with single missed cleavage allowed. Other parameters included methyl-thiol modification of cysteine, iTRAQ labels at the peptide N-terminus and Lysine residues were set as static modifications and oxidation of methionine as variable modification. A false discovery rate (FDR) of 1% was applied to the results. Relative quantitation of proteins was calculated by using reporter ions quantifier node in Proteome Discoverer software. Different bioinformatics tools were used to find out the domain structure, sequence homology and ortholog information as described in earlier section 2.12. The differentially expressed proteins identified were further analyzed for cellular localization and functional categorization based on biological processes and molecular functions using gene ontology annotations available for *An. stephensi* in VectorBase database.

3.3 Results and discussion

In this study, a quantitative proteomic analysis approach was applied to identify differentially expressed proteins in the midgut of female *An. stephensi* after 48hrs of feeding on sugar solution, uninfected/normal blood and *Plasmodium berghei* infected blood. Midguts were dissected from these three mosquito groups, followed by protein extraction. The proteins extracted from the three conditions were individually labeled with iTRAQ reagents and SCX fractionation followed by analysis on high resolution Fourier transform mass spectrometry (LTQ-Orbitrap Velos). The overall workflow for the study is shown in Figure 3.3. From this, a total of 74,042 tandem mass spectra were acquired. These spectra were then searched against a combined protein database of *An. stephensi*, *Plasmodium berghei* and mouse using Mascot and Sequest search engines through the Proteome Discoverer software. A total of 55,328 peptide-spectrum matches (PSMs) were obtained which lead to identification of 11,601 peptides and 2,327 proteins. Of these 2,327 proteins, identification of 1,500 proteins was based on 2 or more peptides, while 827 protein identifications were based on single unique peptide evidence with multiple PSMs.

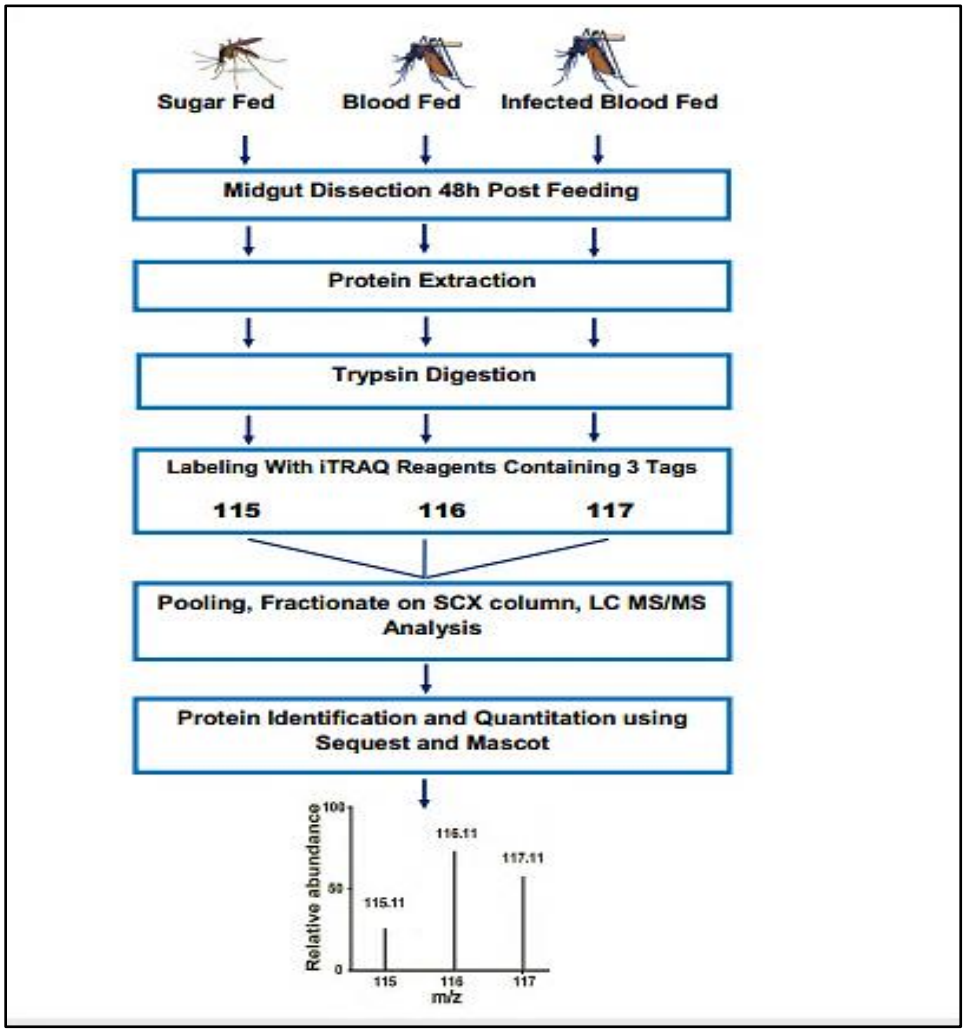


Figure 3.3: Workflow illustrating steps involved in the quantitative proteomic strategy. Midgut tissues from sugar fed, normal blood and *Plasmodium berghei* infected blood fed mosquitoes were used to extract proteins and subjected to digestion with trypsin. Peptides from the three groups were labeled with iTRAQ reagents followed by LC-MS/MS.

3.3.1 Differential expression of proteins between blood and sugar fed mosquitoes

Relative quantification of peptides was done on the basis of relative intensity of reporter ions. The identified protein ratios were calculated as the median of all the peptide ratios corresponding to respective proteins. A fold change of ≥1.5 was considered as over-expressed while ≤0.66 was considered as down-regulated (Zhang *et al*., 2014; Wang *et*

al., 2015). A total of 435 proteins were found to be expressed differentially between blood fed and sugar fed midguts. Of these 435 proteins, 247 proteins were over-expressed by ≥1.5 fold while 188 proteins were down-regulated by ≤0.66 fold in blood fed as compared to sugar fed midgut tissue (Figure 3.4). Of these differentially expressed proteins, 227 were identified with multiple peptides and 208 protein identifications were based on single peptide evidence with multiple PSMs.

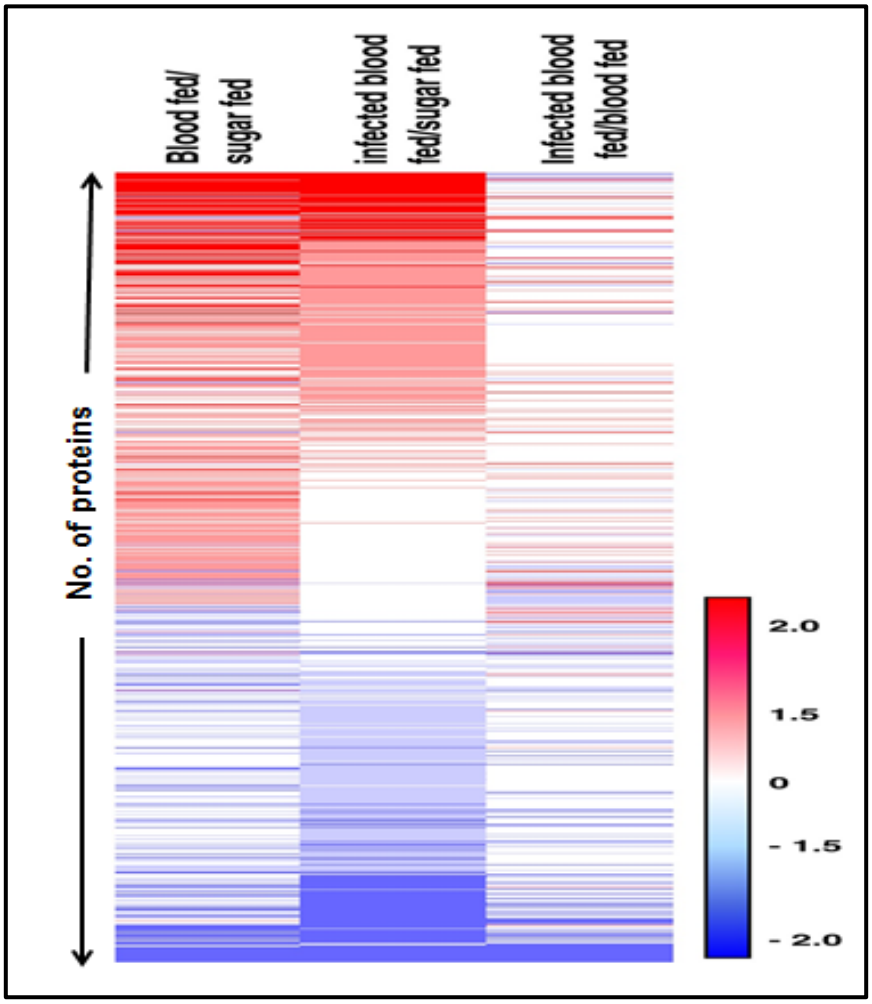


Figure 3.4: Heat map representation of all the proteins identified in the midgut of *An. stephensi* in response to sugar, blood and *P. berghei* infected blood feeding. Protein expression intensities are represented by red, blue and white colour. Red colour indicates proteins over expressed by ≥ 1.5 fold, blue colour indicates proteins down-regulated by ≤ 0.66 fold and white colour indicates no change in fold value.

The proteins which were found to be over-expressed in the blood fed mosquitoes compared to sugar fed included Cathepsin B precursor, protein G12, Bardet-Biedl syndrome 1 protein, vitellogenin, lupus La protein, sterol carrier protein 2 variant 2, Clathrin light chain, ferritin heavy chain, cathepsin D, glutathione S-transferase theta class, calmodulin, innexin 2, ubiquitin-protein ligase E3 C, TOLL pathway signalling NF-kappaB Relish-like transcription factor, carboxylesterase beta esterase, calreticulin, Innexin inx2, catalase, chitinase and serine protease inhibitor (serpin) SRPN10. These proteins were found to be over-expressed by >1.5 fold in blood fed midgut as compare to the sugar fed. The partial list of proteins over-expressed in blood fed midgut as compared to sugar fed is provided in Table 3.1 and remaining proteins are provided in Appendix Table 3 A.

Protein Cathepsin B precursor was found to be over-expressed by>11.6 fold in blood fed midgut as compared to sugar fed. This protein is found in all insects and arthropods and expressed in the fat body of female mosquito during vitellogenesis in response to blood feeding. It is also involved in degradation of yolk proteins during embryonic development in *Aedes aegypti* (Cho *et al*., 1999). Another protein G12 was found over-expressed by >5.76 fold in blood fed midgut. Nolan *et al*., (2011) identified this protein exclusively expressed in female *An. gambiae* midgut after 24 hours of post-blood feeding. The function of this protein is still un-known in mosquitoes, however it was found similar to a number of cockroach allergens. Vitellogenin, which is a major yolk protein was found to be over-expressed by >5.3 fold. This protein acts as nutrient transporter to maturing oocyst in mosquito midgut and also controls the parasite-killing efficiency of the anti-parasitic factor TEP1 (Rono *et al*., 2010). The remaining over-expressed proteins were found to be involved in vector development process, survival, blood digestion or in synthesis of the peritrophic matrix that surround the blood meal, transport and storage of iron within cells and vector immunity. Some of these proteins are also involved in defense against microorganisms ingested along with the blood meal (Crabb, 1998; Monroy *et al*., 2007; Rono *et al*., 2010; Nolan *et al*., 2011). MS/MS and

reporter ion spectra of some of the representative over-expressed proteins are shown in Figure 3.5**.**

A total of 188 proteins were found to be down-regulated by <0.66 fold in blood fed midgut as compared to the sugar fed mosquitoes. Partial list of proteins identified as down-regulated in blood fed midgut as compared to sugar fed are listed in Table 3.2 and remaining proteins are provided in Appendix Table 3 B. The proteins which were found to be down-regulated in blood fed midgut includes solute carrier family 23 (nucleobase transporter), Serine/threonine-protein phosphatase 4 regulatory subunit 2, Cuticular protein RR-1 family (CPR15), Transcription factor BTF3 homolog 4, C-Type Lectin (CTL)-galactose binding, Serine protease SP24D, Odorant binding protein (OBP9) and Serine-type enodpeptidase.

Cuticular protein (CPR15) was found down-regulated by <0.28 fold in blood fed midgut as compared to the sugar fed. This protein belongs to CPR family and highly abundant in insect species. This protein is mainly involved in cuticle synthesis in mosquito species (Marinotti *et al*., 2006; Vannini *et al*., 2014). Another protein C-Type Lectin (CTL) galactose was found to be down-regulated by <0.35 fold in blood fed. This protein has a carbohydrate-binding protein domain known as a lectin. This protein plays crucial role not only in innate immunity but also in promoting disease transmission in mammals (Cheng *et al*., 2014). The remaining down-regulated proteins were found to be involved in transporter activity, protein translation and metabolism. MS/MS and reporter ion spectra of some of the representative down-regulated proteins are shown in Figure 3.6.

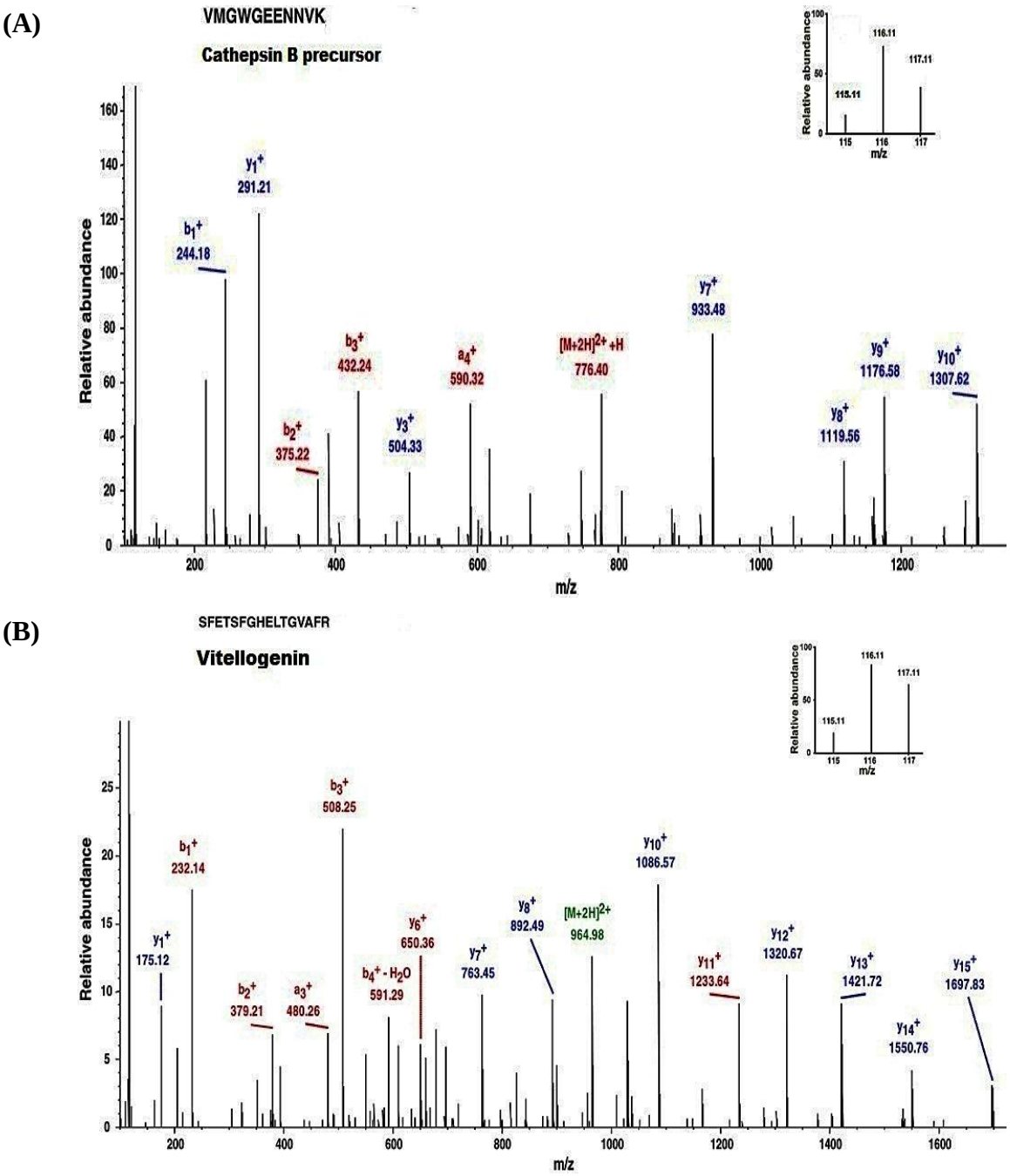


Figure 3.5: Representative MS/MS spectra of over-expressed proteins identified in blood fed midgut as compared to sugar fed midgut tissue. The reporter ions used for quantitation are also shown inside**.** (A) Peptide VMGWGEENNVK identified against cathepsin B precursor protein over-expressed by >11.6 fold in blood fed midgut as compared to sugar fed (B) Peptide SFETSFGHELTGVAFR identified against vitellogenin protein over-expressed by >5.3 fold in blood fed midgut as compared to sugar fed.

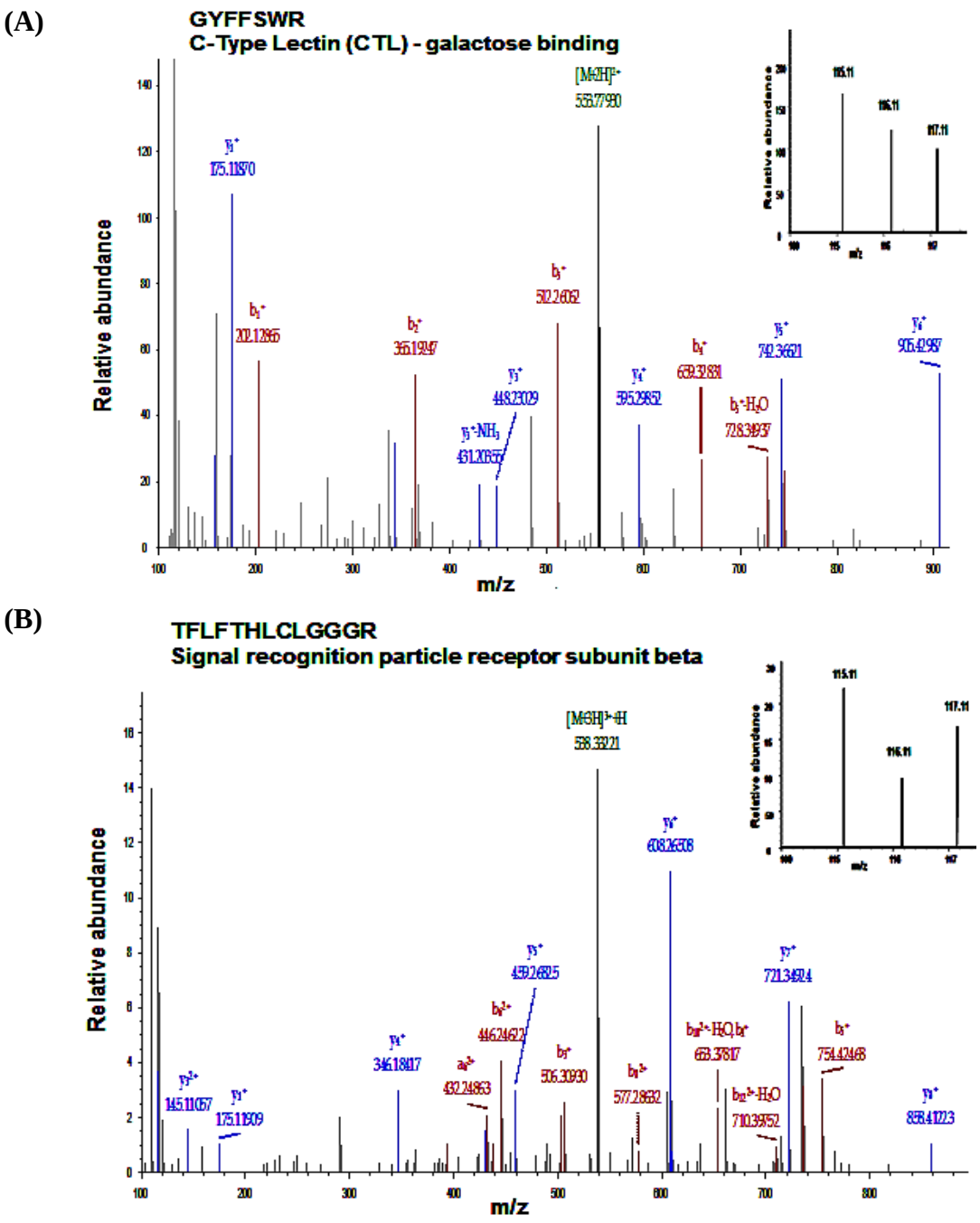


Figure 3.6: Representative MS/MS spectra of down-regulated proteins identified in blood fed midgut as compared to sugar fed midgut tissue. The reporter ions used for quantitation are also shown inside**.** (A) Peptide GYFFSWR identified against protein C-Type Lectin (CTL) galactose binding down-regulated by <0.354 in blood fed midgut as compared to sugar fed (B) Peptide TFLFTHLCLGGGR identified against protein signal recognition particle receptor subunit betadown-regulated by <0.350 in blood fed midgut as compared to sugar fed.

Table 3.1: A partial lists of proteins identified to be over-expressed by ≥ 1.5in blood fed midgut tissue as compared to sugar fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]	Relative fold change Blood/Sugar fed	Protein description as per *An. gambiae*
1	ASTEI04832-PA	1	2	376	41.8	11.653	Cathepsin B precursor
2	ASTEI02220-PA	1	2	277	28.8	7.042	Hypothetical protein
3	ASTEI07890-PA	1	2	204	22.8	5.767	Protein G12
4	ASTEI01388-PA	4	48	111	12.5	5.703	Bardet-Biedl syndrome 1 protein
5	ASTEI11169-PA	36	177	1969	229.9	5.301	Vitellogenin
6	ASTEI00275-PA	1	2	111	12.4	4.808	Sterol carrier protein 2 variant 2
7	ASTEI02154-PA	1	2	197	22.6	4.680	Alpha-crystallin B chain
8	ASTEI07261-PA	1	2	225	25.1	4.277	Charged multivesicular body protein 5
9	ASTEI01945-PA	1	2	405	46.2	4.205	Lupus La protein
10	ASTEI10313-PA	5	32	210	23.8	3.921	Cellular retinaldehyde-binding protein
11	ASTEI11274-PA	4	18	104	11.7	3.731	Probable fatty acid-binding protein
12	ASTEI03362-PA	1	4	90	9.5	3.579	Hypothetical protein
13	ASTEI09210-PA	2	4	829	92.5	3.451	Antizyme inhibitor 1
14	ASTEI08889-PA	8	58	267	29.8	3.318	Hypothetical protein
15	ASTEI11239-PA	7	36	336	36.5	3.282	Fructose-1,6-bisphosphatase I
16	ASTEI08949-PA	1	2	359	40.4	3.277	Hypothetical protein
17	ASTEI11437-PA	1	2	62	7.8	3.205	Clathrin light chain
18	ASTEI01399-PA	2	6	316	35.8	3.133	Methylenetetrahydrofolatereductase (NADPH)
19	ASTEI03112-PA	1	2	403	46.1	3.099	Hypothetical protein
20	ASTEI09982-PA	1	2	768	86.0	2.856	Ubiquitin carboxyl-terminal hydrolase 14
21	ASTEI09051-PA	1	2	84	9.4	2.779	Hypothetical protein
22	ASTEI01563-PA	1	2	239	26.5	2.768	N-acetylneuraminatelyase
23	ASTEI01419-PA	5	13	481	54.3	2.758	Vitellogeniccarboxypeptidase-like protein
24	ASTEI09914-PA	1	2	128	14.1	2.727	Hypothetical protein
25	ASTEI04974-PA	1	2	734	80.1	2.690	Hypothetical protein
26	ASTEI03737-PA	3	8	141	16.4	2.629	Bis(5'-nucleosidyl)-tetraphosphatase

Table 3.2: A partial list of proteins identified to be down-regulated by≤ 0.66 fold in blood fed midgut tissue as compared to sugar fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]	Relative fold change Blood/Sugar fed	Protein description as per *An. gambiae*
1	ASTEI10264-PA	1	8	272	28.7	0.441	Serine protease SP24D
2	ASTEI05053-PA	1	2	95	10.3	0.431	U6 snRNA-associated Sm-like protein LSm8
3	ASTEI04681-PA	1	2	165	19.0	0.422	Adrenodoxin
4	ASTEI05390-PA	2	13	174	19.0	0.417	Succinate dehydrogenase (ubiquinone) cytochrome b560 subunit
5	ASTEI05247-PA	3	8	419	45.5	0.417	Hypothetical protein
6	ASTEI02421-PA	5	17	190	22.8	0.416	60S ribosomal protein L18a (partial)
7	ASTEI07803-PA	1	4	139	15.6	0.414	Odorant binding protein (OBP9)
8	ASTEI10789-PA	1	2	429	47.5	0.410	Karyopherin (Importin) alpha 6
9	ASTEI04041-PA	1	2	251	28.2	0.409	Nat13 protein
10	ASTEI05953-PA	1	1	1418	154.7	0.404	Hypothetical protein
11	ASTEI09202-PA	1	2	316	33.8	0.403	Hypothetical protein
12	ASTEI08252-PA	1	4	224	25.4	0.402	Hypothetical protein
13	ASTEI00772-PA	1	1	1449	160.7	0.395	Bromodomain-containing protein 8
14	ASTEI08686-PA	3	8	977	107.1	0.368	Alkyldi-hydroxyacetonephosphate synthase
15	ASTEI10176-PA	1	1	644	72.0	0.361	Kinase
16	ASTEI10856-PA	1	2	660	76.0	0.359	Alpha-1,6-mannosyltransferase
17	ASTEI02014-PA	2	8	135	15.4	0.359	Glutaredoxin
18	ASTEI10273-PA	1	2	218	24.9	0.354	C-Type Lectin (CTL)galactose binding
19	ASTEI02284-PA	1	2	185	20.8	0.350	Signal recognition particle receptor subunit beta
20	ASTEI02679-PA	2	8	165	17.7	0.347	Transcription factor BTF3 homolog 4
21	ASTEI07952-PA	2	2	729	76.9	0.347	Serine/threonine-protein phosphatase 4 regulatory subunit 2
22	ASTEI07150-PA	1	2	122	14.0	0.333	Hypothetical protein
23	ASTEI02055-PA	2	12	164	17.9	0.309	60S ribosomal protein L28
24	ASTEI06012-PA	1	1	719	76.9	0.306	Hypothetical protein

3.3.2 Differential expression of proteins between *Plasmodium berghei* infected and uninfected blood fed midgut tissue of *An. stephensi*

3.3.2.1 Over-expressed proteins

A total of 231 proteins identified were found to be expressed differentially between *Plasmodium berghei* infected and blood fed midgut. Of these 231proteins, 55 proteins were found to be over-expressed by ≥1.5 fold while 176 proteins were down-regulated by ≤0.66 fold in *Plasmodium berghei* infected as compared to midgut tissue with blood alone (Figure 3.4). The over-expressed proteins identified included tRNA acetyltransferase TAN1, alpha-1,6-mannosyltransferase, Galacto-kinase, Short caspase (CASPS5), oxidation resistance protein, annexin, inhibitor of apoptosis (IAP1), WAS protein family, calmodulin and phagocyte signaling-impaired protein and serine/threonine kinase 3. A partial list of proteins over-expressed in *Plasmodium berghei* infected midgut as compared to blood fed is provided in Table 3.3 and a complete list of proteins can be seen in Appendix Table 3 C**.** Protein short caspase was found to be over-expressed by >2.0 times in *P. berghei* midgut. This protein belongs to caspase family and involved in apoptosis. Caspases are cysteine protease which generally cleave peptides at Asp residues and trigger apoptosis. Knock down of caspase gene in *An. gambiae* showed increased oocyst count upon infection (Ramphul *et al*., 2015). Hence this protein is immunogenic in nature and confers protection against *Plasmodium* infection in vector host. Another protein Annexin B10B was found over-expressed by 1.75 times. This is a calcium dependent phospholipid-binding protein involved in vesicle transport in polarized cells. This protein was reported to be over-expressed in *Aedes aegypti* mosquito in response to chikungunya virus infection (Tchankouo-Nguetcheu *et al*., 2010). MS/MS and reporter ion spectra of some of the representative over-expressed proteins are shown in Figure 3.7.

3.3.2.2 Down-regulated proteins

A total of 176 proteins were found to be down-regulated by ≤ 0.66 fold in *Plasmodium berghei* infected midgut as compared to the blood fed mosquitoes. The down-regulated proteins included carbonic anhydrase, cathepsin B precursor, aquaporin (AQP1), vitellogenin, gelsolin, etc. The partial list of proteins was down-regulated in *Plasmodium berghei* infected midgut as compared to blood fed is provided in Table 3.4 and complete protein list is provided in Appendix Table 3 D.

Protein carbonic anhydrase was found to be down-regulated by less than 0.35 fold in *Plasmodium berghei* infected midgut as compared to blood fed. This protein was expressed both in male and female midgut and has been reported to be involved in the maintenance of pH inside mosquito midgut (Corena *et al*., 2005). Another protein Cathepsin B precursor was found to be down-regulated by less than 0.40 fold in *Plasmodium berghei* infected midgut. The expression level of this protein in blood fed midgut was found to be > 11.05 fold as compared to sugar fed. Interestingly, when its expression level was compared between *Plasmodium berghei* infected and uninfected blood fed midgut, its expression was found to be down-regulated by <0.40 fold in *Plasmodium berghei* infected midgut. This protein is known to be involved in degradation of yolk proteins during embryonic development in *Aedes aegypti* (Cho *et al*., 1999). The MS/MS and reporter ion spectra of some of the representative down-regulated proteins are shown in Figure 3.8. Among the differentially expressed proteins, there were 15 over-expressed and 42 down-regulated proteins identified as hypothetical. The orthologs for these hypothetical proteins were missing in closely related mosquito species. This suggests that the role of these hypothetical proteins have not yet been validated at functional level in any of the mosquito species. It is possible that, these proteins played a major role in vector immunity or parasite development and survival, invasion, etc. in *An. stephensi* which needs confirmation.

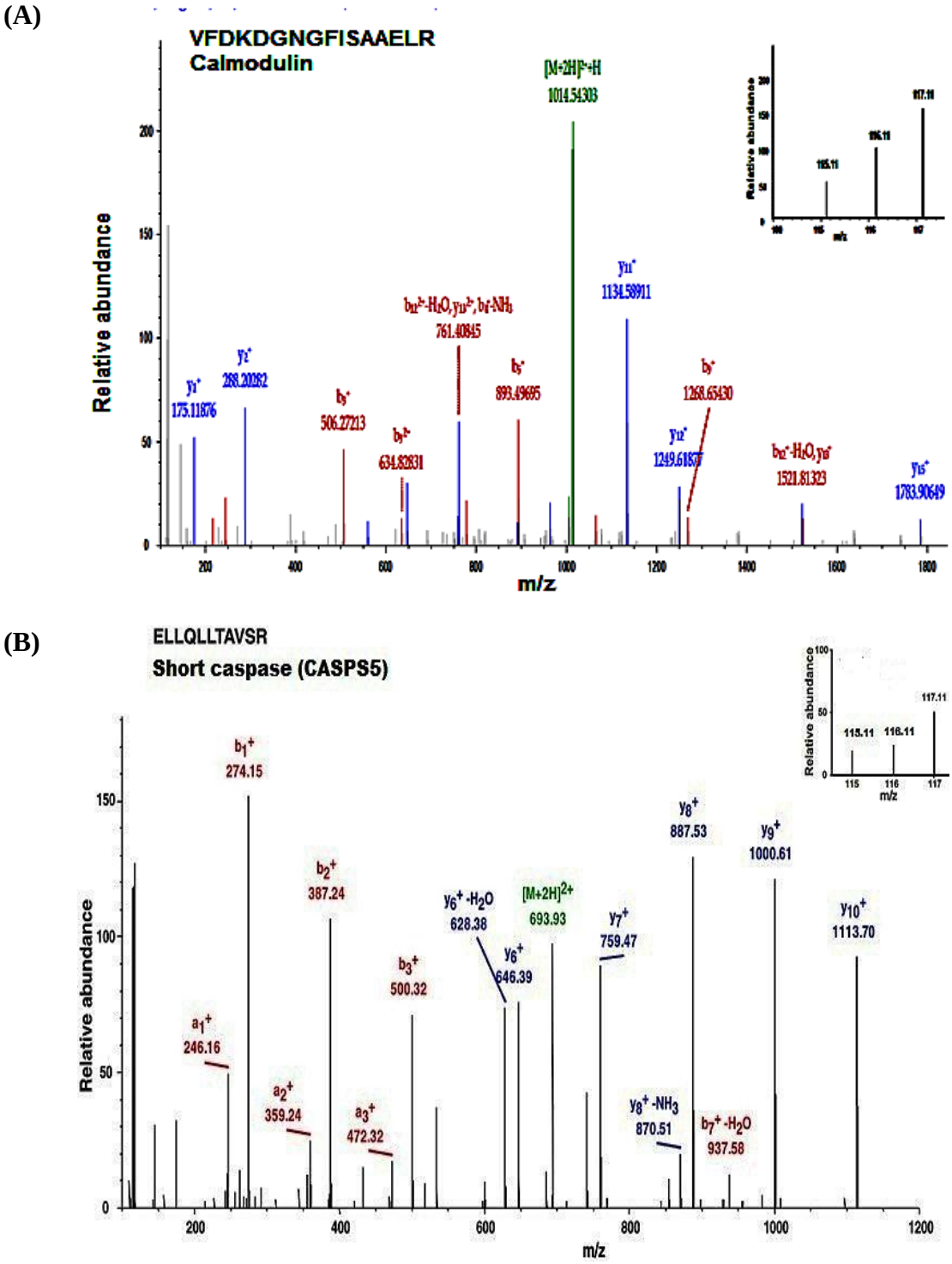


Figure 3.7: Representative MS/MS spectra of over-expressed proteins identified in *P. berghei* infected midgut tissue as compared to blood fed midgut. The reporter ions used for quantitation are also shown inside**.** (A) Peptide VFDKDGNGFISAAELR identified against Calmodulin found to be over-expressed by >1.53 times in *P. berghei* infected midgut tissue as compared to blood fed midgut (B) Peptide ELLQLLTAVSR identified against short caspase found to be over-expressed by> 2.05 times in *P. berghei* infected midgut tissue as compared to blood fed midgut.

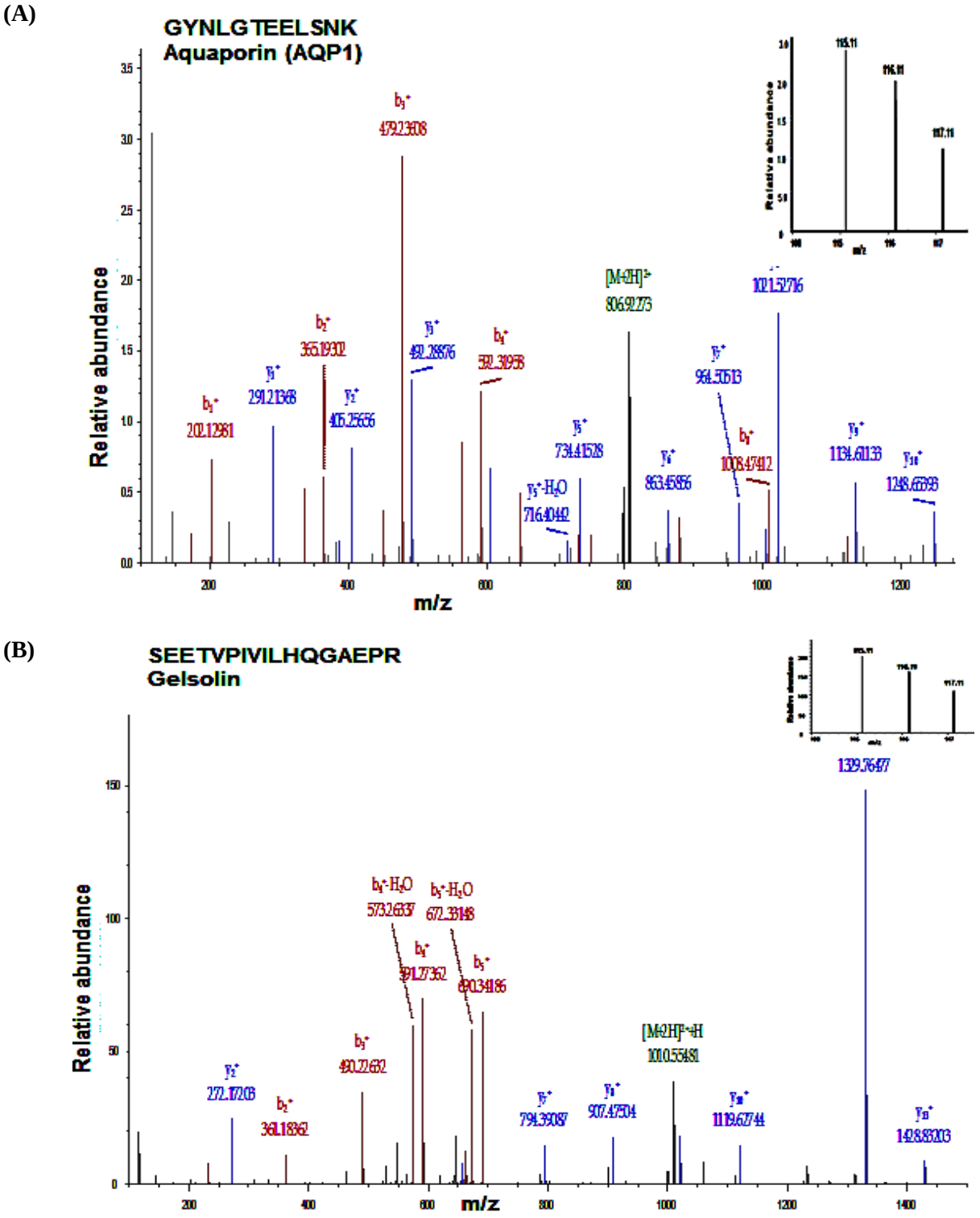


Figure 3.8: Representative MS/MS spectra of down- regulated proteins identified in *P. berghei* infected midgut tissue as compare to blood fed midgut tissue. The reporter ions used for quantitation are also shown inside**.** (A) Peptide GYNLGTEELSNK identified against protein Aquaporin found to be down-regulated by<0.482 times (C) Peptide SEETVPIVILHQGAEPR identified against protein Gelsolin found to be down-regulated by< 0.483 times in *P .berghei* infected midgut tissue as compare to blood fed midgut.

Table 3.3: A partial lists of proteins identified to be over-expressed in *P. berghei* infected midgut tissue as compared to blood fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]	Relative fold change *P.berghei* infected/blood fed	Protein description as per *An. gambiae*
1	ASTEI05273-PA	1	2	648	70.3	3.347	Hypothetical protein
2	ASTEI01559-PA	1	2	294	32.4	3.057	Trnaacetyltransferase TAN1
3	ASTEI01940-PA	1	2	408	43.7	3.004	Galactokinase
4	ASTEI10856-PA	1	2	660	76.0	2.536	Alpha-1,6-mannosyltransferase
5	ASTEI09546-PA	1	1	206	22.7	2.465	Osi15
6	ASTEI01953-PA	2	4	2389	272.3	2.425	Hypothetical protein
7	ASTEI04704-PA	1	1	1001	112.3	2.216	Hypothetical protein
8	ASTEI09946-PA	1	2	471	53.8	2.195	P37NB protein
9	ASTEI06887-PA	1	4	627	73.1	2.109	Pescadillo homolog
10	ASTEI02284-PA	1	2	185	20.8	2.064	Signal recognition particle receptor subunit beta
11	ASTEI09037-PA	2	8	292	32.5	2.053	Short caspase (CASPS5)
12	ASTEI00038-PA	1	1	1189	126.6	2.017	Oxidation resistance protein
13	ASTEI09896-PA	3	10	910	107.3	1.991	Hypothetical protein
14	ASTEI00680-PA	3	6	524	56.5	1.959	Solute carrier family 23 (nucleobase transporter)
15	ASTEI03839-PA	1	1	629	67.9	1.902	Lipoma-prefererred partner
16	ASTEI01616-PA	1	2	287	33.4	1.883	mRNA turnover protein 4
17	ASTEI04531-PA	2	3	191	21.6	1.878	Ubiquinone biosynthesis monooxygenase Coq7
18	ASTEI01928-PA	2	4	2265	261.4	1.868	Hypothetical protein
19	ASTEI07849-PA	1	2	669	77.0	1.847	Hypothetical protein
20	ASTEI04731-PA	1	2	231	26.1	1.810	28S ribosomal protein S7, mitochondrial precursor
21	ASTEI01168-PA	1	1	266	29.0	1.808	Hypothetical protein
22	ASTEI04784-PA	1	2	323	36.6	1.776	Translin-associated factor X
23	ASTEI10355-PA	1	2	925	101.0	1.773	Calcium-independent phospholipase A2
24	ASTEI08657-PA	3	8	620	69.7	1.753	Annexin B10B
25	ASTEI04988-PA	1	2	264	29.1	1.728	Hypothetical protein

Table 3.4: A partial lists of proteins identified to be down-regulated in *P. berghei* infected midgut tissue as compared to blood fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Σ# Peptides	Σ# PSMs	# AAs	MW [kDa]	Relative fold change *P. berghei* infected/blood fed	Protein description as per *An. gambiae*
1	ASTEI02785-PA	1	2	462	50.0	0.489	Solute carrier family 30 (zinc transporter), member 1
2	ASTEI05761-PA	5	20	196	20.6	0.489	Peroxiredoxin 5, atypical 2-Cys peroxiredoxin
3	ASTEI04519-PA	3	12	249	27.5	0.488	Hypothetical protein
4	ASTEI00616-PA	1	2	718	81.7	0.487	Conserved oligomeric Golgi complex subunit 2
5	ASTEI11169-PA	36	177	1969	229.9	0.483	Vitellogenin
6	ASTEI05011-PA	2	6	259	27.5	0.482	Aquaporin (AQP1)
7	ASTEI09989-PA	1	8	208	21.8	0.481	V-type H+-transporting atpase 21kda proteolipid subunit
8	ASTEI07742-PA	20	841	376	41.8	0.479	Actin
9	ASTEI10330-PA	1	2	1120	128.1	0.477	Ubiquitin-protein ligase E3 C
10	ASTEI00846-PA	1	4	136	15.5	0.477	EF-hand domain (C-terminal) containing 2
11	ASTEI09051-PA	1	2	84	9.4	0.476	Hypothetical protein
12	ASTEI03737-PA	3	8	141	16.4	0.473	Bis(5'-nucleosidyl)-tetraphosphatase
13	ASTEI06261-PA	1	2	318	34.5	0.465	Squid
14	ASTEI02454-PA	4	9	844	92.3	0.463	Ubiquitin carboxyl-terminal hydrolase 5/13
15	ASTEI10780-PA	1	7	158	17.9	0.454	Glutathione transferase microsomal
16	ASTEI06336-PA	1	2	606	68.0	0.446	Hypothetical protein
17	ASTEI09101-PA	4	24	160	16.9	0.445	F-type H+-transporting atpase subunit delta
18	ASTEI01004-PA	5	11	253	27.2	0.440	Membrane-associated progesterone receptor component 2
19	ASTEI09857-PA	1	8	154	16.3	0.440	Mitochondrial glutamate carrier 1
20	ASTEI03534-PA	1	1	155	18.4	0.434	Myofilin variant B
21	ASTEI11396-PA	2	6	477	53.1	0.428	Lysosomal Pro-X carboxypeptidase
22	ASTEI04235-PA	1	2	512	54.8	0.414	Acetyl-coa C-acetyltransferase

3.3.3 Differential expression of proteins in *Plasmodium berghei* infected and sugar fed midgut tissue of *An. stephensi*

When protein expression level was compared between *Plasmodium berghei* infected and sugar fed mosquitoes, 494 proteins were found to be expressed differentially. Of these, 204 proteins were over-expressed by ≥1.5 fold and 290 proteins were down-regulated by ≤ 0.66 or lower levels in *Plasmodium berghei* infected midgut. Out of the 204 over-expressed proteins, 54 proteins were found specifically over-expressed in response to *Plasmodium berghei* infected midgut as compared to sugar fed midgut. The remaining proteins were also found over-expressed in blood fed when compared with sugar fed gut. The prominent *Plasmodium berghei* infected midgut specific proteins identified were Symplekin, Selenide, IMD pathway signalling Fas-Associated Death Domain (FADD), V-type proton ATPase subunit F, ATP-dependent RNA helicase DDX1, Solute carrier family 16 (monocarboxylic acid transporters), member 10, Lysosomal alpha-mannosidase, etc. The list of these proteins is provided in Table 3.5 and a complete list can be found in Appendix Table 3 E. The proteins which were down-regulated in *Plasmodium berghei* infected midgut as compared to sugar fed include C-Type Lectin (CTL)-galactose binding, Lipoyl transferase 1, Odorant binding protein (OBP9), Translocon-associated protein subunit beta, Glutamate-rich WD repeat-containing protein 1. These proteins were also found to be down-regulated in blood fed midgut tissue as compared to sugar fed. The partial list of down-regulated proteins is provided in Table 3.6 and a complete list can be found in Appendix Table 3 F.

Table 3.5: A partial list of the proteins specifically over-expressed between *Plasmodium berghei* infected and sugar fed midgut tissue.

S. No	*An. stephensi* VectorBase ID	Protein description as per *An. gambiae*	Relative fold change *P. berghei* infected/sugar fed
1	ASTEI04905-PA	Hypothetical protein	1.828
2	ASTEI04093-PA	Symplekin	1.743
3	ASTEI05435-PA	Hypothetical protein	1.500
4	ASTEI09440-PA	Hypothetical protein	1.694
5	ASTEI02439-PA	Lysophospholipidacyltransferase 5	1.747
6	ASTEI06533-PA	Hypothetical protein	1.779
7	ASTEI01731-PA	Selenide, water dikinase	1.509
8	ASTEI00696-PA	V-type proton ATPase subunit F	1.567
9	ASTEI03367-PA	Hypothetical protein	1.790
10	ASTEI02725-PA	IMD pathway signalling Fas-Associated Death Domain (FADD)	1.567
11	ASTEI03813-PA	Hypothetical protein	1.843
12	ASTEI08112-PA	ATP-dependent RNA helicase DDX1	1.723
13	ASTEI03764-PA	Hypothetical protein	1.520
14	ASTEI01016-PA	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	1.514
15	ASTEI05065-PA	Platelet-activating factor acetylhydrolase IB subunit alpha	1.720
16	ASTEI07899-PA	20S proteasome subunit alpha 6	1.508
17	ASTEI09955-PA	39S ribosomal protein L19, mitochondrial	1.789
18	ASTEI06754-PA	Calumenin	1.736
19	ASTEI01710-PA	Solute carrier family 16 (monocarboxylic acid transporters), member 10	1.651
20	ASTEI10474-PA	Lysosomal alpha-mannosidase	1.552
21	ASTEI00934-PA	Hypothetical protein	1.688
22	ASTEI02521-PA	Hypothetical protein	1.693
23	ASTEI09369-PA	Mesenchymal stem cell protein DSCD75	1.763
24	ASTEI06356-PA	2-acylglycerol O-acyltransferase 2	1.778
25	ASTEI00386-PA	glycerol-3-phosphate O-acyltransferase 3/4	1.506
26	ASTEI08151-PA	N-acetylglucosamine-6-sulfatase	1.756
27	ASTEI08952-PA	Hypothetical protein	1.639
28	ASTEI01441-PA	Heat shock protein 67B2	1.559

Table 3.5 contd: A partial list of the proteins specifically over-expressed between *Plasmodium berghei* infected and sugar fed midgut tissue.

S. No	*An. stephensi* VectorBase ID	Protein description as per *An. gambiae*	Relative fold change *P. berghei* infected/sugar fed
29	ASTEI06240-PA	JAKSTAT pathway signalling	1.590
30	ASTEI00150-PA	Lipophorin	1.512
31	ASTEI06638-PA	Hypothetical protein	1.632
32	ASTEI09553-PA	Leucylaminopeptidase	1.592
33	ASTEI06657-PA	Ubiquitin-conjugating enzyme E2 L3	1.563
34	ASTEI08136-PA	Rhomboid-4, isoform B	1.514
35	ASTEI04209-PA	Alkaline phosphatase	1.596
36	ASTEI05015-PA	Minor histocompatibility antigen H13	1.637
37	ASTEI00811-PA	V-type H+-transporting ATPase S1 subunit	1.514
38	ASTEI10973-PA	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3	1.612
39	ASTEI10548-PA	Plasma glutamate carboxypeptidase	1.537
40	ASTEI04332-PA	ADP,ATP carrier protein 2	1.502
41	ASTEI09117-PA	glucosyl/glucuronosyltransferases	1.696
42	ASTEI02403-PA	Muscle LIM protein at 84B	1.537
43	ASTEI06266-PA	Hypothetical protein	1.509
44	ASTEI09894-PA	Hypothetical protein	1.518
45	ASTEI06346-PA	60S ribosomal protein L7a	1.781
46	ASTEI06048-PA	26S proteasome regulatory subunit N8	1.561
47	ASTEI11212-PA	Hypothetical protein	1.514
48	ASTEI04991-PA	Protein kinase, cgmp-dependent	1.702
49	ASTEI11088-PA	Oxysterol-binding protein 1	1.652
50	ASTEI03043-PA	Trehalose 6-phosphate synthase/phosphatase	1.507
51	ASTEI06955-PA	Hypothetical protein	1.677
52	ASTEI02161-PA	Motile sperm domain-containing protein 2	1.545
53	ASTEI03104-PA	Isopentenyl-diphosphate delta-isomerase	1.629
54	ASTEI07913-PA	Vacuolar protein sorting-associated protein 33A	1.559

Table 3.6: A partial list of the proteins down-regulated between *Plasmodium berghei* infected and sugar fed midgut tissue.

S. No	*An. stephensi* VectorBase ID	Protein description as per *An. gambiae*	Relative fold change *P. berghei* infected/sugar fed
1	ASTEI10273-PA	C-Type Lectin (CTL) - galactose binding	0.337
2	ASTEI09679-PA	Lipoyltransferase 1	0.336
3	ASTEI08624-PA	Nascent polypeptide-associated complex subunit alpha	0.335
4	ASTEI08252-PA	Hypothetical protein	0.333
5	ASTEI02077-PA	Cuticular protein RR-1 family (CPR15)	0.332
6	ASTEI06012-PA	Hypothetical protein	0.322
7	ASTEI06045-PA	60S ribosomal protein L21	0.320
8	ASTEI01142-PA	Hypothetical protein	0.313
9	ASTEI07582-PA	Yellow-e	0.301
10	ASTEI10264-PA	Serine protease SP24D	0.298
11	ASTEI05908-PA	CDGSH iron sulfur domain-containing protein 3, mitochondrial	0.291
12	ASTEI05480-PA	Hypothetical protein	0.276
13	ASTEI07746-PA	40S ribosomal protein S30	0.259
14	ASTEI07098-PA	Yellow-x	0.250
15	ASTEI04952-PA	Hypothetical protein	0.250
16	ASTEI06899-PA	Glutamate-rich WD repeat-containing protein 1	0.234
17	ASTEI00001-PA	Hypothetical protein	0.230
18	ASTEI10176-PA	Kinase	0.211
19	ASTEI03534-PA	Myofilin variant B	0.210
20	ASTEI02529-PA	Hypothetical protein	0.201
21	ASTEI06837-PA	Translocon-associated protein subunit beta	0.200
22	ASTEI10165-PA	Actin, cytoplasmic	0.175
23	ASTEI10327-PA	Envelysin	0.153
24	ASTEI03020-PA	Protein of unknown function, DUF255 family	0.147
25	ASTEI078091-PA	Synaptobrevin homolog YKT6	0.130

3.3.4 Differential expression of proteins involved in vector-*Plasmodium* interaction

The differentially expressed proteins identified in this study were compared with proteins which are previously known to play a role in vector-*Plasmodium* interaction in *An. gambiae* (Sreenivasamurthy *et al*., 2013). A total of 12 agonistic and 10 antagonistic known proteins were identified in this study. Among the agonistic proteins, oxidation resistance protein (OXR1) was found to be 2.01 times over-expressed in *P. berghei* infected midgut as compared to blood fed mosquitoes. The siRNA transfection experiments in *An. gambiae* have shown decrease in rate of oocyst formation upon silencing of its responsible gene (Jaramillo *et al*., 2010). Three agonistic proteins such as gelsolin, Short-chain dehydrogenases/reductases and vitellogenin were found specifically down-regulated by 0.63, 0.59 and 0.48 levels in *P. berghei* infected midgut as compared to blood fed midgut respectively. Knocking down of these three genes in mosquitoes has resulted in decrease in oocyst count signifying their positive role in *Plasmodium* infection in vectors (Vlachou *et al*., 2005; Rono *et al*., 2010; Berois *et al*., 2012). Remaining eight proteins were found unchanged in their expression level. Among the 10 antagonistic proteins identified Wiskott-Aldrich syndrome protein were found to be down-regulated by < 0.41 fold in *P. berghei* infected midgut when compared with that of blood fed mosquitoes. These proteins which have been validated in siRNA transfection studies in *An. gambiae*, have shown increase in rate of oocyst formation upon silencing expression of its responsible gene (Vlachou *et al*., 2005; Mendes *et al*., 2008). The remaining 9 antagonistic proteins were found to be unchanged between *P. berghei* infected and blood fed midgut. The details about the proteins involved in vector-*Plasmodium* interaction are provided in Table 3.7.

3.3.5 Proteins involved in vector immunity

The proteins identified in this study were compared to the vector immunity genes reported by Jiang *et al*., 2014. A total of 11 midgut proteins were identified to be involved in vector immunity. Protein hepatocyte growth factor-regulated tyrosine kinase substrate was found to be 1.98 and 1.84 times over-expressed in both blood fed and *P. berghei* infected midgut tissues respectively as compared to sugar fed mosquitoes respectively. This suggests that this protein is specifically over-expressed in response to blood meal. This protein belongs to VHS domain family and generally

involved in promoting rapid recycling of endocytosed signaling receptors to the plasma membrane. Another protein IMD pathway signalling Fas-Associated Death Domain (FADD) was found over-expressed by >1.56 times in *P. berghei* infected midgut and sugar fed. However its expression level remains unchanged in response to blood feeding. This shows that this protein is specifically over-expressed due to invasion of parasite inside the midgut. This protein has a death domain which mediates apoptotic signaling of all the death domain receptors in mouse (Kuang et al., 2000). Previous studies have also demonstrated that this protein inhibit the development of *Plasmodium falciparum* infection in *An. gambiae* as gene silencing of Imd pathway factor FADD, resulted in decrease in oocyst production in mosquito midgut (Garver *et al*., 2012). Protein tRNA pseudouridine13 synthase was found to be 0.6 and 0.546 times down-regulated in both blood and *P. berghei* infected midgut tissue respectively as compared to sugar fed mosquitoes. This shows that protein tRNA pseudouridine13 synthase supports *Plasmodium* infection in the midgut of *An. stephensi*. Interestingly, it was observed that the expression level of these 11 proteins remains unchanged between *P. berghei* infected and blood fed midgut. The predicted role of these proteins was found to be involved in signal transduction and binding activity. Further details of these vector immunity related proteins are provided in Table 3.8**.**

3.3.6 Classification of differentially expressed proteins between sugar, blood and *P. berghei* infected midgut tissue

A total of 730 proteins were found to be differentially expressed between sugar fed, blood fed and *P. berghei* infected blood fed midgut tissue. The information on cellular localization, biological processes and molecular functions for all the differentially expressed proteins identified were retrieved from *Anopheles stephensi* protein reference database (www.VectorBase.org). As illustrated in Figure 3.9 A, the majority of differentially expressed proteins were found to be membrane bound followed by intracellular, cytoplasm, ribosome, mitochondria and nucleus. Majority of proteins (29%) were found to be unclassified in terms of their cellular localization. Moreover in terms of biological processes, the majority of differentially expressed proteins were involved in transport process followed by metabolic process oxidation reduction process, proteolysis, translation and 16% of proteins were found to be unclassified in terms of their role in different biological process as shown in Figure 3.9 B. Molecular

categorization showed that majority of proteins were involved in binding activity followed oxidoreductase activity, transferase activity, catalytic activity, and hydrolase activity. However, 11% of proteins were found to be unclassified in terms of their molecular function (Figure 3.9 C).

Table 3.7: List of proteins involved in vector parasite interaction and their relative fold change between sugar fed, blood fed and *P. berghei* infected midgut tissue. **Agonistic molecules***:* those that support *Plasmodium* development in mosquito midgut.

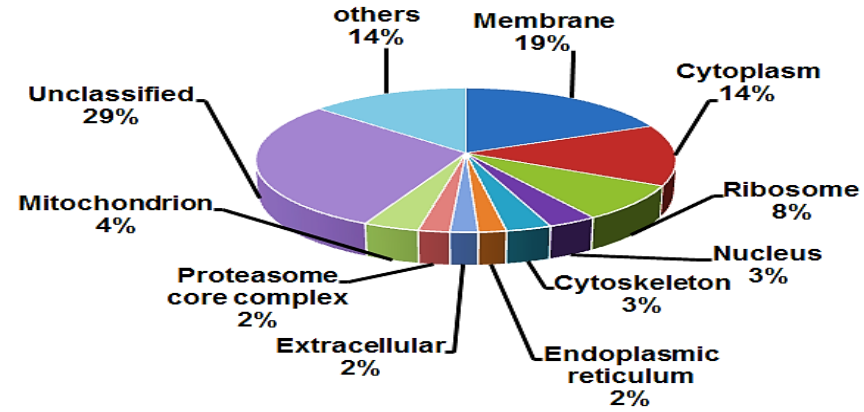
S. No.	*An. stephensi* VectorBase ID	Protein description as per *Anopheles gambiae*	Relative fold change (Infected / blood fed midgut)	Relative fold change(Blood / sugar fed midgut)	Relative fold change (Infected / sugar fed midgut)
1	ASTEI00038-PA	Oxidation Resistance gene 1	2.017	0.753	1.519
2	ASTEI00150-PA	Retinoid and fatty-acid binding glycoprotein, also known as lipophorin or ApoII/I	1.148	1.388	1.512
3	ASTEI04504-PA	Cytochrome P450 reductase	0.998	0.960	1.026
4	ASTEI03111-PA	Actin related 2/3 complex 21 KDa subunit P21	0.956	0.687	0.657
5	ASTEI08607-PA	Adenine nucleotide translocator	0.947	1.032	0.968
6	ASTEI00999-PA	Glutathione-S-transferase theta-1	0.902	1.267	1.222
7	ASTEI02883-PA	*Plasmodium* responsive salivary 1	0.821	0.530	0.589
8	ASTEI01530-PA	Laminin gamma 1	0.819	0.902	0.836
9	ASTEI02221-PA	Arginine kinase	0.777	1.229	1.236
10	ASTEI04270-PA	Gelsolin	0.634	1.129	0.654
11	ASTEI07966-PA	Short-chain dehydrogenases/reductases	0.593	0.913	0.616
12	ASTEI11169-PA	Vitellogenin	0.483	5.301	2.614

Table 3.7 contd: List of proteins involved in vector parasite interaction and their relative fold change between sugar fed, blood fed and *P.berghei* infected midgut tissue. **Antagonistic molecules***:* those that inhibit *Plasmodium* development in mosquito midgut.

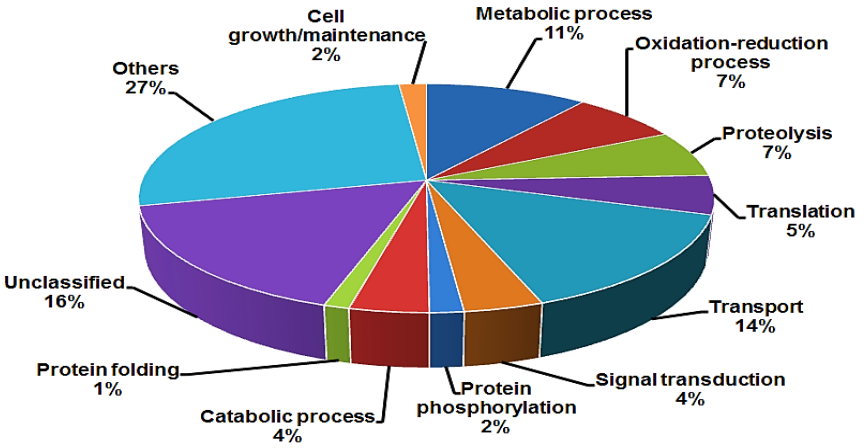
S. No.	*An. stephensi* VectorBase ID	Protein description as in *Anopheles gambiae*	Relative fold change (Infected / blood fed midgut)	Relative fold change(Blood /sugar fed midgut)	Relative fold change(Infected / sugar fed midgut)
1	ASTEI02725-PA	Fas-Associated Death Domain	1.361	1.152	1.567
2	ASTEI06240-PA	Signal Transducers and Activators of Transcription	1.16	1.371	1.590
3	ASTEI01898-PA	Actin related 2/3 complex 41 KDa subunit P41	1.149	0.901	0.971
4	ASTEI07389-PA	Jun N-terminal Kinase	0.883	1.071	0.751
5	ASTEI00731-PA	Cell division cycle 42	0.836	0.820	0.685
6	ASTEI03480-PA	Glutathione-S-transferase theta-1	0.701	1.952	1.873
7	ASTEI03826-PA	Relish 1	0.681	1.647	1.121
8	ASTEI07645-PA	Hsc-3	0.674	1.496	1.045
9	ASTEI01142-PA	Wiskott-Aldrich syndrome protein	0.413	0.759	0.313
10	ASTEI06644-PA	Thioester-containing protein 1	1.104	1.055	1.043

Table 3.8: List of immunity related proteins identified to be differentially expressed between sugar, blood and infected blood fed midgut tissue.

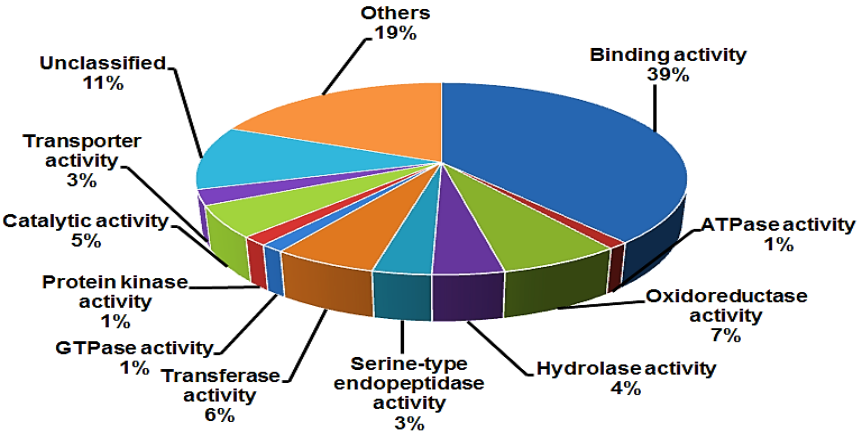
S. No.	*An. stephensi* VectorBase Id	Gene symbol	Protein description as per *An. gambiae*	Relative fold change Blood / sugar fed midgut	Relative fold change Infected / sugar fed midgut	Relative fold change Infected / blood fed midgut	Molecular function	Biological process
1	ASTEI02725-PA	FADD	IMD pathway signalling Fas-Associated Death Domain (FADD).	1.152	1.567	1.361	Protein binding	Signal transduction
2	ASTEI09147-PA	Grb2	Growth factor receptor-binding protein 2	0.934	0.966	1.157	Protein binding	Unknown
3	ASTEI05979-PA	Myd88	TOLL pathway signalling.	1.066	1.168	1.096	Protein binding	Signal transduction
4	ASTEI03942-PA	Raptor	TORC1 growth control complex subunit Kog1	0.763	0.681	1.055	Protein binding	TOR signaling
5	ASTEI02599-PA	Uev1A	Protein arginine N-methyltransferase 7	0.824	0.791	0.960	Unknown	Unknown
6	ASTEI04964-PA	Hrsv1	Hepatocyte growth factor-regulated tyrosine kinase substrate	1.983	1.847	0.931	Metal ion binding	Intracellular protein transport
7	ASTEI07728-PA	Target of ramycin	Autophagy related gene TOR	0.770	0.856	0.928	Protein binding	Unknown
8	ASTEI10386-PA	LRIM26	tRNApseudouridine13 synthase	0.604	0.546	0.904	Protein binding	Unknown
9	ASTEI01297-PA	LRIM17	p70 ribosomal S6 kinase	1.380	1.107	0.802	Protein kinase activity/ATP binding	Protein phosphorylation
10	ASTEI03826-PA	REL1	TOLL pathway signalling NF-kappab Relish-like transcription factor	1.647	1.121	0.681	Sequence-specific DNA binding transcription factor activity	Regulation of transcription
11	ASTEI02289-PA	Insulin Receptor	Insulin receptor	0.636	0.567	0.893	Protein binding/ATP binding	Protein phosphorylation



(A) Subcellular localization



(B) Biological process



(C) Molecular function

Figure 3.9: Classification of differentially expressed proteins between sugar, blood and *P. berghei* infected midgut tissue based on gene ontology annotation (A) Subcellular localization (B) Biological process (C) Molecular function.

3.4 Conclusion:

Mosquito midgut is the first blood reservoir of blood after feeding on infected host and first response site against the malaria parasite mounted by the vector host. Proteins, which are differentially expressed in midgut after a blood meal and *Plasmodium* infection may provide vital information on the sequel of events that would take place in midgut during parasite development and malaria transmission. The present study successfully captured and identified several differentially expressed proteins in response to blood feeding and parasite infection inside the mosquito midgut. Several of the proteins identified were specifically over-expressed in the midgut following normal blood feeding and several others after ingestion of *Plasmodium berghei* infected blood. Some proteins identified have immunogenic function basically involved in cell apoptosis and signaling pathways. Some proteases were found over-expressed which have also been previously reported involved in blood digestion. Several hypothetical proteins were also identified specifically over-expressed in response to *P. berghei* infection. These hypothetical proteins are attractive candidates that can be pursued for unraveling their specific role in parasite invasion, inhibition or survival. The candidate proteins identified in this study may serve as focal starting point for further targeted studies in the context of their potential role in malaria transmission, either promontory or inhibitory. This study thus provides new insights in understanding vector parasite interactions paving way for developing novel strategies for the control of malaria transmission.

Chapter-IV

Application of mass spectrometry-derived proteomic data from midgut in genome annotation of *Anopheles stephensi*

4.1 Introduction

Anopheles stephensi is an important vector of *Plasmodium* species in urban areas of India. The genome sequence of *Anopheles stephensi* (Indian strain) and SDA 500 strain from Pakistan has been recently published and now available in VectorBase (Jiang *et al*., 2014; Neafsey *et al*., 2015). A total of 12,350 genes comprising of 11,789 protein-coding and 571 non-coding genes have been annotated in the genome of Indian strain of *An. stephensi*. The genome of SDA-500 strain has been proposed to encode 13,113 protein-coding and 539 non-coding genes. Biological importance of the completely sequenced genome can be recognized only with precise identification of protein coding genes. This process called 'genome annotation' is partially achieved through computational approaches deployed after genome has been sequenced. One of the initial and crucial steps in understanding the genome is to determine 'protein coding' genes. For this, VectorBase uses a set of gene prediction algorithms to locate genes in the genome (Megy *et al*., 2012). However, most gene prediction algorithms are far from perfect. Computational approaches cannot accurately define protein start sites and exon boundaries for every gene model. In addition, small genes, exact start codons, splice forms and pseudogenes cannot be delineated by existing software. A comprehensive analysis of proteomic data based genome annotation can assist in the identification of protein coding regions and provide a near complete annotation of genome (Brunner *et al*., 2007; Castellana *et al*., 2014).

Proteomic data obtained by mass spectrometry provides unique opportunity to systematically identify protein coding regions in the genome. Proteomic-data driven genome annotation enrich genome analyses much in the line of how cDNA library sequencing based approaches immensely contributed to mammalian genome projects including human (Koga *et al*., 2004; Ota *et al*., 2004). Mass spectrometry-based proteomics methodologies have emerged as a powerful tool for the analysis of protein–protein interactions, post translational modifications (PTMs) and genome annotation of any existing or newly sequenced genomes, which is referred to as proteogenomics (Renuse *et al*., 2011). Proteogenomic analysis validates computationally derived gene models at protein level, allows the identification of novel protein-coding genomic

regions, which are missed by gene prediction algorithms and also provide information to correct existing gene models. Therefore, genome annotation using mass spectrometry derived proteomic data is complementary to other gene prediction methods (Mann & Pandey, 2001). The use of mass spectrometry in the validation of genome annotation has been successfully demonstrated in *Anopheles gambiae* (Kalume *et al*., 2005; Chaerkady *et al*., 2011; Dwivedi *et al*., 2014), *Plasmodium falciparum* (Lasonder *et al*., 2002) and several other organisms.

In VectorBase, approximately 12,350 genes currently listed for *Anopheles stephensi* have been derived using gene prediction algorithms, which include non-coding RNAs as well. The large majority of hypothetically predicted protein coding genes in *Anopheles stephensi* is yet to be validated by any other experimental methods. Until now, no proteogenomic investigation has been carried out in case of *An. stephensi* since its genome has been published. In this study genome annotation of *Anopheles stephensi* Indian strain was carried out using mass spectrometry derived proteomic data obtained from the midgut of female *An. stephensi*. This analysis led to validation of several hypothetically predicted protein coding genes, besides identifying novel protein coding regions in the *An. stephensi* genome that were missed out by the computational predictions. It is expected that a well annotated genome and proteome of *Anopheles stephensi* will provide a better platform for investigating proteins and genes involved in disease transmission and would eventually help in defining and finding new ways to control the transmission of malaria.

4.2 Materials and methods

4.2.1 Creation of databases for genome annotation

The genome sequence for *Anopheles stephensi* Indian strain was downloaded from the VectorBase database (www.VectorBase.org) version AsteI1 and AsteI2 which has free access. Using in-house python scripts, a six-frame translated database was created. Another 3 frame translated RNA-Seq transcripts database was created using RNA-seq data that was generated from *Anopheles stephensi* Indian strain (IOB-NIMR unpublished data). MS/MS data was searched using a unique search workflow on Proteome

Discoverer (Version 1.4.1.14) software (Thermo Fisher Scientific, Bremen, Germany). The first search was carried out against *An. stephensi* protein database, which resulted in the identification of proteins. The unassigned spectra were searched against a combined database of 3 frame translated RNA-Seq-based transcript database and 6 frame translated *An. stephensi* genome databases, respectively as shown in Figure 4.1. The peptides which were obtained after application of 1% FDR at PSM level were selected for proteogenomic analysis. Peptides identified only against six-frame translated genome, but not present in protein database, were used for genome annotation. These peptides are categorized as genome search specific peptides (GSSPs). Those peptides which had unique location in genome were only considered for genome annotation. GSSPs were further subjected to a computational pipeline developed for validation of known genes models, identification of novel protein coding genes and correction of existing gene models of *Anopheles stephensi.* To find out the genome coordinates, tBLASTn was performed of all these GSSPs. Further the GSSPs were categorized in six categories such as (1) peptides mapping to intergenic regions; (2) peptides overlapping with exon–intron junctions; (3) peptides mapping to intronic regions (4) peptides mapping to UTRs; (5) peptides mapping onto gene boundaries and (6) peptides mapping within the exons but translated in a different frame. The protein blast algorithm was used to align GSSPs to protein sequences from other mosquitoes including *Anopheles stephensi* SDA 500 strain and *Anopheles gambiae*, to find ortholog evidence. Peptides spanning the exon splice junctions of known proteins were fetched using in-house scripts and were used for confirmation of the splice junctions of the known protein coding genes. The coding potential of a given genomic region was confirmed by the presence of transcript evidence or ortholog evidence from *An. stephensi* SDA 500 strain.

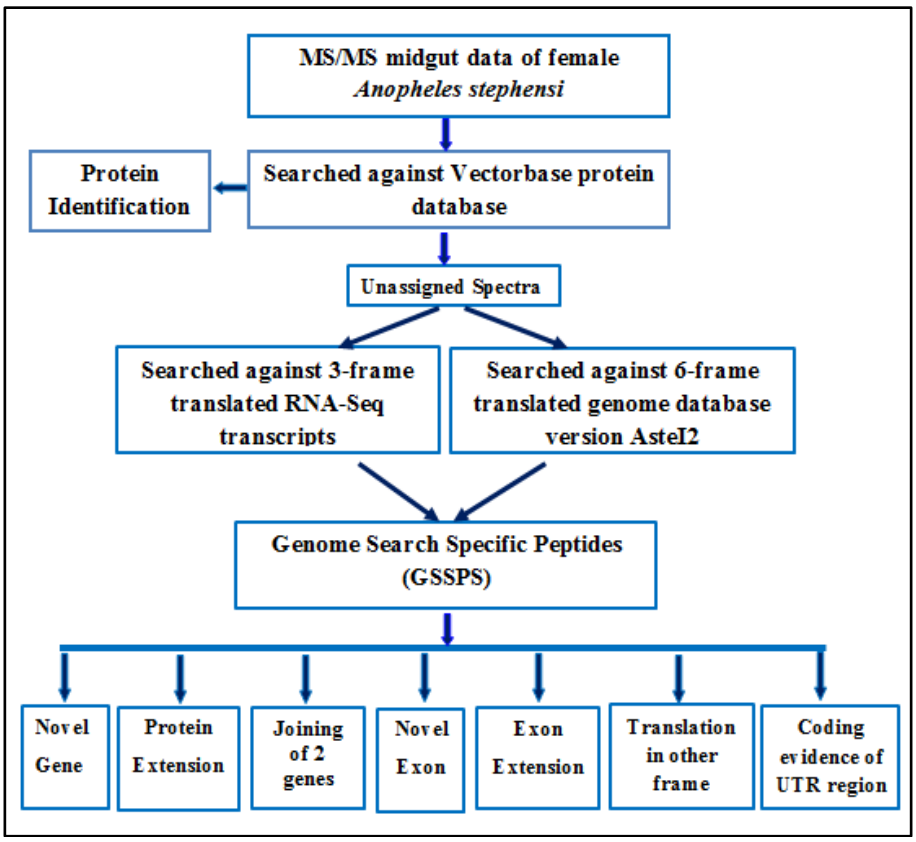


Figure 4.1: Workflow of proteogenomic analysis.

4.2.2 Primer designing strategy

Specific primers were designed for novel gene models using the Primer 3 (v.0.4.0) software (Rozen and Skaletsky, 2000). The primers were designed based on annotation of gene models supported by transcript evidence or based on homology to other mosquito species.

4.2.3 Validation of novel gene models by RT-PCR

Total RNA was isolated from pooled mosquito midgut tissues using QiagenRNeasy mini kit (Cat. No.74104). RNA was treated with DNase to remove any contaminating genomic DNA. The RNA yield was estimated using Nano spectrophotometer Q5000 (Quawell). cDNA synthesis was carried out using Qiagen's QuantiTect Reverse Transcription kit (Cat. No.205311). The resulting cDNA was used for the PCR reactions. The PCR experiment was performed using Taq polymerase (Merck), 50ng of cDNA and 200 nM of each primer pair. The PCR products obtained were observed on agarose gel

electrophoresis. The size of the PCR product was determined using 100bp DNA ladder (Merck). Bands visualized at the expected sizes were purified using Qiagen's Gel Extraction kit (Cat. No.28704). The PCR purified product obtained was outsourced to Bangalore Genei for sequencing. The cDNA sequences obtained have been submitted to dbEST (http://www.ncbi.nlm.nih.gov/nucest?LinkName=biosample_nucest&from_uid=1837904).

4.3 Results and discussion

4.3.1 Genome annotation refinement using GSSPs

Mass spectrometry data was searched against six frame translated genome database of *An. stephensi* latest version AsteI2. The total number of genome search specific peptides exclusively identified from the genome database search results was 1,037, hereafter named as GSSPs. These GSSPs were identified against the latest revised assembly version AsteI2. These GSSPs were further categorized using the proteogenomics pipeline as discussed in the methodology section. Of these 1,037 GSSPs, peptides mapping to the intergenic region were (553), intronic region (305), UTR region (41), peptides spanning exon-intron boundary (94), peptides spanning gene no gene boundary (38) and peptides mapped to coding region in different frame (6). The details of all the GSSPs identified in this study have been provided in the Appendix Table 4 A. A summary of all the novel findings identified in the proteogenomic analysis of *An. stephensi* have been summarized in Table 4.1. The annotation refinement will be discussed in detail with examples for each category.

Table 4.1: The details about all the novel genes annotated/ identified in the proteogenomic analysis of *An. stephensi.*

S. No.	Categories	Number of novel gene models identified/annotated
1	Novel genes	104
2	Joining of two genes	10
3	Novel coding exons	40
4	Exon extensions	45
5	Protein extensions	30
6	Translation in alternate frame	5
7	Alternate protein start site	10

4.3.2 Identification of novel protein-coding regions using intergenic peptides

To identify genes that might have escaped prediction by computational pipelines, the unassigned mass spectra obtained after protein database was searched against customized databases containing 6-frame translation of genome version AsteI2 and 3-frame translation of transcriptome of *An. stephensi*. This analysis led to identification of 553 GSSPs mapped to the intergenic region which in turn lead to the identification of 104 novel protein coding genes and correction of 10 known gene models in *An. stephensi* Indian strain, which are not represented in predicted protein database for *An. stephensi* Indian strain. The details about the peptides and their genomic coordinates are provided in Appendix (Table 4 B). Identification of novel genes is an important finding of any proteogenomic analysis. As an example, 12 unique peptides were found to map to an intergenic region of Scaffold 00003, which suggested a missed out region of protein coding potential in Scaffold 00003. This claim is further validated by the identification of a novel transcript in the same region, which supports its coding potential. As an additional evidence, ortholog gene (ASTE010150) for the novel peptides in the SDA-500 strain was also identified. Considering all the evidences, a novel gene in the intergenic region of Scaffold 00003 is proposed (Figure 4.2). Domain analysis of this novel gene predicted its role in transcription process.

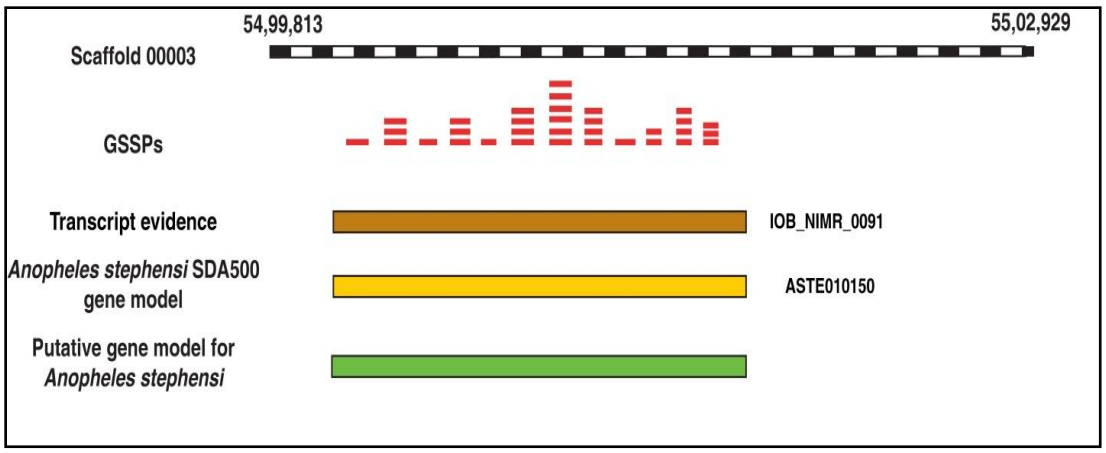


Figure 4.2: Identification of novel genes using peptides identified in intergenic regions. Twelve peptides were identified in the intergenic region of Scaffold00003. Presence of the novel gene is supported by transcript evidence and ortholog protein (ASTE010150) in SDA-500 strain.

Another example of novel gene identification using evidence from unique peptides is shown in Figure 4.3. In this case, six novel peptides were found mapping to an intergenic region in Scaffold00067, which was missed by the annotation tools in Indian strain of *Anopheles stephensi*. A novel transcript was also identified in the same region, which supports its coding potential. The corresponding protein (ASTE008151) for the novel peptides was also found in the SDA-500 strain. Domain analysis of this novel gene predicted an unknown function for this protein.

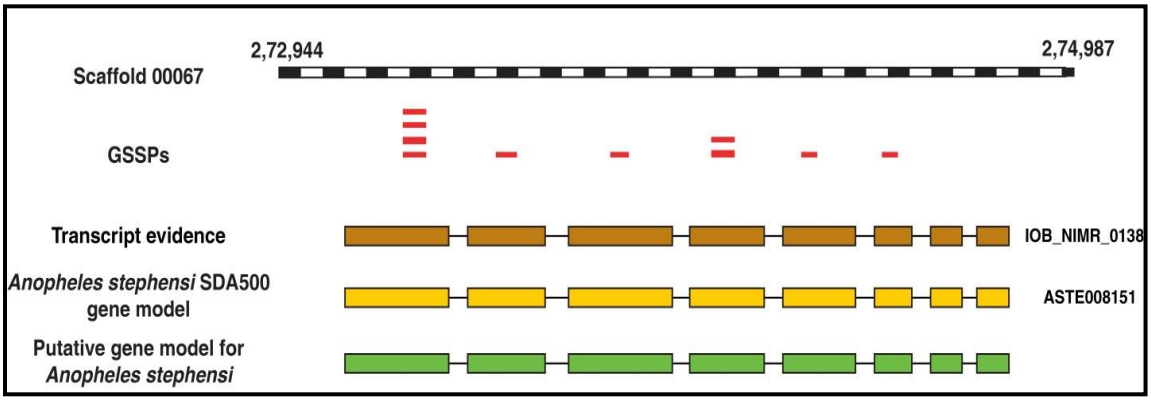


Figure 4.3: Identification of novel gene using peptides identified in intergenic regions. Six peptides were identified in the intergenic region of Scaffold00067. Presence of a novel gene is supported by transcript evidence and corresponding protein (ASTE008151) in SDA-500 strain.

Using intergenic peptides, similar analysis was carried out against previous genome assembly version AsteI1, which resulted in identification of 31 novel gene models. The RT-PCR based validation and cDNA sequencing was carried out for 31 such novel genes and the cDNA sequences obtained were submitted to dbEST (http://www.ncbi.nlm.nih.gov/nucest?LinkName=biosample_nucest&from_uid=1837904). Later, in the current revised assembly version AsteI2, 31of these novel genes are incorporated (Jiang *et al*., 2014). This illustrates the significance of proteomic data in the revision of genome annotation.

4.3.3 Joining of two genes

One of the major limitations in genome annotation using gene prediction methods is their inaccuracy in predicting gene structures by frequently merging two or more genes as one single gene or even mistakenly splitting a single true gene into two genes. This kind of observation was also identified during deep analysis of the midgut dataset. In this study such ten examples were identified, where multiple peptides supported the existence of a single gene but annotation predicted two separate gene models. The details of these unique examples have been illustrated in Appendix Table 4 C. Figure 4.4 shows that five peptides were found to map to the intergenic region between two annotated VectorBase genes ASTEI10980 and ASTEI10981. The peptides provided evidence for extension of both these two genes. However, transcript evidence indicated that the two genes were erroneously annotated as separate genes and could be merged into one single gene. The presence of single orthologous gene ASTE009796 in SDA500 strain provides further support for the revised gene structures. Further blast analysis showed alignment with MP dependent ligase protein in *An. gambiae*.

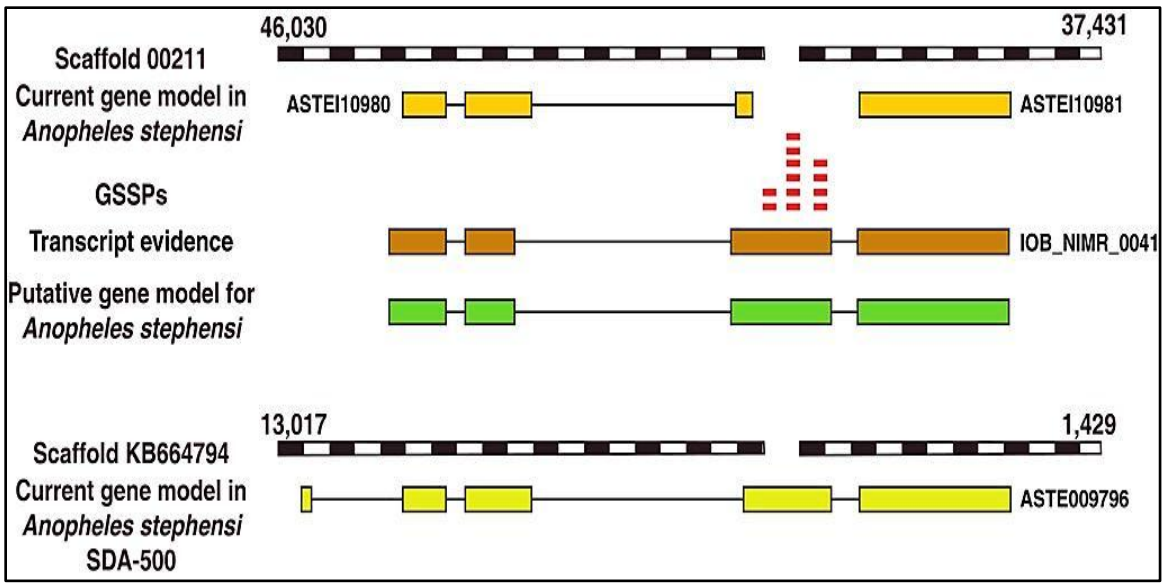


Figure 4.4: An example of joining of genes ASTEI10980 and ASTEI10981 based on transcript, intergenic peptides and ortholog gene in SDA 500 strain. A single gene structure was identified in place of two separate predicted genes using 3 intergenic peptides mapped between the two annotated genes, ASTEI10980 and ASTEI10981.

4.3.4 Identification of novel coding exons using peptides mapping to intronic region

Peptides that are mapping in intronic regions of an annotated gene can lead to either correction of gene structure by extending an exon or identifying a novel exon. Altogether 305 GSSPs were identified mapping to intronic region of 40 genes in *An. stephensi* Indian strain. For example, a peptide (GTVNCDDEQTNPAPPQAQVCVCVVHNK) was identified mapping to the intronic region of a gene ASTEI06691 in scaffold00043. A transcript evidence, which supports the coding potential of this novel exon was identified as shown in Figure 4.5 A. Similarly, another peptide YIDDAVYEMIR mapped to intronic region of gene ASTEI09052 and the presence of a novel exon was also supported by transcript evidence (Figure 4.5 B). In this study such 40 examples were identified, where multiple peptides identified against the intronic region, which leads to the identification of novel coding exons in the annotated gene models. The details of all the novel coding exons identified from the analysis have been provided in Appendix Table 4 D.

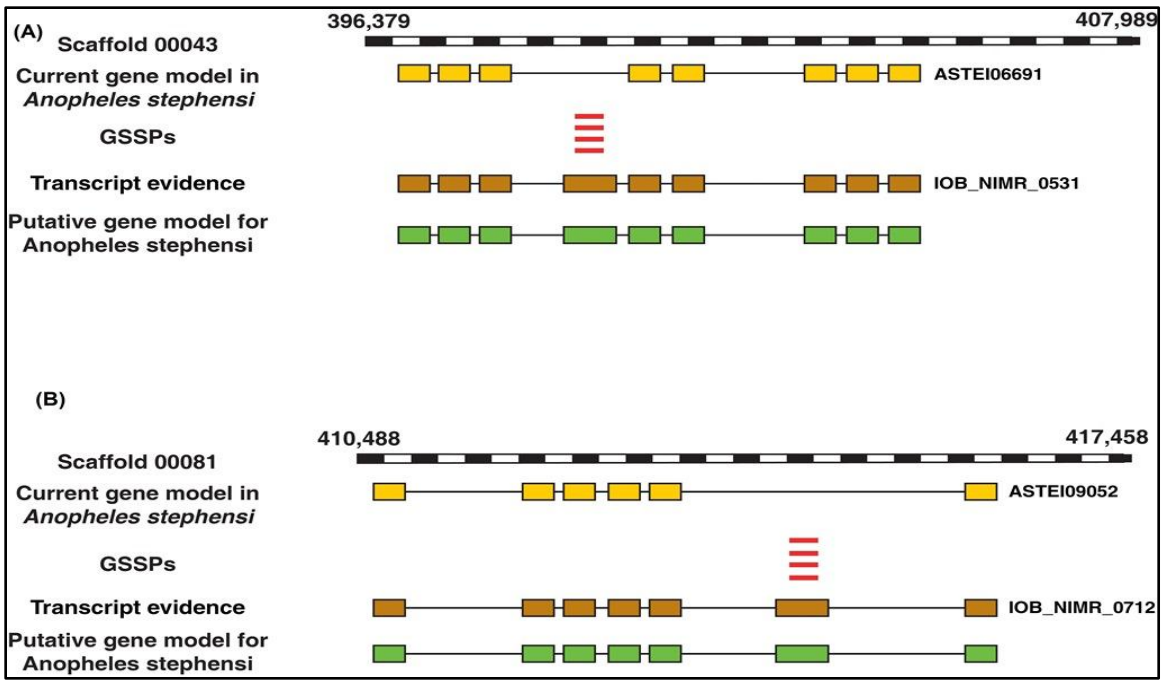


Figure 4.5 A and B: Correction in the gene structure using peptide mapping to an intron of an annotated gene. Peptides were identified in introns of the two genes annotated as ASTEI06691 and ASTEI09052. These intronic peptides support a novel coding exon in these two gene models, which is currently missing in existing gene model.

4.3.5 Identification of exon extension using peptides overlapping to exon-intron boundaries

In this study, 94 peptides were identified which map partially to annotated exons which clearly suggest a change in exon structure by extending them. These unique peptides spanned the exon-intron boundaries which indicate changes in 45 gene structures. Figure 4.6 A and B illustrated an example of exon extension based on a peptide overlapping exon-intron boundary. The presence of exon extension is also supported by transcript evidence. The peptide that maps onto the exon was verified whether it is translated in the same frame as the annotated exon itself. The details about the GSSPs mapping to exon-intron junctions of VectorBase genes leading to exon extension are provided in Appendix table 4 E.

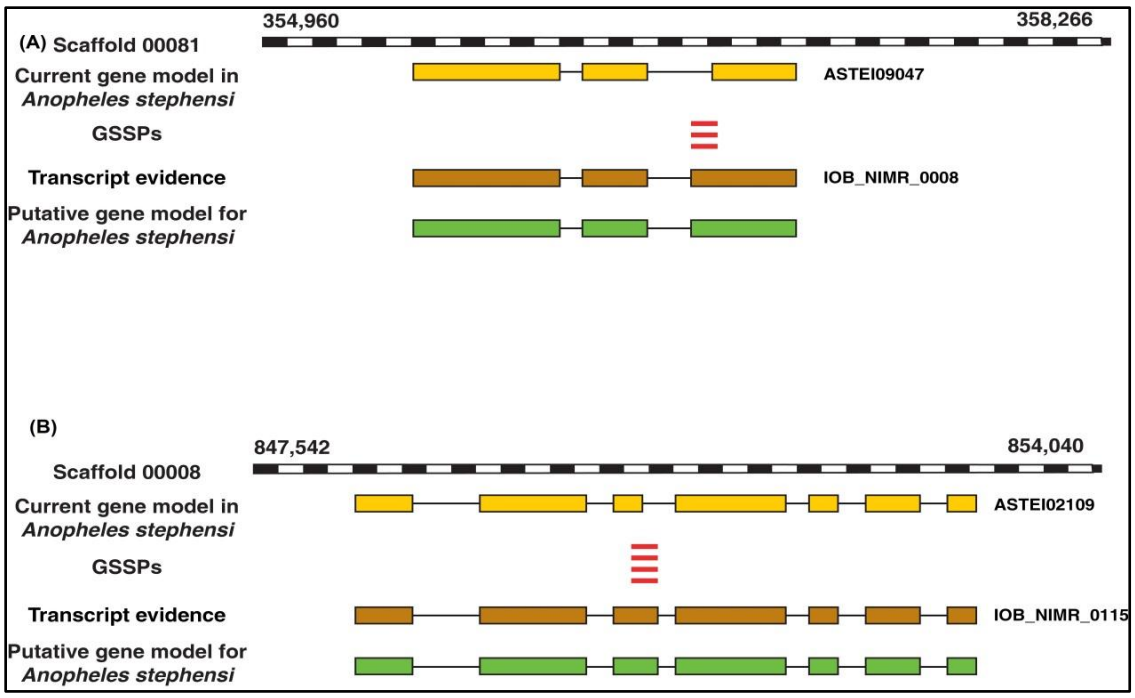


Figure 4.6 A and B: Identification of peptides mapping to exon-intron boundary of annotated genes ASTEI09047 and ASTEI02109. Presence of transcript and peptides support extension of existing exons in these two gene models.

4.3.6 Identification of protein extension

During this study, 79 GSSPs were identified that mapped to 5' and 3' boundary of existing 30 genes in *An. stephensi* which suggests refinement by an N-terminal or C-terminal extension of the existing gene model. One such example is shown in Figure 4.7 A where two GSSPs were found to map N-terminal boundary of existing gene ASTEI02407 in *An. stephensi*. The presence of protein extension is also supported by transcript evidence and ortholog gene ASTE001573 in SDA 500 strain. Another four peptides were found to map to C-terminal extension boundary of existing gene ASTEI00095 in *An. stephensi* Figure 4.7 B. The presence of protein extension is also supported by transcript evidence and ortholog gene ASTE009977in SDA 500 strain also supports the longer gene model. A total of 20N-terminal and 10 C-terminal extension of the existing gene models were identified in this study (Appendix Table 4 F).

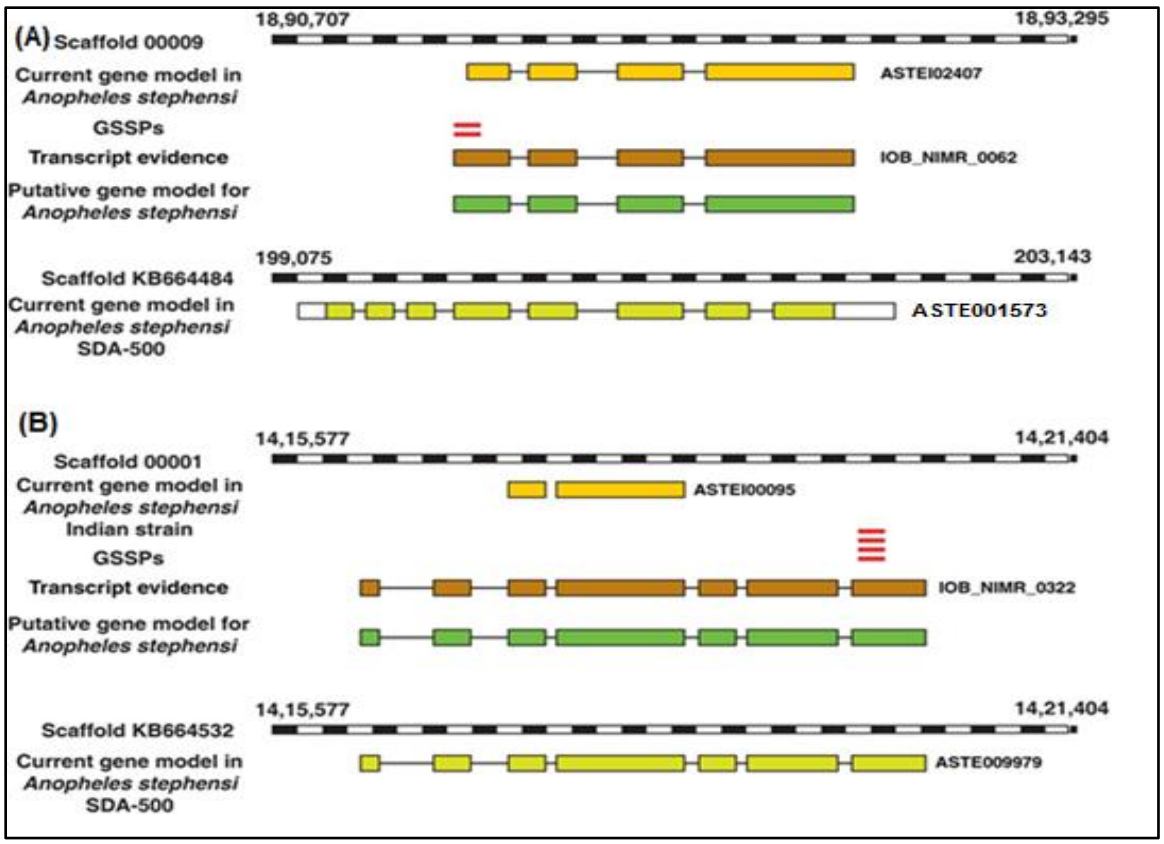


Figure 4.7 A and B: N-terminal extension of a protein using peptides mapping upstream to an annotated start site. Peptides were mapped to the N-terminal region of gene ASTEI02407 and ASTEI00095 suggest much longer gene extending N-terminal and C-terminal of the existing gene model respectively. Transcript and ortholog gene in SDA 500 strain also supports the N- terminal and C-terminal extension.

4.3.7 Identification of an alternate frame of translation

In the genome annotation, the biggest challenge is determination of the correct reading frame of translation for a particular protein coding gene. Due to inaccurate prediction by gene finding programs or due to some biological events reading frames can be wrongly assigned. Previous studies have reported several cases of translational frame shifts, where two translational reading frames are expressed as a single protein (Farabaugh, 1996; Chaerkady *et al*., 2011; Dwivedi *et al*., 2014). It is very difficult to find out the exact reason behind the frame changing events. However, mass spectrometry based proteomics data can provide experimental evidence of the correct translational frame through peptide data. Such 5 events were identified in this study and the details of which are provided in Appendix Table 4 G. In this study 6 peptides were found mapping to annotated exons but

in a different frame of translation. Figure 4.8 shows an example where peptide EETPAAVPAETVGDVEGEPK was found to map within the fourth exon of the ASTEI06745 gene but in different translation frame than the existing one. The predicted protein product of the transcript IOB-NIMR 0584 and orthologous gene ASTE006676 of SDA500 strain, showed the presence of the identified novel peptide. Blastp analysis of this protein showed alignment with Solute carrier family 15 member 1 protein in *An. gambiae.*

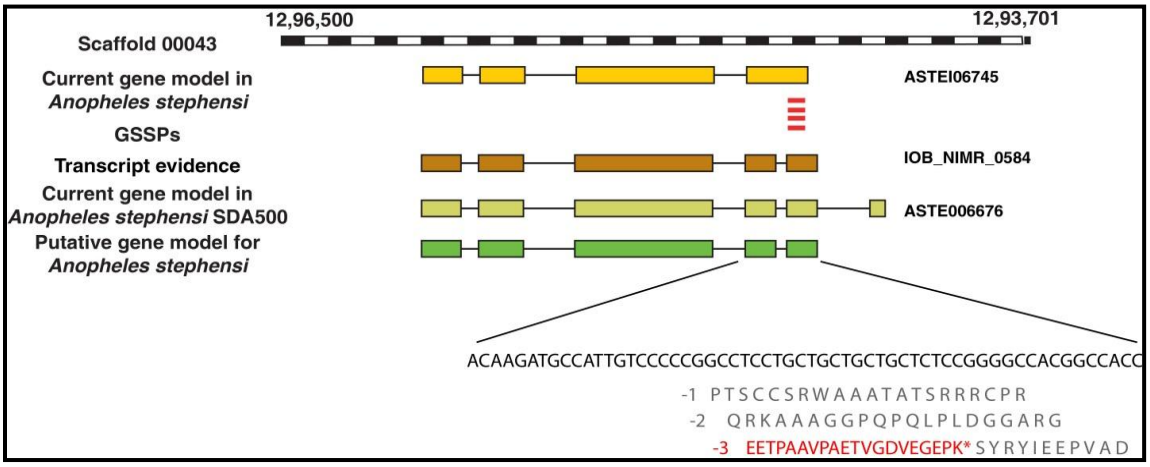


Figure 4.8: Identification of peptides translated in a different frame than the annotated protein sequence. Peptide SETPAAVPAETVGDVEGEPK mapped with in the co-ordinates of the existing gene ASTEI06745, but fell on a translational reading frame different from the predicted polypeptide.

4.3.8 Identification of alternate translational start site using N-terminally acetylated peptides

Alternative start sites can result in smaller or longer isoforms that may potentially have different functions. Studies in *Drosophila* have shown the presence of non-canonical start codon and impact of alternative start sites on protein function (Beerman and Jongens, 2011). Gene prediction algorithms are error prone and can go wrong in precise identification of protein start site. Most of the translational start sites are predicted based on the longest open reading frame. Using proteomic data, one can provide confirmatory evidence for determination of translational start sites. To identify alternate start sites, MS/MS data was searched against a customized protein sequence databases of upstream and downstream peptide sequence starting with methionine as previously described by

(Chaerkady *et al*., 2011). In addition, N-acetylome analysis of *An. stephensi* was also carried out. As most of the proteins are N-term acetylated, N-term acetylated peptides from midgut tissue were enriched and analyzed using high resolution mass spectrometer. This analysis led to the identification of 10 nos. of acetylated peptides, which resulted in correction of start sites in 10 no. of proteins (Appendix Table 4 H). Figure 4.9 A and B demonstrates acetylated peptide-based confirmation of alternate start site downstream of gene ASTEI08087 and ASTEI08114.

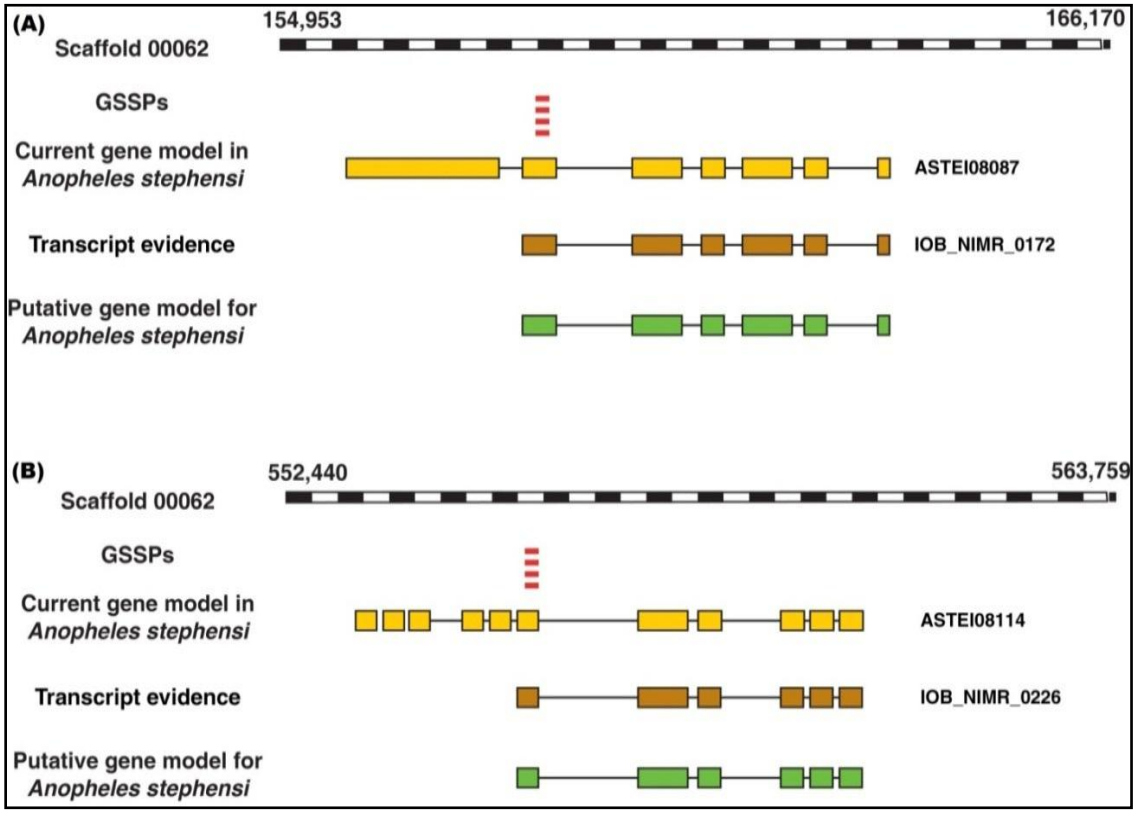


Figure 4.9 A and B: Identification of an N-terminally acetylated peptide indicated an alternate start site for ASTEI08087 gene and ASTEI08114.

4.3.9 RT-PCR validation of subset of novel genes annotated against *An. stephensi* genome assembly version AsteI1

RT-PCR validation was done for the 31 novel gene models identified against the first genome assembly of *An. stephensi* (AsteI1) submitted to VectorBase in June 2012. Figure 4.10 shows the amplified products for the RT-PCR experiments performed for the subset of novel gene models. The cDNA sequences have been submitted to NCBI dbEST (http://www.ncbi.nlm.nih.gov/nucest?LinkName=biosample_nucest&from_uid=1837904). Later, in the current revised assembly AsteI2, 31 of these novel gene models have been incorporated. Also, it is important to note that the RT-PCR experiments were performed before the release of the latest genome assembly version AsteI2. The details of sequences submitted to GenBank including GSSPs and primers are provided in Appendix Table 4 I.

.

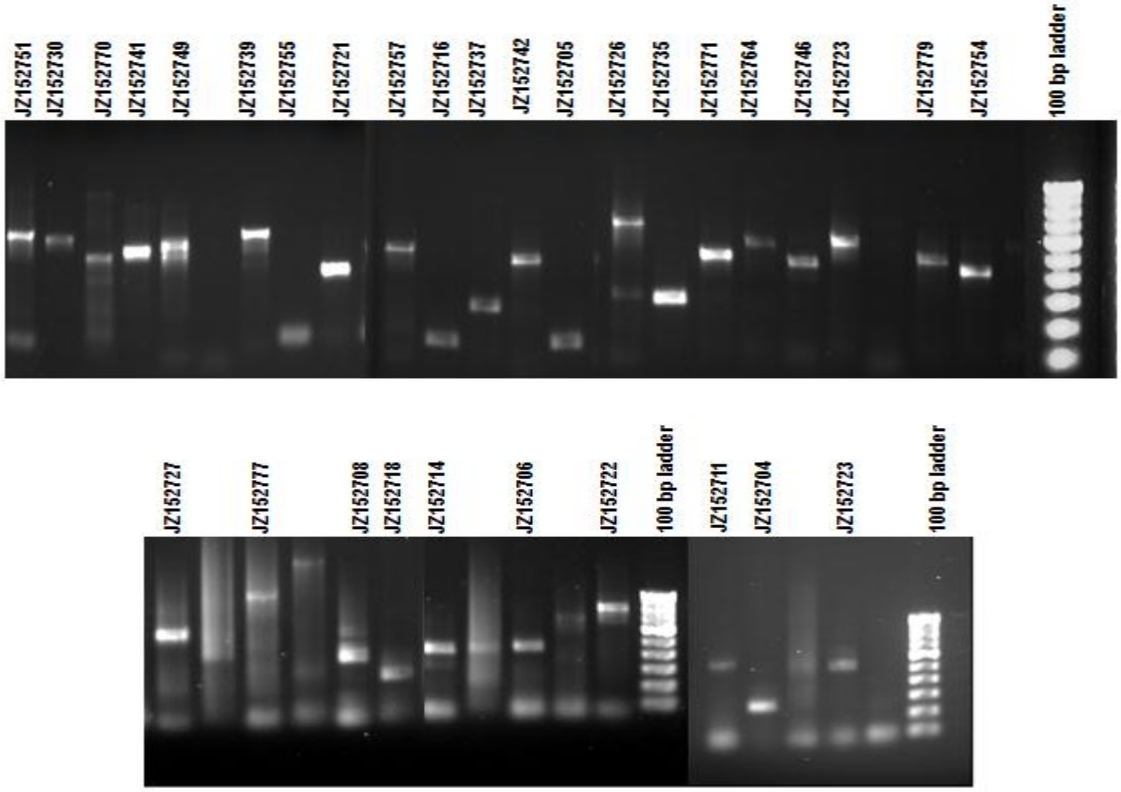


Figure 4.10: Gel image of RT-PCR validation of the novel gene models identified against first genome assembly AsteI1 of *An. stephensi* Indian strain. RT-PCR products were sequenced and the resulting sequences were submitted to GenBank. Accession numbers provided by NCBI GenBank are indicated above each lane.

4.4 Conclusion

In conclusion, the findings of the study illustrates the application of high-resolution mass spectrometry data for analysis and annotation of the *An. stephensi* genome. The peptide data served as corroboratory evidence in the validation of known and novel protein-coding regions as well as in correction of the existing gene models. Overall, this study demonstrates the power of high mass accuracy mass spectrometry–derived proteomic data to complement other approaches for genome annotation. Using proteomic data, large number of genome assembly errors in *An. stephensi* genome of Indian strain were identified and corrected. Similar analysis with previous database i.e. Astel1 resulted in identification of 31 novel gene models, which improved gene annotations and were included in the revised database version Astel2. This illustrates the significance of proteomic data in revision of genome annotation. The pipeline adopted in this study for mapping proteomic data over genome sequences can provide a framework for developing more automated workflows for genome annotation. Accurate genome annotation along with proteomic landscape will provide a better interpretation of a sequenced genome and aid in further investigations on finding novel specific targets and approaches for disease control. The study has also identified several novel genes which could be studied further as targets for malaria control strategies.

Summary

&

Future Prospects

Summary

Midgut of mosquito is a vital tissue, involved in blood digestion and constitutes the first physical barrier to Plasmodium invasion. Therefore, molecular understanding of midgut proteome is necessary to appreciate its role in vector-parasite interaction and blood digestion. The present investigation was carried out in order to characterize midgut proteome of female *Anopheles stephensi* using mass spectrometry. In this study multiple fractionation methods as well as high-resolution and high mass accuracy mass spectrometer were used. The three different protein/peptide fractionation strategies were used namely, In-gel digestion, bRPLC and Off-gel fractionation which resulted in reduction of sample complexity and facilitated deep proteome coverage of midgut tissues. The fractionated samples were analyzed on high resolution mass spectrometer (LTQ-Orbitrap Elite and LTQ-OrbitrapVelos) which generated over 7,16,932 high resolution tandem mass spectra from 58 LC-MS/MS runs. Two search engines i.e. Sequest and Mascot were used for searching the mass spectrometry derived data against the protein database of Indian strain of *An. stephensi* available on VectorBase, which resulted in identification of 5,43,028 peptide spectral matches (PSMs) corresponding to 63,525 high quality peptides.These peptides provided experimental evidence for 5,434 computationally predicted proteins, out of 11,785 total numbers of proteins, which comprises ~ 46% of the total reported proteome of the malaria vector *An. stephensi*. This is the largest protein catalogue of midgut in any mosquito species reported so far. This proteomic profiling of midgut tissue has provided translation evidence for thousands of genes for which there were no experimental evidence thus far.

For the entire peptide data set, the distribution of mass error in parts per million was calculated. This analysis shows that nearly 95% of the peptides were within ±5 ppm mass error, confirming high accuracy of peptide data obtained from the mass spectrometer. To obtain relative expression of these 5,434 identified proteins within in the tissue, intensity based quantitative expression was carried out. The quantitation was based on spectral count identified for each protein and the variability in the length of the proteins was accounted for to normalize the abundance values. This analysis shows that most of the house-keeping genes including cytoskeleton proteins, glucose metabolism and proteins

involved in translation and embryonic development were among the most abundant found in midgut tissue. These proteins were identified with high coverage, multiple unique peptides and multiple number of PSMs. Qualitative and quantitative information of protein expression will lead to better understanding of the vector life cycle. Further the identified proteins were categorized according to the subcellular localization, their involvement in different biological processes and molecular functions based on gene ontology annotation. For all the identified proteins, ortholog information from the closely related mosquito species *An. gambiae* was fetched using in-paranoid tool. This analysis resulted in identification of 43 known proteins similar to *Anopheles gambiae*. These proteins mediate the interaction of the *Plasmodium* parasites with mosquito tissues and regulate their maturation within their host vectors (Sreenivasamurthy et al., 2013). Another, 49 proteins have been identified in this study which play an important role in immune responses in *Anopheles stephensi,* as previously reported by Jiang *et al.,* (2014). For the proteins, whose ortholog information was missing in closely related species, domain analysis was carried out using SMART domain prediction in order to predict their function in midgut tissue.

Comparative proteomic analysis of male and female midgut revealed differential expression of proteinsbetween two sexes. By employing iTRAQ based quantitative approach, a total of 3,316 proteins were identified. Further relative quantitation of these identified proteins was achieved by using reporter ions quantifier node in Proteome Discoverer software to determine relative quantity of a protein in male and female midgut tissue. This analysis resulted in identification of 758 proteins found to be differentially expressed in female midgut with respect to the male counterpart. As expected, proteins involved in blood digestion, glucose metabolism and detoxification processes were found to be over-expressed in female midgut. Thirty two proteins involved in vector parasite interaction were also identified. Further, the differentially expressed proteins identified in this study were compared with sex-specific genes of *An. stephensi* Indian strain and found a strong correlation (Jiang *et al*., 2014). Fifty female specific proteins identified in this study were found to be over-expressed by >2 fold in female midgut as compare to male. These female specific proteins were mainly involved in proteolysis and metabolic

processes relevant to blood meal digestion, vector immunity and some are involved in parasite interaction. Another, 19 male specific proteins identified in this study were reported earlier as male biased by Jiang *et al*., (2014). This quantitative proteomic information generated clearly explains the functional diversity between the male and female mosquito, and also explains why female mosquitoes are the vector for this debilitating disease while a male mosquito does not possess vectorial capability. Further functional analyses of these female-specific proteins are necessary for better understanding the molecular differences between male and female *An. stephensi.*

Comparative proteomic analysis of midgut of sugar, blood fed and *Plasmodium berghei* infected mosquitoes provides information about the proteins differentially expressed in response to blood meal and parasite infection in midgut tissue of female *An. stephensi*. This analysis led to identification of 2,327 proteins. Relative quantification of proteins identified between blood fed versus sugar fed and infected blood fed versus blood fed was carried out on the basis of relative intensity of reporter ions for each of the peptides identified. This analysis showed that, 435 proteins expressed differentially between blood fed and sugar fed midguts. Of these, 247 proteins were up-regulated by >1.5 fold while 188 proteins were down-regulated by <0.66 fold in blood fed as compare to sugar fed midgut tissue. These proteins were found to be involved in vector development process, survival, blood digestion or in the synthesis of the peritrophic matrix that surround the blood meal, transport and storage of iron within cells, vector immunity and defense against microorganisms ingested along with the blood meal (Nolan *et al*., 2011; Rono*et al.*, 2010). Similarily, 231 proteins were identified expressed differentially between *Plasmodium berghei* infected and blood fed midgut. Of these 231 proteins, 55 proteins were found to be up-regulated by >1.5 fold while 176 proteins were down-regulated by <0.66 fold in *Plasmodium berghei* infected as compare to blood midgut tissue. Interestingly one agonistic protein, oxidation resistance protein (OXR1) was found to be 2.01 times up-regulated in *P. berghei* infected midgut as compare to blood fed mosquitoes. Another protein, Wiskott - Aldrich syndrome which is a antagonistic protein was found < 0.41 times down-regulated in *P. berghei* infected midgut as compared to blood fed mosquitoes. Among the 55 up-regulated proteins, 15 hypothetical proteins were

identified for which orthologous information was missing in closed related mosquito species *An. gambiae*. These hypothetical proteins are anticipated to be of vital importance to *Anopheles stephensi* and can be used for the targeted expression of anti-Plasmodium effector molecules. These proteins can be explored for functional studies on mosquito-parasite interaction by transgenic expression of these, which can coincide with midgut stages of the Plasmodium parasite. Further the role of these proteins needs to be confirmed by conducting functional analysis using gene knock down.

The proteogenomic analysis was carried out using the peptides identified from the midgut tissue. The novel peptides, which could not be mapped to the known proteins, were searched against 3 frame translated transcript and 6 frame translated genome database. This analysis led to identification of 1,077 genome search specific peptides (GSSPs). The proteogenomic analyses of GSSPs lead to the identification of 104 novel gene models that were missed during genome assembly and correction of over 140 existing gene models through discovery of 40 novel exons, 75 protein extensions, 10 novel protein start sites, 5 novel translational frames and 10 events of joining of adjacent genes as a single gene. The similar proteogenomic analysis was carried out against the first genome assembly of *An. stephensi* (AsteI1) submitted to VectorBase in June 2012. Analysis against first predicted protein and translated genome database (Astel1) led to identification of over 31 novel protein coding regions. These novel gene models were validated by performing RT-PCR and subsequent sequencing of cDNA to support the existence of novel gene models discovered in this analysis. The cDNA sequences thus obtained were submitted as ESTs to NCBI Genbank. All the 31novel genes identified previously were annotated in the next revised database i.e. Astel2. This shows the significance of proteomic data in refinement of genome annotation.

This is the first in-depth proteogenomic analysis of *An. stephensi* Indian strain since its genome sequence has become available. This study suggests that the genome annotation process should comprise of both computational and experimental methods. Several of the novel genes identified in the study are interesting candidates for intervention studies. This study provides a platform for further investigations especially those focused on preventing malaria transmission and vector development.

Future Prospects

In the present study tandem mass spectrometry-based proteomic profiling of *An. stephensi* was carried out to achieve broad coverage of midgut proteome. This study provides valuable information on sex-specific protein expression in female and male midgut. Over-expressed proteins in the female midgut are attractive candidates for targeted studies to test their role in vector development, malaria transmission or inhibition process.

Quantitative proteomic analysis of *Anopheles stephensi* midgut infected with *Plasmodium berghei* allowed us to obtain the biologically relevant midgut proteins that are specifically induced in response to malaria parasite infection. Attractive candidates among these can be further pursued for their role in parasite invasion, inhibition or survival using dsRNA interference based gene silencing assays, paving way for candidates for the development of malaria transmission blocking vaccines. The transmission-blocking vaccines (TBVs) can be explored for interruption of transmission of parasites via suppression of parasite development in vector mosquitoes.

The major highlight of the study has been the use of high-resolution and high mass accuracy mass spectrometer for more accurate annotation of the protein coding genes in *Anopheles stephensi*. This study clearly demonstrates that mass spectrometry derived proteome data can be used to accurately annotate protein-coding genes and can be effectively used to validate hypothetical proteins. Novel genes identified exclusive to midgut can be further investigated for their role in malaria transmission by gene silencing and expression studies.

Appendix

Appendix Table 1 A: The table includes the molecules that aid or inhibit the oocyst formation in mosquito midgut. The mosquito proteins that aid oocyst formation are known to be agonistic and those that inhibit the oocyst formation are known to be antagonistic in nature.

Agonistic molecules: those that aid *Plasmodium* development in mosquito (oocyst number decreases upon knock-down)				
S. No.	**Protein**	**Protein name**	**VectorBase ID**	**References**
1	ANT	Adenine nucleotide translocator	AGAP006782	*Oliveira, JH et al., 2011.*
2	ArgK	Arginine kinase	AGAP005627	*Jaramillo-Gutierrez, G et al., 2009.* *Brandt, SM et al., 2008.*
3	CACT	Cactus	AGAP007938	*Garver, LS et al., 2009.* *Zou, Z et al., 2011.*
4	Caspar	Caspar	AGAP006473	*Garver, LS et al., 2009.* *Garver, LS et al., 2012.*
5	CLIPA2	CLIP-domain serine protease subfamily A 2	AGAP011790	*Volz, J et al., 2006.*
6	CLIPA5	CLIP-domain serine protease subfamily A 5	AGAP011787	*Volz, J et al., 2006.*
7	CLIPA8	CLIP-domain serine protease subfamily A 8	AGAP010731	*Volz, J et al., 2006.*
8	CLIPB3	CLIP-domain serine protease subfamily B 3	AGAP003249	*Volz, J et al., 2006.*
9	CLSP2	Serine protease, putative	AAEL011616	*Shin, SW et al., 2011.*
10	CP	F-actin capping protein	AGAP007864	*Vlachou, D et al., 2005.*
11	CPR	Cytochrome P450 reductase	AGAP000500	*Félix, RC & Silveira, H, 2011.*
12	CTL4	C-type lectin 4	AGAP005335	*Volz, J et al., 2006.*
13	CTLMA2	CTL mannose binding 2	AGAP005334	*Cohuet, A et al., 2006.*
14	DUOX	Dual oxidase	AGAP009978	*Kumar, S et al., 2010.*
15	ESP	*Anopheles gambiae* epithelial serine protease	AGAP010240	*Rodrigues, J et al., 2012.*
16	Gelsolin	Gelsolin	AGAP011369	*Vlachou, D et al., 2005.*
17	GPRNPY3	Putative neuropeptide Y receptor	AGAP012378	*Mendes, AM et al., 2011.*
18	GSTT1	Glutathione-S-transferase theta-1	AGAP000761	*Jaramillo-Gutierrez, G et al., 2009.*
19	IMPer	Immunomodulatory peroxidase	AGAP013327	*Kumar, S et al., 2010.*
20	LANB2	Laminin gamma 1	AGAP007629	*Arrighi, RB et al., 2005.*
21	LYSC-1	Lysozyme c-1	AGAP007347	*Lapcharoen, P et al., 2012.* *Kajla, MK et al., 2011.*
22	OXR1	Oxidation Resistance gene 1	AGAP001751	*Jaramillo-Gutierrez, G et al., 2010.* *Brandt, SM et al., 2008.*
23	OXT1	Peptide-O-xylosyltransferase 1	AGAP005811	*Dinglasan, RR et al., 2007.*
24	PIAS	Protein Inhibitor of Activated STAT	AGAP005031	*Zou, Z et al., 2011.*

S. No.	Protein	Protein name	VectorBase ID	References
25	PRS1	*Plasmodium* responsive salivary 1	AGAP006102	*Chertemps, T et al., 2010.*
26	RFABG	Retinoid and fatty-acid binding glycoprotein, also known as lipophorin or ApoII/I	AGAP001826	*Mendes, AM et al., 2008. Vlachou, D et al., 2005. Rono, MK et al., 2010. Cheon, HM et al., 2006.*
27	SCRBQ2	Croquemort ortholog	AGAP010133	*González-Lázaro, M et al., 2009.*
28	SDR1	Short-chain dehydrogenases/reductases	AGAP002521	*Berois, M et al., 2012.*
29	SRPN2	Serine protease inhibitor 2 (also known as serpin 2)	AGAP006911	*Michel, K et al., 2005. Michel, K et al., 2006.*
30	Vg	Vitellogenin	AGAP004203	*Rono, MK et al., 2010.*
31	AgaP_ AGAP004016	AgaP_ AGAP004016	AGAP004016	*Pinto, SB et al., 2009.*
32	AgaP_ AGAP006914	AgaP_ AGAP006914	AGAP006914	*Pinto, SB et al., 2009.*
33	AgaP_ AGAP013527	AgaP_ AGAP013527	AGAP013527	*Pinto, SB et al., 2009.*
Antagonistic molecules: those that prevent *Plasmodium* development in mosquito (oocyst number increases upon knock-down)				
S. No.	**Protein**	**Protein name**	**VectorBase ID**	**References**
1	APL1B	*Anopheles Plasmodium*-responsive Leucine-rich repeat protein 1B	AGAP007035	*Garver, LS et al., 2012.*
2	APL1C	*Anopheles Plasmodium*-responsive Leucine-rich repeat protein 1C	AGAP007033	*Riehle, MM et al., 2008. Mitri, C et al., 2009. Garver, LS et al., 2012.*
3	APOD	Apo lipoprotein	AGAP002593	*Dong, Y et al., 2006.*
4	ApoLp-III	Apolipophorin-III	AGAP013365	*Gupta, L et al., 2010.*
5	CASPL1	Caspase long class	AGAP011693	*Garver, LS et al., 2012.*
6	Ciboulot	Beta thymosin family	AGAP000235	*Vlachou, D et al., 2005.*
7	CLIP1	CLIP-domain serine protease 1	ENSANGT00000 020158	*Dong, Y et al., 2006.*
8	CLIPB14	CLIP-domain serine protease subfamily B14	AGAP010833	*Volz, J et al., 2005.*
9	CLIPB15	CLIP-domain serine protease subfamily B15	AGAP009844	*Volz, J et al., 2005.*
10	FADD	Fas-Associated Death Domain	AGAP007173	*Garver, LS et al., 2012.*
11	FBN6	Fibrinogen domain immunolectin 6	AGAP011231	*Dong, Y et al., 2009.*

S. No.	Protein	Protein name	VectorBase ID	References
12	FBN8	Fibrinogen domain immunolectin 8	AGAP011223	*Dong, Y et al., 2006.* *Dong, Y et al., 2009.*
13	FBN9	Fibrinogen domain immunolectin 9	AGAP011197	*Dong, Y et al., 2006.* *Garver, LS et al., 2009.* *Dong, Y et al., 2009.*
14	FBN39	Fibrinogen domain immunolectin 39	AGAP000806	*Dong, Y et al., 2006.* *Dong, Y et al., 2009.*
15	GAM1	Gambicin	AGAP008645	*Dong, Y et al., 2006.*
16	GNBPA2	Gram-negative bacteria binding protein A2	AGAP012409	*Warr, E et al., 2008.*
17	GNBPB1	Gram-negative bacteria binding protein B1	AGAP004455	*Dong, Y et al., 2006.*
18	GNBPB3	Gram-negative bacteria binding protein B3	AGAP002799	*Warr, E et al., 2008.*
19	GNBPB4	Gram-negative bacteria binding protein B4	AGAP002796	*Warr, E et al., 2008.*
20	GPGRP1	Gastrin/bombesin receptor 1	AGAP011452	*Mendes, AM et al., 2011.*
21	GPGRP2	Gastrin/bombesin receptor 2	AGAP003631	*Mendes, AM et al., 2011.*
22	GSTT2	Glutathione-S-transferase theta-2	AGAP000888	*Jaramillo-Gutierrez, G et al., 2009.*
23	IKK-gamma	Inhibitor of kappa B kinase gamma	AGAP005933	*Garver, LS et al., 2012.*
24	IMD	Immunodeficiency domain	AGAP004959	*Meister, S et al., 2005.* *Garver, LS et al., 2012.*
25	IRSP1	Infection responsive secreted peptide 1	AGAP006421	*Dong, Y et al., 2006.*
26	IRSP5	Infection responsive secreted peptide 5	AGAP000151	*Dong, Y et al., 2006.*
27	JNK	Jun N-terminal Kinase	AGAP009461	*Jaramillo-Gutierrez, G et al., 2011.*
28	LRIM1	Leucine-Rich Immune Molecule 1	AGAP006348	*Habtewold, T et al., 2008.* *Jaramillo-Gutierrez, G et al., 2009.* *Riehle, MM et al., 2006.* *Povelones, M et al., 2009.* *Osta, MA et al., 2004.* *Cohuet, A et al., 2006.* *Garver, LS et al., 2012.*
29	LRIM2	Leucine-Rich Immune Molecule 2 also known as APL2 or LRRD7	AGAP005693	*Habtewold, T et al., 2008.* *Jaramillo-Gutierrez, G et al., 2009.* *Riehle, MM et al., 2006.* *Povelones, M et al., 2009.* *Dong, Y et al., 2006.* *Garver, LS et al., 2009.* *Garver, LS et al., 2012.*
30	MC1	Mitochondrial carrier 1	AGAP001297	*Gonçalves, RL et al., 2012.*
31	MDL1	MD2-like receptor 1	AGAP012352	*Dong, Y et al., 2006.*
32	PGRPLC	PGN Recognition Protein LC	AGAP005203	*Meister, S et al., 2009.*

S. No.	Protein	Protein name	VectorBase ID	References
33	REL2	Relish	AGAP006747	*Meister, S et al., 2005.* *Garver, LS et al., 2012.*
34	Sp SNAKE like	Serine protease SNAKE like	AGAP012946	*Harris, C et al., 2010.*
35	SRPN6	Serine protease inhibitor 6 (also known as serpin 6)	AGAP009212	*Abraham, EG et al., 2005.*
36	STAT	Signal Transducers and Activators of Transcription	AGAP000099	*Bahia, AC et al., 2011.*
37	TEP1	Thioester-containing protein 1	AGAP010815	*Dong, Y et al., 2006.* *Habtewold, T et al., 2008.* *Jaramillo-Gutierrez, G et al., 2009.* *Garver, LS et al., 2009.* *Dong, Y et al., 2009.* *Rono, MK et al., 2010.* *Molina-Cruz, A et al., 2012.* *Garver, LS et al., 2012.* *Blandin, S et al., 2004.*
38	Tetrasp.	Tetraspanin	AGAP005233	*Jaramillo-Gutierrez, G et al., 2009.* *Brandt, SM et al., 2008.*
39	WASP	Wiskott-Aldrich syndrome protein	AGAP001081	*Vlachou, D et al., 2005.* *Mendes, AM et al., 2008.*
40	LL3	LITAF-like 3	AGAP009053	*Smith, RC et al., 2012.*
41	Kto	Kohtalo (also known as Med12 or TRAP230)	AGAP002523	*Chen, Y et al., 2012.*
42	Skd	Skuld (also known as Med13 or TRAP240)	AGAP006436	*Chen, Y et al., 2012.*
43	AgaP_ AGAP005761	AgaP_ AGAP005761	AGAP005761	*Pinto, SB et al., 2009.*
44	AgaP_ AGAP009642	AgaP_ AGAP009642	AGAP009642	*Pinto, SB et al., 2009.*
45	AgaP_ AGAP010325	AgaP_ AGAP010325	AGAP010325	*Pinto, SB et al., 2009.*

Appendix Table 2 A: List of proteins over-expressed by ≥ 2.0 fold in female midgut as compared to male midgut tissue.

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
1	ASTEI10274-PA	44.163	F-box and leucine-rich repeat protein 6
2	ASTEI00645-PA	13.695	Calcineurin beta subunit
3	ASTEI00772-PA	11.125	Bromodomain-containing protein 8
4	ASTEI10988-PA	10.243	Piggybac-derived 2
5	ASTEI09499-PA	9.948	Hypothetical protein
6	ASTEI03820-PA	9.809	Hypothetical protein
7	ASTEI06278-PA	8.575	Dnaj-like protein subfamily C
8	ASTEI07509-PA	7.638	Hypothetical protein
9	ASTEI03729-PA	7.346	Hypothetical protein
10	ASTEI02407-PA	7.269	Sodium-independent sulfate anion transporter
11	ASTEI11298-PA	7.207	Hypothetical protein
12	ASTEI08015-PA	7.206	Hypothetical protein
13	ASTEI025761-PA	6.677	Nuclear transport factor 2
14	ASTEI05236-PA	6.619	DNA-directed RNA polymerase II subunit RPB11
15	ASTEI08033-PA	6.599	Serine protease
16	ASTEI04077-PA	6.364	Elongation factor 1 alpha-like protein
17	ASTEI06639-PA	6.058	Hypothetical protein
18	ASTEI08034-PA	6.023	Serine protease
19	ASTEI09575-PA	6.004	Hypothetical protein
20	ASTEI11463-PA	6.003	Hypothetical protein
21	ASTEI05455-PA	5.985	Hypothetical protein
22	ASTEI03153-PA	5.625	Hypothetical protein
23	ASTEI01570-PA	5.463	Anaphase-promoting complex subunit 3
24	ASTEI11423-PA	5.389	Hypothetical protein
25	ASTEI09278-PA	5.364	Ccdc1Hypothetical protein9a
26	ASTEI02774-PA	5.364	Pyridoxine kinase
27	ASTEI07669-PA	5.358	Hypothetical protein
28	ASTEI03928-PA	5.356	Hypothetical protein
29	ASTEI02547-PA	5.296	Hypothetical protein
30	ASTEI10087-PA	5.260	Hypothetical protein
31	ASTEI08929-PA	5.207	N-myc downstream regulated
32	ASTEI10850-PA	5.149	Mitochondrial ribosomal protein L4Hypothetical protein
33	ASTEI11289-PA	4.992	Hypothetical protein
34	ASTEI04335-PA	4.990	Lipase 3
35	ASTEI06369-PA	4.985	Tumor susceptibility 1Hypothetical protein1 protein
36	ASTEI03226-PA	4.979	DNA excision repair protein ERCC-4
37	ASTEI08359-PA	4.958	Hypothetical protein
38	ASTEI04073-PA	4.934	Trna nucleotidyltransferase (CCA-adding enzyme)
39	ASTEI06230-PA	4.917	Cytochrome P45Hypothetical protein (CYP4H24)

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
40	ASTEI10642-PA	4.916	Hypothetical protein
41	ASTEI05414-PA	4.847	U3 small nucleolar RNA-associated protein 5
42	ASTEI00678-PA	4.830	Hypothetical protein
43	ASTEI05273-PA	4.826	Hypothetical protein
44	ASTEI00941-PA	4.766	Hypothetical protein
45	ASTEI09370-PA	4.765	Nucleotide exchange factor SIL1
46	ASTEI04369-PA	4.754	Hypothetical protein
47	ASTEI09414-PA	4.714	Hypothetical protein
48	ASTEI11180-PA	4.705	Ionotropic receptor IR75d
49	ASTEI06444-PA	4.704	IQ motif and SEC7 domain-containing protein
50	ASTEI06153-PA	4.700	ADP-ribosylation factor
51	ASTEI10796-PA	4.638	Multiprotein-bridging factor 1
52	ASTEI04703-PA	4.584	DNA-directed RNA polymerases I and III subunit RPAC2
53	ASTEI07672-PA	4.577	5Hypothetical proteins ribosomal protein L4
54	ASTEI07979-PA	4.538	Poly(ADP-ribose) glycohydrolase ARH3
55	ASTEI08885-PA	4.533	Hypothetical protein
56	ASTEI06988-PA	4.518	Oxygen-dependent protoporphyrinogen oxidase
57	ASTEI04974-PA	4.515	Hypothetical protein
58	ASTEI03054-PA	4.477	GSG7 salivary protein
59	ASTEI07506-PA	4.451	Hypothetical protein
60	ASTEI07375-PA	4.413	Excitatory amino acid transporter 3
61	ASTEI10094-PA	4.365	Mrna-capping enzyme
62	ASTEI07024-PA	4.357	Helix-loop-helix transcription factor
63	ASTEI03866-PA	4.324	Dynein intermediate chain 2, axonemal
64	ASTEI09597-PA	4.256	U4/U6 small nuclear ribonucleoprotein SNU13
65	ASTEI00465-PA	4.252	COP9 signalosome complex subunit 5
66	ASTEI11251-PA	4.238	Aminopeptidase N
67	ASTEI10276-PA	4.168	Hypothetical protein
68	ASTEI03028-PA	4.156	Hypothetical protein
69	ASTEI10866-PA	4.155	Metaxin-2
70	ASTEI04562-PA	4.153	Hypothetical protein
71	ASTEI05085-PA	4.116	Nucleolar protein 4
72	ASTEI05011-PA	4.113	Aquaporin (AQP1)
73	ASTEI01417-PA	4.091	Ras-related protein Rab-1A
74	ASTEI01273-PA	4.043	Hypothetical protein
75	ASTEI05179-PA	4.031	TIM21-like protein, mitochondrial precursor
76	ASTEI11212-PA	3.979	Hypothetical protein
77	ASTEI09036-PA	3.978	Short caspase (CASPS6)
78	ASTEI11037-PA	3.967	ATP-dependent RNA helicase DDX31/DBP7
79	ASTEI04208-PA	3.962	Alkaline phosphatase
80	ASTEI10881-PA	3.961	Hypothetical protein
81	ASTEI02680-PA	3.948	Eukaryotic translation initiation factor 3 subunit J
82	ASTEI03324-PA	3.934	Hypothetical protein
83	ASTEI08379-PA	3.905	Hypothetical protein
84	ASTEI09364-PA	3.890	Sugar transporter ERD6-like 7

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
85	ASTEI08095-PA	3.878	Hypothetical protein
86	ASTEI01304-PA	3.850	Solute carrier family 2 (facilitated glucose transporter), member 4
87	ASTEI08491-PA	3.848	DNA replication licensing factor MCM7
88	ASTEI05365-PA	3.837	WD repeat-containing protein
89	ASTEI08374-PA	3.824	Anoctamin 5
90	ASTEI04993-PA	3.823	Female reproductive tract protease GLEANR_896
91	ASTEI05012-PA	3.822	Aquaporin
92	ASTEI03526-PA	3.818	Hypothetical protein
93	ASTEI09168-PA	3.812	Hypothetical protein
94	ASTEI11004-PA	3.802	Hypothetical protein
95	ASTEI10981-PA	3.756	Hypothetical protein
96	ASTEI01155-PA	3.726	Hypothetical protein
97	ASTEI05683-PA	3.656	Ribosomal RNA-processing protein 7
98	ASTEI05935-PA	3.649	Protease m1 zinc metalloprotease
99	ASTEI10714-PA	3.625	ATP-dependent RNA helicase DDX18/HAS1
100	ASTEI00395-PA	3.597	Hypothetical protein
101	ASTEI10564-PA	3.583	Hypothetical protein
102	ASTEI06575-PA	3.582	Hypothetical protein
103	ASTEI00402-PA	3.577	Hypothetical protein
104	ASTEI07797-PA	3.555	Cytochrome P45Hypothetical protein
105	ASTEI10823-PA	3.555	Hypothetical protein
106	ASTEI10262-PA	3.517	Chitinase-3-like protein 1
107	ASTEI00972-PA	3.498	Component of oligomeric golgi complex 3
108	ASTEI01523-PA	3.481	Hypothetical protein
109	ASTEI04739-PA	3.475	Glutaryl-coa dehydrogenase
110	ASTEI10269-PA	3.462	Cyanogenic beta-glucosidase
111	ASTEI10377-PA	3.458	Hypothetical protein
112	ASTEI05709-PA	3.456	Triosephosphate isomerase (TIM)
113	ASTEI08588-PA	3.447	6Hypothetical proteins ribosomal protein l27a
114	ASTEI05597-PA	3.409	Hypothetical protein
115	ASTEI02266-PA	3.400	Hypothetical protein
116	ASTEI03798-PA	3.398	Endoplasmic reticulum lectin 1
117	ASTEI06636-PA	3.394	Mannose-P-dolichol utilization defect 1
118	ASTEI06165-PA	3.390	Hypothetical protein
119	ASTEI10851-PA	3.388	Hypothetical protein
120	ASTEI02827-PA	3.363	Ribosome biogenesis protein BMS1
121	ASTEI08341-PA	3.362	Gamma-glutamylcyclotransferase
122	ASTEI09761-PA	3.357	39S ribosomal protein L13, mitochondrial
123	ASTEI00788-PA	3.352	Hypothetical protein
124	ASTEI01559-PA	3.340	Trna acetyltransferase TAN1
125	ASTEI03481-PA	3.328	Coactosin-like protein
126	ASTEI07458-PA	3.319	Aminopeptidase N
127	ASTEI07872-PA	3.306	Hypothetical protein
128	ASTEI10368-PA	3.295	Ribose-phosphate pyrophosphokinase
129	ASTEI09723-PA	3.286	Ribosome biogenesis protein

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
130	ASTEI04374-PA	3.286	Transportin-3
131	ASTEI06122-PA	3.266	Tetratricopeptide repeat protein 4
132	ASTEI00856-PA	3.260	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex assembly factor 3
133	ASTEI07693-PA	3.259	Xanthine dehydrogenase/oxidase
134	ASTEI09899-PA	3.250	Alpha-aminoadipic semialdehyde synthase
135	ASTEI02661-PA	3.224	Mitotic spindle assembly checkpoint protein MAD1
136	ASTEI09615-PA	3.213	Class B Scavenger Receptor (CD36 domain).
137	ASTEI00982-PA	3.197	Hypothetical protein
138	ASTEI02465-PA	3.158	Hypothetical protein
139	ASTEI06668-PA	3.138	Nardilysin
140	ASTEI00567-PA	3.130	Yip1 interacting factor homolog B
141	ASTEI08150-PA	3.116	Hypothetical protein
142	ASTEI06798-PA	3.091	Exosome complex component RRP4
143	ASTEI02737-PA	3.085	Hypothetical protein
144	ASTEI04542-PA	3.083	Histone deacetylase 1/2
145	ASTEI10887-PA	3.081	Juvenile hormone esterase binding protein
146	ASTEI11069-PA	3.049	Hypothetical protein
147	ASTEI00080-PA	3.046	Ubiquitin carboxyl-terminal hydrolase 8
148	ASTEI09598-PA	3.045	La-related protein 7
149	ASTEI02958-PA	3.042	4Hypothetical proteins ribosomal protein S5
150	ASTEI10180-PA	3.027	Nucleosome assembly protein 1-like 1
151	ASTEI09679-PA	3.026	Lipoyltransferase 1
152	ASTEI07863-PA	3.013	Proliferating cell nuclear antigen
153	ASTEI00896-PA	3.009	Hypothetical protein
154	ASTEI07834-PA	3.005	Bifunctional pyrimidine biosynthesis protein (pyrabcn)
155	ASTEI04732-PA	2.996	28S ribosomal protein S7, mitochondrial precursor
156	ASTEI03367-PA	2.991	Hypothetical protein
157	ASTEI10352-PA	2.983	Protease m1 zinc metalloprotease
158	ASTEI01137-PA	2.974	Hypothetical protein
159	ASTEI00273-PA	2.965	Replication factor C subunit 2/4
160	ASTEI04731-PA	2.951	28S ribosomal protein S7, mitochondrial precursor
161	ASTEI05655-PA	2.943	Molecular chaperone grpe
162	ASTEI06516-PA	2.939	Hypothetical protein
163	ASTEI09644-PA	2.931	Hypothetical protein
164	ASTEI02664-PA	2.918	Hypothetical protein
165	ASTEI11020-PA	2.913	Ubiquitin carboxyl-terminal hydrolase 16/45
166	ASTEI07320-PA	2.905	U3 small nucleolar RNA-associated protein 1Hypothetical protein
167	ASTEI04234-PA	2.888	Acetyl-coa C-acetyltransferase
168	ASTEI10342-PA	2.885	Ribose 5-phosphate isomerase A
169	ASTEI09868-PA	2.884	Hypothetical protein
170	ASTEI06363-PA	2.872	Hypothetical protein
171	ASTEI04209-PA	2.866	Hypothetical protein

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
172	ASTEI00863-PA	2.863	Hypothetical protein
173	ASTEI04886-PA	2.858	Hypothetical protein
174	ASTEI07111-PA	2.851	Kinesin-like protein unc-1Hypothetical protein4
175	ASTEI02150-PA	2.841	Hypothetical protein
176	ASTEI08656-PA	2.834	Hypothetical protein
177	ASTEI09354-PA	2.811	ATP-dependent RNA helicase SUPV3L1, mitochondrial
178	ASTEI03251-PA	2.810	Alpha-centractin
179	ASTEI04391-PA	2.800	DDRGK domain-containing protein 1
180	ASTEI01351-PA	2.796	Hypothetical protein
181	ASTEI10310-PA	2.795	Cytochrome P45Hypothetical protein
182	ASTEI09250-PA	2.794	Exosome complex exonuclease DIS3/RRP44
183	ASTEI09336-PA	2.793	4Hypothetical proteins ribosomal protein s3a
184	ASTEI11122-PA	2.786	6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase
185	ASTEI095322-PA	2.781	Beta-1,4-N-acetylglucosaminyltransferase
186	ASTEI06482-PA	2.767	Translocation protein SEC62
187	ASTEI00644-PA	2.756	Hypothetical protein
188	ASTEI06289-PA	2.754	Programmed cell death protein 4
189	ASTEI06517-PA	2.753	NADH dehydrogenase (ubiquinone) Fe-S protein 5
190	ASTEI10035-PA	2.752	Serine/threonine-protein kinase OSR1/STK39
191	ASTEI11219-PA	2.736	Transducin (beta)-like 1
192	ASTEI02615-PA	2.731	Mitochondrial ribosomal protein, S34
193	ASTEI00997-PA	2.729	Peptide-O-fucosyltransferase
194	ASTEI03205-PA	2.725	ATP-binding cassette, subfamily F, member 2
195	ASTEI04191-PA	2.724	Eukaryotic translation initiation factor 5A
196	ASTEI03565-PA	2.722	Beta-mannosidase
197	ASTEI07746-PA	2.707	4Hypothetical proteins ribosomal protein S3Hypothetical protein
198	ASTEI01945-PA	2.706	Lupus La protein
199	ASTEI10789-PA	2.705	Karyopherin (Importin) alpha 6
200	ASTEI05711-PA	2.698	Cystinosin
201	ASTEI01393-PA	2.697	FK5Hypothetical protein6-binding nuclear protein
202	ASTEI04948-PA	2.696	28S ribosomal protein S24, mitochondrial
203	ASTEI02424-PA	2.689	Eukaryotic translation initiation factor 3 subunit M
204	ASTEI06017-PA	2.686	Ribosomal RNA methyltransferase Nop2
205	ASTEI08277-PA	2.684	Stress-induced-phosphoprotein 1
206	ASTEI05891-PA	2.683	Cytosolic Fe-S cluster assembly factor NUBP1 homolog
207	ASTEI09878-PA	2.683	Hypothetical protein
208	ASTEI08292-PA	2.681	Hypothetical protein
209	ASTEI05877-PA	2.675	Hypothetical protein
210	ASTEI00843-PA	2.673	Hypothetical protein
211	ASTEI01422-PA	2.672	Hypothetical protein

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
212	ASTEI04774-PA	2.668	DNA-directed RNA polymerases I and III subunit RPAC1
213	ASTEI06343-PA	2.666	Hypothetical protein
214	ASTEI07459-PA	2.662	Aminopeptidase (APN5)
215	ASTEI06938-PA	2.653	Mitochondrial inner membrane translocase subunit (TIM17)
216	ASTEI02725-PA	2.652	IMD pathway signalling Fas-Associated Death Domain (FADD).
217	ASTEI06098-PA	2.645	Actin-related protein 1Hypothetical protein
218	ASTEI07661-PA	2.643	Uncharacterized protein kiaahypothetical protein152
219	ASTEI03253-PA	2.642	Deoxyhypusine monooxygenase
220	ASTEI00152-PA	2.640	Hypothetical protein
221	ASTEI03941-PA	2.633	Hypothetical protein
222	ASTEI06078-PA	2.632	4Hypothetical proteins ribosomal protein S26
223	ASTEI08042-PA	2.632	Chitin synthase
224	ASTEI10215-PA	2.628	ADP-ribosylation factor-like protein 2
225	ASTEI00791-PA	2.627	Peptide deformylase
226	ASTEI07007-PA	2.627	4Hypothetical proteins ribosomal protein S1Hypothetical protein
227	ASTEI10434-PA	2.621	Mitochondrial ribosomal protein L53
228	ASTEI02604-PA	2.613	Importin-4
229	ASTEI08458-PA	2.612	Anaphase-promoting complex subunit 2
230	ASTEI01665-PA	2.612	Sodium/potassium-transporting atpase subunit beta
231	ASTEI03036-PA	2.603	THO complex subunit 2
232	ASTEI06261-PA	2.592	Squid
233	ASTEI08944-PA	2.591	39S ribosomal protein L45
234	ASTEI08739-PA	2.589	Protease m1 zinc metalloprotease
235	ASTEI11520-PA	2.588	6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase
236	ASTEI02835-PA	2.587	N-alpha-acetyltransferase 35, natc auxiliary subunit
237	ASTEI04671-PA	2.580	Regulator of ribosome biosynthesis
238	ASTEI07898-PA	2.579	6Hypothetical proteins ribosomal protein L23
239	ASTEI04808-PA	2.576	U3 small nucleolar RNA-associated protein 21
240	ASTEI09249-PA	2.574	Protein KTI12
241	ASTEI02287-PA	2.567	Hypothetical protein
242	ASTEI09035-PA	2.564	6Hypothetical proteins ribosomal protein L15
243	ASTEI05409-PA	2.562	Periodic tryptophan protein 2
244	ASTEI11138-PA	2.560	Hypothetical protein
245	ASTEI05904-PA	2.559	Hypothetical protein
246	ASTEI06116-PA	2.558	Hypothetical protein
247	ASTEI07116-PA	2.557	Activator of 9Hypothetical protein kda heat shock protein atpase
248	ASTEI06857-PA	2.554	Hypothetical protein
249	ASTEI10999-PA	2.553	Hypothetical protein
250	ASTEI06641-PA	2.552	Hypothetical protein
251	ASTEI09335-PA	2.549	Hypothetical protein

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
252	ASTEI05379-PA	2.546	Mitochondrial 39S ribosomal protein L39
253	ASTEI01468-PA	2.541	Chaperone dnaj protein
254	ASTEI10331-PA	2.541	Leucine-rich repeat-containing protein 47
255	ASTEI05276-PA	2.539	Pre-rrna-processing protein TSR3
256	ASTEI06441-PA	2.539	Hypothetical protein
257	ASTEI02040-PA	2.534	Protein PTCD3-like protein, mitochondrial
258	ASTEI03378-PA	2.531	Hypothetical protein
259	ASTEI09638-PA	2.531	Hypothetical protein
260	ASTEI01478-PA	2.528	Hypothetical protein
261	ASTEI09034-PA	2.527	Fibrinogen
262	ASTEI00511-PA	2.526	Ribosomal RNA-processing protein 12
263	ASTEI09716-PA	2.523	Ribosome biogenesis protein WDR12 homolog
264	ASTEI00450-PA	2.522	Sodium/potassium-transporting atpase subunit alpha
265	ASTEI03371-PA	2.512	Hypothetical protein
266	ASTEI00057-PA	2.507	Bystin
267	ASTEI01343-PA	2.506	Midasin
268	ASTEI05833-PA	2.505	Phosphatidylinositol-4-phosphate 3-kinase
269	ASTEI03543-PA	2.503	Interferon-related developmental regulator 1
270	ASTEI07754-PA	2.503	28S ribosomal protein s18c, mitochondrial precursor
271	ASTEI00099-PA	2.502	E3 ubiquitin-protein ligase EDD1
272	ASTEI06703-PA	2.499	39S ribosomal protein L23, mitochondrial
273	ASTEI05845-PA	2.496	5' nucleotidase, ecto
274	ASTEI01428-PA	2.495	28S ribosomal protein S22, mitochondrial
275	ASTEI00899-PA	2.495	4Hypothetical proteins ribosomal protein S2
276	ASTEI02879-PA	2.493	Hypothetical protein
277	ASTEI06890-PA	2.493	Hypothetical protein
278	ASTEI02495-PA	2.492	H/ACA ribonucleoprotein complex subunit 2
279	ASTEI05032-PA	2.492	N-myc downstream regulated
280	ASTEI02990-PA	2.491	Leucyl-trna synthetase
281	ASTEI02603-PA	2.483	ATP-dependent RNA helicase DDX51/DBP6
282	ASTEI04676-PA	2.481	Kinesin family member 18/19
283	ASTEI05140-PA	2.479	Hypothetical protein
284	ASTEI07854-PA	2.479	Translocase of inner mitochondrial membrane 23 homolog
285	ASTEI02545-PA	2.479	Origin recognition complex subunit 4
286	ASTEI01711-PA	2.476	E3 UFM1-protein ligase 1 homolog
287	ASTEI07951-PA	2.471	Methylmalonate-semialdehyde dehydrogenase (acylating), mitochondrial
288	ASTEI01997-PA	2.471	Hypothetical protein
289	ASTEI02048-PA	2.460	Aminoacyl trna synthase complex-interacting multifunctional protein 2
290	ASTEI00139-PA	2.456	Hypothetical protein
291	ASTEI02998-PA	2.454	Long form D7 salivary protein
292	ASTEI06995-PA	2.452	Allatostatin 1
293	ASTEI09495-PA	2.452	Hypothetical protein
294	ASTEI06979-PA	2.452	Signal recognition particle subunit SRP54

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
295	ASTEI07254-PA	2.449	Proliferation-associated 2G4
296	ASTEI02544-PA	2.448	Hypothetical protein
297	ASTEI05996-PA	2.443	Hypothetical protein
298	ASTEI00477-PA	2.434	Putative metabotropic glutamate receptor 4
299	ASTEI00562-PA	2.434	NADH dehydrogenase (ubiquinone) flavoprotein 2
300	ASTEI04102-PA	2.433	S-formylglutathione hydrolase
301	ASTEI04007-PA	2.426	Ribosome production factor 1
302	ASTEI05583-PA	2.426	ATP-dependent RNA helicase DDX27
303	ASTEI08957-PA	2.424	Mitochondrial ribosomal protein L37
304	ASTEI08981-PA	2.422	Low molecular weight phosphotyrosine protein phosphatase
305	ASTEI00846-PA	2.419	EF-hand domain (C-terminal) containing 2
306	ASTEI04420-PA	2.419	C-factor
307	ASTEI09324-PA	2.417	Eukaryotic translation elongation factor 1 alpha 1
308	ASTEI09153-PA	2.417	Hypothetical protein
309	ASTEI00732-PA	2.415	Hypothetical protein
310	ASTEI06507-PA	2.415	Cdk5rap3
311	ASTEI04412-PA	2.408	PDGF- and VEGF-related factor 1
312	ASTEI02064-PA	2.402	Ribosome biogenesis protein MAK21
313	ASTEI06526-PA	2.401	Voltage-dependent anion-selective channel protein 2
314	ASTEI09229-PA	2.398	Cysteinyl-trna synthetase
315	ASTEI03969-PA	2.389	Coiled-coil domain-containing protein 47
316	ASTEI03335-PA	2.387	Hypothetical protein
317	ASTEI04450-PA	2.386	Rhythmically expressed gene 2 protein
318	ASTEI08347-PA	2.381	Ribosome biogenesis protein NSA1
319	ASTEI10066-PA	2.378	Ribosome biogenesis protein BRX1
320	ASTEI00898-PA	2.376	Importin beta-3
321	ASTEI01001-PA	2.374	Alkylated DNA repair protein alkb homolog 7
322	ASTEI10005-PA	2.373	Hypothetical protein
323	ASTEI02374-PA	2.369	Structure-specific recognition protein 1
324	ASTEI03904-PA	2.368	Large subunit ribosomal protein L7/L12
325	ASTEI06140-PA	2.366	Hypothetical protein
326	ASTEI04302-PA	2.359	Hypothetical protein
327	ASTEI01883-PA	2.354	Peptidyl-prolyl cis-trans isomerase-like 4
328	ASTEI00231-PA	2.352	Sulfide:quinone oxidoreductase
329	ASTEI05168-PA	2.351	Very-long-chain enoyl-coa reductase
330	ASTEI08476-PA	2.344	Serine-threonine kinase receptor-associated protein
331	ASTEI11151-PA	2.344	Hypothetical protein
332	ASTEI01682-PA	2.342	ATP-dependent DNA helicase PIF1
333	ASTEI01149-PA	2.342	Hypothetical protein
334	ASTEI10488-PA	2.341	C-type lectin (CTL)
335	ASTEI04409-PA	2.341	AMP deaminase
336	ASTEI02288-PA	2.340	Insulin receptor
337	ASTEI08978-PA	2.339	Hypothetical protein

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
338	ASTEI11141-PA	2.339	Hypothetical protein
339	ASTEI04646-PA	2.337	Trpm
340	ASTEI04889-PA	2.335	4Hypothetical proteins ribosomal protein S25
341	ASTEI08948-PA	2.334	Trna pseudouridine38-4Hypothetical protein synthase
342	ASTEI06937-PA	2.332	Hypothetical protein
343	ASTEI01391-PA	2.331	Hypothetical protein
344	ASTEI01150-PA	2.330	Hypothetical protein
345	ASTEI055272-PA	2.329	Nucleolar protein 56
346	ASTEI02013-PA	2.328	Elongation factor G, mitochondrial
347	ASTEI05457-PA	2.327	Glyoxylate/hydroxypyruvate reductase
348	ASTEI07300-PA	2.321	Cytosolic Fe-S cluster assembly factor agaphypothetical proteinhypothetical protein9hypothetical protein23
349	ASTEI06657-PA	2.320	Ubiquitin-conjugating enzyme E2 L3
350	ASTEI00047-PA	2.320	Protein phosphatase 3, regulatory subunit
351	ASTEI06894-PA	2.319	Nucleoside-diphosphate kinase
352	ASTEI06775-PA	2.318	Rrna 2'-O-methyltransferase fibrillarin
353	ASTEI10732-PA	2.318	Fructose-bisphosphate aldolase, class I
354	ASTEI05710-PA	2.317	Hypothetical protein
355	ASTEI07744-PA	2.316	Xenotropic and polytropic retrovirus receptor 1
356	ASTEI06997-PA	2.314	Mitogen-activated protein kinase kinase kinase 5
357	ASTEI06736-PA	2.313	4Hypothetical proteins ribosomal protein S18
358	ASTEI06447-PA	2.313	4Hypothetical proteins ribosomal protein S16
359	ASTEI04339-PA	2.312	Exosome complex component RRP45
360	ASTEI07858-PA	2.309	Hypothetical protein
361	ASTEI11090-PA	2.306	Signal recognition particle receptor subunit alpha
362	ASTEI06745-PA	2.305	Solute carrier family 15 member 1
363	ASTEI05605-PA	2.304	Guanylate kinase
364	ASTEI00587-PA	2.303	Hypothetical protein
365	ASTEI09265-PA	2.302	Spermine synthase
366	ASTEI09497-PA	2.300	ER degradation enhancer, mannosidase alpha-like 3
367	ASTEI00445-PA	2.300	Niemann-pick type c-2
368	ASTEI00240-PA	2.296	FK5Hypothetical protein6-binding protein 2
369	ASTEI09951-PA	2.293	DNA replication licensing factor MCM3
370	ASTEI09308-PA	2.291	Kelch domain-containing protein 1Hypothetical protein
371	ASTEI07454-PA	2.291	Glycerol-3-phosphate O-acyltransferase 1/2
372	ASTEI06457-PA	2.290	Casein kinase II subunit beta
373	ASTEI05566-PA	2.289	Dihydroorotate dehydrogenase
374	ASTEI09443-PA	2.289	Structural maintenance of chromosome 3 (chondroitin sulfate proteoglycan 6)
375	ASTEI02849-PA	2.289	Cohesin complex subunit SA-1/2
376	ASTEI07629-PA	2.283	Serine/threonine-protein kinase/endoribonuclease IRE1
377	ASTEI09207-PA	2.283	Translation initiation factor 2D

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
378	ASTEI09449-PA	2.283	ADP-dependent glucokinase
379	ASTEI02995-PA	2.282	Hypothetical protein
380	ASTEI04359-PA	2.282	Protein arginine methyltransferase rmtb
381	ASTEI05576-PA	2.280	4Hypothetical proteins ribosomal protein S12
382	ASTEI07743-PA	2.279	39S ribosomal protein L38, mitochondrial
383	ASTEI11160-PA	2.266	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase
384	ASTEI09964-PA	2.263	U3 small nucleolar RNA-associated protein 12
385	ASTEI00968-PA	2.261	Hypothetical protein
386	ASTEI01713-PA	2.258	6Hypothetical proteins ribosomal protein LP1
387	ASTEI05130-PA	2.256	39S ribosomal protein L15, mitochondrial precursor
388	ASTEI01450-PA	2.254	Thioredoxin peroxidase
389	ASTEI05898-PA	2.253	ATP-binding cassette, subfamily F, member 3
390	ASTEI05439-PA	2.249	4Hypothetical proteins ribosomal protein S23
391	ASTEI07343-PA	2.246	CCR4-NOT transcription complex subunit 1Hypothetical protein
392	ASTEI05334-PA	2.241	4Hypothetical proteins ribosomal protein S2Hypothetical protein
393	ASTEI03834-PA	2.241	6Hypothetical proteins ribosomal protein L9
394	ASTEI09926-PA	2.240	Protein-tyrosine phosphatase
395	ASTEI03295-PA	2.238	Mesencephalic astrocyte-derived neurotrophic factor homolog
396	ASTEI06364-PA	2.235	Hypothetical protein
397	ASTEI01104-PA	2.234	Htra serine peptidase 2
398	ASTEI01140-PA	2.231	Hypothetical protein
399	ASTEI10265-PA	2.227	Venom allergen
400	ASTEI02606-PA	2.227	DNA repair and recombination protein RAD54B
401	ASTEI03618-PA	2.225	Replication factor A1
402	ASTEI03494-PA	2.225	Alanine transaminase
403	ASTEI01943-PA	2.222	Hypothetical protein
404	ASTEI09106-PA	2.218	Carboxypeptidase B
405	ASTEI08754-PA	2.215	39S ribosomal protein L17, mitochondrial
406	ASTEI09959-PA	2.213	Hypothetical protein
407	ASTEI06324-PA	2.208	Glycolipid transfer protein domain-containing protein 1
408	ASTEI00475-PA	2.208	NADH dehydrogenase (ubiquinone) 1 subcomplex unknown 2
409	ASTEI05308-PA	2.208	Poly(U)-binding-splicing factor PUF6Hypothetical protein
410	ASTEI03499-PA	2.207	Dynactin 4
411	ASTEI06569-PA	2.205	Hypothetical protein
412	ASTEI00540-PA	2.205	Protein phosphatase PTC7 homolog fig
413	ASTEI03980-PA	2.203	4Hypothetical proteins ribosomal protein S11
414	ASTEI00289-PA	2.203	Glutamyl aminopeptidase
415	ASTEI03972-PA	2.200	Uncharacterized protein KIAA2Hypothetical protein13-like protein
416	ASTEI02676-PA	2.198	Armadillo repeat-containing protein 6
417	ASTEI04327-PA	2.198	Hypothetical protein

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
418	ASTEI08071-PA	2.192	Phosphatidylserine decarboxylase
419	ASTEI08159-PA	2.192	Arginyl-trna synthetase
420	ASTEI01924-PA	2.188	Threonyl-trna synthetase
421	ASTEI08652-PA	2.178	Cyclin G-associated kinase
422	ASTEI00431-PA	2.177	Hypothetical protein
423	ASTEI051271-PA	2.175	Short caspase (CASPS4)
424	ASTEI05342-PA	2.175	Hypothetical protein
425	ASTEI05097-PA	2.174	Hypothetical protein
426	ASTEI06899-PA	2.174	Glutamate-rich WD repeat-containing protein 1
427	ASTEI03455-PA	2.171	Choline-phosphate cytidylyltransferase
428	ASTEI03397-PA	2.170	Haloacid dehalogenase-like hydrolase domain-containing protein 1A
429	ASTEI05482-PA	2.170	Proline synthetase co-transcribed homolog
430	ASTEI08383-PA	2.170	Exosome complex component RRP42
431	ASTEI10920-PA	2.170	Hypothetical protein
432	ASTEI03409-PA	2.169	Hypothetical protein
433	ASTEI05218-PA	2.168	Protein transport protein SEC61 subunit alpha
434	ASTEI07544-PA	2.167	Hypothetical protein
435	ASTEI08252-PA	2.163	Hypothetical protein
436	ASTEI08025-PA	2.161	Hypothetical protein
437	ASTEI11137-PA	2.160	Guanine nucleotide-binding protein subunit beta-2-like 1 protein
438	ASTEI02336-PA	2.156	Protein aurora borealis
439	ASTEI09930-PA	2.154	Sorting nexin family member 27 isoform 1
440	ASTEI07914-PA	2.151	Thousand and one amino acid protein kinase
441	ASTEI06125-PA	2.149	Hypothetical protein
442	ASTEI02180-PA	2.148	Aspartyl-trna synthetase
443	ASTEI04256-PA	2.147	Glucosamine--fructose-6-phosphate aminotransferase (isomerizing)
444	ASTEI09692-PA	2.145	Hypothetical protein
445	ASTEI10039-PA	2.141	Hypothetical protein
446	ASTEI07218-PA	2.140	Augmenter of liver regeneration
447	ASTEI05736-PA	2.138	Biotin--protein ligase
448	ASTEI03842-PA	2.138	Proto-oncogene C-crk
449	ASTEI01727-PA	2.138	Hypothetical protein
450	ASTEI06177-PA	2.138	Class A Scavenger Receptor (SRCR domain) with C-Type Lectin domain.
451	ASTEI05773-PA	2.138	Hypothetical protein
452	ASTEI02980-PA	2.137	Phosphoglucomutase
453	ASTEI05764-PA	2.137	Golgi apparatus protein 1
454	ASTEI00049-PA	2.135	ADP-ribosylation factor
455	ASTEI04364-PA	2.132	Phosphate carrier, mitochondrial
456	ASTEI03309-PA	2.130	Pyruvate dehydrogenase E1 component subunit alpha
457	ASTEI00666-PA	2.129	Pumilio homology domain family member 6
458	ASTEI04133-PA	2.128	Inhibitor of apoptosis (IAP6)
459	ASTEI03948-PA	2.127	Phospholipid-translocating atpase

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
460	ASTEI10825-PA	2.126	Adaptor-related protein complex 3, delta 1 subunit
461	ASTEI08982-PA	2.125	DNA topoisomerase I
462	ASTEI00509-PA	2.125	Squamous cell carcinoma antigen recognized by T-cells 3
463	ASTEI05355-PA	2.123	Elongation factor 1-beta
464	ASTEI10371-PA	2.122	4Hypothetical proteins ribosomal protein S17
465	ASTEI11142-PA	2.121	Translationally-controlled tumor protein homolog
466	ASTEI09305-PA	2.120	Kelch domain containing 4
467	ASTEI09485-PA	2.118	Enoyl-coa hydratase domain-containing protein 3, mitochondrial
468	ASTEI06424-PA	2.118	Hypothetical protein
469	ASTEI07800-PA	2.118	Hypothetical protein
470	ASTEI06385-PA	2.117	Clock
471	ASTEI10327-PA	2.116	Envelysin
472	ASTEI00563-PA	2.116	Nucleolar protein 58
473	ASTEI10568-PA	2.115	Inositol hexakisphosphate/diphosphoinositol-pentakisphosphate kinase
474	ASTEI04334-PA	2.114	EVI5-like protein
475	ASTEI06806-PA	2.112	Hypothetical protein
476	ASTEI01958-PA	2.112	Nuclear pore complex protein Nup214
477	ASTEI05314-PA	2.111	Solute carrier family 7 (cationic amino acid transporter, y+ system)
478	ASTEI05386-PA	2.110	ATP-dependent RNA helicase DDX46/PRP5
479	ASTEI05909-PA	2.110	Muskelin
480	ASTEI00558-PA	2.110	A-kinase anchor protein 1, mitochondrial
481	ASTEI05666-PA	2.110	Prefoldin
482	ASTEI02878-PA	2.108	Hypothetical protein
483	ASTEI10246-PA	2.108	CCR4-NOT transcription complex subunit 3
484	ASTEI00683-PA	2.106	Glutaredoxin
485	ASTEI10579-PA	2.105	Hypothetical protein
486	ASTEI03931-PA	2.102	Hypothetical protein
487	ASTEI05713-PA	2.101	Procollagen-lysine,2-oxoglutarate 5-dioxygenase, invertebrate
488	ASTEI03793-PA	2.099	Sulfotransferase (Sult)
489	ASTEI08950-PA	2.099	Mitochondrial ribosomal protein S18A
490	ASTEI03215-PA	2.097	Ribosomal RNA-processing protein 1
491	ASTEI09014-PA	2.096	Porphobilinogen synthase
492	ASTEI08966-PA	2.096	6Hypothetical proteins ribosomal protein L1Hypothetical protein-2
493	ASTEI06044-PA	2.095	Ubiquitin carboxyl-terminal hydrolase 12/46
494	ASTEI02729-PA	2.093	Hypothetical protein
495	ASTEI00221-PA	2.091	Transcription elongation factor B polypeptide 2
496	ASTEI06826-PA	2.090	Cyclin-dependent kinase 1Hypothetical protein
497	ASTEI08124-PA	2.087	Translation initiation factor 4G
498	ASTEI05332-PA	2.086	Mitochondrial ribosomal protein L3Hypothetical protein
499	ASTEI06434-PA	2.084	CCR4-NOT transcription complex subunit 7/8

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
500	ASTEI00127-PA	2.083	Glycylpeptide N-tetradecanoyltransferase
501	ASTEI09046-PA	2.083	Galectin (GALE3)
502	ASTEI05654-PA	2.082	Pre-mrna-splicing factor ATP-dependent RNA helicase DHX15/PRP43
503	ASTEI08943-PA	2.081	Basic leucine zipper and W2 domain-containing protein 1
504	ASTEI07113-PA	2.079	Manganese-iron (Mn-Fe) superoxide dismutase
505	ASTEI08445-PA	2.079	4Hypothetical proteins ribosomal protein S9
506	ASTEI00446-PA	2.078	Niemann-Pick C2 protein
507	ASTEI05897-PA	2.075	Hypothetical protein
508	ASTEI07680-PA	2.075	Ralgapb
509	ASTEI10974-PA	2.074	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3
510	ASTEI07105-PA	2.074	Nucleolar complex protein 2
511	ASTEI07532-PA	2.073	Translation initiation factor 6
512	ASTEI05067-PA	2.073	Protein kiaahypothetical protein664 homolog
513	ASTEI06357-PA	2.071	Kynurenine 3-monooxygenase
514	ASTEI06346-PA	2.071	6Hypothetical proteins ribosomal protein l7a
515	ASTEI08378-PA	2.069	Protein crossbronx homolog
516	ASTEI07793-PA	2.069	Calcium binding protein
517	ASTEI09230-PA	2.069	Proteasome component ECM29
518	ASTEI01601-PA	2.069	Hypothetical protein
519	ASTEI04259-PA	2.068	Methylglutaconyl-coa hydratase
520	ASTEI01377-PA	2.068	DNA topoisomerase 2-associated protein PAT1
521	ASTEI00345-PA	2.068	Hypothetical protein
522	ASTEI01594-PA	2.067	Hypothetical protein
523	ASTEI08201-PA	2.064	Hypothetical protein
524	ASTEI02772-PA	2.063	Hypothetical protein
525	ASTEI11312-PA	2.058	Tubulin--tyrosine ligase-like protein 12
526	ASTEI00574-PA	2.058	Signal recognition particle subunit SRP68
527	ASTEI11403-PA	2.056	Ceramide synthetase
528	ASTEI01562-PA	2.056	Protein CLP1 homolog
529	ASTEI07520-PA	2.056	Hypothetical protein
530	ASTEI07753-PA	2.054	Saposin
531	ASTEI00112-PA	2.053	P21-activated kinase 1
532	ASTEI06887-PA	2.053	Pescadillo homolog
533	ASTEI08224-PA	2.052	6Hypothetical protein kda heat shock protein, mitochondrial precursor
534	ASTEI06653-PA	2.050	Nuclear pore complex protein Nup16Hypothetical protein
535	ASTEI04989-PA	2.049	Hypothetical protein
536	ASTEI01448-PA	2.048	Dnaj homolog subfamily C member 2
537	ASTEI00132-PA	2.048	Dnaj homolog subfamily B
538	ASTEI00404-PA	2.047	Isoleucyl-trna synthetase
539	ASTEI01907-PA	2.047	6Hypothetical proteins ribosomal protein L7
540	ASTEI04628-PA	2.046	4Hypothetical proteins ribosomal protein S4
541	ASTEI04048-PA	2.046	Hypothetical protein
542	ASTEI11367-PA	2.045	Tubulin alpha-4A chain

S. No	*An. stephensi* VectorBase ID	Relative fold change female/male midgut	Protein description as per *An. gambiae*
543	ASTEI05230-PA	2.042	SPRY domain-containing SOCS box protein 3
544	ASTEI06748-PA	2.039	Alanine-glyoxylate aminotransferase
545	ASTEI09104-PA	2.039	Carboxypeptidase B
546	ASTEI04740-PA	2.038	Glutaryl-coa dehydrogenase
547	ASTEI08818-PA	2.037	Lysyl-trna synthetase, class II
548	ASTEI05112-PA	2.037	Hypothetical protein
549	ASTEI05210-PA	2.035	Fatty acid synthase, animal type
550	ASTEI07603-PA	2.031	Protein phosphatase 5
551	ASTEI05533-PA	2.030	Cell cycle progression
552	ASTEI04300-PA	2.030	5'-3' exoribonuclease 2
553	ASTEI08573-PA	2.030	6Hypothetical proteins ribosomal protein L27
554	ASTEI04461-PA	2.027	Hypothetical protein
555	ASTEI01406-PA	2.027	Hypothetical protein
556	ASTEI10055-PA	2.023	DEAD box polypeptide 5
557	ASTEI05117-PA	2.023	Mitochondrial ribosomal protein S31
558	ASTEI07219-PA	2.022	Cytochrome c oxidase subunit vib polypeptide 1
559	ASTEI10877-PA	2.022	Bromodomain and WD repeat domain containing protein 1/3
560	ASTEI03617-PA	2.022	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase
561	ASTEI00769-PA	2.021	Proline-, glutamic acid- and leucine-rich protein 1
562	ASTEI02556-PA	2.019	Nucleolar GTP-binding protein
563	ASTEI01161-PA	2.019	MAX-like protein X
564	ASTEI01959-PA	2.019	Nuclear pore complex protein Nup214
565	ASTEI03797-PA	2.019	AAA family atpase
566	ASTEI10332-PA	2.018	Large subunit gtpase 1
567	ASTEI05767-PA	2.017	Acetyl-coa C-acetyltransferase
568	ASTEI05080-PA	2.016	Hypothetical protein
569	ASTEI06663-PA	2.013	Eukaryotic peptide chain release factor subunit
570	ASTEI06358-PA	2.013	4Hypothetical proteins ribosomal protein S13
571	ASTEI08952-PA	2.011	Hypothetical protein
572	ASTEI05639-PA	2.001	Inosine-5'-monophosphate dehydrogenase
573	ASTEI10278-PA	2.001	Hypothetical protein
574	ASTEI09007-PA	2.001	Molecular chaperone dnak
575	ASTEI09182-PA	2.000	Nonspecific nucleoside hydrolase

Appendix Table 2 B: List of proteins over-expressed by ≥ 2.0 fold in male midgut as compared to female midgut tissue.

S. No.	*An. stephensi* VectorBase ID	Relative fold change male/female midgut	Protein description as per *An. gambiae*
1	ASTEI08968-PA	2.003	Solute carrier family 35 (GDP-fucose transporter), member C1
2	ASTEI11530-PA	2.005	Hypothetical protein
3	ASTEI10622-PA	2.009	Hypothetical protein
4	ASTEI01510-PA	2.027	Hypothetical protein
5	ASTEI08959-PA	2.029	Hypothetical protein
6	ASTEI09257-PA	2.036	Solute carrier family 35
7	ASTEI017980-PA	2.037	Sensory appendage protein
8	ASTEI10017-PA	2.037	Hypothetical protein
9	ASTEI04026-PA	2.048	Glutathione transferase microsomal
10	ASTEI00212-PA	2.050	Carnitine O-octanoyltransferase
11	ASTEI00850-PA	2.051	Cytochrome P45Hypothetical protein
12	ASTEI08091-PA	2.053	Pyruvate dehydrogenase E1 component subunit alpha
13	ASTEI08908-PA	2.062	Hypothetical protein
14	ASTEI02745-PA	2.069	Hypothetical protein
15	ASTEI07208-PA	2.069	Atrophin-1 interacting protein 5 (WW domain containing E3 ubiquitin protein li
16	ASTEI03781-PA	2.073	Zinc finger protein 271
17	ASTEI09117-PA	2.077	Glucosyl/glucuronosyl transferases
18	ASTEI07417-PA	2.086	Hypothetical protein
19	ASTEI00091-PA	2.086	Glutathione S-transferase zeta class
20	ASTEI08006-PA	2.090	14.5 kda salivary peptide
21	ASTEI04814-PA	2.104	Gamma-interferon-inducible lysosomal thiol reductase
22	ASTEI10271-PA	2.108	ATP-binding cassette, sub-family C (CFTR/MRP), member 4
23	ASTEI06229-PA	2.109	Hypothetical protein
24	ASTEI00232-PA	2.112	Insulin-related peptide binding protein
25	ASTEI06691-PA	2.113	Aquaporin
26	ASTEI05189-PA	2.115	AMP-binding domain protein
27	ASTEI04331-PA	2.117	ADP,ATP carrier protein 2
28	ASTEI00793-PA	2.126	Hypothetical protein
29	ASTEI09560-PA	2.133	Hypothetical protein
30	ASTEI08229-PA	2.143	Hypothetical protein
31	ASTEI04618-PA	2.146	Hypothetical protein
32	ASTEI09660-PA	2.157	Hypothetical protein
33	ASTEI07227-PA	2.157	Hypothetical protein
34	ASTEI00901-PA	2.157	Hypothetical protein
35	ASTEI02217-PA	2.157	Uncharacterized protein C45G9.7
36	ASTEI00095-PA	2.157	Hypothetical protein
37	ASTEI07705-PA	2.161	Leucine aminopeptidase-like protein
38	ASTEI01906-PA	2.162	Gamma-glutamyltranspeptidase
39	ASTEI02949-PA	2.171	Spohypothetical proteinb-associated GTP-binding protein
40	ASTEI01099-PA	2.174	Thymosin

S. No.	*An. stephensi* VectorBase ID	Relative fold change male/female midgut	Protein description as per *An. gambiae*
41	ASTEI03188-PA	2.185	4-coumarate:coa ligase isoform 2
42	ASTEI02744-PA	2.192	Hypothetical protein
43	ASTEI05604-PA	2.199	Hypothetical protein
44	ASTEI052402-PA	2.209	Dimeric dihydrodiol dehydrogenase
45	ASTEI11668-PA	2.215	Hypothetical protein
46	ASTEI04206-PA	2.233	Hypothetical protein
47	ASTEI01168-PA	2.234	Hypothetical protein
48	ASTEI04482-PA	2.239	Cyanogenic beta-glucosidase
49	ASTEI01333-PA	2.241	L-gulonate 3-dehydrogenase
50	ASTEI04338-PA	2.246	Troponin T, fast skeletal muscle
51	ASTEI09992-PA	2.253	Troponin C isoform 4"
52	ASTEI00004-PA	2.256	26S proteasome regulatory subunit T2
53	ASTEI02539-PA	2.268	Iduronate 2-sulfatase
54	ASTEI01563-PA	2.270	N-acetylneuraminate lyase
55	ASTEI03577-PA	2.278	Neutral ceramidase
56	ASTEI04333-PA	2.289	Glutamine synthetase
57	ASTEI02766-PA	2.292	Trna (guanine(37)-N1)-methyltransferase
58	ASTEI05197-PA	2.295	AMP dependent ligase
59	ASTEI10123-PA	2.296	Hypothetical protein
60	ASTEI02077-PA	2.297	Cuticular protein RR-1 family (CPR15)
61	ASTEI10605-PA	2.304	Hypothetical protein
62	ASTEI08207-PA	2.304	Hypothetical protein
63	ASTEI04368-PA	2.307	D-xylulose reductase A
64	ASTEI06054-PA	2.315	Hypothetical protein
65	ASTEI06314-PA	2.351	Hypothetical protein
66	ASTEI05393-PA	2.357	Leucine-rich immune protein (Coil-less)
67	ASTEI07236-PA	2.362	Trypsin-alpha
68	ASTEI01181-PA	2.363	Vesicle-associated membrane protein 7
69	ASTEI04852-PA	2.369	Hypothetical protein
70	ASTEI02403-PA	2.392	Muscle LIM protein at 84B
71	ASTEI06525-PA	2.409	Short caspase (CASPS8)
72	ASTEI01312-PA	2.414	Hypothetical protein
73	ASTEI03817-PA	2.421	Hypothetical protein
74	ASTEI02311-PA	2.426	Maltase
75	ASTEI018591-PA	2.428	U6 snrna-associated Sm-like protein lsm7
76	ASTEI07420-PA	2.429	Terribly reduced optic lobes, isoform B
77	ASTEI07429-PA	2.437	Envelysin
78	ASTEI08186-PA	2.438	Hypothetical protein
79	ASTEI06808-PA	2.449	Hypothetical protein
80	ASTEI11406-PA	2.466	Annexin B9
81	ASTEI04064-PA	2.476	Apolipoprotein D
82	ASTEI01952-PA	2.477	Maltase
83	ASTEI00143-PA	2.491	V-type H+-transporting atpase subunit G
84	ASTEI10573-PA	2.493	2-oxoisovalerate dehydrogenase E2 component (dihydrolipoyl transacylase)
85	ASTEI02736-PA	2.502	Alpha-crystallin chain B
86	ASTEI03074-PA	2.531	Actin, cytoplasmic

S. No.	*An. stephensi* VectorBase ID	Relative fold change male/female midgut	Protein description as per *An. gambiae*
87	ASTEI04951-PA	2.551	Mitochondrial atpase inhibitor
88	ASTEI08678-PA	2.559	Hypothetical protein
89	ASTEI01723-PA	2.568	Choline/ethanolamine kinase
90	ASTEI05065-PA	2.568	Platelet-activating factor acetylhydrolase IB subunit alpha
91	ASTEI04742-PA	2.579	Hypothetical protein
92	ASTEI05980-PA	2.589	Potassium voltage-gated channel, subfamily H (eag-related), member 4
93	ASTEI10558-PA	2.595	Eupolytin
94	ASTEI10813-PA	2.614	Transmembrane emp24 domain-containing protein 1Hypothetical protein precursor
95	ASTEI00815-PA	2.623	Peptidyl-trna hydrolase, PTH2 family
96	ASTEI02537-PA	2.626	Pigment dispersing hormone
97	ASTEI04340-PA	2.629	N-acetylglucosamine-6-phosphate deacetylase
98	ASTEI10663-PA	2.638	Prophenoloxidase
99	ASTEI11577-PA	2.652	Cytochrome P45Hypothetical protein
100	ASTEI03849-PA	2.661	Hypothetical protein
101	ASTEI07235-PA	2.665	Prss3
102	ASTEI07221-PA	2.677	FAS-associated factor 1
103	ASTEI03263-PA	2.692	Hypothetical protein
104	ASTEI10004-PA	2.696	Adenylate cyclase 3
105	ASTEI03057-PA	2.744	Hypothetical protein
106	ASTEI02412-PA	2.753	Cytochrome P45Hypothetical protein (CYP9J5)
107	ASTEI02654-PA	2.759	Oatp58Dc
108	ASTEI08865-PA	2.761	Solute carrier family 12 (sodium/potassium/chloride transporter), member 2
109	ASTEI02820-PA	2.761	Coiled-coil domain-containing protein 6
110	ASTEI09440-PA	2.766	Hypothetical protein
111	ASTEI09149-PA	2.769	Ethanolaminephosphotransferase
112	ASTEI04056-PA	2.779	Hypothetical protein
113	ASTEI11637-PA	2.786	Glutathione S-transferase epsilon class
114	ASTEI09815-PA	2.805	Vacuolar protein sorting-associated protein 16-like protein
115	ASTEI10890-PA	2.811	Proton-coupled amino acid transporter
116	ASTEI01575-PA	2.818	Tubulointerstitial nephritis antigen
117	ASTEI04027-PA	2.828	Dihydrodiol dehydrogenase / D-xylose 1-dehydrogenase (NADP)
118	ASTEI04799-PA	2.835	Oviductin
119	ASTEI02529-PA	2.843	Hypothetical protein
120	ASTEI02148-PA	2.847	Queuine trna-ribosyltransferase subunit QTRTD1 homolog
121	ASTEI06732-PA	2.852	Solute carrier family 17 (anion/sugar transporter), member 5
122	ASTEI10601-PA	2.861	Uridine phosphorylase
123	ASTEI07604-PA	2.891	Protein phosphatase 5
124	ASTEI06605-PA	2.931	Cytochrome P45Hypothetical protein

S. No.	*An. stephensi* VectorBase ID	Relative fold change male/female midgut	Protein description as per *An. gambiae*
125	ASTEI10972-PA	2.955	3-glucan binding protein (bgbp)
126	ASTEI02650-PA	2.961	Hypothetical protein
127	ASTEI06608-PA	2.961	Cytochrome P45Hypothetical protein
128	ASTEI11456-PA	3.035	Hypothetical protein
129	ASTEI00312-PA	3.037	Xeroderma pigmentosum group C-complementing protein
130	ASTEI10872-PA	3.046	Phosphatidate phosphatase
131	ASTEI02571-PA	3.058	Leucine-rich immune protein (Short)
132	ASTEI04057-PA	3.067	Hypothetical protein
133	ASTEI01211-PA	3.100	Cytochrome c oxidase subunit 6a, mitochrondrial
134	ASTEI09466-PA	3.106	Limbic system-associated membrane protein
135	ASTEI07811-PA	3.173	Endoplasmic reticulum protein
136	ASTEI04311-PA	3.174	Hypothetical protein
137	ASTEI02610-PA	3.179	Tubulin polymerization-promoting protein
138	ASTEI02597-PA	3.203	Hypothetical protein
139	ASTEI07112-PA	3.215	Hypothetical protein
140	ASTEI08052-PA	3.217	Hypothetical protein
141	ASTEI02644-PA	3.263	Peroxisomal 3,2-trans-enoyl-coa isomerase
142	ASTEI05261-PA	3.268	Ubiquitin thioesterase CYLD
143	ASTEI01311-PA	3.287	Hypothetical protein
144	ASTEI10751-PA	3.302	Troponin C
145	ASTEI02758-PA	3.307	Ganglioside-induced differentiation-associated-protein 1
146	ASTEI05223-PA	3.351	Glutathione S-transferase epsilon class
147	ASTEI00766-PA	3.381	Synapse-associated protein
148	ASTEI01662-PA	3.391	Regucalcin protein
149	ASTEI06003-PA	3.413	Solute carrier family 2, facilitated glucose transporter member 3
150	ASTEI01072-PA	3.419	Cytochrome P45Hypothetical protein
151	ASTEI08061-PA	3.424	Cleavage and polyadenylation specificity factor subunit 3
152	ASTEI06135-PA	3.430	Hypothetical protein
153	ASTEI09063-PA	3.456	Ser/Thr protein phosphatase/nucleotidase
154	ASTEI11776-PA	3.459	Hypothetical protein
155	ASTEI00413-PA	3.476	Solute carrier family 31 (copper transporter), member 1
156	ASTEI01231-PA	3.651	Amino acid transporter protein
157	ASTEI09807-PA	3.662	Hypothetical protein
158	ASTEI08181-PA	3.698	Fatty acyl-coa reductase
159	ASTEI10161-PA	3.769	Hypothetical protein
160	ASTEI00353-PA	3.787	Dynein heavy chain, axonemal
161	ASTEI01741-PA	3.931	Fas-associated factor
162	ASTEI04034-PA	3.953	Ankyrin
163	ASTEI00680-PA	3.973	Solute carrier family 23 (nucleobase transporter)
164	ASTEI10118-PA	3.980	Hypothetical protein

S. No.	*An. stephensi* VectorBase ID	Relative fold change male/female midgut	Protein description as per *An. gambiae*
165	ASTEI07367-PA	4.223	4Hypothetical proteins ribosomal protein S21
166	ASTEI05843-PA	4.472	Cytochrome P45Hypothetical protein (CYP6AF1)
167	ASTEI07626-PA	4.483	Glutathione S-transferase delta class
168	ASTEI09665-PA	4.811	Hypothetical protein
169	ASTEI06325-PA	5.008	Pyrimidine nucleoside transport protein
170	ASTEI04241-PA	5.024	Hypothetical protein
171	ASTEI09681-PA	5.102	FK5Hypothetical protein6-binding protein 1
172	ASTEI04365-PA	5.238	Phosphate carrier, mitochondrial
173	ASTEI09441-PA	5.754	Transcription factor AP-1
174	ASTEI11712-PA	5.829	Innexin inx2
175	ASTEI06317-PA	6.112	Cuticular protein RR-1 family (CPR24)
176	ASTEI05983-PA	6.472	Potassium voltage-gated channel, subfamily H (eag-related), member 4
177	ASTEI08483-PA	6.694	Hypothetical protein
178	ASTEI10959-PA	6.846	Sodium-coupled monocarboxylate transporter 1
179	ASTEI09506-PA	9.460	Translocon-associated protein subunit gamma
180	ASTEI08704-PA	9.645	Hypothetical protein
181	ASTEI10790-PA	11.219	Hypothetical protein
182	ASTEI07019-PA	18.422	Hypothetical protein
183	ASTEI06544-PA	19.453	Gustatory receptor

Appendix Table 3 A: Lists of proteins identified to be over-expressed by ≥ 1.5 fold in blood fed midgut tissue as compared to sugar fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
1	ASTEI04832-PA	11.653	Cathepsin B precursor
2	ASTEI02220-PA	7.042	Hypothetical protein
3	ASTEI07890-PA	5.767	Protein G12
4	ASTEI01388-PA	5.703	Bardet-Biedl syndrome 1 protein
5	ASTEI11169-PA	5.301	Vitellogenin
6	ASTEI00275-PA	4.808	Sterol carrier protein 2 variant 2
7	ASTEI02154-PA	4.680	Alpha-crystallin B chain
8	ASTEI07261-PA	4.277	Charged multivesicular body protein 5
9	ASTEI01945-PA	4.205	Lupus La protein
10	ASTEI10313-PA	3.921	Cellular retinaldehyde-binding protein
11	ASTEI11274-PA	3.731	Probable fatty acid-binding protein
12	ASTEI03362-PA	3.579	Hypothetical protein
13	ASTEI09210-PA	3.451	Antizyme inhibitor 1
14	ASTEI08889-PA	3.318	Hypothetical protein
15	ASTEI11239-PA	3.282	Fructose-1,6-bisphosphatase I
16	ASTEI08949-PA	3.277	Hypothetical protein
17	ASTEI11437-PA	3.205	Clathrin light chain
18	ASTEI01399-PA	3.133	Methylenetetrahydrofolate reductase (NADPH)
19	ASTEI03112-PA	3.099	Hypothetical protein
20	ASTEI09982-PA	2.856	Ubiquitin carboxyl-terminal hydrolase 14
21	ASTEI09051-PA	2.779	Hypothetical protein
22	ASTEI01563-PA	2.768	N-acetylneuraminate lyase
23	ASTEI01419-PA	2.758	Vitellogenic carboxypeptidase-like protein
24	ASTEI09914-PA	2.727	Hypothetical protein
25	ASTEI04974-PA	2.690	Hypothetical protein
26	ASTEI00384-PA	2.655	Hypothetical protein
27	ASTEI03737-PA	2.629	Bis(5'-nucleosidyl)-tetraphosphatase
28	ASTEI07379-PA	2.612	S-adenosylmethionine synthetase
29	ASTEI05141-PA	2.571	Hypothetical protein
30	ASTEI00613-PA	2.539	Phosphoenolpyruvate carboxykinase (GTP)
31	ASTEI10621-PA	2.502	Alpha-tocopherol transfer protein-like protein
32	ASTEI11547-PA	2.488	High mobility group protein B1
33	ASTEI02362-PA	2.480	Echinoderm microtubule associated protein like 1
34	ASTEI06252-PA	2.442	Benzodiazapine receptor
35	ASTEI07893-PA	2.438	Valacyclovir hydrolase
36	ASTEI07584-PA	2.434	Cytoplasmic actin
37	ASTEI06523-PA	2.433	Hypothetical protein
38	ASTEI00143-PA	2.408	V-type H+-transporting atpase subunit G

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
39	ASTEI01565-PA	2.398	Syntaxin 1A
40	ASTEI09203-PA	2.394	Acyl-coa oxidase
41	ASTEI03295-PA	2.372	Mesencephalic astrocyte-derived neurotrophic factor homolog
42	ASTEI04803-PA	2.334	Angiotensin-converting enzyme 9
43	ASTEI07289-PA	2.320	Hypothetical protein
44	ASTEI05005-PA	2.311	D-3-phosphoglycerate dehydrogenase
45	ASTEI00054-PA	2.280	Epidermal retinal dehydrogenase 2
46	ASTEI11456-PA	2.278	Thioredoxin
47	ASTEI00706-PA	2.272	Ferritin heavy chain
48	ASTEI04243-PA	2.272	Hypothetical protein
49	ASTEI00381-PA	2.267	C-4 methylsterol oxidase
50	ASTEI06288-PA	2.248	Hypothetical protein
51	ASTEI04169-PA	2.244	Vanin-like protein 1
52	ASTEI06380-PA	2.238	Gamma-glutamyl hydrolase
53	ASTEI03600-PA	2.212	Septin 4
54	ASTEI07547-PA	2.199	Optic atrophy 3 family protein
55	ASTEI11067-PA	2.184	Hypothetical protein
56	ASTEI04293-PA	2.173	Sterol carrier protein-2
57	ASTEI03122-PA	2.167	Sphinganine-1-phosphate aldolase
58	ASTEI02400-PA	2.165	Ornithine decarboxylase
59	ASTEI06909-PA	2.147	Palmitoyl-protein thioesterase
60	ASTEI05763-PA	2.145	Hypothetical protein
61	ASTEI06976-PA	2.144	Phosphotransferase LOC123688
62	ASTEI05529-PA	2.143	ATP-binding cassette, subfamily D (ALD), member 2
63	ASTEI07496-PA	2.133	Bleomycin hydrolase
64	ASTEI01105-PA	2.110	N-acetylglucosamine kinase
65	ASTEI01329-PA	2.106	Glucuronosyltransferase
66	ASTEI03226-PA	2.090	DNA excision repair protein ERCC-4
67	ASTEI04134-PA	2.088	Hypothetical protein
68	ASTEI10939-PA	2.044	Rhythmically expressed gene 2 protein
69	ASTEI06878-PA	2.034	Farnesyl diphosphate synthase
70	ASTEI10510-PA	2.024	Solute carrier family 12 (potassium/chloride transporters), member 8
71	ASTEI07409-PA	2.022	Translation initiation factor eif-2B subunit alpha
72	ASTEI01540-PA	2.016	Adenylate kinase 2, mitochondrial
73	ASTEI10756-PA	2.016	Hypothetical protein
74	ASTEI08755-PA	1.993	Cytochrome P450 (CYP4H15)
75	ASTEI07267-PA	1.989	Cytoplasmic actin
76	ASTEI01405-PA	1.986	Hypothetical protein
77	ASTEI04964-PA	1.983	Hepatocyte growth factor-regulated tyrosine kinase substrate
78	ASTEI08132-PA	1.974	Hypothetical protein
79	ASTEI08867-PA	1.973	Cathepsin D
80	ASTEI05578-PA	1.964	Hypothetical protein

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
81	ASTEI03480-PA	1.952	Glutathione S-transferase theta class
82	ASTEI02634-PA	1.951	Isocitrate dehydrogenase
83	ASTEI07648-PA	1.945	Solute carrier family 5 (sodium-coupled monocarboxylate transporter), member 8
84	ASTEI11367-PA	1.943	Tubulin alpha-4A chain
85	ASTEI07600-PA	1.939	Exocyst complex component 5
86	ASTEI07794-PA	1.937	Prostaglandin reductase 1
87	ASTEI09571-PA	1.935	Hydroxysteroid (17-beta) dehydrogenase 4
88	ASTEI01594-PA	1.927	Hypothetical protein
89	ASTEI05179-PA	1.925	TIM21-like protein, mitochondrial precursor
90	ASTEI00464-PA	1.918	Cathepsin F
91	ASTEI07734-PA	1.913	Hypothetical protein
92	ASTEI09386-PA	1.909	S-phase kinase-associated protein 1
93	ASTEI03478-PA	1.904	Palmitoyltransferase
94	ASTEI08938-PA	1.886	Hypothetical protein
95	ASTEI08299-PA	1.884	Carnitine O-palmitoyltransferase 1
96	ASTEI10458-PA	1.881	Hypothetical protein
97	ASTEI01553-PA	1.879	Hypothetical protein
98	ASTEI05856-PA	1.878	Estrogen sulfotransferase
99	ASTEI07390-PA	1.876	ATP-binding cassette sub-family G member 1
100	ASTEI02731-PA	1.876	Lethal(2)denticleless
101	ASTEI00830-PA	1.874	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 7
102	ASTEI08356-PA	1.868	Rhoa activator c11orf59 homolog
103	ASTEI03541-PA	1.867	Hypothetical protein
104	ASTEI09025-PA	1.866	Calmodulin
105	ASTEI05728-PA	1.861	Innexin 2
106	ASTEI07863-PA	1.853	Proliferating cell nuclear antigen
107	ASTEI11264-PA	1.844	Hypothetical protein
108	ASTEI00490-PA	1.841	Solute carrier family 27 (fatty acid transporter), member 1/4
109	ASTEI10759-PA	1.831	Maltase
110	ASTEI00707-PA	1.830	Secreted ferritin G subunit
111	ASTEI02436-PA	1.826	Hypothetical protein
112	ASTEI02767-PA	1.823	Trna (guanine(37)-N1)-methyltransferase
113	ASTEI01517-PA	1.822	Cytosolic nonspecific dipeptidase
114	ASTEI10006-PA	1.820	Peptide chain release factor subunit 3
115	ASTEI01896-PA	1.820	Brefeldin A-inhibited guanine nucleotide-exchange protein
116	ASTEI05925-PA	1.814	V-type H+-transporting atpase subunit G
117	ASTEI07687-PA	1.812	Sterol carrier protein 2
118	ASTEI02048-PA	1.809	Aminoacyl trna synthase complex-interacting multifunctional protein 2
119	ASTEI02586-PA	1.809	Hypothetical protein
120	ASTEI07933-PA	1.808	Epidermal retinal dehydrogenase
121	ASTEI04257-PA	1.807	Phosphomannomutase
122	ASTEI01456-PA	1.802	Hypothetical protein

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
123	ASTEI07021-PA	1.801	Proteasome 26S non-atpase subunit 10
124	ASTEI00367-PA	1.796	Dynein heavy chain 1, cytosolic
125	ASTEI08277-PA	1.776	Stress-induced-phosphoprotein 1
126	ASTEI00067-PA	1.771	Galactokinase
127	ASTEI07838-PA	1.771	Histone H2A
128	ASTEI07135-PA	1.759	Peroxisomal biogenesis factor 12
129	ASTEI10590-PA	1.758	Chloride intracellular channel
130	ASTEI00278-PA	1.758	Hypothetical protein
131	ASTEI00008-PA	1.752	Hypothetical protein
132	ASTEI05818-PA	1.748	Thioredoxin peroxidase
133	ASTEI10783-PA	1.745	Lipid storage droplets surface-binding protein 2
134	ASTEI01359-PA	1.745	Elongation factor 1-alpha
135	ASTEI06711-PA	1.738	Copper-zinc (Cu-Zn) superoxide dismutase
136	ASTEI04251-PA	1.730	Arylsulfatase b
137	ASTEI10330-PA	1.728	Ubiquitin-protein ligase E3 C
138	ASTEI02553-PA	1.726	Hypothetical protein
139	ASTEI09657-PA	1.722	6-phosphogluconolactonase
140	ASTEI00947-PA	1.716	Battenin
141	ASTEI10314-PA	1.715	Cellular retinaldehyde-binding protein
142	ASTEI04927-PA	1.714	5'-3' exoribonuclease 1
143	ASTEI06540-PA	1.713	ATP-binding cassette sub-family G member 4
144	ASTEI01149-PA	1.709	Hypothetical protein
145	ASTEI03010-PA	1.705	Seryl-trna synthetase
146	ASTEI07979-PA	1.703	Poly(ADP-ribose) glycohydrolase ARH3
147	ASTEI08026-PA	1.701	Hypothetical protein
148	ASTEI02649-PA	1.696	Hypothetical protein
149	ASTEI11435-PA	1.695	Karyopherin (Importin) alpha 6
150	ASTEI00366-PA	1.694	Dynein heavy chain 1, cytosolic
151	ASTEI07881-PA	1.692	Microsomal triglyceride transfer protein large subunit
152	ASTEI06446-PA	1.690	20S proteasome subunit beta 2
153	ASTEI051573-PA	1.689	Very-long-chain enoyl-coa reductase
154	ASTEI09047-PA	1.687	Hypothetical protein
155	ASTEI04256-PA	1.685	Glucosamine--fructose-6-phosphate aminotransferase (isomerizing)
156	ASTEI08666-PA	1.685	Filamin
157	ASTEI02789-PA	1.664	Hypothetical protein
158	ASTEI04349-PA	1.662	Hypothetical protein
159	ASTEI02678-PA	1.661	FK506-binding protein 8
160	ASTEI10964-PA	1.659	UNC93-like protein MFSD11
161	ASTEI05231-PA	1.658	Eukaryotic translation initiation factor 3 subunit H
162	ASTEI01004-PA	1.658	Membrane-associated progesterone receptor component 2
163	ASTEI01233-PA	1.655	Hypothetical protein
164	ASTEI01140-PA	1.655	60S ribosomal protein L37
165	ASTEI06609-PA	1.653	Cytochrome P450

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
166	ASTEI03550-PA	1.650	Xaa-Pro aminopeptidase
167	ASTEI03036-PA	1.649	THO complex subunit 2
168	ASTEI03826-PA	1.647	TOLL pathway signalling NF-kappab Relish-like transcription factor
169	ASTEI00221-PA	1.645	Transcription elongation factor B polypeptide 2
170	ASTEI09696-PA	1.644	Hypothetical protein
171	ASTEI01564-PA	1.642	GTP-binding nuclear protein Ran
172	ASTEI10069-PA	1.641	Carboxylesterase beta esterase
173	ASTEI04310-PA	1.637	Methionyl-trna synthetase
174	ASTEI10796-PA	1.627	Multiprotein-bridging factor 1
175	ASTEI09762-PA	1.627	Leukotriene-A4 hydrolase
176	ASTEI01523-PA	1.627	Hypothetical protein
177	ASTEI00578-PA	1.626	Very low-density lipoprotein receptor
178	ASTEI08274-PA	1.625	Intraflagellar transport 80 homolog
179	ASTEI05161-PA	1.622	20S proteasome subunit beta 5
180	ASTEI06954-PA	1.621	Hypothetical protein
181	ASTEI09546-PA	1.617	Osi15
182	ASTEI08611-PA	1.616	Glycerol kinase
183	ASTEI01972-PA	1.614	Glutamine synthetase
184	ASTEI06324-PA	1.610	Glycolipid transfer protein domain-containing protein 1
185	ASTEI05911-PA	1.605	Hypothetical protein
186	ASTEI00352-PA	1.596	Hypothetical protein
187	ASTEI08308-PA	1.595	Cytochrome b5-related
188	ASTEI06492-PA	1.595	Protein kinase, cgmp-dependent
189	ASTEI06894-PA	1.588	Nucleoside-diphosphate kinase
190	ASTEI07418-PA	1.587	Terribly reduced optic lobes, isoform B
191	ASTEI07320-PA	1.584	U3 small nucleolar RNA-associated protein 10
192	ASTEI04147-PA	1.584	Hypothetical protein
193	ASTEI03473-PA	1.582	Aldehyde dehydrogenase family 7 member A1
194	ASTEI05524-PA	1.582	Myosin alkali light chain 1
195	ASTEI06801-PA	1.582	Spermine oxidase
196	ASTEI02575-PA	1.581	Glucosyl/glucuronosyl transferases
197	ASTEI10842-PA	1.577	Hypothetical protein
198	ASTEI01829-PA	1.572	Cellular nucleic acid-binding protein
199	ASTEI05206-PA	1.571	Prolyl oligopeptidase
200	ASTEI03698-PA	1.570	Calreticulin
201	ASTEI05740-PA	1.569	Innexin inx2
202	ASTEI01030-PA	1.568	4a-hydroxytetrahydrobiopterin dehydratase
203	ASTEI08015-PA	1.567	Hypothetical protein
204	ASTEI05314-PA	1.566	Solute carrier family 7 (cationic amino acid transporter, y+ system)
205	ASTEI00491-PA	1.565	Hypothetical protein
206	ASTEI11452-PA	1.562	ATP-binding cassette, subfamily D (ALD), member 3
207	ASTEI10644-PA	1.560	Catalase
208	ASTEI00716-PA	1.557	Chitinase

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
209	ASTEI06534-PA	1.557	Hypothetical protein
210	ASTEI05673-PA	1.554	60S ribosomal protein L11
211	ASTEI03907-PA	1.554	Hypothetical protein
212	ASTEI09992-PA	1.554	Troponin C isoform 4"
213	ASTEI03861-PA	1.551	Transcription elongation factor B, polypeptide 1
214	ASTEI00797-PA	1.543	Cytosol aminopeptidase
215	ASTEI00084-PA	1.543	Hypothetical protein
216	ASTEI04102-PA	1.542	S-formylglutathione hydrolase
217	ASTEI03526-PA	1.542	Hypothetical protein
218	ASTEI05905-PA	1.539	Chromodomain-helicase-DNA-binding protein 4
219	ASTEI00249-PA	1.538	Hypothetical protein
220	ASTEI07805-PA	1.538	Epidermal retinal dehydrogenase
221	ASTEI05380-PA	1.534	Hypothetical protein
222	ASTEI04735-PA	1.533	Glyoxylate/hydroxypyruvate reductase
223	ASTEI00463-PA	1.530	Cathepsin F
224	ASTEI05989-PA	1.529	Serine protease inhibitor (serpin) SRPN10
225	ASTEI08196-PA	1.525	Long-chain acyl-coa synthetase
226	ASTEI08974-PA	1.525	Acid phosphatase
227	ASTEI02802-PA	1.525	Sterile alpha and TIR motif-containing protein 1
228	ASTEI01312-PA	1.525	Hypothetical protein
229	ASTEI09582-PA	1.523	Juvenile hormone epoxide hydrolase
230	ASTEI08941-PA	1.522	Hypothetical protein
231	ASTEI00708-PA	1.521	Ubiquitin associated and SH3 domain-containing protein B
232	ASTEI05234-PA	1.520	Extracellular signal-regulated kinase 1/2
233	ASTEI02082-PA	1.519	Hypothetical protein
234	ASTEI01231-PA	1.519	Amino acid transporter protein
235	ASTEI00540-PA	1.515	Protein phosphatase PTC7 homolog fig
236	ASTEI09088-PA	1.512	Protein phosphatase
237	ASTEI08700-PA	1.508	Aldose 1-epimerase
238	ASTEI00843-PA	1.507	Hypothetical protein
239	ASTEI01533-PA	1.507	Solute carrier family 36 (proton-coupled amino acid transporter)
240	ASTEI10685-PA	1.506	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase
241	ASTEI03761-PA	1.506	Hypothetical protein
242	ASTEI10933-PA	1.504	20S proteasome subunit alpha 3
243	ASTEI09268-PA	1.504	Ribonuclease UK114
244	ASTEI02336-PA	1.503	Protein aurora borealis
245	ASTEI03532-PA	1.502	Dihydropyrimidine dehydrogenase (NADP+)
246	ASTEI10364-PA	1.501	Translation initiation factor 5B
247	ASTEI03368-PA	1.501	Grx3p

Appendix Table 3 B: Lists of proteins identified to be down-regulated by ≤ 0.66 fold in blood fed midgut tissue as compared to sugar fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
1	ASTEI06613-PA	0.659	39S ribosomal protein L24, mitochondrial precursor
2	ASTEI03297-PA	0.659	Serine/threonine-protein phosphatase 6 catalytic subunit
3	ASTEI00584-PA	0.659	Isocitrate dehydrogenase (NAD+)
4	ASTEI00388-PA	0.658	Sorting and assembly machinery component 5Hypothetical protein homolog A
5	ASTEI02576-PA	0.655	Glucosyl/glucuronosyl transferases
6	ASTEI10276-PA	0.655	Hypothetical protein
7	ASTEI09348-PA	0.654	Phyhd1 protein
8	ASTEI09371-PA	0.652	Mesenchymal stem cell protein DSCD75
9	ASTEI03775-PA	0.652	4Hypothetical proteins ribosomal protein S15a
10	ASTEI10045-PA	0.650	Carboxylesterase
11	ASTEI11075-PA	0.648	Phosphatidylserine synthase 1
12	ASTEI08285-PA	0.647	Obelix
13	ASTEI00823-PA	0.646	Solute carrier family 29 (equilibrative nucleoside transporter), member 1/2/3
14	ASTEI01484-PA	0.643	Golgi-associated plant pathogenesis-related protein 1
15	ASTEI09761-PA	0.641	39S ribosomal protein L13, mitochondrial
16	ASTEI05658-PA	0.641	Importin 9
17	ASTEI07792-PA	0.640	Frataxin homolog, mitochondrial
18	ASTEI07996-PA	0.640	Hypothetical protein
19	ASTEI10020-PA	0.638	Fumarylacetoacetate hydrolase domain containing 2A
20	ASTEI00941-PA	0.638	Hypothetical protein
21	ASTEI05654-PA	0.638	Pre-mrna-splicing factor ATP-dependent RNA helicase DHX15/PRP43
22	ASTEI01559-PA	0.636	Trna acetyltransferase TAN1
23	ASTEI02289-PA	0.636	Insulin receptor
24	ASTEI02110-PA	0.635	Phospholipid scramblase 1
25	ASTEI08034-PA	0.635	Serine protease
26	ASTEI01448-PA	0.633	Dnaj homolog subfamily C member 2
27	ASTEI09499-PA	0.633	Hypothetical protein
28	ASTEI06038-PA	0.633	ATP-binding cassette sub-family A member 7
29	ASTEI09930-PA	0.633	Sorting nexin family member 27 isoform 1
30	ASTEI10350-PA	0.632	Galectin (GALE7)
31	ASTEI09365-PA	0.632	Solute carrier family 2, facilitated glucose transporter member 8
32	ASTEI04993-PA	0.632	Female reproductive tract protease GLEANR_896
33	ASTEI04540-PA	0.631	Adenosine 3'-phospho 5'-phosphosulfate transporter 2
34	ASTEI04331-PA	0.631	ADP,ATP carrier protein 2

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
35	ASTEI09324-PA	0.629	Eukaryotic translation elongation factor 1 alpha 1
36	ASTEI11298-PA	0.627	Hypothetical protein
37	ASTEI05196-PA	0.626	Hypothetical protein
38	ASTEI05274-PA	0.625	WW domain-containing oxidoreductase
39	ASTEI10589-PA	0.624	Hypothetical protein
40	ASTEI01662-PA	0.623	Regucalcin protein
41	ASTEI04215-PA	0.623	Nuclear GTP-binding protein
42	ASTEI04697-PA	0.622	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 13
43	ASTEI01269-PA	0.622	Inhibitor of apoptosis (IAP1)
44	ASTEI06101-PA	0.621	Calpain, invertebrate
45	ASTEI04058-PA	0.621	Hypothetical protein
46	ASTEI09969-PA	0.618	Polymerase delta-interacting protein 2
47	ASTEI09148-PA	0.618	Protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha
48	ASTEI00770-PA	0.617	Proline-, glutamic acid- and leucine-rich protein 1
49	ASTEI01072-PA	0.617	Cytochrome P45Hypothetical protein
50	ASTEI09013-PA	0.613	Hypothetical protein
51	ASTEI02728-PA	0.611	Lethal(2)denticleless
52	ASTEI06664-PA	0.611	Hypothetical protein
53	ASTEI08273-PA	0.610	Dedicator of cytokinesis protein 7
54	ASTEI08515-PA	0.609	Myo-inositol-1(or 4)-monophosphatase
55	ASTEI08379-PA	0.606	Hypothetical protein
56	ASTEI10409-PA	0.606	Serine/threonine kinase 3
57	ASTEI03591-PA	0.604	Phosphoribosyl pyrophosphate synthetase-associated protein 2
58	ASTEI04229-PA	0.604	Fibulin 2
59	ASTEI10386-PA	0.604	Trna pseudouridine13 synthase
60	ASTEI06332-PA	0.604	Retinol dehydrogenase 13 (all-trans/9-cis)
61	ASTEI03729-PA	0.602	Hypothetical protein
62	ASTEI03360-PA	0.602	Catenin alpha
63	ASTEI00535-PA	0.602	Hypothetical protein
64	ASTEI02538-PA	0.599	Chloride channel
65	ASTEI02332-PA	0.596	Hypothetical protein
66	ASTEI08234-PA	0.596	U3 small nucleolar RNA-associated protein 22
67	ASTEI07242-PA	0.596	Hypothetical protein
68	ASTEI06837-PA	0.595	Translocon-associated protein subunit beta
69	ASTEI02686-PA	0.595	Eukaryotic translation initiation factor 3 subunit I
70	ASTEI05272-PA	0.595	All-trans/9-cis
71	ASTEI02990-PA	0.592	Leucyl-trna synthetase
72	ASTEI00943-PA	0.592	Alpha-1,3-glucosyltransferase
73	ASTEI04482-PA	0.591	Cyanogenic beta-glucosidase
74	ASTEI07532-PA	0.590	Translation initiation factor 6
75	ASTEI01198-PA	0.590	Alpha 1,3-glucosidase
76	ASTEI10268-PA	0.590	Cyanogenic beta-glucosidase
77	ASTEI11181-PA	0.587	Hypothetical protein

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
78	ASTEI07098-PA	0.587	Yellow-x
79	ASTEI04369-PA	0.587	Hypothetical protein
80	ASTEI11355-PA	0.586	Hypothetical protein
81	ASTEI08617-PA	0.585	Hypothetical protein
82	ASTEI03611-PA	0.583	WD4Hypothetical protein repeat-containing protein SMU1
83	ASTEI03935-PA	0.582	Hypothetical protein
84	ASTEI07837-PA	0.581	Sparc
85	ASTEI04784-PA	0.569	Translin-associated factor X
86	ASTEI00985-PA	0.568	3-dehydrosphinganine reductase
87	ASTEI01104-PA	0.568	Htra serine peptidase 2
88	ASTEI04254-PA	0.566	4-nitrophenyl phosphatase
89	ASTEI09265-PA	0.564	Spermine synthase
90	ASTEI04234-PA	0.562	Acetyl-coa C-acetyltransferase
91	ASTEI01543-PA	0.561	Hypothetical protein
92	ASTEI06635-PA	0.560	Mitochondrial ribosomal protein L28
93	ASTEI10573-PA	0.560	2-oxoisovalerate dehydrogenase E2 component (dihydrolipoyl transacylase)
94	ASTEI06668-PA	0.559	Nardilysin
95	ASTEI01134-PA	0.554	Synaptic vesicle protein
96	ASTEI03980-PA	0.553	4Hypothetical proteins ribosomal protein S11
97	ASTEI09457-PA	0.550	Succinate dehydrogenase (ubiquinone) flavoprotein subunit
98	ASTEI09055-PA	0.545	Heat shock 7Hypothetical proteinkda protein 1/8
99	ASTEI09063-PA	0.545	Ser/Thr protein phosphatase/nucleotidase
100	ASTEI09679-PA	0.543	Lipoyltransferase 1
101	ASTEI08424-PA	0.541	Heme peroxidase
102	ASTEI10632-PA	0.539	Hypothetical protein
103	ASTEI08885-PA	0.539	Hypothetical protein
104	ASTEI04175-PA	0.535	Glucose-6-phosphate 1-dehydrogenase
105	ASTEI00139-PA	0.534	Hypothetical protein
106	ASTEI04704-PA	0.534	Hypothetical protein
107	ASTEI08201-PA	0.534	Hypothetical protein
108	ASTEI02883-PA	0.530	Hypothetical protein
109	ASTEI10033-PA	0.528	Serine-type enodpeptidase
110	ASTEI11283-PA	0.527	ATP-binding cassette, subfamily D (ALD), member 3
111	ASTEI00923-PA	0.521	Glycine dehydrogenase
112	ASTEI09044-PA	0.520	Solute carrier family 6 (neurotransmitter transporter), invertebrate
113	ASTEI06887-PA	0.520	Pescadillo homolog
114	ASTEI07480-PA	0.519	TBC1 domain family, member 1Hypothetical protein
115	ASTEI03044-PA	0.519	Trehalose 6-phosphate phosphatase
116	ASTEI01026-PA	0.517	Cytosolic iron-sulfur protein assembly protein Ciao1
117	ASTEI04415-PA	0.517	Hypothetical protein

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
118	ASTEI06556-PA	0.516	Dolichol-phosphate mannosyltransferase
119	ASTEI03928-PA	0.511	Hypothetical protein
120	ASTEI07201-PA	0.511	Tmco1
121	ASTEI04936-PA	0.509	Zinc finger protein
122	ASTEI08011-PA	0.509	Nuclear pore complex protein Nup88
123	ASTEI03099-PA	0.509	Fatty-acid amide hydrolase 2
124	ASTEI00264-PA	0.508	Protein phosphatase 1, catalytic subunit
125	ASTEI04731-PA	0.507	28S ribosomal protein S7, mitochondrial precursor
126	ASTEI04026-PA	0.507	Glutathione transferase microsomal
127	ASTEI07669-PA	0.505	Starch phosphorylase
128	ASTEI02791-PA	0.504	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1
129	ASTEI00291-PA	0.504	Hypothetical protein
130	ASTEI05653-PA	0.504	39S ribosomal protein L14, mitochondrial
131	ASTEI05350-PA	0.501	Mitochondrial import inner membrane translocase subunit Tim8 B
132	ASTEI07285-PA	0.499	Mesoderm development candidate 2
133	ASTEI11301-PA	0.498	Nicotinate phosphoribosyltransferase
134	ASTEI10540-PA	0.498	Vacuolar protein sorting-associated protein 28 homolog
135	ASTEI09305-PA	0.494	Kelch domain containing 4
136	ASTEI03647-PA	0.493	ELMO domain-containing protein 2
137	ASTEI05480-PA	0.485	Hypothetical protein
138	ASTEI00793-PA	0.484	Hypothetical protein
139	ASTEI03534-PA	0.483	Myofilin variant B
140	ASTEI04830-PA	0.481	3-ketoacyl-coa reductase
141	ASTEI08573-PA	0.472	6Hypothetical proteins ribosomal protein L27
142	ASTEI07186-PA	0.472	Transglutaminase
143	ASTEI01960-PA	0.468	Protein cornichon homolog
144	ASTEI00001-PA	0.465	Hypothetical protein
145	ASTEI01940-PA	0.464	Galactokinase
146	ASTEI03758-PA	0.464	Alkylated DNA repair protein alkb homolog 4
147	ASTEI08521-PA	0.462	Hypothetical protein
148	ASTEI04172-PA	0.458	Hypothetical protein
149	ASTEI01953-PA	0.448	Hypothetical protein
150	ASTEI10264-PA	0.441	Serine protease SP24D
151	ASTEI05053-PA	0.431	U6 snrna-associated Sm-like protein lsm8
152	ASTEI04681-PA	0.422	Adrenodoxin
153	ASTEI05390-PA	0.417	Succinate dehydrogenase (ubiquinone) cytochrome b56Hypothetical protein subunit
154	ASTEI05247-PA	0.417	Hypothetical protein
155	ASTEI02421-PA	0.416	6Hypothetical proteins ribosomal protein L18a (partial)
156	ASTEI07803-PA	0.414	Odorant binding protein (OBP9)
157	ASTEI10789-PA	0.410	Karyopherin (Importin) alpha 6
158	ASTEI04041-PA	0.409	Nat13 protein
159	ASTEI05953-PA	0.404	Hypothetical protein

S. No.	*An. stephensi* VectorBase ID	Relative fold change Blood fed/sugar fed	Protein description as per *An. gambiae*
160	ASTEI09202-PA	0.403	Hypothetical protein
161	ASTEI08252-PA	0.402	Hypothetical protein
162	ASTEI05273-PA	0.400	Hypothetical protein
163	ASTEI00772-PA	0.395	Bromodomain-containing protein 8
164	ASTEI08686-PA	0.368	Alkyldihydroxyacetonephosphate synthase
165	ASTEI10176-PA	0.361	Kinase
166	ASTEI10856-PA	0.359	Alpha-1,6-mannosyltransferase
167	ASTEI02014-PA	0.359	Glutaredoxin
168	ASTEI10273-PA	0.354	C-Type Lectin (CTL) - galactose binding
169	ASTEI02284-PA	0.350	Signal recognition particle receptor subunit beta
170	ASTEI02679-PA	0.347	Transcription factor BTF3 homolog 4
171	ASTEI07952-PA	0.347	Serine/threonine-protein phosphatase 4 regulatory subunit 2
172	ASTEI07150-PA	0.333	Hypothetical protein
173	ASTEI02055-PA	0.309	6Hypothetical proteins ribosomal protein L28
174	ASTEI06012-PA	0.306	Hypothetical protein
175	ASTEI06639-PA	0.294	Hypothetical protein
176	ASTEI02077-PA	0.289	Cuticular protein RR-1 family (CPR15)
177	ASTEI06660-PA	0.289	Protease m1 zinc metalloprotease
178	ASTEI04988-PA	0.288	Hypothetical protein
179	ASTEI06899-PA	0.273	Glutamate-rich WD repeat-containing protein 1
180	ASTEI09946-PA	0.263	P37NB protein
181	ASTEI07730-PA	0.257	Mitochondrial import receptor subunit TOM4Hypothetical protein
182	ASTEI01928-PA	0.249	Hypothetical protein
183	ASTEI07582-PA	0.242	Yellow-e
184	ASTEI01168-PA	0.241	Hypothetical protein
185	ASTEI00680-PA	0.236	Solute carrier family 23 (nucleobase transporter)
186	ASTEI10327-PA	0.232	Envelysin
187	ASTEI10165-PA	0.167	Actin, cytoplasmic
188	ASTEI03020-PA	0.092	Protein of unknown function, DUF255 family

Appendix Table 3 C: Lists of proteins identified to be over-expressed by ≥ 1.5 fold in *P. berghei* infected midgut tissue as compared to blood fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Relative fold change *P. berghei* infected /blood fed	Protein description as per *An. gambiae*
1	ASTEI05273-PA	3.347	Hypothetical protein
2	ASTEI01559-PA	3.057	tRNA acetyltransferase TAN1
3	ASTEI01940-PA	3.004	Galactokinase
4	ASTEI10856-PA	2.536	Alpha-1,6-mannosyltransferase
5	ASTEI09546-PA	2.465	Osi15
6	ASTEI01953-PA	2.425	Hypothetical protein
7	ASTEI04704-PA	2.216	Hypothetical protein
8	ASTEI09946-PA	2.195	P37NB protein
9	ASTEI06887-PA	2.109	Pescadillo homolog
10	ASTEI02284-PA	2.064	Signal recognition particle receptor subunit beta
11	ASTEI09037-PA	2.053	Short caspase (CASPS5)
12	ASTEI00038-PA	2.017	Oxidation resistance protein
13	ASTEI09896-PA	1.991	Hypothetical protein
14	ASTEI00680-PA	1.959	Solute carrier family 23 (nucleobase transporter)
15	ASTEI03839-PA	1.902	Lipoma-prefererred partner
16	ASTEI01616-PA	1.883	Mrna turnover protein 4
17	ASTEI04531-PA	1.878	Ubiquinone biosynthesis monooxygenase Coq7
18	ASTEI01928-PA	1.868	Hypothetical protein
19	ASTEI07849-PA	1.847	Hypothetical protein
20	ASTEI04731-PA	1.810	28S ribosomal protein S7, mitochondrial precursor
21	ASTEI01168-PA	1.808	Hypothetical protein
22	ASTEI04784-PA	1.776	Translin-associated factor X
23	ASTEI10355-PA	1.773	Calcium-independent phospholipase A2
24	ASTEI08657-PA	1.753	Annexin B1Hypothetical proteinb
25	ASTEI04988-PA	1.728	Hypothetical protein
26	ASTEI02511-PA	1.722	Hypothetical protein
27	ASTEI01269-PA	1.674	Inhibitor of apoptosis (IAP1)
28	ASTEI04041-PA	1.672	Nat13 protein
29	ASTEI04750-PA	1.665	WAS protein family
30	ASTEI01960-PA	1.658	Protein cornichon homolog
31	ASTEI07055-PA	1.654	Myosin regulatory light chain, invertebrate
32	ASTEI11107-PA	1.645	Hypothetical protein
33	ASTEI04026-PA	1.631	Glutathione transferase microsomal
34	ASTEI00949-PA	1.626	Hypothetical protein
35	ASTEI03020-PA	1.610	Protein of unknown function, DUF255 family
36	ASTEI03517-PA	1.599	Insulin-like growth factor 2 receptor
37	ASTEI09376-PA	1.596	Laccase-like multicopper oxidase 2
38	ASTEI07481-PA	1.573	Phagocyte signaling-impaired protein

S. No.	*An. stephensi* VectorBase ID	Relative fold change *P. berghei* infected /blood fed	Protein description as per *An. gambiae*
39	ASTEI09013-PA	1.564	Hypothetical protein
40	ASTEI08244-PA	1.559	Hypothetical protein
41	ASTEI01601-PA	1.553	Hypothetical protein
42	ASTEI07243-PA	1.551	RNA polymerase II subunit A C-terminal domain phosphatase
43	ASTEI03647-PA	1.547	ELMO domain-containing protein 2
44	ASTEI04482-PA	1.546	Cyanogenic beta-glucosidase
45	ASTEI04297-PA	1.540	28S ribosomal protein S5
46	ASTEI09914-PA	1.536	Hypothetical protein
47	ASTEI09025-PA	1.530	Calmodulin
48	ASTEI07730-PA	1.530	Mitochondrial import receptor subunit TOM4Hypothetical protein
49	ASTEI06660-PA	1.519	Protease m1 zinc metalloprotease
50	ASTEI11436-PA	1.516	Peritrophin 1
51	ASTEI02014-PA	1.510	Glutaredoxin
52	ASTEI03418-PA	1.507	ATP-dependent RNA helicase DHX57
53	ASTEI03360-PA	1.503	Catenin alpha
54	ASTEI10409-PA	1.503	Serine/threonine kinase 3
55	ASTEI02451-PA	1.502	Inorganic phosphate cotransporter

Appendix Table 3 D: Lists of proteins identified to be down-regulated by ≤ 0.66 fold in *P. berghei* infected midgut tissue as compared to blood fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Relative fold change *P. berghei* infected /blood fed	Protein description as per *An. gambiae*
1	ASTEI10327-PA	0.658	Envelysin
2	ASTEI04871-PA	0.657	CDGSH iron sulfur domain-containing protein 2 homolog
3	ASTEI04951-PA	0.657	Mitochondrial atpase inhibitor
4	ASTEI048280-PA	0.656	Hypothetical protein
5	ASTEI08379-PA	0.656	Hypothetical protein
6	ASTEI08277-PA	0.655	Stress-induced-phosphoprotein 1
7	ASTEI10403-PA	0.655	Axonemal dynein intermediate chain inner arm i1
8	ASTEI02622-PA	0.654	Molecular chaperone htpg
9	ASTEI071955-PA	0.652	ATP-dependent RNA helicase UAP56/SUB2
10	ASTEI01948-PA	0.652	Ubiquinol-cytochrome c reductase iron-sulfur subunit
11	ASTEI07428-PA	0.652	CDC-like kinase 2
12	ASTEI02436-PA	0.651	Hypothetical protein
13	ASTEI10866-PA	0.649	Metaxin-2
14	ASTEI04263-PA	0.648	Arrestin kurtz-like
15	ASTEI02586-PA	0.648	Hypothetical protein
16	ASTEI07289-PA	0.647	Hypothetical protein
17	ASTEI02911-PA	0.647	B9 domain-containing protein 1
18	ASTEI07881-PA	0.647	Microsomal triglyceride transfer protein large subunit
19	ASTEI04868-PA	0.646	SNX5 protein
20	ASTEI02813-PA	0.646	Coronin homolog
21	ASTEI02215-PA	0.645	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 5
22	ASTEI07373-PA	0.645	Elongation factor 2
23	ASTEI10693-PA	0.641	Iron-regulated short-chain dehydrogenase/reductase
24	ASTEI11276-PA	0.639	Tubulin alpha
25	ASTEI03832-PA	0.639	Malate dehydrogenase
26	ASTEI05212-PA	0.639	Reticulon 4 interacting protein 1
27	ASTEI09681-PA	0.639	FK5Hypothetical protein6-binding protein 1
28	ASTEI00535-PA	0.638	Hypothetical protein
29	ASTEI11547-PA	0.636	High mobility group protein B1
30	ASTEI05775-PA	0.635	Ubiquitin fusion degradation protein 1
31	ASTEI05655-PA	0.635	Molecular chaperone grpe
32	ASTEI09992-PA	0.635	Troponin C isoform 4"
33	ASTEI09193-PA	0.634	4Hypothetical proteins ribosomal protein SA
34	ASTEI04270-PA	0.634	Gelsolin
35	ASTEI00881-PA	0.633	Splicing factor, proline- and glutamine-rich
36	ASTEI06523-PA	0.629	Hypothetical protein
37	ASTEI05323-PA	0.629	Phosphatidylinositol-bisphosphatase
38	ASTEI04562-PA	0.628	Hypothetical protein
39	ASTEI10017-PA	0.628	Hypothetical protein
40	ASTEI02283-PA	0.627	Vacuolar fusion protein CCZ1 homolog
41	ASTEI03146-PA	0.626	3-methylcrotonyl-coa carboxylase beta subunit

42	ASTEI00845-PA	0.625	Hypothetical protein
43	ASTEI04650-PA	0.625	Acetyl-coa acyltransferase 2
44	ASTEI02644-PA	0.624	Peroxisomal 3,2-trans-enoyl-coa isomerase
45	ASTEI10361-PA	0.624	Receptor expression-enhancing protein 5
46	ASTEI10621-PA	0.622	Alpha-tocopherol transfer protein-like protein
47	ASTEI02223-PA	0.622	Transitional endoplasmic reticulum atpase
48	ASTEI10065-PA	0.619	Protein BAT5
49	ASTEI10756-PA	0.617	Hypothetical protein
50	ASTEI10184-PA	0.616	Hypothetical protein
51	ASTEI05222-PA	0.615	Hypothetical protein
52	ASTEI05439-PA	0.614	4Hypothetical proteins ribosomal protein S23
53	ASTEI03649-PA	0.614	Secernin
54	ASTEI11274-PA	0.613	Probable fatty acid-binding protein
55	ASTEI03226-PA	0.613	DNA excision repair protein ERCC-4
56	ASTEI03179-PA	0.611	
57	ASTEI00540-PA	0.611	Protein phosphatase PTC7 homolog fig
58	ASTEI05086-PA	0.611	Mitochondrial 28S ribosomal protein S2
59	ASTEI03783-PA	0.611	Hypothetical protein
60	ASTEI07890-PA	0.609	Protein G12
61	ASTEI05769-PA	0.608	Hypothetical protein
62	ASTEI04267-PA	0.608	Hypothetical protein
63	ASTEI012841-PA	0.607	Hypothetical protein
64	ASTEI09597-PA	0.607	U4/U6 small nuclear ribonucleoprotein SNU13
65	ASTEI07532-PA	0.604	Translation initiation factor 6
66	ASTEI06804-PA	0.602	Eukaryotic translation initiation factor
67	ASTEI01565-PA	0.602	Syntaxin 1A
68	ASTEI04024-PA	0.601	Oxidoreductase GLYR1 homolog
69	ASTEI00788-PA	0.601	Hypothetical protein
70	ASTEI11264-PA	0.600	Hypothetical protein
71	ASTEI03778-PA	0.599	Syntenin-1 isoform 1
72	ASTEI02678-PA	0.598	FK5Hypothetical protein6-binding protein 8
73	ASTEI11367-PA	0.598	Tubulin alpha-4A chain
74	ASTEI02960-PA	0.597	Dnaj homolog subfamily B member 11
75	ASTEI11222-PA	0.595	Nicotinate phosphoribosyltransferase
76	ASTEI07966-PA	0.593	Hydroxybutyrate dehydrogenase
77	ASTEI07189-PA	0.592	Hypothetical protein
78	ASTEI06279-PA	0.590	Hypothetical protein
79	ASTEI10176-PA	0.585	Kinase
80	ASTEI07893-PA	0.585	Valacyclovir hydrolase
81	ASTEI06958-PA	0.585	Scaffold protein salvador
82	ASTEI04310-PA	0.585	Methionyl-trna synthetase
83	ASTEI08300-PA	0.583	6Hypothetical proteins ribosomal protein L38
84	ASTEI03793-PA	0.582	Sulfotransferase (Sult)
85	ASTEI04057-PA	0.582	Hypothetical protein
86	ASTEI01116-PA	0.581	Oligopeptidase A
87	ASTEI05179-PA	0.580	TIM21-like protein, mitochondrial precursor
88	ASTEI04536-PA	0.578	Protein phosphatase inhibitor 2
89	ASTEI00381-PA	0.576	C-4 methylsterol oxidase

90	ASTEI11239-PA	0.576	Fructose-1,6-bisphosphatase I
91	ASTEI10723-PA	0.576	Cytochrome P45Hypothetical protein
92	ASTEI05878-PA	0.574	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2Hypothetical protein
93	ASTEI02400-PA	0.573	Ornithine decarboxylase
94	ASTEI06933-PA	0.573	Electron transfer flavoprotein subunit beta, mitochondrial
95	ASTEI06041-PA	0.571	Hypothetical protein
96	ASTEI05480-PA	0.570	Hypothetical protein
97	ASTEI10910-PA	0.570	Histidine triad nucleotide binding protein 1
98	ASTEI07414-PA	0.570	Aldehyde dehydrogenase (NAD+)
99	ASTEI01359-PA	0.570	Elongation factor 1-alpha
100	ASTEI08817-PA	0.570	Hypothetical protein
101	ASTEI08107-PA	0.568	Ras-related protein Rap-1b precursor
102	ASTEI00316-PA	0.566	2Hypothetical proteins proteasome subunit alpha 7
103	ASTEI07398-PA	0.564	Intron-binding protein aquarius
104	ASTEI03205-PA	0.561	ATP-binding cassette, subfamily F, member 2
105	ASTEI08949-PA	0.561	Hypothetical protein
106	ASTEI10677-PA	0.560	Histone acetyltransferase 1
107	ASTEI07646-PA	0.559	Dynactin 1
108	ASTEI09153-PA	0.555	Hypothetical protein
109	ASTEI00774-PA	0.554	Acyl-coa thioesterase 9.1
110	ASTEI03112-PA	0.554	Hypothetical protein
111	ASTEI03744-PA	0.551	Hypothetical protein
112	ASTEI03402-PA	0.550	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 2
113	ASTEI04915-PA	0.547	Spectrin beta
114	ASTEI07261-PA	0.546	Charged multivesicular body protein 5
115	ASTEI09383-PA	0.541	Mitochondrial cytochrome c oxidase subunit 5B isoform 1
116	ASTEI10364-PA	0.541	Translation initiation factor 5B
117	ASTEI09293-PA	0.540	Transcription elongation factor SPT5
118	ASTEI07559-PA	0.533	39S ribosomal protein L55, mitochondrial
119	ASTEI02055-PA	0.531	6Hypothetical proteins ribosomal protein L28
120	ASTEI09982-PA	0.531	Ubiquitin carboxyl-terminal hydrolase 14
121	ASTEI10759-PA	0.531	Maltase
122	ASTEI09239-PA	0.529	Enoyl-coa hydratase
123	ASTEI08719-PA	0.524	Malate dehydrogenase
124	ASTEI00384-PA	0.523	Hypothetical protein
125	ASTEI02154-PA	0.517	Alpha-crystallin B chain
126	ASTEI00008-PA	0.517	Hypothetical protein
127	ASTEI07702-PA	0.514	Spink6
128	ASTEI11435-PA	0.513	Karyopherin (Importin) alpha 6
129	ASTEI01540-PA	0.513	Adenylate kinase 2, mitochondrial
130	ASTEI08700-PA	0.511	Aldose 1-epimerase
131	ASTEI02679-PA	0.511	Transcription factor BTF3 homolog 4
132	ASTEI01727-PA	0.510	Sarcoplasmic calcium-binding protein
133	ASTEI07979-PA	0.506	Poly(ADP-ribose) glycohydrolase ARH3
134	ASTEI03328-PA	0.497	Ribosome production factor 2

135	ASTEI06045-PA	0.496	6Hypothetical proteins ribosomal protein L21
136	ASTEI08978-PA	0.496	Hypothetical protein
137	ASTEI00001-PA	0.493	Hypothetical protein
138	ASTEI01823-PA	0.492	Hypothetical protein
139	ASTEI02785-PA	0.489	Solute carrier family 3Hypothetical protein (zinc transporter), member 1
140	ASTEI05761-PA	0.489	Peroxiredoxin 5, atypical 2-Cys peroxiredoxin
141	ASTEI04519-PA	0.488	Hypothetical protein
142	ASTEI00616-PA	0.487	Conserved oligomeric Golgi complex subunit 2
143	ASTEI11169-PA	0.483	Vitellogenin
144	ASTEI05011-PA	0.482	Aquaporin (AQP1)
145	ASTEI09989-PA	0.481	V-type H+-transporting atpase 21kda proteolipid subunit
146	ASTEI07742-PA	0.479	Actin
147	ASTEI10330-PA	0.477	Ubiquitin-protein ligase E3 C
148	ASTEI00846-PA	0.477	EF-hand domain (C-terminal) containing 2
149	ASTEI09051-PA	0.476	Hypothetical protein
150	ASTEI03737-PA	0.473	Bis(5'-nucleosidyl)-tetraphosphatase
151	ASTEI06261-PA	0.465	Squid
152	ASTEI02454-PA	0.463	Ubiquitin carboxyl-terminal hydrolase 5/13
153	ASTEI10780-PA	0.454	Glutathione transferase microsomal
154	ASTEI06336-PA	0.446	Hypothetical protein
155	ASTEI09101-PA	0.445	F-type H+-transporting atpase subunit delta
156	ASTEI01004-PA	0.440	Membrane-associated progesterone receptor component
157	ASTEI09857-PA	0.440	Mitochondrial glutamate carrier 1
158	ASTEI03534-PA	0.434	Myofilin variant B
159	ASTEI11396-PA	0.428	Lysosomal Pro-X carboxypeptidase
160	ASTEI07098-PA	0.426	Yellow-x
161	ASTEI02220-PA	0.424	Hypothetical protein
162	ASTEI04235-PA	0.414	Acetyl-coa C-acetyltransferase
163	ASTEI01142-PA	0.413	Hypothetical protein
164	ASTEI04832-PA	0.407	Cathepsin B precursor
165	ASTEI08624-PA	0.403	Nascent polypeptide-associated complex subunit alpha
166	ASTEI01150-PA	0.398	Hypothetical protein
167	ASTEI02954-PA	0.390	Replication factor A2
168	ASTEI07746-PA	0.374	4Hypothetical proteins ribosomal protein S3Hypothetical protein
169	ASTEI05908-PA	0.366	CDGSH iron sulfur domain-containing protein 3, mitochondrial
170	ASTEI05673-PA	0.359	6Hypothetical proteins ribosomal protein L11
171	ASTEI08877-PA	0.357	Carbonic anhydrase
172	ASTEI03333-PA	0.355	Small glutamine-rich tetratricopeptide repeat-containing protein alpha
173	ASTEI04478-PA	0.345	26S proteasome regulatory subunit T5
174	ASTEI04952-PA	0.243	Hypothetical protein
175	ASTEI00830-PA	0.211	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 7
176	ASTEI078091-PA	0.170	Synaptobrevin homolog YKT6

Appendix Table 3 E: Lists of proteins identified to be over-expressed by ≥ 1.5 fold in *P. berghei* infected midgut tissue as compared to sugar fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Relative fold change *P. berghei* infected /sugar fed	Protein description as per *An. gambiae*
1	ASTEI04832-PA	4.740	Cathepsin B precursor
2	ASTEI01388-PA	4.301	Bardet-Biedl syndrome 1 protein
3	ASTEI00275-PA	4.217	Sterol carrier protein 2 variant 2
4	ASTEI09914-PA	4.188	Hypothetical protein
5	ASTEI09546-PA	3.986	Osi15
6	ASTEI10313-PA	3.555	Cellular retinaldehyde-binding protein
7	ASTEI11437-PA	3.552	Clathrin light chain
8	ASTEI07890-PA	3.515	Protein G12
9	ASTEI01945-PA	3.223	Lupus La protein
10	ASTEI09210-PA	3.137	Antizyme inhibitor 1
11	ASTEI02220-PA	2.989	Hypothetical protein
12	ASTEI08889-PA	2.770	Hypothetical protein
13	ASTEI06976-PA	2.755	Phosphotransferase LOC123688
14	ASTEI01419-PA	2.677	Vitellogenic carboxypeptidase-like protein
15	ASTEI04803-PA	2.652	Angiotensin-converting enzyme 9
16	ASTEI04169-PA	2.647	Vanin-like protein 1
17	ASTEI11169-PA	2.614	Vitellogenin
18	ASTEI08938-PA	2.590	Hypothetical protein
19	ASTEI09025-PA	2.589	Calmodulin
20	ASTEI03362-PA	2.570	Hypothetical protein
21	ASTEI00613-PA	2.550	Phosphoenolpyruvate carboxykinase (GTP)
22	ASTEI01399-PA	2.526	Methylenetetrahydrofolate reductase (NADPH)
23	ASTEI02362-PA	2.428	Echinoderm microtubule associated protein like 1
24	ASTEI02154-PA	2.421	Alpha-crystallin B chain
25	ASTEI04927-PA	2.412	5'-3' exoribonuclease 1
26	ASTEI04293-PA	2.350	Sterol carrier protein-2
27	ASTEI07261-PA	2.335	Charged multivesicular body protein 5
28	ASTEI07021-PA	2.294	Proteasome 26S non-atpase subunit 1Hypothetical protein
29	ASTEI03295-PA	2.286	Mesencephalic astrocyte-derived neurotrophic factor homolog
30	ASTEI09203-PA	2.270	Acyl-coa oxidase
31	ASTEI00278-PA	2.268	Hypothetical protein
32	ASTEI06909-PA	2.221	Palmitoyl-protein thioesterase
33	ASTEI08657-PA	2.211	Annexin B1Hypothetical proteinb
34	ASTEI08941-PA	2.182	Hypothetical protein
35	ASTEI11107-PA	2.179	Hypothetical protein
36	ASTEI06801-PA	2.158	Spermine oxidase
37	ASTEI05578-PA	2.154	Hypothetical protein

38	ASTEI01105-PA	2.137	N-acetylglucosamine kinase
39	ASTEI06252-PA	2.118	Benzodiazapine receptor
40	ASTEI04134-PA	2.113	Hypothetical protein
41	ASTEI07379-PA	2.111	S-adenosylmethionine synthetase
42	ASTEI07547-PA	2.104	Optic atrophy 3 family protein
43	ASTEI09047-PA	2.100	Hypothetical protein
44	ASTEI04531-PA	2.085	Ubiquinone biosynthesis monooxygenase Coq7
45	ASTEI10314-PA	2.081	Cellular retinaldehyde-binding protein
46	ASTEI08974-PA	2.069	Acid phosphatase
47	ASTEI01523-PA	2.068	Hypothetical protein
48	ASTEI07863-PA	2.043	Proliferating cell nuclear antigen
49	ASTEI06711-PA	2.029	Copper-zinc (Cu-Zn) superoxide dismutase
50	ASTEI00054-PA	2.025	Epidermal retinal dehydrogenase 2
51	ASTEI09781-PA	2.018	Lipopolysaccharide-induced TNF-alpha factor
52	ASTEI11274-PA	2.017	Probable fatty acid-binding protein
53	ASTEI01594-PA	2.003	Hypothetical protein
54	ASTEI11456-PA	2.002	Thioredoxin
55	ASTEI02731-PA	2.000	Lethal(2)denticleless
56	ASTEI00490-PA	1.991	Solute carrier family 27 (fatty acid transporter), member 1/4
57	ASTEI05161-PA	1.977	2Hypothetical proteins proteasome subunit beta 5
58	ASTEI00463-PA	1.963	Cathepsin F
59	ASTEI01616-PA	1.958	Mrna turnover protein 4
60	ASTEI05234-PA	1.957	Extracellular signal-regulated kinase 1/2
61	ASTEI09386-PA	1.950	S-phase kinase-associated protein 1
62	ASTEI04251-PA	1.946	Arylsulfatase b
63	ASTEI01559-PA	1.944	Trna acetyltransferase TAN1
64	ASTEI02649-PA	1.931	Hypothetical protein
65	ASTEI02048-PA	1.919	Aminoacyl trna synthase complex-interacting multifunctional protein 2
66	ASTEI06894-PA	1.917	Nucleoside-diphosphate kinase
67	ASTEI03600-PA	1.906	Septin 4
68	ASTEI04974-PA	1.901	Hypothetical protein
69	ASTEI00706-PA	1.899	Ferritin heavy chain
70	ASTEI07267-PA	1.894	Cytoplasmic actin
71	ASTEI03480-PA	1.873	Glutathione S-transferase theta class
72	ASTEI09582-PA	1.871	Juvenile hormone epoxide hydrolase
73	ASTEI01563-PA	1.862	N-acetylneuraminate lyase
74	ASTEI00366-PA	1.851	Dynein heavy chain 1, cytosolic
75	ASTEI04964-PA	1.847	Hepatocyte growth factor-regulated tyrosine kinase substrate
76	ASTEI03813-PA	1.843	Hypothetical protein
77	ASTEI05141-PA	1.840	Hypothetical protein
78	ASTEI09696-PA	1.840	Hypothetical protein
79	ASTEI08949-PA	1.838	Hypothetical protein
80	ASTEI01553-PA	1.829	Hypothetical protein

81	ASTEI04905-PA	1.828	Hypothetical protein
82	ASTEI07390-PA	1.811	ATP-binding cassette sub-family G member 1
83	ASTEI01517-PA	1.808	Cytosolic nonspecific dipeptidase
84	ASTEI02336-PA	1.792	Protein aurora borealis
85	ASTEI07584-PA	1.792	Cytoplasmic actin
86	ASTEI03367-PA	1.790	Hypothetical protein
87	ASTEI07409-PA	1.789	Translation initiation factor eif-2B subunit alpha
88	ASTEI09955-PA	1.789	39S ribosomal protein L19, mitochondrial
89	ASTEI03698-PA	1.786	Calreticulin
90	ASTEI06346-PA	1.781	6Hypothetical proteins ribosomal protein l7a
91	ASTEI06533-PA	1.779	Hypothetical protein
92	ASTEI06356-PA	1.778	2-acylglycerol O-acyltransferase 2
93	ASTEI06288-PA	1.767	Hypothetical protein
94	ASTEI07794-PA	1.766	Prostaglandin reductase 1
95	ASTEI09369-PA	1.763	Mesenchymal stem cell protein DSCD75
96	ASTEI05728-PA	1.758	Innexin 2
97	ASTEI08867-PA	1.757	Cathepsin D
98	ASTEI08151-PA	1.756	N-acetylglucosamine-6-sulfatase
99	ASTEI02439-PA	1.747	Lysophospholipid acyltransferase 5
100	ASTEI04093-PA	1.743	Symplekin
101	ASTEI07648-PA	1.743	Solute carrier family 5 (sodium-coupled monocarboxylate transporter), member 8
102	ASTEI10458-PA	1.742	Hypothetical protein
103	ASTEI08299-PA	1.737	Carnitine O-palmitoyltransferase 1
104	ASTEI06754-PA	1.736	Calumenin
105	ASTEI08112-PA	1.723	ATP-dependent RNA helicase DDX1
106	ASTEI03010-PA	1.721	Seryl-trna synthetase
107	ASTEI00367-PA	1.720	Dynein heavy chain 1, cytosolic
108	ASTEI05065-PA	1.720	Platelet-activating factor acetylhydrolase IB subunit alpha
109	ASTEI04257-PA	1.719	Phosphomannomutase
110	ASTEI03112-PA	1.716	Hypothetical protein
111	ASTEI06380-PA	1.712	Gamma-glutamyl hydrolase
112	ASTEI07600-PA	1.709	Exocyst complex component 5
113	ASTEI04991-PA	1.702	Protein kinase, cgmp-dependent
114	ASTEI09117-PA	1.696	Glucosyl/glucuronosyl transferases
115	ASTEI09571-PA	1.695	Hydroxysteroid (17-beta) dehydrogenase 4
116	ASTEI09440-PA	1.694	Hypothetical protein
117	ASTEI02521-PA	1.693	Hypothetical protein
118	ASTEI00934-PA	1.688	Hypothetical protein
119	ASTEI03036-PA	1.683	THO complex subunit 2
120	ASTEI09037-PA	1.682	Short caspase (CASPS5)
121	ASTEI03839-PA	1.678	Lipoma-prefererred partner
122	ASTEI06955-PA	1.677	Hypothetical protein
123	ASTEI03541-PA	1.673	Hypothetical protein
124	ASTEI07496-PA	1.671	Bleomycin hydrolase
125	ASTEI11067-PA	1.664	Hypothetical protein

126	ASTEI10842-PA	1.657	Hypothetical protein
127	ASTEI06540-PA	1.657	ATP-binding cassette sub-family G member 4
128	ASTEI02634-PA	1.657	Isocitrate dehydrogenase
129	ASTEI07849-PA	1.654	Hypothetical protein
130	ASTEI09896-PA	1.653	Hypothetical protein
131	ASTEI11088-PA	1.652	Oxysterol-binding protein 1
132	ASTEI01710-PA	1.651	Solute carrier family 16 (monocarboxylic acid transporters), member 1Hypothetical protein
133	ASTEI10069-PA	1.645	Carboxylesterase beta esterase
134	ASTEI07055-PA	1.643	Myosin regulatory light chain, invertebrate
135	ASTEI08952-PA	1.639	Hypothetical protein
136	ASTEI01456-PA	1.637	Hypothetical protein
137	ASTEI05015-PA	1.637	Minor histocompatibility antigen H13
138	ASTEI04256-PA	1.635	Glucosamine--fructose-6-phosphate aminotransferase (isomerizing)
139	ASTEI10510-PA	1.633	Solute carrier family 12 (potassium/chloride transporters), member 8
140	ASTEI06638-PA	1.632	Hypothetical protein
141	ASTEI03104-PA	1.629	Isopentenyl-diphosphate delta-isomerase
142	ASTEI06954-PA	1.629	Hypothetical protein
143	ASTEI00949-PA	1.627	Hypothetical protein
144	ASTEI10685-PA	1.622	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase
145	ASTEI03122-PA	1.622	Sphinganine-1-phosphate aldolase
146	ASTEI05989-PA	1.618	Serine protease inhibitor (serpin) SRPN1Hypothetical protein
147	ASTEI00707-PA	1.616	Secreted ferritin G subunit
148	ASTEI00143-PA	1.616	V-type H+-transporting atpase subunit G
149	ASTEI10973-PA	1.612	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3
150	ASTEI01231-PA	1.610	Amino acid transporter protein
151	ASTEI04243-PA	1.596	Hypothetical protein
152	ASTEI04209-PA	1.596	Alkaline phosphatase
153	ASTEI09553-PA	1.592	Leucyl aminopeptidase
154	ASTEI02511-PA	1.592	Hypothetical protein
155	ASTEI06240-PA	1.590	JAKSTAT pathway signalling
156	ASTEI07805-PA	1.587	Epidermal retinal dehydrogenase
157	ASTEI11547-PA	1.582	High mobility group protein B1
158	ASTEI06324-PA	1.582	Glycolipid transfer protein domain-containing protein 1
159	ASTEI05740-PA	1.579	Innexin inx2
160	ASTEI08308-PA	1.576	Cytochrome b5-related
161	ASTEI10621-PA	1.574	Alpha-tocopherol transfer protein-like protein
162	ASTEI02725-PA	1.567	IMD pathway signalling Fas-Associated Death Domain (FADD).
163	ASTEI00696-PA	1.567	V-type proton atpase subunit F
164	ASTEI07135-PA	1.564	Peroxisomal biogenesis factor 12

165	ASTEI06657-PA	1.563	Ubiquitin-conjugating enzyme E2 L3
166	ASTEI06048-PA	1.561	26S proteasome regulatory subunit N8
167	ASTEI05529-PA	1.560	ATP-binding cassette, subfamily D (ALD), member 2
168	ASTEI01441-PA	1.559	Heat shock protein 67B2
169	ASTEI07913-PA	1.559	Vacuolar protein sorting-associated protein 33A
170	ASTEI07687-PA	1.555	Sterol carrier protein 2
171	ASTEI08132-PA	1.552	Hypothetical protein
172	ASTEI10474-PA	1.552	Lysosomal alpha-mannosidase
173	ASTEI05005-PA	1.549	D-3-phosphoglycerate dehydrogenase
174	ASTEI02161-PA	1.545	Motile sperm domain-containing protein 2
175	ASTEI00947-PA	1.543	Battenin
176	ASTEI03761-PA	1.539	Hypothetical protein
177	ASTEI09088-PA	1.538	Protein phosphatase
178	ASTEI10548-PA	1.537	Plasma glutamate carboxypeptidase
179	ASTEI02403-PA	1.537	Muscle LIM protein at 84B
180	ASTEI06523-PA	1.531	Hypothetical protein
181	ASTEI08274-PA	1.529	Intraflagellar transport 8Hypothetical protein homolog
182	ASTEI08666-PA	1.527	Filamin
183	ASTEI07734-PA	1.525	Hypothetical protein
184	ASTEI06878-PA	1.524	Farnesyl diphosphate synthase
185	ASTEI04735-PA	1.523	Glyoxylate/hydroxypyruvate reductase
186	ASTEI03764-PA	1.520	Hypothetical protein
187	ASTEI00038-PA	1.519	Oxidation resistance protein
188	ASTEI09894-PA	1.518	Hypothetical protein
189	ASTEI00384-PA	1.518	Hypothetical protein
190	ASTEI09982-PA	1.516	Ubiquitin carboxyl-terminal hydrolase 14
191	ASTEI01016-PA	1.514	Protein-L-isoaspartate(D-aspartate) O-methyltransferase
192	ASTEI08136-PA	1.514	Rhomboid-4, isoform B
193	ASTEI11212-PA	1.514	Hypothetical protein
194	ASTEI00811-PA	1.514	V-type H+-transporting atpase S1 subunit
195	ASTEI00150-PA	1.512	Lipophorin
196	ASTEI01731-PA	1.509	Selenide, water dikinase
197	ASTEI01140-PA	1.509	6Hypothetical proteins ribosomal protein L37
198	ASTEI06266-PA	1.509	Hypothetical protein
199	ASTEI07899-PA	1.508	2Hypothetical proteins proteasome subunit alpha 6
200	ASTEI03043-PA	1.507	Trehalose 6-phosphate synthase/phosphatase
201	ASTEI00386-PA	1.506	Glycerol-3-phosphate O-acyltransferase 3/4
202	ASTEI03418-PA	1.506	ATP-dependent RNA helicase DHX57
203	ASTEI07289-PA	1.502	Hypothetical protein
204	ASTEI04332-PA	1.502	ADP,ATP carrier protein 2

Appendix Table 3 F: Lists of proteins identified to be down-regulated by ≤ 0.66 fold in *P. berghei* infected midgut tissue as compared to sugar fed mosquitoes.

S. No.	*An. stephensi* VectorBase ID	Relative fold change *P. berghei* infected /sugar fed	Protein description as per *An. gambiae*
1	ASTEI01026-PA	0.659	Cytosolic iron-sulfur protein assembly protein Ciao1
2	ASTEI00702-PA	0.659	6Hypothetical proteins ribosomal protein L26
3	ASTEI02576-PA	0.658	Glucosyl/glucuronosyl transferases
4	ASTEI05368-PA	0.657	Hypothetical protein
5	ASTEI03111-PA	0.657	Actin related protein 2/3 complex, subunit 3
6	ASTEI04200-PA	0.657	Defensin anti-microbial peptide
7	ASTEI09394-PA	0.656	Protein-tyrosine phosphatase
8	ASTEI04270-PA	0.654	Gelsolin
9	ASTEI09647-PA	0.654	Ribosome assembly protein SQT1
10	ASTEI04175-PA	0.654	Glucose-6-phosphate 1-dehydrogenase
11	ASTEI07139-PA	0.653	N-acetyltransferase 1Hypothetical protein
12	ASTEI02332-PA	0.653	Hypothetical protein
13	ASTEI05222-PA	0.653	Hypothetical protein
14	ASTEI09438-PA	0.653	Oligosaccharyltransferase complex subunit beta
15	ASTEI08667-PA	0.653	Hypothetical protein
16	ASTEI10045-PA	0.653	Carboxylesterase
17	ASTEI08096-PA	0.652	Tubulin-specific chaperone D
18	ASTEI05597-PA	0.652	Hypothetical protein
19	ASTEI10392-PA	0.652	Glucosyl/glucuronosyl transferases
20	ASTEI05761-PA	0.651	Peroxiredoxin 5, atypical 2-Cys peroxiredoxin
21	ASTEI00037-PA	0.651	Stromal cell-derived factor 2
22	ASTEI09383-PA	0.649	Mitochondrial cytochrome c oxidase subunit 5B isoform 1
23	ASTEI04882-PA	0.647	STE24 endopeptidase
24	ASTEI01072-PA	0.647	Cytochrome P45Hypothetical protein
25	ASTEI01548-PA	0.647	Solute carrier family 39 (zinc transporter), member 11
26	ASTEI04318-PA	0.646	Hypothetical protein
27	ASTEI07781-PA	0.645	Breast carcinoma amplified sequence 3
28	ASTEI04263-PA	0.644	Arrestin kurtz-like
29	ASTEI02791-PA	0.643	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1
30	ASTEI07018-PA	0.643	Translocase of inner mitochondrial membrane 44 homolog (yeast)
31	ASTEI10554-PA	0.643	3'-phosphoadenosine 5'-phosphosulfate synthase
32	ASTEI05967-PA	0.643	Protein transport protein SEC24
33	ASTEI10573-PA	0.643	2-oxoisovalerate dehydrogenase E2 component (dihydrolipoyl transacylase)
34	ASTEI01083-PA	0.642	Xylulokinase
35	ASTEI06864-PA	0.642	Hypothetical protein
36	ASTEI10539-PA	0.641	Malate synthase
37	ASTEI11673-PA	0.639	Cell division control protein 42
38	ASTEI07342-PA	0.639	CCR4-NOT transcription complex subunit

			1Hypothetical protein
39	ASTEI09396-PA	0.636	Juvenile hormone-inducible protein
40	ASTEI01139-PA	0.636	6Hypothetical proteins ribosomal protein L1Hypothetical protein-1
41	ASTEI04515-PA	0.635	39S ribosomal protein L22, mitochondrial
42	ASTEI08291-PA	0.635	Protein phosphatase 2 (formerly 2A), catalytic subunit
43	ASTEI05713-PA	0.633	Procollagen-lysine,2-oxoglutarate 5-dioxygenase, invertebrate
44	ASTEI03729-PA	0.633	Hypothetical protein
45	ASTEI04198-PA	0.633	Protein flightless-1
46	ASTEI05350-PA	0.633	Mitochondrial import inner membrane translocase subunit Tim8 B
47	ASTEI04234-PA	0.632	Acetyl-coa C-acetyltransferase
48	ASTEI09899-PA	0.630	Alpha-aminoadipic semialdehyde synthase
49	ASTEI05212-PA	0.629	Reticulon 4 interacting protein 1
50	ASTEI09467-PA	0.629	Pyruvate dehydrogenase E1 component subunit beta
51	ASTEI09840-PA	0.629	Hypothetical protein
52	ASTEI02421-PA	0.628	6Hypothetical proteins ribosomal protein L18a (partial)
53	ASTEI02728-PA	0.627	Lethal(2)denticleless
54	ASTEI02538-PA	0.627	Chloride channel
55	ASTEI03778-PA	0.625	Syntenin-1 isoform 1
56	ASTEI06989-PA	0.624	Set1/Ash2 histone methyltransferase complex subunit ASH2
57	ASTEI09193-PA	0.623	4Hypothetical proteins ribosomal protein SA
58	ASTEI06042-PA	0.623	Ankyrin repeat-rich membrane spanning protein
59	ASTEI09612-PA	0.622	Dipeptidyl-peptidase 4
60	ASTEI06668-PA	0.622	Nardilysin
61	ASTEI05272-PA	0.622	All-trans/9-cis
62	ASTEI03784-PA	0.620	Myotubularin related protein 2
63	ASTEI10940-PA	0.619	G protein pathway suppressor 1
64	ASTEI06457-PA	0.619	Casein kinase II subunit beta
65	ASTEI10780-PA	0.619	Glutathione transferase microsomal
66	ASTEI11301-PA	0.619	Nicotinate phosphoribosyltransferase
67	ASTEI05533-PA	0.618	Cell cycle progression
68	ASTEI06049-PA	0.617	HIV-1 Vpr-binding protein
69	ASTEI09821-PA	0.616	Acyl-coa-binding protein
70	ASTEI05583-PA	0.616	ATP-dependent RNA helicase DDX27
71	ASTEI10403-PA	0.616	Axonemal dynein intermediate chain inner arm i1
72	ASTEI07966-PA	0.616	Hydroxybutyrate dehydrogenase
73	ASTEI03415-PA	0.613	Gtpase kras
74	ASTEI11403-PA	0.613	Ceramide synthetase
75	ASTEI05228-PA	0.613	Collagen, type IV, alpha
76	ASTEI06268-PA	0.613	Rho-associated protein kinase
77	ASTEI10184-PA	0.612	Hypothetical protein
78	ASTEI00794-PA	0.612	Gamma-tubulin complex component 3
79	ASTEI05294-PA	0.611	Hypothetical protein
80	ASTEI00700-PA	0.611	AP2-associated kinase

81	ASTEI11186-PA	0.610	Extended synaptotagmin-like protein 2a
82	ASTEI01662-PA	0.610	Regucalcin protein
83	ASTEI05576-PA	0.610	4Hypothetical proteins ribosomal protein S12
84	ASTEI09055-PA	0.608	Heat shock 7Hypothetical proteinkda protein 1/8
85	ASTEI06017-PA	0.607	Ribosomal RNA methyltransferase Nop2
86	ASTEI03758-PA	0.605	Alkylated DNA repair protein alkb homolog 4
87	ASTEI04229-PA	0.601	Fibulin 2
88	ASTEI04993-PA	0.600	Female reproductive tract protease GLEANR_896
89	ASTEI04287-PA	0.600	39S ribosomal protein L43
90	ASTEI04337-PA	0.600	Nuclear pore complex protein Nup98-Nup96
91	ASTEI03673-PA	0.600	T-complex protein 1 subunit alpha
92	ASTEI11150-PA	0.598	Eukaryotic translation initiation factor 4E-binding protein 1
93	ASTEI01484-PA	0.598	Golgi-associated plant pathogenesis-related protein 1
94	ASTEI00923-PA	0.597	Glycine dehydrogenase
95	ASTEI08091-PA	0.596	Pyruvate dehydrogenase E1 component subunit alpha
96	ASTEI04254-PA	0.595	4-nitrophenyl phosphatase
97	ASTEI00003-PA	0.595	28S ribosomal protein S6
98	ASTEI08959-PA	0.593	Hypothetical protein
99	ASTEI03269-PA	0.592	Hypothetical protein
100	ASTEI10972-PA	0.592	3-glucan binding protein (bgbp)
101	ASTEI01948-PA	0.592	Ubiquinol-cytochrome c reductase iron-sulfur subunit
102	ASTEI09969-PA	0.591	Polymerase delta-interacting protein 2
103	ASTEI09761-PA	0.591	39S ribosomal protein L13, mitochondrial
104	ASTEI08617-PA	0.591	Hypothetical protein
105	ASTEI00291-PA	0.591	Hypothetical protein
106	ASTEI08234-PA	0.590	U3 small nucleolar RNA-associated protein 22
107	ASTEI03309-PA	0.590	Pyruvate dehydrogenase E1 component subunit alpha
108	ASTEI02883-PA	0.589	Hypothetical protein
109	ASTEI09035-PA	0.588	6Hypothetical proteins ribosomal protein L15
110	ASTEI05681-PA	0.588	Zinc/iron regulated transporter-related protein 1
111	ASTEI10124-PA	0.588	Inositol-1,4,5-trisphosphate 5-phosphatase
112	ASTEI08521-PA	0.587	Hypothetical protein
113	ASTEI06038-PA	0.585	ATP-binding cassette sub-family A member 7
114	ASTEI05247-PA	0.584	Hypothetical protein
115	ASTEI04331-PA	0.584	ADP,ATP carrier protein 2
116	ASTEI10335-PA	0.583	Hypothetical protein
117	ASTEI09857-PA	0.583	Mitochondrial glutamate carrier 1
118	ASTEI10789-PA	0.583	Karyopherin (Importin) alpha 6
119	ASTEI04936-PA	0.582	Zinc finger protein
120	ASTEI07996-PA	0.579	Hypothetical protein
121	ASTEI09946-PA	0.578	P37NB protein
122	ASTEI00665-PA	0.576	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1
123	ASTEI03482-PA	0.575	28S ribosomal protein S29, mitochondrial
124	ASTEI09239-PA	0.575	Enoyl-coa hydratase

125	ASTEI02454-PA	0.575	Ubiquitin carboxyl-terminal hydrolase 5/13
126	ASTEI05913-PA	0.575	L-lactate dehydrogenase
127	ASTEI08379-PA	0.575	Hypothetical protein
128	ASTEI10632-PA	0.574	Hypothetical protein
129	ASTEI06279-PA	0.573	Hypothetical protein
130	ASTEI04951-PA	0.573	Mitochondrial atpase inhibitor
131	ASTEI03783-PA	0.572	Hypothetical protein
132	ASTEI03591-PA	0.571	Phosphoribosyl pyrophosphate synthetase-associated protein 2
133	ASTEI05281-PA	0.570	Hypothetical protein
134	ASTEI02041-PA	0.569	Hypothetical protein
135	ASTEI09489-PA	0.569	Abhydrolase domain-containing protein FAM1Hypothetical protein8b1
136	ASTEI02289-PA	0.567	Insulin receptor
137	ASTEI08260-PA	0.567	Glucosyl/glucuronosyl transferases
138	ASTEI06556-PA	0.567	Dolichol-phosphate mannosyltransferase
139	ASTEI10391-PA	0.566	Glucosyl/glucuronosyl transferases
140	ASTEI09293-PA	0.566	Transcription elongation factor SPT5
141	ASTEI07742-PA	0.566	Actin
142	ASTEI09989-PA	0.565	V-type H+-transporting atpase 21kda proteolipid subunit
143	ASTEI11283-PA	0.565	ATP-binding cassette, subfamily D (ALD), member 3
144	ASTEI09265-PA	0.562	Spermine synthase
145	ASTEI01134-PA	0.561	Synaptic vesicle protein
146	ASTEI08719-PA	0.561	Malate dehydrogenase
147	ASTEI05635-PA	0.561	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 1Hypothetical protein
148	ASTEI07559-PA	0.558	39S ribosomal protein L55, mitochondrial
149	ASTEI05673-PA	0.558	6Hypothetical proteins ribosomal protein L11
150	ASTEI04830-PA	0.557	3-ketoacyl-coa reductase
151	ASTEI02051-PA	0.557	Mitochondrial inner membrane protein OXA1L
152	ASTEI10020-PA	0.557	Fumarylacetoacetate hydrolase domain containing 2A
153	ASTEI00985-PA	0.557	3-dehydrosphinganine reductase
154	ASTEI07186-PA	0.555	Transglutaminase
155	ASTEI01205-PA	0.555	Mitochondrial 39S ribosomal protein L3
156	ASTEI08336-PA	0.554	Cytochrome c oxidase assembly protein subunit 11
157	ASTEI10637-PA	0.553	ATP-dependent RNA helicase DDX5/DBP2
158	ASTEI10968-PA	0.552	Integrin alpha-PS1
159	ASTEI06635-PA	0.551	Mitochondrial ribosomal protein L28
160	ASTEI10866-PA	0.551	Metaxin-2
161	ASTEI08201-PA	0.550	Hypothetical protein
162	ASTEI03297-PA	0.550	Serine/threonine-protein phosphatase 6 catalytic subunit
163	ASTEI08817-PA	0.550	Hypothetical protein
164	ASTEI04027-PA	0.550	Dihydrodiol dehydrogenase / D-xylose 1-dehydrogenase (NADP)
165	ASTEI02022-PA	0.549	Ruvb-like protein 2
166	ASTEI06336-PA	0.549	Hypothetical protein

167	ASTEI10386-PA	0.546	tRNA pseudouridine13 synthase
168	ASTEI04697-PA	0.542	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 13
169	ASTEI02014-PA	0.542	Glutaredoxin
170	ASTEI10017-PA	0.542	Hypothetical protein
171	ASTEI02643-PA	0.541	Ubiquitin carboxyl-terminal hydrolase 1Hypothetical protein
172	ASTEI02750-PA	0.541	Capping protein (actin filament) muscle Z-line, alpha 1
173	ASTEI04267-PA	0.540	Hypothetical protein
174	ASTEI06803-PA	0.537	Aubergine
175	ASTEI01198-PA	0.535	Alpha 1,3-glucosidase
176	ASTEI00897-PA	0.531	39S ribosomal protein L51, mitochondrial
177	ASTEI10795-PA	0.530	2Hypothetical proteins proteasome subunit beta 6
178	ASTEI00616-PA	0.529	Conserved oligomeric Golgi complex subunit 2
179	ASTEI09681-PA	0.525	FK5Hypothetical protein6-binding protein 1
180	ASTEI06707-PA	0.525	Hypothetical protein
181	ASTEI03508-PA	0.525	Platelet-activating factor acetylhydrolase IB subunit beta/gamma
182	ASTEI03205-PA	0.522	ATP-binding cassette, subfamily F, member 2
183	ASTEI05439-PA	0.521	4Hypothetical proteins ribosomal protein S23
184	ASTEI11217-PA	0.520	Glycine C-acetyltransferase
185	ASTEI07511-PA	0.520	Beta-catenin-like protein 1
186	ASTEI10652-PA	0.520	Trypsin type 4
187	ASTEI07368-PA	0.518	Arylformamidase
188	ASTEI07792-PA	0.517	Frataxin homolog, mitochondrial
189	ASTEI08515-PA	0.517	Myo-inositol-1(or 4)-monophosphatase
190	ASTEI04478-PA	0.515	26S proteasome regulatory subunit T5
191	ASTEI04681-PA	0.515	Adrenodoxin
192	ASTEI02055-PA	0.514	6Hypothetical proteins ribosomal protein L28
193	ASTEI07480-PA	0.512	TBC1 domain family, member 1Hypothetical proteina
194	ASTEI09348-PA	0.509	Phyhd1 protein
195	ASTEI07730-PA	0.505	Mitochondrial import receptor subunit TOM4Hypothetical protein
196	ASTEI09371-PA	0.503	Mesenchymal stem cell protein DSCD75
197	ASTEI03146-PA	0.502	3-methylcrotonyl-coa carboxylase beta subunit
198	ASTEI00379-PA	0.501	Hypothetical protein
199	ASTEI03611-PA	0.500	WD4Hypothetical protein repeat-containing protein SMU1
200	ASTEI02371-PA	0.499	Translocase of outer mitochondrial membrane 2Hypothetical protein homolog
201	ASTEI07201-PA	0.499	Tmco1
202	ASTEI04988-PA	0.498	Hypothetical protein
203	ASTEI06261-PA	0.498	Squid
204	ASTEI05053-PA	0.498	U6 snrna-associated Sm-like protein lsm8
205	ASTEI11396-PA	0.497	Lysosomal Pro-X carboxypeptidase
206	ASTEI10345-PA	0.496	Hypothetical protein
207	ASTEI09499-PA	0.494	Hypothetical protein
208	ASTEI04057-PA	0.488	Hypothetical protein

209	ASTEI11222-PA	0.488	Nicotinate phosphoribosyltransferase
210	ASTEI05655-PA	0.486	Molecular chaperone grpe
211	ASTEI01116-PA	0.485	Oligopeptidase A
212	ASTEI00941-PA	0.484	Hypothetical protein
213	ASTEI07837-PA	0.483	Sparc
214	ASTEI00772-PA	0.482	Bromodomain-containing protein 8
215	ASTEI05390-PA	0.474	Succinate dehydrogenase (ubiquinone) cytochrome b56Hypothetical protein subunit
216	ASTEI03980-PA	0.474	4Hypothetical proteins ribosomal protein S11
217	ASTEI06933-PA	0.467	Electron transfer flavoprotein subunit beta, mitochondrial
218	ASTEI09457-PA	0.466	Succinate dehydrogenase (ubiquinone) flavoprotein subunit
219	ASTEI05653-PA	0.466	39S ribosomal protein L14, mitochondrial
220	ASTEI08011-PA	0.466	Nuclear pore complex protein Nup88
221	ASTEI01928-PA	0.466	Hypothetical protein
222	ASTEI04058-PA	0.463	Hypothetical protein
223	ASTEI09597-PA	0.463	U4/U6 small nuclear ribonucleoprotein SNU13
224	ASTEI00680-PA	0.463	Solute carrier family 23 (nucleobase transporter)
225	ASTEI03928-PA	0.462	Hypothetical protein
226	ASTEI03333-PA	0.461	Small glutamine-rich tetratricopeptide repeat-containing protein alpha
227	ASTEI07952-PA	0.458	Serine/threonine-protein phosphatase 4 regulatory subunit 2
228	ASTEI04519-PA	0.458	Hypothetical protein
229	ASTEI07669-PA	0.456	Starch phosphorylase
230	ASTEI09305-PA	0.448	Kelch domain containing 4
231	ASTEI02679-PA	0.448	Transcription factor BTF3 homolog 4
232	ASTEI01823-PA	0.444	Hypothetical protein
233	ASTEI04172-PA	0.443	Hypothetical protein
234	ASTEI03179-PA	0.440	
235	ASTEI06660-PA	0.439	Protease m1 zinc metalloprotease
236	ASTEI05953-PA	0.438	Hypothetical protein
237	ASTEI01168-PA	0.436	Hypothetical protein
238	ASTEI07702-PA	0.434	Spink6
239	ASTEI06664-PA	0.432	Hypothetical protein
240	ASTEI08300-PA	0.428	6Hypothetical proteins ribosomal protein L38
241	ASTEI07285-PA	0.427	Mesoderm development candidate 2
242	ASTEI01150-PA	0.424	Hypothetical protein
243	ASTEI03744-PA	0.423	Hypothetical protein
244	ASTEI08877-PA	0.420	Carbonic anhydrase
245	ASTEI03099-PA	0.419	Fatty-acid amide hydrolase 2
246	ASTEI10589-PA	0.417	Hypothetical protein
247	ASTEI00846-PA	0.414	EF-hand domain (C-terminal) containing 2
248	ASTEI02785-PA	0.411	Solute carrier family 3Hypothetical protein (zinc transporter), member 1
249	ASTEI10350-PA	0.406	Galectin (GALE7)
250	ASTEI00830-PA	0.400	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 7
251	ASTEI00584-PA	0.400	Isocitrate dehydrogenase (NAD+)

252	ASTEI07150-PA	0.396	Hypothetical protein
253	ASTEI02686-PA	0.395	Eukaryotic translation initiation factor 3 subunit I
254	ASTEI05011-PA	0.394	Aquaporin (AQP1)
255	ASTEI08573-PA	0.391	6Hypothetical proteins ribosomal protein L27
256	ASTEI06639-PA	0.390	Hypothetical protein
257	ASTEI08107-PA	0.376	Ras-related protein Rap-1b precursor
258	ASTEI00535-PA	0.370	Hypothetical protein
259	ASTEI09101-PA	0.370	F-type H+-transporting atpase subunit delta
260	ASTEI09202-PA	0.365	Hypothetical protein
261	ASTEI08978-PA	0.364	Hypothetical protein
262	ASTEI02954-PA	0.361	Replication factor A2
263	ASTEI07532-PA	0.357	Translation initiation factor 6
264	ASTEI04235-PA	0.355	Acetyl-coa C-acetyltransferase
265	ASTEI10033-PA	0.352	Serine-type enodpeptidase
266	ASTEI07803-PA	0.345	Odorant binding protein (OBP9)
267	ASTEI10273-PA	0.337	C-Type Lectin (CTL) - galactose binding
268	ASTEI09679-PA	0.336	Lipoyltransferase 1
269	ASTEI08624-PA	0.335	Nascent polypeptide-associated complex subunit alpha
270	ASTEI08252-PA	0.333	Hypothetical protein
271	ASTEI02077-PA	0.332	Cuticular protein RR-1 family (CPR15)
272	ASTEI06012-PA	0.322	Hypothetical protein
273	ASTEI06045-PA	0.320	6Hypothetical proteins ribosomal protein L21
274	ASTEI01142-PA	0.313	Hypothetical protein
275	ASTEI07582-PA	0.301	Yellow-e
276	ASTEI10264-PA	0.298	Serine protease SP24D
277	ASTEI05908-PA	0.291	CDGSH iron sulfur domain-containing protein 3, mitochondrial
278	ASTEI05480-PA	0.276	Hypothetical protein
279	ASTEI07746-PA	0.259	4Hypothetical proteins ribosomal protein S3Hypothetical protein
280	ASTEI07098-PA	0.250	Yellow-x
281	ASTEI04952-PA	0.250	Hypothetical protein
282	ASTEI06899-PA	0.234	Glutamate-rich WD repeat-containing protein 1
283	ASTEI00001-PA	0.230	Hypothetical protein
284	ASTEI10176-PA	0.211	Kinase
285	ASTEI03534-PA	0.210	Myofilin variant B
286	ASTEI06837-PA	0.200	Translocon-associated protein subunit beta
287	ASTEI10165-PA	0.175	Actin, cytoplasmic
288	ASTEI10327-PA	0.153	Envelysin
289	ASTEI03020-PA	0.147	Protein of unknown function, DUF255 family
290	ASTEI078091-PA	0.130	Synaptobrevin homolog YKT6

Appendix Table 4 A: List of Genome Search Specific Peptides (GSSPs) identified in this study.

Peptide sequence	Genome coordinates of identified peptide	Peptide sequence	Genome coordinates of identified peptide
AAEIGGESGGANTADSKDDEYR	scaffold00003:-:2535824-2535889	HSSLPNVPEEIFR	scaffold00038:+:70312-70350
AGGPGLEHGEQGVPAEFNVWTR	scaffold01145:-:2538-2603	IDFFPHLETQCSVR	scaffold00038:-:877322-877363
AGHQPIALVGGATGLIGDPSGR	scaffold00004:+:4378785-4378850	INCTGESLAEAGGQETSSVIVETVAK	scaffold01145:-:1582-1659
ALHEEPTPPVIQPGQGGK	scaffold00003:-:5500981-5501034	INVDQAFHELVR	scaffold00014:-:193140-193175
AMASSCCLLLLHCIEEPHK	scaffold00002:+:5294975-5295031	ITAEDVQNVAR	scaffold00006:+:3380060-3380092
APEDIFADVEK	scaffold00108:+:378770-378802	KEPFFHGHDNYDQLVR	scaffold00008:+:1853617-1853664
AQEEKPEPETTTAAPVEEK	scaffold00052:+:1260107-1260163	KFQLSERPLVDEK	scaffold00014:-:193092-193130
ASDPDGTTINVPSK	scaffold00006:+:3378641-3378682	KGEGIDAGYGEVDWICNR	scaffold00176:+:46148-46201
CNICQSSVSNSCR	scaffold00067:+:278305-278343	LEDLSDAIVQDFALMR	scaffold00390:-:8379-8426
DADDAVYELNGK	scaffold00176:-:25139-25174	LIEEDSADVR	scaffold00122:+:273655-273684
DELIDDLLR	scaffold00111:-:306792-306818	LLILIRPRVLPWVGG	scaffold00070:+:139053-139097
DLKPENLLIDSQGYLK	scaffold00055:-:1041268-1041315	LSEDEVELAR	scaffold00006:+:3379100-3379129
DMGSVDELLPK	scaffold00001:-:1422336-1422368	LTQLFENLSHDELR	scaffold00018:+:1042576-1042617
DMVLATGPAESER	scaffold00001:-:1420929-1420967	MQGLPPTDAAEEK	scaffold00062:-:623749-623787
EAGGGTLAISVEGPSK	scaffold01145:-:2490-2537	NAANADLVPAEQANVK	scaffold00062:-:815673-815720
EALENLEDDQPIVAGGSYQK	scaffold00086:+:489371-489430	NALNNTYPGYR	scaffold00332:-:5676-5708
ECTSESCGAGVFMAAHEDR	scaffold00006:-:2100651-2100707	NFGSYVTHETR	scaffold00327:-:11495-11527
EFLDQAKEDFEEK	scaffold00055:-:1041724-1041762	NINYCVDAGIAK	scaffold00006:-:1950612-1950647
EGTFQGLHTFPVTIGLR	mitocondorial:-:10024-10074	NLLQVDLTK	scaffold00055:-:1040974-1041000
EHGVDTVDSPGVVHVVHSER	scaffold00043:-:191122-191181	NPTNTAALDDFER	scaffold00055:-:1041676-1041714
EIEIDIEPTDKVDR	scaffold00057:-:812388-812429	QCFLTPVYVDFITK	scaffold00003:-:3636594-3636635
ELADEIEEVNER	scaffold00114:-:288315-288350	QEHAQQHPSQGPPPPQDEPSPDEDSVHQR	scaffold01194:-:1200-1286
ENVQISCGCFHDVK	scaffold00003:-:5501044-5501085	QFVAETDEIEPGTEWER	scaffold05080:-:1121-1171
EPFFHGHDNYDQLVR	scaffold00008:+:1853620-1853664	QPGAPDVQEEQVDSEKPIEK	scaffold00081:-:345605-345664
EPSSIDDTNTPPKY	scaffold00059:+:1100324-1100365	QSIDTHIEIVR	scaffold00057:+:381379-381411
EQELEEVLTK	scaffold01124:-:2413-2442	QTEHTIFVGNLPTTSNR	scaffold00003:+:67024-67074
EYDYSEYYPAK	scaffold00006:-:2909207-2909239	RECTSESCGAGVFMAAHEDR	scaffold00006:-:2100651-2100710
EYLDMDLTLSQEK	scaffold00057:+:381039-381077	RNPTNTAALDDFER	scaffold00055:-:1041676-1041717
FARPASTAAPVAFDNRPSSSGLR	scaffold00058:+:814768-814836	SEGFDDYMK	scaffold06797:-:598-624
FDAQVDHDVNGVNYAK	scaffold00003:-:5501140-5501187	SIVTAPAVSAAIMEENYDTTHPK	scaffold00028:+:1673350-1673418
FDCAAHNFDPSTIPK	scaffold00039:+:250208-250252	SLFAGVPSLGENKPK	scaffold00166:+:94472-94516
FDSDGFLYLIDR	scaffold00211:-:44509-44544	SNFGTVAMDPTVR	scaffold05547:-:745-783
FQLSERPLVDEK	scaffold00014:-:193092-193127	TAFDLLDR	scaffold00111:-:306888-306911
FVHSENQHLVSPEGLDFLDK	scaffold00008:+:1853782-1853841	TDGTESPIITESDADTDKPR	scaffold00016:+:1504753-1504812

GGAFDGVEDVVSPFGYR	scaffold00176:+:46097-46147	TDPIFESLQPVDGK	scaffold00176:+:46214-46255
GGVIFVDQLPMTASGK	scaffold00211:-:44194-44241	TEQHSVQQQHQQQQVLQQQSHQVTQQR	scaffold00005:+:2642327-2642407
GLTAGINHEVTEGGENLSVGQR	scaffold00041:-:813668-813733	TGQGFNIPQGTFNSPTSGSAASK	scaffold01930:-:1403-1471
GRFDCAAHNFDPSTIPK	scaffold00039:+:250202-250252	TGSNIHATTTTSSTTTNCSISER	scaffold00003:-:3335933-3336001
GSYSLVEPDGTLR	scaffold00014:+:1655090-1655128	TGVILPIFK	scaffold01145:-:1534-1560
HEDNLKPEGAFTVPER	scaffold00025:-:490526-490573	THRPDYTDEQLEDIVDSVMR	scaffold00046:+:895296-895355
TPALIFEHVNNTDFK	scaffold00008:+:1853254-1853298	IDNDSRPSGATASTDTLAAFNR	scaffold00003:-:5502034-5502099
TPCPKPSIVCPTEETK	scaffold00001:+:3508684-3508731	IKVGKDDVDPAAVHASGK	scaffold01145:-:1885-1938
TTQASYEEMIAQCEK	scaffold00059:+:373380-373424	ILLPGVLR	scaffold00062:-:698521-698544
TVFYTADPVNGFNAVVQR	scaffold00014:+:1655129-1655182	IPYIECSAK	scaffold00014:-:193182-193208
VAMDCTEVEEGYK	scaffold01145:-:1744-1782	ISTGNPCPICR	scaffold00006:-:2909423-2909455
VEETEKPVTSTTAEDASSSEER	scaffold00052:+:1261577-1261642	ISTGNPCPICRDEYLVLDHNNIDLLK	scaffold00006:-:2909378-2909455
VGFNGLLQPFSDAVK	mitocondorial:+:2526-2570	ITYEGKDYQLQELGQVVR	scaffold00024:-:2127977-2128030
VGKDDVDPAAVHASGK	scaffold01145:-:1885-1932	IVFLFYVSLLMSCLFLLLLFSLHTAIVLIYLNVLYVLLFSVR	scaffold00042:+:208725-208850
VHNVAAPVDSYAGK	scaffold00001:-:3940812-3940853	KLGSGQFGDVWEGLWNNTTPVAIK	scaffold00078:+:700122-700193
VKDRDEFPMLMVGNK	scaffold00014:-:193275-193319	KLYPDAPSSTTFYK	scaffold00006:-:1951113-1951154
VLGTEDLFAYLDK	scaffold00008:+:1853674-1853712	KVDAAESVR	scaffold00055:-:1041763-1041789
VNGDEADPLYK	scaffold00016:-:1249993-1250025	KYGLNICR	scaffold00031:+:1038424-1038447
VPDSACAGTAFLTGVK	scaffold05547:-:784-831	LAANDFFR	scaffold00105:+:307731-307754
VPVLGFSETAQSVFR	scaffold00030:-:1974167-1974211	LFDEFGLNLCEAAIK	scaffold00006:-:1951485-1951529
VYISPAMGEAHK	scaffold01145:-:2262-2297	LIDWGLAEFYHPGQEYNVR	scaffold00008:+:1853449-1853505
WGDDHIPGSPFK	scaffold01145:-:1194-1229	LIHVLCSSHFSAR	scaffold00065:+:910319-910357
YSNSLEELLLDANHIR	scaffold00038:+:70351-70398	LKPLEVDLK	scaffold00390:-:8430-8456
ADGVELNPWLVYYSR	scaffold00001:-:2460313-2460357	LKPLEVDLKR	scaffold00390:-:8427-8456
AGGYGIQGLGGNFVER	scaffold00090:-:336293-336340	LKQNRPIPQWIR	scaffold00085:-:176797-176832
ALDYCHSLGIMHR	scaffold00008:+:1853359-1853397	LLETLFVHR	scaffold10320:+:176-202
ALLAISGYR	scaffold02626:-:770-796	LLNNLLLKR	scaffold00062:-:923510-923536
ALPVVAGVK	scaffold00004:-:3195014-3195040	LPQYQFLWK	scaffold00006:-:1950846-1950872
ALYDYEGAESDELTFK	scaffold01259:-:1039-1086	LSDYDQDGFLDIEEFALAMHLINVK	scaffold00176:+:46737-46811
AMALYEQHELIR	scaffold00107:+:135431-135466	LTLPSHLLSISMEFVPQKK	scaffold20480:+:96-152
ARTDPIFESLQPVDGK	scaffold00176:+:46208-46255	LVVENLSTR	scaffold00176:-:24719-24745
AVVPVAAGPAVFAPAPVAR	scaffold00014:+:1655201-1655257	LVVVGGGGVGK	scaffold00014:-:195411-195443
AYEELTEMIKNEEELK	scaffold00014:+:381399-381446	MDGNELPTALPLHLVPPSK	scaffold00176:+:46812-46868
CHLLHSAVHLNLPESICVHVFVQEAVR	scaffold00032:+:1077264-1077344	MDGNELPTALPLHLVPPSKR	scaffold00176:+:46812-46871

CHLTLVYSK	scaffold00006:-:2100609-2100635	MKLLNNLLLK	scaffold00062:-:923513-923542
CISLTELSFYCHFLTYFLCNLR	scaffold00021:-:1126772-1126837	MKLLNNLLLKR	scaffold00062:-:923510-923542
DCVWNPYTSLTK	scaffold00067:+:278101-278136	MNVDPFNNFSDDQVWK	scaffold00029:+:1925578-1925625
DINDVYSYTSISHK	scaffold00101:-:411617-411658	MQGLPPTDAAEEKR	scaffold00062:-:623746-623787
DKDTTTSSSSSTTADQTDASGVAK	scaffold00020:+:1229033-1229104	MYVLNDVPTLTEQLK	scaffold00059:+:373335-373379
DSLVVEIDQHHPVR	scaffold00007:+:1470728-1470769	NIGVFWVSCEAGSYIR	scaffold00062:-:698806-698853
DSSVNAVCYGAK	scaffold00062:-:698545-698580	NLLQVDLTKR	scaffold00055:-:1040971-1041000
ECTSESCGAGVFMAAHEDRHYCGK	scaffold00006:-:2100636-2100707	NQYTVNYLVR	scaffold01145:-:1260-1289
EFFLSLLK	scaffold00025:+:1306450-1306473	NVLSFVKPSAAGSK	scaffold00001:-:2460427-2460468
EGYDIDDANFGK	scaffold00007:+:390684-390719	NVPVLHHYTQPARAR	scaffold00080:+:108244-108288
EILQLLTAVSR	scaffold00028:+:1673989-1674021	AAADVEYPAYGVTGPNR	scaffold00162:-:80489-80539
FESDESKLPMPVPK	scaffold00006:-:1950804-1950845	AALENGTEGFVR	scaffold00211:-:44563-44598
FPSHFGSELKDLLR	scaffold00055:-:1041001-1041042	AAMGAEECDDLNDVCYTR	scaffold00067:+:273972-274025
FSPELLDSLR	scaffold00019:+:655259-655288	AASASASAAAR	scaffold00080:+:52503-52535
GALFQRPPLISAVK	scaffold00062:-:698914-698955	AATATITNPTTGVR	scaffold00003:-:3535935-3535976
GEGIDAGYGEVDWICNR	scaffold00176:+:46151-46201	ACATVILPNGHTYR	scaffold00067:+:278158-278199
GFANLWYSHPR	scaffold00031:+:1038159-1038191	ACPSLIAYVR	scaffold00003:+:67171-67200
GFDMITDAVFESHNK	scaffold00062:-:923414-923458	ACSNNHGMIR	scaffold00031:+:1038394-1038423
GGAFDGVEDVVSPFGYRK	scaffold00176:+:46097-46150	ADENENSCDKFDDVKPDK	scaffold00008:-:1622360-1622413
GGTNIITLLAVVK	scaffold00008:+:1853200-1853238	AEAEGGSEQDDVSFLR	scaffold00021:+:983940-983987
GGTNIITLLAVVKDPVSR	scaffold00008:+:1853200-1853253	AEDGDFPDETSR	scaffold00033:-:471913-471948
GKYSEVFEAIK	scaffold00008:+:1853074-1853106	AEFEQQIDR	scaffold00002:-:2434968-2434994
GMLTFDVPFR	scaffold00006:-:2909240-2909269	AEGYFGMYR	scaffold00105:+:307656-307682
GMLTFDVPFREYDYSEYYPAK	scaffold00006:-:2909207-2909269	AFTEMQVNK	scaffold00059:+:373170-373196
GPGDTSNFDDYEEETLR	scaffold00055:-:1040803-1040853	AFTEMQVNKIESTK	scaffold00059:+:373170-373211
GSAVNILLITPEK	scaffold00105:+:307683-307721	AGGGDESKLTGLSR	scaffold00062:-:167816-167857
GVDTSVAQYTGGSK	scaffold00006:+:3379442-3379483	AGKFDDIMYFYESPR	scaffold00003:-:5502730-5502774
HDFNATADDELSFRK	scaffold00084:+:68299-68343	AGTININEGTVR	scaffold00059:-:868989-869024
HIPDSPFK	scaffold01145:-:2298-2321	AGTTESGVPTAAAR	scaffold00003:-:5500060-5500101
HTQLLVAVER	scaffold00006:-:2909282-2909311	AGTTHSASDYPVQR	scaffold00196:-:7389-7430
HVGQCVSIHLK	scaffold00065:-:469554-469586	AHSEDPDIDPPSER	scaffold00057:-:959764-959805
IDNDSRPSGATASTDTLAAFNR	scaffold00003:-:5502034-5502099	AIAQSDEELLK	scaffold12641:-:501-533
MQDVVEDVSNR	scaffold00038:+:1578069-1578101	AIDELLAAK	scaffold00108:+:378803-378829
MSAGNQPPPPPPPPGLR	scaffold00006:-:3080815-3080865	AIEAYESDFK	scaffold00003:+:1611349-1611378
MSAVTASTGTGGK	scaffold00039:-:1629960-1629998	AINPSPTINEK	scaffold00137:+:91283-91315
MVLAGVGVPHDELVR	scaffold00006:+:3379325-3379369	AKLEESDYDSEAK	scaffold00058:+:1121620-1121658
NAATNAVAPVNDNADDGAEEVERR	scaffold00002:+:2440495-2440566	ALPIRIDNDSRPSGATASTDTLAAFNR	scaffold00003:-:5502034-5502114

NCLSSLEDTEQQK	scaffold00067:+:274677-274715	ALTEQAVIDHCAR	scaffold00211:-:44284-44322
NDPESYISVEQFR	scaffold00032:-:1917186-1917224	ANDTAAAPAGSPSSK	scaffold00079:+:184020-184064
NDPIPNDPIVK	scaffold03440:-:1296-1328	ANSDAALGK	scaffold00020:+:5932-5958
NEANFYNVVLK	scaffold04388:-:152-184	APSDEEVDTYVR	scaffold00028:+:704894-704929
NEDVELPDSDESLYWCK	scaffold01723:+:160-210	AQCLTVLPALR	scaffold00159:+:22266-22298
NEINEMR	scaffold00027:+:1505408-1505428	ASEDDTIFER	scaffold00034:+:417087-417116
NHSIYSGQIK	scaffold00025:-:1799164-1799193	ASVTATEQSR	scaffold00021:-:1312049-1312078
NIVNIIDVYENSYSGNR	scaffold01239:-:247-297	ASVTATEQSRK	scaffold00021:-:1312046-1312078
NLNPADQQQLR	scaffold00038:-:1247144-1247176	ATASPWPGTTIR	scaffold00343:+:5195-5230
NLPEPVAQPR	scaffold00034:+:417687-417716	ATTQALQDFAR	scaffold00022:+:1585519-1585551
NNNNNINVTDEPR	scaffold00036:+:1119237-1119275	AVHMSWFQK	scaffold00006:-:1950387-1950413
NNQVLESEFR	scaffold00046:+:1174849-1174878	AVQDDEDSLETYK	scaffold00002:-:2434863-2434901
NPAEVLSPSVHILSNR	scaffold00004:+:4379805-4379852	AVQDDEDSLETYKQTEAR	scaffold00002:-:2434848-2434901
NQMILPDDKNPK	scaffold00044:+:699234-699269	AVWEGKK	scaffold06797:-:640-660
NRPETSANSGIESDNHSILA	scaffold00008:+:3535979-3536038	AYAGWVQEQYR	scaffold00003:-:5501743-5501775
NSAFYSMAK	scaffold00041:-:813321-813347	AYDVEPASADINDPYAR	scaffold00003:+:1610112-1610162
NSTALSITNSTGNNR	scaffold01259:-:1415-1459	CDSSQSSVECAVCSR	scaffold00067:+:278482-278526
NVLVGDNNIVK	scaffold00078:+:700635-700667	CESDSSWSDCSASAR	scaffold00020:-:1817250-1817294
NVNCSVDESIAR	scaffold00090:-:209637-209672	CGHTNNLRPK	scaffold00057:-:631069-631098
QAEETAEFER	scaffold00097:+:301437-301466	CGPLDNSSPITK	scaffold00003:-:5499616-5499651
QAGEVTYADAHK	scaffold00176:-:20232-20267	CGPLDNSSPITKK	scaffold00003:-:5499613-5499651
QCFPDTVATVPCR	scaffold00066:-:720454-720492	CIDVYTPK	scaffold02078:-:1876-1899
QCFREYAKDIGFR	scaffold00031:+:1038448-1038486	CKGPGDTSNFDDYEEETLR	scaffold00055:-:1040803-1040859
QCVIDDTPAK	scaffold00014:-:193552-193581	CNFGTTANISTR	scaffold00282:+:1789-1824
QDGAAVVEGLK	scaffold02563:+:179-211	CSSDASTAAFCDDPFTQDIITEQQR	scaffold06372:+:916-990
QDGAAVVEGLKGPSAR	scaffold02563:+:179-226	CTDPHDLSCVVCEEPGCNQNPWTK	scaffold00067:+:274716-274787
QEAAWELPNSTLPR	scaffold00041:-:814200-814241	CVACSEEGCNDLPK	scaffold00067:+:275714-275755
QEFSQQQQFSSSQQR	scaffold00005:+:2642270-2642314	CYNDLATALIVR	scaffold00211:-:44353-44388
QEGYIPSNYVAK	scaffold00078:+:631302-631337	DAEEAPEGGDEGEEQPMDQTVDGDEEPKK	scaffold00062:-:697972-698058
QEPDAEIENLGELQK	scaffold00062:-:699391-699435	DAETNEIVAVK	scaffold00020:+:1854162-1854194
QFISPQTGEVLSYR	scaffold00006:-:2909336-2909377	DALEVVDGEGNQPK	scaffold00081:+:346897-346938
QGEVTPVPPSGVLPWQDAR	scaffold00165:+:105151-105207	DAPNLLCTPHAAFYSEAATTELR	scaffold06271:-:600-668
QIDEREEIPPNQQR	scaffold00001:+:5040619-5040660	DDQYQQFTTNNTFVTSPVR	scaffold00036:+:570680-570736
QIEYVDNR	scaffold00038:+:70288-70311	DDVDPAAVHASGK	scaffold01145:-:1885-1923
QIVEENVATVK	scaffold00063:-:146364-146396	DEEGDNELGEGALR	scaffold00025:+:1306561-1306602
QLEALEPSCK	scaffold00036:+:1226934-1226963	DEVAHTMAESR	scaffold00002:+:3354518-3354550

QLKIPYIECSAK	scaffold00014:-:193182-193217	DGEECDKAAMGAEECDDLNDVCYTR	scaffold00067:+:273951-274025
QMLQFVAAAK	scaffold00025:+:1306525-1306554	DGMVTMHDVLDAQYLYENHKDESMLR	scaffold00062:-:698638-698715
QNRPIPQWIR	scaffold00085:-:176797-176826	DISEICADK	scaffold00006:+:2906539-2906565
QQALEQNQHAA	scaffold00021:-:2216504-2216536	DIVTPLPPLSQPIPDLPPVQYARPGDQK	scaffold00006:+:3378683-3378766
QQELNIIEK	scaffold00002:-:2442329-2442355	DLDASNQTILDLNAK	scaffold15230:-:185-229
QQLSCAQEQTDDGSVQYCR	scaffold00067:+:278833-278889	DLEFEIEPR	scaffold00121:-:269990-270016
QQNCPVMGAMK	scaffold00026:+:1730206-1730238	DLERFFKGYGR	scaffold00176:-:25311-25343
QQQQQFVQR	scaffold00005:+:2642243-2642269	DLLDELDK	scaffold00028:+:1673563-1673586
QQTNGTCFIQAK	scaffold00003:-:5502694-5502729	DLLGAAETGSGK	scaffold00625:-:5095-5130
QSAELENTMTSVER	scaffold00121:-:270170-270211	DMVEATGPSESESLKR	scaffold00096:+:9219-9266
QSAEVANFMMSVER	scaffold00001:-:1421814-1421855	DNTLGFPK	scaffold00006:+:3379220-3379243
QSDATNGSVKDLEQMK	scaffold00007:-:457211-457258	DQQIELDEYGK	scaffold00029:+:667550-667582
QSFGDDAPR	scaffold00006:+:10101-10127	DQTDGNAEALQAEIDR	scaffold15230:-:335-382
QSPSVCALVGR	scaffold00015:+:849501-849533	DSSSSSFTPSINK	scaffold00130:+:282879-282917
QSQSTVPQPNR	scaffold00036:-:494823-494855	DTAIETVTTCATVCPQVFDPR	scaffold00012:+:2109029-2109091
QSTETDPEKIEK	scaffold00003:-:4538004-4538039	DTFVYAASADSR	scaffold00006:+:3379013-3379048
QTTESIQQMTAVER	scaffold00096:+:8345-8386	DTIVLDPTSSDIR	scaffold00010:+:2763792-2763830
QTYQDDPVWK	scaffold00006:-:2909528-2909557	DVKPHNVMIDHENR	scaffold00008:+:1853398-1853439
RFPPLVHR	scaffold00018:+:425912-425935	DVSTIPCFR	scaffold00057:-:959651-959677
RLEDLSDAIVQDFALMR	scaffold00390:-:8379-8429	DWLNSYTDFNR	scaffold00062:-:698224-698256
RPANQFIPR	scaffold02078:-:1759-1785	DWPEQGMVEFR	scaffold00041:-:814167-814199
SADAELSSILNRR	scaffold00017:-:1810681-1810719	DYQLQELGQVVR	scaffold00024:-:2127977-2128012
SAGNAVWLSDER	scaffold00004:+:4379331-4379366	EAAEDCVER	scaffold23100:-:142-168
SASTSATTGAAKPPASK	scaffold00006:+:3380982-3381032	EAFWEAYQTQAK	scaffold00003:-:5502142-5502177
SCASVLTSK	scaffold00015:-:1273451-1273477	EAHVDEFVR	scaffold00013:-:2306896-2306922
SEPYIGMIDPR	scaffold00120:-:129646-129678	EALSYREEYQK	scaffold00001:+:5289529-5289561
SFSSCVQCSSMDESSDCNK	scaffold00067:+:274479-274535	EAPCWVNEEGAPFR	scaffold00003:-:3103380-3103421
SGFAAEAQR	scaffold00033:-:1185061-1185087	EATVSGFGR	scaffold00059:-:868806-868832
SGIQSEKDGMVTMHDVLDAQYLYENHKDESMLR	scaffold00062:-:698638-698736	EEISASVVDAEER	scaffold00005:+:2443845-2443883
SGMNLNPQQDGTTLFIPVPK	scaffold00024:-:2127839-2127898	EETKPDNAATDEQDAK	scaffold02563:+:26-73
SISFYQTSVR	scaffold00001:-:1422474-1422503	EFDAPHVLLQQEGGILK	scaffold00096:+:9168-9218
SIVKDIDKLLEA	scaffold00016:-:1249837-1249872	EFEEETGLSK	scaffold00142:+:17288-17317
SLADEMDSTFAELAGY	scaffold00001:+:1788171-1788218	EFLDQAKEDFEEKWK	scaffold00055:-:1041718-1041762
SLLSGLLVK	scaffold00002:+:3358764-3358790	EINKSDVEFEILDT	scaffold00055:+:92847-92888
SLNYLLATNSNPSEATER	scaffold00055:-:1132789-1132842	EITGCSIQVASEMLPNSTER	scaffold15831:+:110-169

SLTGSGPMYVGK	scaffold00071:-:240832-240867	EIVADSIANSIK	scaffold00105:+:68224-68259
SLVEVITR	scaffold00006:+:3379790-3379813	ENLDCPELIK	scaffold00053:-:936009-936038
SNLLVMEDLTSAGYR	scaffold00057:-:347042-347086	ENLNPWDFK	scaffold00165:+:103107-103133
SPQSVEVLSDPTPR	scaffold00034:+:416709-416750	EPSSIDDTNTPPK	scaffold00059:+:1100324-1100362
SQLETLLSETQK	scaffold00013:+:2776492-2776527	EQIFDPPLDAPK	scaffold00036:+:551194-551229
SQTIWSDSLR	scaffold00003:-:5499745-5499774	EQPGGNNPWEAEMVR	scaffold00058:+:1121347-1121391
SSDGSDAVIVENGQNISSSTPPMETHK	scaffold00011:-:699575-699655	ESVSDIEQNR	scaffold00038:+:1583652-1583681
SSSGDFGNPLR	scaffold00021:-:2200505-2200537	ETPFIEEIEAR	scaffold00002:+:3204530-3204562
SSSGDFGNPLRK	scaffold00021:-:2200502-2200537	EVFGLTLPLITNEEGDKFGK	scaffold00004:+:4379271-4379330
SSYWEPPPQAIAR	scaffold00034:+:416162-416200	EVQISQSMTSTR	scaffold00009:-:3394833-3394868
STINDESFFR	scaffold00196:-:7104-7133	EYDHLFK	scaffold00014:-:1005787-1005807
STLSCVQCDSR	scaffold00067:+:278800-278832	EYWDYENYIVDWVNQDDYQLVR	scaffold00008:+:1852996-1853061
STPDSTSGQYLR	scaffold00038:+:1584232-1584267	FAGANPAAQPITSIEK	scaffold00132:-:209876-209923
STTGSIETLR	scaffold00024:-:2128031-2128060	FDAQVDHDVNGVNYAKIHQWYVQR	scaffold00003:-:5501116-5501187
SVINASSEYR	scaffold00002:-:2442773-2442802	FIITNETEGEVEK	scaffold00002:-:2442815-2442853
SYSWDQR	scaffold00006:+:1479626-1479646	FLEVLQNDFR	scaffold00019:+:1276744-1276773
TAPTVPFFTAETHQYSR	scaffold00003:-:3103638-3103688	FNGVHIPGSPYR	scaffold01145:-:1939-1974
TASENDDENSQLLK	scaffold00046:+:903860-903901	FQDVMDAAK	scaffold00090:-:209610-209636
TAVQCEQQEK	scaffold00003:-:5501086-5501115	FREAFWEAYQTQAK	scaffold00003:-:5502142-5502183
TAVQCEQQEKENVQISCGCFHDVK	scaffold00003:-:5501044-5501115	FTEETLDEIVR	scaffold04388:-:395-427
TCECNLQNSQNR	scaffold00059:+:796525-796560	FVQEYGVDTSK	scaffold00159:+:22542-22574
TCECNLQNSQNRR	scaffold00059:+:796525-796563	FVSFAHTK	scaffold00057:+:391688-391711
TDADGSLQR	scaffold00067:+:278923-278949	FYDTNPSGR	scaffold00001:-:1422390-1422416
TEDRIELITR	scaffold00114:-:290562-290591	GCLSDLSDPEVIASCEQGDVDCAVCDGSGCNAGFRPK	scaffold00067:+:278689-278799
TEELNQETR	scaffold00013:+:2776564-2776590	GCLSQLTATNQK	scaffold00067:+:276437-276472
TGHSGTLDPK	scaffold00062:-:699106-699135	GCSAESEAICEGNNR	scaffold00067:+:273321-273365
TGMTQAVASPR	scaffold00034:+:417654-417686	GDAYLSELYR	scaffold04388:-:308-337
TGTGTATGETPATTSK	scaffold00090:+:214250-214297	GDRLEQFSGHHLTADAWSR	scaffold00003:-:5501302-5501358
TIAIDTEPEATVESVK	scaffold00001:+:5040568-5040615	GGGGGPWAGSNSSQR	scaffold00016:+:408815-408859
TIAIDTEPEATVESVKK	scaffold00001:+:5040568-5040618	GKYSEVFEAIKMTSNEK	scaffold00008:+:1853074-1853124
TIQAAVCR	scaffold00137:+:91480-91503	GPCDEVFIK	scaffold01145:-:1302-1328
TLGTGSFGR	scaffold00055:-:1041643-1041669	GQIVEFAPPAELLQSK	scaffold00041:-:813348-813395
TLLAAGAAVNENTLR	scaffold00196:-:4908-4952	GTLPAQNCPQEDDVCFSR	scaffold00067:+:272944-272997
TNHYTPLPFGSSPLNR	scaffold00062:-:699244-699291	GYSSEENTWEPEENLDCDDLIQAFK	scaffold10785:+:685-759
TQYSNVDQR	scaffold00049:-:437150-437176	GYSVENQTLGLSYEQSK	scaffold00003:+:1611067-1611117

TRLQNQQVGPNGEK	scaffold00105:+:307059-307100	HAVSGAVESATVR	scaffold00088:-:447545-447583
TSDGSQATSNVVR	scaffold00008:-:2926736-2926774	HDFNATADDELSFR	scaffold00084:+:68299-68340
TSVNSVTSQSFR	scaffold00009:-:3394797-3394832	HISSLAPDTR	scaffold00059:+:373284-373313
TVDGEGVER	scaffold00014:+:71164-71190	HMPLPTK	scaffold00002:+:2027787-2027807
TVDWLAGESDR	scaffold00022:-:744262-744294	HSLEQSDDGEGVEVIENK	scaffold00026:+:1143685-1143738
TVDWLAGESDRTTV	scaffold00022:-:744253-744294	IAAPLAYPAVAK	scaffold00014:-:1641207-1641242
TVLIGNVDDR	scaffold00052:+:229864-229893	IAARPFVNSSSSSNAQK	scaffold00003:-:3103920-3103970
TVPAQAPFTVAPLPER	scaffold00036:+:1480437-1480484	IAIEDDQSR	scaffold08424:+:80-106
TVPSYDDHHVAK	scaffold00064:+:871375-871410	IAQFICEK	scaffold00065:-:655013-655036
TWTLCGTPEYLAPEIILSK	scaffold00055:-:1041172-1041228	ICQSSSSVACHTCEGNGCNTPK	scaffold00067:+:276473-276538
VAAPVVAK	scaffold00014:-:1641557-1641580	IDIVHQNIEWQR	scaffold00057:+:391643-391678
VAIASNPPTSPPQPK	scaffold00034:+:417246-417290	IEAPFIPR	scaffold00055:-:1040860-1040883
VAMDCTEVEEGYKVR	scaffold01145:-:1738-1782	IELINQLSETGLR	scaffold00090:-:209956-209994
VASTTSSPPSSPIHVVPITPIPR	scaffold00034:+:416751-416819	IEQDSSATSEPTR	scaffold00105:+:68266-68304
VAYQYQSNVPDNLAMDAMK	scaffold00028:+:1674025-1674081	IEQDSSATSEPTRR	scaffold00105:+:68266-68307
VDAAESVREFLDQAKEDFEEK	scaffold00055:-:1041724-1041786	IGDFGLATAK	scaffold00062:-:1079882-1079911
VDEKGFLVR	scaffold00067:+:278662-278688	IHELQSELDNVREQLGDAK	scaffold00002:-:2441627-2441683
VDGEQVFEPEQPHVALR	scaffold00065:+:910415-910465	IIGPGAQNNDGAK	scaffold00003:+:3418144-3418182
VERLGWTPFEIPR	scaffold00053:+:827517-827555	IKQEPDAEIENLGELQK	scaffold00062:-:699391-699441
VFVPALPEEQR	scaffold00027:+:1163849-1163881	IKTLGTGSFGR	scaffold00055:-:1041643-1041675
VGQVTNDVQYI	scaffold00027:+:1029094-1029126	ILAQKYNCDK	scaffold00057:-:631162-631191
VIPVETSIR	scaffold00006:-:2909582-2909608	ILIPVTSVGPTVEVASGK	scaffold00008:+:651542-651595
VKDLELSTR	scaffold00010:-:2139651-2139677	ILSEVVLRPK	scaffold00006:+:3379070-3379099
VLAEPTEASKR	scaffold03498:-:492-524	ILVMDAGEAR	scaffold00096:+:9138-9167
VLEQQIEQLNK	scaffold00013:+:2776738-2776770	INAEDQALDQIAK	scaffold00004:-:2679301-2679339
VLGTEDLFAYLDKYNIELDPR	scaffold00008:+:1853674-1853736	INLDPFNAQSDDDIWK	scaffold00041:-:813852-813899
VLGYIQIR	mitocondorial:+:2487-2510	INTTGGDYK	scaffold00047:+:883-909
VLVTGGQDR	scaffold00015:-:117605-117631	IQDPQLAK	scaffold00114:-:288795-288818
VNLSANPR	scaffold00001:+:1433374-1433397	IQSTNEEFEVAR	scaffold00002:-:2434245-2434280
VPVDAIPQGR	scaffold00003:-:3103608-3103637	ISADQSNQLVQMALESK	scaffold16294:-:526-576
VSFSNVNPIIQYK	scaffold00006:-:1951065-1951103	ITEDDFER	scaffold00067:+:274347-274370
VSPEELEQR	scaffold00053:+:1177731-1177757	ITFDGIPSAPER	scaffold00034:+:415954-415989
VSSAASASPMSGRR	scaffold00039:+:990562-990603	ITGPAGGLGSGGATGGGAGGGGQQK	scaffold00007:-:1547741-1547815
VTDFGFAK	scaffold00055:-:1041244-1041267	ITILECLSDGDPSIQR	scaffold00002:-:4298465-4298512
VVAPSNNEDDCFIQLIDQDQYSVR	scaffold01145:-:2017-2088	ITPPFKPQVTSDTDTR	scaffold00002:+:3358893-3358940
VVEYDNVDPEPALEAPADK	scaffold00121:-:270113-270169	IVDHVDIAFEEPSEK	scaffold00026:+:1144296-1144340
VVNIYTEFHEFSR	scaffold00017:-:1810588-1810626	IVFPEAER	scaffold00058:+:1125819-1125842
VVSLEEAQQLSR	scaffold00014:-:193218-193253	IVGGSTVKEMEIK	scaffold00220:+:37603-37641

VVVLGSGGVGK	scaffold00062:-:425137-425169	KINISALEK	scaffold00072:+:512298-512324
VVYPIEELK	scaffold00024:-:612617-612643	KNPAAGGSPAATTKPTPEK	scaffold00016:+:2311508-2311564
VYADVNSHKPR	scaffold00008:+:1852963-1852995	KPANPFKDIDR	scaffold04388:-:92-124
VYISGMSGNK	scaffold00005:-:1223344-1223373	KQEEVLQESMMMIPDCQR	scaffold00014:+:381243-381296
VYVGGLPYGVR	scaffold00176:-:25350-25382	KQLIASGKLDK	scaffold00062:-:698287-698319
WAPDCLLAR	scaffold00057:-:347087-347113	KWGLGPK	scaffold00062:-:698332-698352
WGDDHIPGSPFKVEV	scaffold01145:-:1185-1229	LAFQSTSEYGER	scaffold00006:+:3378917-3378952
WNCNTATVR	scaffold00067:+:278527-278553	LAPEEEMAHFVDQSR	scaffold01221:-:2398-2442
WVPIGNQSTVYTEAAVGPIDDDILISK	scaffold00025:-:1799083-1799163	LCDFNPK	scaffold05080:-:1091-1111
WVPQMGDNAEVLK	scaffold00090:-:209884-209922	LCDGTSDCYMGSDELR	scaffold00019:-:519150-519197
YAPTTSPK	scaffold00016:-:1249873-1249896	LDANSDRPNDDWGLEGK	scaffold00125:-:72542-72592
YDLGTCNLITELGNVSPER	scaffold00052:+:229343-229399	LEDEDEQGWCK	scaffold01259:-:391-423
YDYDDHDQLIK	scaffold09901:+:490-522	LEESDYDSEAK	scaffold00058:+:1121626-1121658
YEPVNLEK	scaffold00674:+:4107-4130	LIEEDSADVRK	scaffold00122:+:273655-273687
YFAQGSATWENEK	scaffold00006:+:3379382-3379420	LIKEDEYEAR	scaffold00078:+:700757-700786
YFNGETMR	scaffold00062:-:167792-167815	LKQVEHTLNEK	scaffold00055:-:1041535-1041567
YGGFVAPK	scaffold00038:+:1584397-1584420	LKQVEHTLNEKR	scaffold00055:-:1041532-1041567
YGKPNENTPK	scaffold00062:-:698257-698286	LLDIGSGDGR	scaffold00001:-:2460397-2460426
YGLNICR	scaffold00031:+:1038427-1038447	LLDVYPFGER	scaffold00001:+:2049109-2049138
YKSLADEMDSTFAELAGY	scaffold00001:+:1788165-1788218	LLDYNDQR	scaffold00062:-:698854-698877
YLDDAHLKGFEK	scaffold00106:-:134963-134998	LNDSLYLCVR	scaffold00003:-:5499652-5499681
YLQAQPEFIFPEGASK	scaffold00059:+:369933-369980	LPPSENYEGFVER	scaffold00137:+:90900-90938
YLSALQYPR	scaffold00001:-:2050017-2050043	LPVKDGQYSVLNDR	scaffold23207:+:361-402
YNPTWHCIVGR	scaffold00327:-:11528-11560	LQNQQVGPNGEK	scaffold00105:+:307065-307100
YNQTHAGYAQTFK	scaffold00014:-:195444-195482	LSEHAFAPTK	scaffold00065:-:655376-655405
YSNSLEELLLDANHIRDLPK	scaffold00038:+:70351-70410	LSIGDTTTADDSMAAAGEGEK	scaffold00062:-:698086-698148
YVDNHEASTSSEVAR	scaffold00196:-:51662-51706	LSNQIACIK	scaffold00030:+:638879-638905
YVVEQSEAK	scaffold00025:+:1306402-1306428	LSPGIVPK	scaffold00047:+:859-882
YYSENNGLDGYK	scaffold00020:-:1147183-1147218	LTIPLVLDETSAGLNAK	scaffold06255:-:1051-1101
ACALVTDFEQLPDGDR	scaffold00001:-:1423504-1423551	LVHGEAGLEK	scaffold00004:+:4379544-4379573
QNILFGQPMEK	scaffold00001:-:1423576-1423608	LYPDAPSSTTFYK	scaffold00006:-:1951113-1951151
IAVCSVIYR	scaffold00001:-:1424710-1424736	MACHIGYYTPK	scaffold00006:+:2906566-2906598
ILVLDEATANVDPDTDR	scaffold00001:+:1437572-1437622	MDFTEVFR	scaffold00065:+:910370-910393
NLDPFEAYPDAALWR	scaffold00001:+:1437377-1437421	MDTDVIK	scaffold00166:-:39462-39482
EIGDDLDTAFVELILK	scaffold00001:+:1775827-1775874	MEETENTTPVATAEESPAEDVPFDKEFK	scaffold02563:+:74-157
EIGDDLDTAFVELILKE	scaffold00001:+:1775827-1775877	MEHINTPK	scaffold00139:-:119821-119844
YREIGDDLDTAFVELILKE	scaffold00001:+:1775821-1775877	MEKSEGFDDYMK	scaffold06797:-:598-633

TNMEDDRVAILEAQLAQAK	scaffold00001:+:1758539-1758595	MFCLMTGDPHQR	scaffold00142:+:47497-47532
VAILEAQLAQAK	scaffold00001:+:1758560-1758595	MLSYEDQQQAR	scaffold00196:-:7266-7298
ATQREESYGGQVR	scaffold00001:+:1768497-1768535	MMDLNNDGFVDYSEYR	scaffold00046:+:895356-895403
EESYGGQVR	scaffold00001:+:1768509-1768535	MNKEAAEDCVER	scaffold23100:-:142-177
LIAEEADKK	scaffold00001:+:1758596-1758622	MPVYGTAMDK	scaffold00002:+:3354264-3354293
MLQEEMEATLHDIQNM	scaffold00001:+:1772316-1772363	LGVKPIEVI	scaffold00010:+:3460521-3460547
SAGIEVAR	scaffold00001:+:3957859-3957882	LLTYNFLTSK	scaffold00010:+:3461079-3461108
EVEVLPCEVCGK	scaffold00001:-:4244601-4244636	YSISNLLDCMEEDEERHDGIGLV SLAQR	scaffold00011:+:94794-94877
FIQATQLR	scaffold00001:-:4244226-4244249	TMEAEALGR	scaffold00011:+:506338-506364
DVYGNEVQR	scaffold00001:+:5233750-5233776	SVQEQYDTEANLTECAR	scaffold00011:+:752892-752942
EKDVYGNEVQR	scaffold00001:+:5233744-5233776	YGPAVTCDQCK	scaffold00011:+:826865-826897
QLTEATIR	scaffold00002:+:1110840-1110863	SLSEQSEENATGK	scaffold00012:-:764017-764055
NPFTDKLVNER	scaffold00002:+:1395112-1395144	NCLVGSENTVK	scaffold00012:-:940354-940386
RPPRDPFNDKLVNER	scaffold00002:+:1393430-1393474	VGTMSEDDFIEEAK	scaffold00012:-:940775-940816
DLGEYAKDEAR	scaffold00002:-:1935947-1935979	AVNVPTINQEECDK	scaffold00012:-:1266507-1266548
QYEMGECTR	scaffold00002:+:2519574-2519600	AVNVPTINQEECDKAYR	scaffold00012:-:1266498-1266548
CDGEPDCPDGSDER	scaffold00002:-:3923485-3923526	HASDGQVAR	scaffold00012:-:1266762-1266788
GVPYQNLVIGVPK	scaffold00003:-:53528-53566	HNCGGSVLSSK	scaffold00012:-:1261578-1261610
SSCTTPAPSIEAPECIK	scaffold00003:-:2992149-2992199	LSFEPVAATLAPK	scaffold00012:+:2908389-2908427
DAGFESNSDINGAR	scaffold00003:+:3483167-3483208	KLSFEPVAATLAPK	scaffold00012:+:2908386-2908427
DYDSWAELGNANWGWR	scaffold00003:+:3482963-3483010	FGFSSTIPR	scaffold00012:+:2908359-2908385
EDVEAVLR	scaffold00003:+:3483986-3484009	SFFVEALDAIPSVPESEQR	scaffold00013:+:1584669-1584725
GLDEYDFVIVGAGSAGSVVANR	scaffold00003:+:3478728-3478793	DVILQDDDVDCTMTEK	scaffold00013:-:1857291-1857338
GSDLIKEDYDYVGSK	scaffold00003:+:3486121-3486165	IVDYTADDVR	scaffold00014:-:559469-559498
IESGYLEAR	scaffold00003:+:3483959-3483985	SYQLDGNTGSVALSELIQK	scaffold00014:+:1303232-1303288
IVDTPTFR	scaffold00003:+:3484031-3484054	HQIDSVPTVLFLR	scaffold00014:-:1477326-1477364
LLDTENFGYHELK	scaffold00003:+:3485818-3485856	QITDVMTELAK	scaffold00014:-:1477434-1477466
MGPEGDQSAVVDSR	scaffold00003:+:3485974-3486015	SFAAPVAYSSPVAYTAPVVAR	scaffold00014:+:1662316-1662378
TSDPYDAPIINANYLDDQR	scaffold00003:+:3485719-3485775	DIPVAVATPAVAVAAK	scaffold00014:+:1662805-1662852
VGNFPVNHPLAEVMLQAFK	scaffold00003:+:3483110-3483166	AAPIAAAAYTAPVVAAAAAPAV VAAR	scaffold00014:+:1662433-1662510
VIDASIMPDIVSGNTNAPTIMIGEK	scaffold00003:+:3486046-3486120	AAPVVAAAPVAAAPVVAAR	scaffold00014:+:1662511-1662567
VTLATENFNEFNPPR	scaffold00003:+:3483914-3483958	AAYTAAAPVAYSAPVVAR	scaffold00014:+:1662379-1662432
VVDGSIMPNIVSGNTNAPIMMIGE K	scaffold00003:+:3484259-3484333	GSYSLIEPDGSR	scaffold00014:+:1662712-1662747
YMSTTIYHPTGTAK	scaffold00003:+:3485932-3485973	GSYSLIEPDGSRR	scaffold00014:+:1662712-1662750
AEYHAQGGLLK	scaffold00003:+:3484876-3484908	IVNYYADPINGFNAVVQK	scaffold00014:+:1662751-1662804
ASDMIKEDWGVGPK	scaffold00003:+:3484334-3484375	RIVNYYADPINGFNAVVQK	scaffold00014:+:1662748-1662804

DGAFWPR	scaffold00003:+:3482879-3482899	TQEETRDGDVVK	scaffold00014:+:1662676-1662711
DYGPTALAR	scaffold00003:+:3478701-3478727	TVVAPAVAAK	scaffold00014:+:1662853-1662882
GRVTLATENFNEFNPPR	scaffold00003:+:3483908-3483958	VSGGNATATEK	scaffold00015:-:147900-147932
HMTLTLYHPVGTAK	scaffold00003:+:3484145-3484186	LGYYEGDAYK	scaffold00015:-:1374694-1374723
IAECDQLEYDSDSYWECYAR	scaffold00003:+:3485872-3485931	LGYYEGDAYKQEL	scaffold00015:-:1374685-1374723
LLDTENFGYHELKDFHLK	scaffold00003:+:3485818-3485871	VGNENSTANSAVANEAAVK	scaffold00015:-:1907075-1907131
LSECDQQTYDSDAYWECYSR	scaffold00003:+:3484085-3484144	TLCGTPNYIAPEILNK	scaffold00016:-:1448302-1448349
LSENPDWK	scaffold00003:+:3478794-3478817	SYTEAEVSDICR	scaffold00017:-:472148-472183
ASTSPEGCFWPR	scaffold00003:+:3479155-3479190	LHTGETPYQCTYCQK	scaffold00017:-:972766-972810
DYDSWAELGNANWGWRDVLPYFK	scaffold00003:+:3482963-3483031	LHTGETPYQCSYCQK	scaffold00017:-:972165-972209
EDWGVGPK	scaffold00003:+:3484352-3484375	IHTGETPYR	scaffold00017:-:984470-984496
EDYDYVGSKDSTTHTEL	scaffold00003:+:3486139-3486189	GENSYGIDIFCK	scaffold00017:+:1024997-1025032
EFGISVVSDVPVGR	scaffold00003:+:3479806-3479847	VPPLKVEDLIQDLK	scaffold00017:+:1151952-1151993
FHASGGYLK	scaffold00003:+:3483083-3483109	LLVQVNIPHKAAEEQAALER	scaffold00017:+:1194608-1194667
GSKDGAFWPR	scaffold00003:+:3482870-3482899	AAEEQAALER	scaffold00017:+:1194638-1194667
HGSYWPR	scaffold00003:+:3484681-3484701	TYAEVVAGQEAER	scaffold00017:+:1194839-1194877
IESGYLEAREDVEAVLR	scaffold00003:+:3483959-3484009	LSAELEELR	scaffold00017:+:2154357-2154383
KSEDNNDPELLR	scaffold00003:+:3483032-3483067	ALEDITAACIVDQFQNK	scaffold00017:+:2399407-2399457
LLEAAQESGYEYLEDFNGEDHVGFGR	scaffold00003:+:3479434-3479511	LYDDNIWQQEQQQK	scaffold00017:+:2425669-2425710
MGPDSDKGAVVDGR	scaffold00003:+:3484187-3484228	AWGQGQQQQQGPEGGQQQQKPK	scaffold00018:-:271374-271439
MGPSEDPDAVVDSR	scaffold00003:+:3480481-3480522	SLLGYGDTEMER	scaffold00019:-:318363-318398
SEDNNDPELLR	scaffold00003:+:3483035-3483067	YNTDDSISIEDTLEETGWK	scaffold00019:+:411284-411340
SEDNNDPELLRQDGGK	scaffold00003:+:3483035-3483082	TTGCVEGECVR	scaffold00019:-:506542-506574
SGIAGQWLR	scaffold00003:+:3480079-3480105	FEVPNQSDER	scaffold00020:+:735459-735488
STAVDFDLNR	scaffold00003:+:3479896-3479925	YADASKFEVPNQSDER	scaffold00020:+:735441-735488
TDADWENYAER	scaffold00003:+:3482819-3482851	DLKPENILLDDNLNVK	scaffold00020:+:1860404-1860451
TSDPYDAPIINANYLDDQRDVK	scaffold00003:+:3485719-3485784	ITDFGFAR	scaffold00020:+:1860452-1860475
VGNFPVNHPLAEVMLQAFKDAGFESNSDINGAR	scaffold00003:+:3483110-3483208	TMYLEDR	scaffold00021:-:2185703-2185723
VKGVEGLR	scaffold00003:+:3484235-3484258	LLFLGLDNAGK	scaffold00022:-:350247-350279
VKGVEGLRVVDGSIMPNIVSGNTNAPIMMIGEK	scaffold00003:+:3484235-3484333	NNAGPDDDGWTDVK	scaffold00022:-:1024535-1024576
VTLATENFNEFNPPRIESGYLEAR	scaffold00003:+:3483914-3483985	GGIDESEQFFDLK	scaffold00022:-:1085120-1085158
VVDASVMPDIVSGNTNAPVIMIAEK	scaffold00003:+:3480553-3480627	EWGNPLPDMEGFQPEHSDH	scaffold00022:-:1352903-1352959
YINKIVDTPTFR	scaffold00003:+:3484019-3484054	HVIITTWK	scaffold00023:-:475713-475736
VNSFMSNEMTK	scaffold00003:+:3484909-3484941	EVIEQIER	scaffold00024:-:21295-21318

TNLINELCK	scaffold00003:-:3973431-3973457	NVLIGENNVAK	scaffold00024:-:21523-21555
DYSNISSPAPK	scaffold00003:+:3983628-3983660	QEGLIPWNFVAEDR	scaffold00024:-:22465-22506
NNPEPWEEYR	scaffold00003:+:3983363-3983392	VIADDEYCPK	scaffold00024:-:21469-21498
FYSPVRDYSNISSPAPK	scaffold00003:+:3983610-3983660	GQYSLVEPDGNVR	scaffold00024:-:1566970-1567008
KNNPEPWEEYR	scaffold00003:+:3983360-3983392	TVDYTADDVNGFNAVVSK	scaffold00024:-:1743467-1743520
LALRSPEVTWSR	scaffold00003:+:3983324-3983359	SFYAMEGKPVCPGCVGADEDE	scaffold00025:+:1288107-1288169
SPEVTWSR	scaffold00003:+:3983336-3983359	TFDLPDVK	scaffold00026:+:1080945-1080968
MQGLTMASLK	scaffold00003:+:3983030-3983059	MQVYAGCSDGTVQLVK	scaffold00026:+:1961235-1961282
MQGLTMASLKK	scaffold00003:+:3983030-3983062	TVLQNAPLMCK	scaffold00027:-:73422-73454
QGLTMASLKK	scaffold00003:+:3983033-3983062	NIDALKTVLQNAPLMCK	scaffold00027:-:73422-73472
FSTAGPSGTSVSNTMK	scaffold00004:-:1921183-1921230	TAASSAVAATASVNAHGFASSDR	scaffold00027:+:467556-467624
GEIEEDFWWCIDK	scaffold00004:-:2420055-2420093	CEQSATASVPR	scaffold00027:+:467670-467702
VPAYLNVVDIAGLVK	scaffold00004:-:2701277-2701321	YLSTLSGAK	scaffold00027:+:552426-552452
TYDLLFK	scaffold00004:-:2692670-2692690	NVLVQTPSCVK	scaffold00027:+:1591402-1591434
IRVEMSSGR	scaffold00005:-:535266-535292	ITDFGLAK	scaffold00027:+:1591435-1591458
CACGGGNYAPSEDGSQK	scaffold00005:+:2492671-2492721	KGEQCEDECPQDFYANEDTR	scaffold00027:+:1590499-1590558
ECALKVESTGDV	scaffold00005:+:2494909-2494944	LEVLDPIAPMLVSGK	scaffold00027:+:1767519-1767563
SEQASASFVPVQSHTVVTVEANEK	scaffold00005:+:2684761-2684832	TTLILEYLK	scaffold00027:+:1767711-1767737
DEEGYNTYKPHDPTDTHEYR	scaffold00005:+:2728219-2728278	DGVPVQPDENSNR	scaffold00027:+:1767657-1767695
DEEGYNTYKPHDPTDTHEYRK	scaffold00005:+:2728219-2728281	IMVNEDETTVAVK	scaffold00027:+:1768191-1768229
STSSSSSSWSSSNVSDK	scaffold00005:+:2728150-2728200	MEDLGTFECR	scaffold00027:+:1767738-1767767
STSSSSSSWSSSNVSDKHTQGAR	scaffold00005:+:2728150-2728218	NILLCDDNVVK	scaffold00027:+:1770648-1770680
NIVWDDLLLR	scaffold00005:+:3040689-3040718	SAINEVVTR	scaffold00028:-:748462-748488
TGTFCGTPEFLAPEVLTETSYTR	scaffold00005:+:3040219-3040287	LYTLVTYVPVPTFK	scaffold00028:-:748109-748150
YPASLDK	scaffold00005:-:3069217-3069237	NDDEDSPNKLYTLVTYVPVPTFK	scaffold00028:-:748109-748177
EGTAVTEQPSR	scaffold00005:-:3312638-3312670	ECTINLNR	scaffold00028:-:748438-748461
DDQREEATHHTLHSTAVDGSGGYQK	scaffold00005:-:3487913-3487987	ECTINLNRR	scaffold00028:-:748435-748461
TQYESLHEQPQQSYHVPSHHPDPHR	scaffold00005:-:3488150-3488224	FAEKEMGTSDVR	scaffold00028:-:748342-748377
SSGGTDGGYIDR	scaffold00005:-:3488114-3488149	LNKAIWHR	scaffold00028:-:748306-748329
STNVGSIYVADVANQTLGR	scaffold00005:-:3488438-3488494	RNDDEDSPNKLYTLVTYVPVPTFK	scaffold00028:-:748109-748180
NMPSDDAINFVVK	scaffold00005:-:3720197-3720235	SFMDSDGDGVGDLNGIASR	scaffold00028:-:1562451-1562507
TVSNHLTTQDQER	scaffold00005:+:4195325-4195363	AAFYQIYPR	scaffold00028:-:1562508-1562534
NGQESADTIALLK	scaffold00006:-:154598-154636	DWWQKAAFYQIYPR	scaffold00028:-:1562508-1562549
GLAEVSPLMCK	scaffold00006:+:2119369-2119401	SNGIVPSGATSSLTVAAFPGPGDGTAGRR	scaffold00029:-:166131-166217
FNQVSHCSASR	scaffold00006:+:2119402-2119434	ELNELQAIVGR	scaffold00029:-:1148259-1148291

ASGVTDSVLSR	scaffold00006:-:2915931-2915963	IEPNPDFVK	scaffold00029:+:1229805-1229831
NDEGGAGEEVPDRDR	scaffold00006:+:3711197-3711241	ISPFEHQTYCQR	scaffold00029:-:1276504-1276539
TIVISTR	scaffold00007:+:3580560-3580580	YYTQSPSVR	scaffold00029:-:1382400-1382426
VQIHELTIKPVGESV	scaffold00008:+:852221-852265	ELHNNLLESSMR	scaffold00029:+:1911699-1911734
VQVHEITIKPVGESM	scaffold00008:+:850585-850629	SYSDDLVWK	scaffold00029:+:1912802-1912828
ESGELALENGSLGR	scaffold00008:-:1182036-1182077	FLNSAELDPSNVTNHK	scaffold00029:+:1922773-1922820
TDDSGEADGFTIK	scaffold00008:-:1188860-1188898	GLAAGLEHPISENGENLSVGQR	scaffold00029:+:1918935-1919000
INFNEGGINVHPQYVGSLYDIATIR	scaffold00008:+:1968066-1968140	GAVHVAPLPGHSVSQTQLNLAP APGSEE	scaffold00029:+:2056833-2056916
TFLVWCNEEDHLR	scaffold00008:+:2460662-2460700	STLEDLEER	scaffold00030:-:1182253-1182279
GIYHNDNKTFLVWCNEEDHLR	scaffold00008:+:2460638-2460700	STLEDLEERR	scaffold00030:-:1182250-1182279
HQDQQQQQQQFQQPK	scaffold00008:-:2949595-2949639	AMNDLLGMDLVPEPK	scaffold00032:-:1792967-1793011
DMADYCVQR	scaffold00008:+:3557511-3557537	KAMNDLLGMDLVPEPK	scaffold00032:-:1792967-1793014
LQSVDNSTSSESFLGK	scaffold00008:+:3562307-3562354	LNDYALAVR	scaffold00032:-:1792907-1792933
IGDFGMTR	scaffold00009:-:14956-14979	CGDKTNEIYPYLLQEIRPTLTELG IDTPEELGYDQPELALK	scaffold00032:-:1792760-1792882
NCMVAEDLTVK	scaffold00009:-:14980-15012	FLEGVKDK	scaffold00032:-:1792883-1792906
MKDEYNK	scaffold00009:+:1094335-1094355	IIVSALKACR	scaffold00032:-:1792937-1792966
QMQDFER	scaffold00009:+:1094356-1094376	RLNDYALAVR	scaffold00032:-:1792907-1792936
TSSDTQLDKVQIK	scaffold00009:+:1890666-1890704	SVAAAVQDIIKM	scaffold00032:+:1799480-1799515
ITTTDTGIHFK	scaffold00010:-:1284654-1284686	NVQLSETDIR	scaffold00032:+:1866954-1866983
EAPGAAINQAGIEEFKK	scaffold00056:+:683513-683563	FQYGGLNSGRPVTPPR	scaffold00032:+:1869707-1869754
DLSEAQEWCR	scaffold00057:-:1099452-1099481	KFQYGGLNSGRPVTPPR	scaffold00032:+:1869704-1869754
SHSVVITR	scaffold00058:+:1105294-1105317	EIFLSQPILLELEAPLK	scaffold00032:+:1867005-1867055
TTTTTTGTSHVVLNTGYLR	scaffold00058:+:1105318-1105374	QVAGMASTLDTWMSSITSMEQL IESR	scaffold00033:+:151663-151740
ALCSGQGECTCGR	scaffold00059:+:796925-796963	FDYSPDVER	scaffold00033:+:219920-219946
SYTQETYLAK	scaffold00060:-:74888-74917	AGNNVVYDVK	scaffold00033:+:292230-292259
TCVPDMSETVLCAQLAEETLGGK	scaffold00061:+:489939-490007	IKAGNNVVYDVK	scaffold00033:+:292224-292259
TCVPDMSETVLCAQLAEETLGGK VEVCK	scaffold00061:+:489939-490022	NTQTGEHVPLAVK	scaffold00033:+:458587-458625
GSQITDDEIR	scaffold00062:-:537550-537579	VLEGTVVAVGPGAR	scaffold00033:+:458545-458586
VTEPPRGFK	scaffold00062:+:684537-684563	VLLPEYGGTK	scaffold00033:+:458638-458667
SNVCNGDLGAGFVVSER	scaffold00062:+:691272-691322	EADILAK	scaffold00033:+:458707-458727
CVGIYQITTPNHFCAEDVVQR	scaffold00062:+:691209-691271	EADILAKIE	scaffold00033:+:458707-458733
VPAPGGDSASVCPVPATPK	scaffold00065:-:809333-809389	GGIVIPEK	scaffold00033:+:458509-458532
ERTQYFAELNDAR	scaffold00065:-:807244-807282	GGIVIPEKAQSK	scaffold00033:+:458509-458544
IGCDQLSNEK	scaffold00065:-:807214-807243	LLPLLDRVLIQR	scaffold00033:+:458449-458484
QPPPASLPATTNAK	scaffold00065:-:809390-809431	TKGGIVIPEK	scaffold00033:+:458503-458532

TQYFAELNDAR	scaffold00065:-:807244-807276	TKGGIVIPEKAQSK	scaffold00033:+:458503-458544
CAQVIMDAAGLDSELYR	scaffold00065:-:813278-813328	VELGDSKEYHLFR	scaffold00033:+:458668-458706
IMLKGDNITLIQNVQN	scaffold00066:-:377240-377287	VLLPEYGGTKVELGDSK	scaffold00033:+:458638-458688
VVGAYDKVTADK	scaffold00066:-:701260-701295	SPNYPTVGVQPDIISSNDIK	scaffold00033:-:730601-730660
IVGGQYAEDTQIPYQIALFYR	scaffold00068:-:300895-300957	VIGICTIK	scaffold00033:-:1058845-1058868
EAFNAISK	scaffold00073:+:860464-860487	VSDDDPIVNSVTVK	scaffold00033:-:1743729-1743770
TVDYTADDHNGFNAVVHK	scaffold00078:+:37311-37364	ICPESSAFGHPHHPFEK	scaffold00034:-:53283-53333
YSFNYGIKDPHTGDIK	scaffold00078:+:37188-37235	ICPESSAFGHPHHPFEKEE	scaffold00034:-:53277-53333
YAVNTEK	scaffold00079:+:201677-201697	LDNVLLDHEGHIK	scaffold00034:-:1474566-1474604
DFFTCSSGDYLEQNPEVR	scaffold00081:+:357330-357383	ALVTDDEDIDWVQTEK	scaffold00034:-:1474935-1474982
YIDDAVYEMIR	scaffold00081:+:413870-413902	LTDYGMCK	scaffold00034:-:1474542-1474565
LLGNCVDVLLR	scaffold00081:+:413819-413851	QYSLNDFELIR	scaffold00034:-:1475150-1475182
SPVFQGVDGR	scaffold00082:+:233213-233242	KQTSIEIEQLNAR	scaffold00035:-:110367-110405
HVNVEYSTR	scaffold00083:+:589723-589749	QTSIEIEQLNAR	scaffold00035:-:110367-110402
AEQSGWGVGVAYTK	scaffold00083:+:590858-590899	STSTVDVSQAAEEASTQYK	scaffold00035:-:112771-112827
LNEQTFSK	scaffold00083:+:590900-590923	ELQDISTDGSPK	scaffold00035:+:486908-486943
SVPVDTSLNTFIR	scaffold00083:+:599260-599298	HQQMVYFVGR	scaffold00035:+:1235166-1235195
VVTCESMVGNEGSYLR	scaffold00083:+:632640-632687	LDHPFLYVLLMGSSR	scaffold00035:+:1234566-1234610
ENMDDYGESTGLLEQPVASVR	scaffold00084:+:178050-178112	YVKPDQSDVTEVNR	scaffold00035:+:1234632-1234673
ENGQLGLIAACAAGGQGVAMLLERHPEANAE	scaffold00085:+:707040-707132	AAAPVIIPSVAK	scaffold00036:+:320508-320543
GELKPIDEPVSWIEPANVEVPK	scaffold00085:+:714813-714878	SGEGFYSYKK	scaffold00036:+:606041-606070
TGQLAAIK	scaffold00087:+:363523-363546	SEIYCHGDLLK	scaffold00036:+:1228055-1228087
MAGAIIGK	scaffold00088:+:34310-34333	LPSVWNPAAGIGR	scaffold00037:+:285085-285123
EASLILGVSPSASK	scaffold00089:-:153484-153525	ALVSDYER	scaffold00038:+:770857-770880
TPEELAAEAEQEDLDDFTSK	scaffold00089:+:485725-485784	DSVPPVVPELSSDDDTR	scaffold00038:+:1195062-1195112
MNPAELNQYQPVATFEKLESSHFYK	scaffold00089:-:596808-596882	SSSAVGTPDYISPEVLQFQGEK	scaffold00038:+:1194704-1194769
MNPAELNQYQPVATFEK	scaffold00089:-:596832-596882	TGCLTYQEDPTR	scaffold00038:-:1395982-1396017
TNAQIQAILGGKK	scaffold00089:-:603118-603156	PDISQDEFEAK	scaffold00038:-:1432286-1432318
ENEWCEEK	scaffold00090:-:184537-184560	HVSSIMSSEK	scaffold00041:-:431335-431364
NDFTPAEEEQVR	scaffold00090:-:184564-184599	AVYNDADVYFLDDPLSAVDSHVGK	scaffold00041:-:825255-825326
NDFTPAEEEQVRK	scaffold00090:-:184561-184599	VIEACALKPDIEMLPGGDMTEIGEK	scaffold00041:-:825375-825449
NDFTPAEEEQVRKENEWCEEK	scaffold00090:-:184537-184599	EGEISESGTYQELMDK	scaffold00041:-:825093-825140
TFNIKNDFTPAEEEQVRK	scaffold00090:-:184561-184614	EGEISESGTYQELMDKK	scaffold00041:-:825090-825140
LDPLGINSADLDDKTPPELLYSSYR	scaffold00092:-:7094-7168	NQIVAVVGTVGSGK	scaffold00041:-:825615-825656
LTLNVGNGQR	scaffold00092:-:532599-532628	TELLTSGGVTLAGQR	scaffold00041:-:1341248-1341292

VVNGEAATLGQFPYQVR	scaffold00092:-:532629-532679	IVQGFDKTELLTSGGVTLAGQR	scaffold00041:-:1341248-1341313
IAVVSYEPDDDIK	scaffold00093:+:533065-533103	IVQGFDK	scaffold00041:-:1341293-1341313
IAVVSYEPDDDIKEK	scaffold00093:+:533065-533109	YIADENGFQPQADHLPVAPSV	scaffold00041:-:1401230-1401292
LVTLNNEVIPFYLEK	scaffold00093:+:533113-533157	TINAEHVLEALDR	scaffold00043:-:354971-355009
SDLVANHANLQR	scaffold00093:+:533380-533415	GTVNCDDEQTNPAPPQAQVCVC VVHNK	scaffold00043:-:403308-403388
VVDNVTSIDSIR	scaffold00093:+:533416-533451	LACGVIGLCKV	scaffold00043:+:826300-826332
KLVTLNNEVIPFYLEK	scaffold00093:+:533110-533157	ELILTDKQIEEIMR	scaffold00044:+:726796-726837
KLVTLNNEVIPFYLEKLDDIAR	scaffold00093:+:533110-533175	LIGELYHPDDSQL	scaffold00045:+:674621-674659
LVTLNNEVIPFYLEKLDDIAR	scaffold00093:+:533113-533175	AGGAGAQGQNLK	scaffold00046:+:1317709-1317744
SWIDKRPK	scaffold00093:+:533452-533475	EEEQVNLLGEEGIR	scaffold00046:+:1317667-1317708
VHQSVAMSR	scaffold00093:+:532536-532562	TNETQVYVASAHK	scaffold00047:+:202810-202848
YLANQVGLAGADDWENLMIDTV VDTVNDFR	scaffold00093:+:532563-532652	YIGKDPFQK	scaffold00047:+:740649-740675
YAPEQNAITWTIK	scaffold00095:+:320875-320913	TLSISCPAGTVSPLEPR	scaffold00049:-:1198419-1198469
DGAYVFIEHCK	scaffold00102:+:354436-354468	TSVAEEEEEPTK	scaffold00051:+:24170-24205
TQLADTPELTK	scaffold00102:+:358287-358319	ITGLDPANPCFNEGESLSGIQR	scaffold00051:+:524288-524353
DLNYYVER	scaffold00103:+:317717-317740	DFNDADMER	scaffold00051:-:1038498-1038524
TANTNNLDVNR	scaffold00106:-:130743-130775	VQFGDTEFEVPDR	scaffold00053:-:1219959-1219997
SLDYGAPGVPGFEK	scaffold00106:+:175474-175515	YINPNLVAIATTGPDNIHK	scaffold00054:-:258619-258675
YCGYGYVSR	scaffold00106:+:244624-244650	TALVHDPSHSGNR	scaffold00054:+:488604-488642
APVAVSIVEEIVR	scaffold00108:-:426640-426678	EQTDVSNFIER	scaffold00054:+:1185097-1185129
DTEQCLAVAVQDAATEFNDMLVK	scaffold00108:-:426517-426585	QSAELENQMVSVER	scaffold00054:+:1185913-1185954
VSPCFAEVPQEIR	scaffold00108:-:426939-426977	SNLDPFDQR	scaffold00054:+:1186321-1186347
LVGVVNFGVPCGLGYPDGFAR	scaffold00110:-:214795-214857	VNEYTNLPSEPPLETAPK	scaffold00054:+:1185955-1186008
SFGGFDNQNADNK	scaffold00111:-:253801-253839	LVVSGSSDNSIR	scaffold00055:+:1383-1418
KAQDEDEEEEDDEDEEEDEEDDEE EVAQKPK	scaffold00111:-:254610-254702	YAKEPDNASK	scaffold00055:+:158018-158047
AQDEDEEEEDDEDEEEDEEDDEEE VAQKPK	scaffold00111:-:254610-254699	NGLGPQAGYIGR	scaffold00055:+:1102345-1102380
ANVADQPCMSDIDMHKS	scaffold00113:+:254156-254206	TLQEDSSTTATTVSNTGGSK	scaffold00056:+:30219-30278
YTLEPFPAPPLETDDEKPTQSQ	scaffold00120:+:352734-352799	TLQEDSSTTATTVSNTGGSKDSL P	scaffold00056:+:30219-30290
SIEKDQQTLR	scaffold00126:+:234538-234567	YPHVAAWYENIR	scaffold00056:+:683474-683509
SIQLFTDKPR	scaffold00126:+:234508-234537	EAPGAAINQAGIEEFK	scaffold00056:+:683513-683560
TLATQEITR	scaffold00126:+:234481-234507	ELPACEICNLPR	scaffold00001:+:5234396-5234431
DGAEASVQK	scaffold00127:-:103005-103031	NGEVMASAQGR	scaffold00005:-:1136750-1136782
EAVGGLTETEVQDIR	scaffold00140:-:40058-40102	LFNLHERPAEEEDVEAE	scaffold00006:+:2913043-2913093
QVEINTIASGFGHLGPSSK	scaffold00141:+:159466-159522	SMEKHEVVPDVVPVAPTEVAK	scaffold00007:-:27497-27559
AFNAYCCWK	scaffold00141:+:159439-159465	IMGCAQSAEER	scaffold00010:-:132891-132923

CDDECSGR	scaffold00141:+:159409-159432	IVECSGTSQVSVMGR	scaffold00020:-:1817205-1817249
SDLMLETR	scaffold00141:+:159385-159408	NMQPQIDPSYNIIDTASGPQGPPSK	scaffold00029:-:1195682-1195756
STPYSFPGGTSQIVNK	scaffold00154:-:142900-142947	VSCTGTMDGQGPPSEATLAQWEK	scaffold00053:-:884179-884247
VVDASVMPK	scaffold00156:-:55091-55117	SADAEQVEEMNGFEK	scaffold00053:-:935775-935819
VTCGEIAAVQESK	scaffold00159:-:25255-25293	SISNSTMQPNTK	scaffold00073:-:876719-876754
NAFNAFDQEK	scaffold00160:-:52089-52118	DLMVGDEASQLR	scaffold00076:-:498224-498259
NAFNAFDQEKK	scaffold00160:-:52086-52118	SGSNDPIPGSPMKDDQAFVFDR	scaffold00098:-:303464-303529
AGQLVLLAPK	scaffold00162:+:95308-95337	AMVVSQEMLENMK	scaffold00165:+:103044-103082
FDASNAQEMSIR	scaffold00162:+:95272-95307	NMNMDEALSDFR	scaffold03801:-:1426-1461
AGGENQLSLR	scaffold00165:+:83049-83078	EDSSGFFSGVVIKGPVAEVITAPIFTR	scaffold08087:-:207-287
NCLVGDNNLVK	scaffold00165:+:84342-84374	VVYADYASSGR	scaffold00001:+:4391981-4392013
GAIDHYVTPVK	scaffold00165:+:102918-102950	ACEMNGQMVGENR	scaffold00003:+:67228-67266
IEAPWQPVTK	scaffold00165:+:102855-102884	TDSDMAVETKEPTTDEEGIVESPVLQDAR	scaffold00005:+:3624702-3624788
LRNYDETCQLV	scaffold00176:+:17565-17597	GPSGSGGMEANSSGATSEPK	scaffold00012:+:1981459-1981518
INSLEEEKYDLEYVVK	scaffold00224:+:10885-10932	ISEIEIGNIMER	scaffold00039:+:250253-250288
YWEAIVPVAR	scaffold00279:+:5473-5502	MTSDVETNIVAVER	scaffold00041:-:814266-814307
TNLPVFK	scaffold00286:-:16216-16236	EETPAAVPAETVGDVEGEPK	scaffold00043:+:1293865-1293924
LSDQSAFGR	scaffold00709:+:1686-1712	MVYESVLDDGLEGIEHQER	scaffold00051:-:865392-865448
GHICGGSLITTNTVLTAAHCLVDSR	scaffold00709:+:1713-1787	ELFEQCVAPNAGDELR	scaffold00059:+:796564-796611
IITTPQATAIGVAGSSGK	scaffold00731:-:3589-3642	ILNMEDDMNWYR	scaffold00084:+:94791-94826
DSHDDNRYLLQR	scaffold00880:+:3146-3181	LEMVNDNFTELR	scaffold00127:+:295828-295863
WIDKNTGAVAK	scaffold00880:+:2466-2498	LQLSEMAK	scaffold00144:-:154165-154188
SVPMPEEEIR	scaffold00001:-:4215342-4215371	ESSDNMISLGWK	scaffold00412:+:944-979

Appendix Table 4 B: A partial list of GSSPS mapping intergenic region of VectorBase genes leading to identification of novel protein coding genes.

S. No.	Peptide sequence	Genomic coordinates of identified peptide (Indian strain)	Mascot Score	m/z [Da]	Charge	Modifications	No. of PSMs	Transcript evidence
1	LNDSLYLCVR	scaffold00003:-:5499652-5499681	64	626.82117	2	C8(Carbamidomethyl)	1	Yes
	CGPLDNSSPITKK	scaffold00003:-:5499613-5499651	67	472.90988	3	C1(Carbamidomethyl)	1	
	SQTIWSDSLR	scaffold00003:-:5499745-5499774	70	596.80237	2	-	3	
	ALHEEPTPPVIQPGQGGK	scaffold00003:-:5500981-5501034	97	927.98883	2	-	14	
	ENVQISCGCFHDVK	scaffold00003:-:5501044-5501085	51	564.92072	3	C7(Carbamidomethyl); C9(Carbamidomethyl)	4	
	TAVQCEQQEK	scaffold00003:-:5501086-5501115	51	610.78125	2	C5(Carbamidomethyl)	1	
	FDAQVDHDVNGVNYAK	scaffold00003:-:5501140-5501187	89	896.41827	2	-	4	
	GDRLEQFSGHHLTADAWSR	scaffold00003:-:5501302-5501358	50	546.51483	4	-	1	
	AYAGWVQEQYR	scaffold00003:-:5501743-5501775	58	685.82666	2	-	1	
	EAFWEAYQTQAK	scaffold00003:-:5502142-5502177	62	736.34583	2	-	2	
	FREAFWEAYQTQAK	scaffold00003:-:5502142-5502183	46	592.28912	3	-	1	
	QQTNGTCFIQAK	scaffold00003:-:5502694-5502729	72	698.33728	2	C7(Carbamidomethyl)	3	

2	GTLPAQNCPQEDDVCFSR	scaffold00067:+:272944-272997	91	1047.45569	2	C8(Carbamidomethyl); C15(Carbamidomethyl)	4	Yes
	GCSAESEAICEGNNR	scaffold00067:+:273321-273365	89	827.33313	2	C2(Carbamidomethyl); C10(Carbamidomethyl)	1	
	DGEECDKAAMGAEECDDLNDVCYTR	scaffold00067:+:273951-274025	74	975.0426	3	C5(Carbamidomethyl); C15(Carbamidomethyl); C22(Carbamidomethyl)	1	
	SFSSCVQCSSMDESSDCNK	scaffold00067:+:274479-274535	120	1107.90747	2	C5(Carbamidomethyl); C8(Carbamidomethyl); C17(Carbamidomethyl)	2	
	ITEDDFER	scaffold00067:+:274347-274370	-	512.73267	2	-	1	
	CTDPHDLSCVVCEEPGCNQNPWTK	scaffold00067:+:274716-274787	35	968.39911	3	C1(Carbamidomethyl); C9(Carbamidomethyl); C12(Carbamidomethyl); C17(Carbamidomethyl)	1	
3	GYSSEENTWEPEENLDCDDLIQAFK	scaffold10785:+:685-759	74	1495.141	2	C17(Carbamidomethyl)	5	No
4	VKDLELSTR	scaffold00010:-:2139651-2139677	53	354.2047	3	-	5	No
5	TASENDDENSQLLK	scaffold00046:+:903860-903901	92	782.3603	2	-	3	Yes
6	LSPGIVPK	scaffold00047:+:859-882	54	405.7572	2	-	11	Yes
	INTTGGDYK	scaffold00047:+:883-909	59	484.7378	2	-	4	

	INTTGGDYK	scaffold00047:+:883-909	59	484.7378	2	-	4	
7	YDLGTCNLITELGNVSP ER	scaffold00052:+:229343-229399	132	1076.02	2	C6(Carbamido methyl)	2	Yes
	TVLIGNVDDR	scaffold00052:+:229864-229893	65	551.2984	2	-	2	
8	SEGFDDYMK	scaffold06797:-:598-624	48	554.218	2	M8(Oxidation)	22	Yes
	MEKSEGFDDYMK	scaffold06797:-:598-633	72	740.3106	2	-	4	
	AVWEGKK	scaffold06797:-:640-660	39	430.2375	2	N-Term(Acetyl)	4	
9	NEDVELPDSDESLYWCK	scaffold01723:+:160-210	75	1049.941	2	C16(Carbamidomethyl)	16	Yes
10	LAPEEEMAHFVDQSR	scaffold01221:-:2398-2442	90	586.9443	3	-	14	Yes
	LAPEEEMAHFVDQSR	scaffold01221:-:2398-2442	90	586.9443	3	-	14	
11	QNRPIPQWIR	scaffold00085:-:176797-176826	52	436.5827	3	-	43	Yes
12	QPGAPDVQEEQVDSEKPIEK	scaffold00081:-:345605-345664	79	1112.042	2	-	9	Yes
13	EQELEEVLTK	scaffold01124:-:2413-2442	64	609.3166	2	-	10	Yes
	EQELEEVLTK	scaffold01124:-:2413-2442	64	609.3166	2	-	10	
14	DVSTIPCFR	scaffold00057:-:959651-959677	30	547.769	2	C7(Carbamidomethyl)	4	Yes
15	IVGGSTVKEMEIK	scaffold00220:+:37603-37641	46	469.5894	3	M10(Oxidation)	2	Yes
	IVGGSTVKEMEIK	scaffold00220:+:37603-37641	46	469.5894	3	M10(Oxidation)	2	
16	EINKSDVEFEILDT	scaffold00055:+:92847-92888	45	826.406	2	-	2	Yes
17	MQGLPPTDAAEEK	scaffold00062:-:623749-623787	70	693.8314	2	-	33	Yes
18	TCVPDMSETVLCAQLAEETLGGK	scaffold00061:+:489939-490007	124	1263.082	2	C2(Carbamidomethyl); M6(Oxidation); C12(Carbamid	42	Yes

						omethyl)		
	TCVPDMSETVLCAQLAEETLGGKVEVCK	scaffold00061:+:489939-490022	71	1042.16	3	C2(Carbamidomethyl); C12(Carbamidomethyl); C27(Carbamidomethyl)	6	
19	APSDEEVDTYVR	scaffold00028:+:704894-704929	63	690.8166	2	-	2	Yes
20	SESFYPYYASCFTR	scaffold00016:+:54729-54770	-	889.3713	2	C11(Carbamidomethyl)	1	Yes
21	ILIPVTSVGPTVEVASGK	scaffold00008:+:651542-651595	76	884.0151	2	-	3	Yes
22	TSDGSQATSNVVR	scaffold00008:-:2926736-2926774	108	661.3201	2	-	6	Yes
23	EQPGGNNPWEAEMVR	scaffold00058:+:1121347-1121391	64	857.3861	2	-	1	Yes
	AKLEESDYDSEAK	scaffold00058:+:1121620-1121658	78	742.8402	2	-	2	
	LEESDYDSEAK	scaffold00058:+:1121626-1121658	58	643.2759	2	-	1	
	IVFPEAER	scaffold00058:+:1125819-1125842	47	480.7614	2	-	1	
24	EATVSGFGR	scaffold00059:-:868806-868832	46	462.233	2	-	2	Yes
	AGTININEGTVR	scaffold00059:-:868989-869024	82	622.832	2	-	4	
25	SLTGSGPMYVGK	scaffold00071:-:240832-240867	49	598.8021	2	-	1	No
26	HVGQCVSIHLK	scaffold00065:-:469554-469586	-	426.5641	3	C5(Carbamidomethyl)	1	Yes
27	VIVSYIPFYGSK	scaffold00075:-:785209-785244	81	686.8804	2	-	17	Yes
28	TGTGTATGETPATTSK	scaffold00090:+:214250-214297	102	740.8595	2	-	10	Yes
30	ISADQSNQLVQMALESK	scaffold16294:-:526-576	141	931.4694	2	-	2	Yes

31	DQTDGNAEALQAEIDR	scaffold15230:-:335-382	96	873.4002	2	-	1	Yes
	DLDASNQTILDLNAK	scaffold15230:-:185-229	60	815.9181	2	-	1	
32	VPDSACAGTAFLTGVK	scaffold05547:-:784-831	117	797.4005	2	C6(Carbamido methyl)	48	Yes
33	NEANFYNVVLK	scaffold04388:-:152-184	65	655.8402	2	-	14	Yes
	GDAYLSELYR	scaffold04388:-:308-337	56	593.7905	2	-	10	
	KPANPFKDIDR	scaffold04388:-:92-124	50	434.2389	3	-	1	
34	AATATITNPTTGVR	scaffold00003:-:3535935-3535976	64	687.3697	2	-	3	Yes
35	VHNVAAPVDSYAGK	scaffold00001:-:3940812-3940853	88	714.3693	2	-	20	Yes

Appendix Table 4 C: List of GSSPs mapping to intergenic region leading to joining of two gene structures.

S. No.	Genes involved (in Indian)	Revised genomic coordinates of transcript (Indian)	Peptide sequence	Genomic coordinates of identified peptide (Indian)	Transcript evidence	Mascot score	m/z	Charge	Modification(s)	PSMs
1	ASTEI02533; ASTEI02535	scaffold00010:102533-154069	IMGCAQSAEER	scaffold00010:-:132891-132923	Yes	76	626.27	2	C4(Carbamidomethyl)	5
2	ASTEI08831; ASTEI08832	scaffold00076:494888-504492	DLMVGDEASQLR	scaffold00076:498224-498259	Yes	87	667.32	2		13
			DLMVGDEASQLR	scaffold00076:498224-498259		65	747.37	2	N-Term(iTRAQ4plex); M3(Oxidation)	1
			DLMVGDEASQLR	scaffold00076:498224-498259		74	675.32	2	M3(Oxidation)	7
3	ASTEI10981; ASTEI10980	scaffold00211:36062-45969	GGVIFVDQLPMTASGK	scaffold00211:-:44194-44241	Yes	75	810.42	2		9
			GGVIFVDQLPMTASGK	scaffold00211:-:44194-44241		15	818.42	2	M11(Oxidation)	1
			ALTEQAVIDHCAR	scaffold00211:-:44284-44322		69	742.36	2	C11(Carbamidomethyl)	5
			CYNDLATALIVR	scaffold00211:-:44353-44388		64	704.86	2	C1(Carbamidomethyl)	1
			FDSDGFLYLIDR	scaffold00211:-:44509-44544		106	730.85	2		11
			AALENGTEGFVR	scaffold00211:-:44563-44598		72	632.31	2		6
4	ASTEI05417; ASTEI05418	scaffold00030:1,961,502-1,985,715	VPVLGFSETAQSVFR	scaffold00030:-:1974167-1974211	Yes	102	818.93	2		4
5	ASTEI04708; ASTEI04709	scaffold00025:481863-495284	HEDNLKPEGAFTVPER	scaffold00025:-:490526-490573	Yes	76	919.95	2		4

6	ASTEI08332; ASTEI08333	scaffold00066:375198-379912	GDNITLIQNVQN	scaffold00066:-:377240-377275	Yes	64	736.89	2	N-Term(iTRAQ4plex)	2
			IMLKGDNITLIQNVQN	scaffold00066:-:377240-377287		61	907.49	2		6
7	ASTEI07878; ASTEI07879	scaffold00059:793864-798400	TCECNLQNSQNR	scaffold00059:+:796525-796560	Yes	56	762.32	2	C2(Carbamidomethyl); C4(Carbamidomethyl)	2
			TCECNLQNSQNRR	scaffold00059:+:796525-796563		31	560.58	3	C2(Carbamidomethyl); C4(Carbamidomethyl)	1
			ELFEQCVAPNAGDELR	scaffold00059:+:796564-796611		103	924.43	2	C6(Carbamidomethyl)	12
8	ASTEI05285; ASTEI05286	scaffold00029:1,884,610-1,926,405	ELHNNLLESSMR	scaffold00029:+:1911699-1911734	Yes	33	721.85	2		1
			SYSDDLVWK	scaffold00029:+:1912802-1912828		34	556.76	2		1
			GLAAGLDHEIAENGENLSVGQR	scaffold00029:+:1912865-1912930	Yes	85	1197.6	2	N-Term(iTRAQ4plex)	2
			GLAAGLEHPISENGENLSVGQR	scaffold00029:+:1918935-1919000		104	1124.5	2		2
			FLNSAELDPSNVTNHK	scaffold00029:+:1922773-1922820		54	692.03	3	N-Term(iTRAQ4plex); K16(iTRAQ4plex)	1
			FLNSAELDPSNVTNHK	scaffold00029:+:1922773-1922820	yes	75	893.44	2		6
			IVEAAGGQIIIDGLDISK	scaffold00029:+:1925449-1925502		124	906.51	2		5
			IVEAAGGQIIIDGLDISK	scaffold00029:+:1925449-1925502		91	1050.6	2	N-Term(iTRAQ4plex);	1

									K18(iTRAQ4plex)	
			GLTAGLEHEIAENGENLSVGQR	scaffold00029:+:1925662-1925727		111	765.38	3		6
			GLTAGLEHEIAENGENLSVGQR	scaffold00029:+:1925662-1925727		52	813.41	3	N-Term(iTRAQ4plex)	1
			DSIFYSMAK	scaffold00029:+:1926049-1926075		40	675.35	2	N-Term(iTRAQ4plex); K9(iTRAQ4plex)	1
9	ASTEI03474; ASTEI03475	scaffold00015:143959-149082	VSGGNATATEK	scaffold00015:-:147900-147932	Yes	65	661.86	2	N-Term(iTRAQ4plex); K11(iTRAQ4plex)	2
10	ASTEI10980; ASTEI10981	scaffold00211:-:36062-47775	GGVIFVDQLPMTASGK	scaffold00211:-:44194-44241	yes	75	810.42	2		9
			GGVIFVDQLPMTASGK	scaffold00211:-:44194-44241		15	818.42	2	M11(Oxidation)	1
			ALTEQAVIDHCAR	scaffold00211:-:44284-44322		69	742.36	2	C11(Carbamidomethyl)	5
			CYNDLATALIVR	scaffold00211:-:44353-44388		64	704.86	2	C1(Carbamidomethyl)	1
			FDSDGFLYLIDR	scaffold00211:-:44509-44544		106	730.85	2		11
			AALENGTEGFVR	scaffold00211:-:44563-44598		72	632.31	2		6

Appendix Table 4 D: A partial list of GSSPs mapping to intronic region of VectorBase genes leading to identification of novel coding exons.

S.No.	Gene symbol (Indian)	Peptide sequence	Genomic coordinates of identified peptide (Indian)	Genome coordinates of novel exon	Transcript evidence	Mascot score	Sequest score	m/z [Da]	Charge	Modifications	PSMs
1	ASTEI06691	GTVNcDDEQTNPAPPQAQVcVcVVHNK	scaffold00043:-:403308-403388	scaffold00043:403259-403454	Yes	54	4.57	1013.12	3	C5(Carbamidomethyl); C20(Carbamidomethyl); C22(Carbamidomethyl)	2
2	ASTEI09052	YIDDAVYEMIR	scaffold00081:+:413870-413902	Scaffold00081:413,794-414,477	Yes	82		694.33	3.1		11
		LLGNcVDVLLR	scaffold00081:+:413819-413851			89			3.37	C5(Carbamidomethyl)	8
3	ASTEI06191	TGDDSQEQPAAAHPHHHHLTESVKPSGSEPK	scaffold00038:+:94449-94541	scaffold00038:94423-94913	Yes	28	3.16	663.11	5	-	1
		TVLMQNVNNK	scaffold00038:+:94731-94760			56	-	580.80	2	-	3
4	ASTEI06528	dVDVLDNTLPQVLR	scaffold00041:-:822883-822924	scaffold00041:822879-823079	Yes	67	4.99	870.98	2	N-Term(iTRAQ4plex)	2
5	ASTEI06582	IRDASANDGK	scaffold00042:184226-184255	scaffold00042:184202-184264	Yes	57	-	523.76	2	-	1
6	ASTEI06698	DSKPNDSSGEGWTQR	scaffold00043:620927-620971	scaffold00043:620906-620998	Yes	60	3.13	832.36	2	-	2
7	ASTEI06650	SVVQYDSAGASNNNR	scaffold00043:-:68235-68279	scaffold00043:68020-69063	Yes	84	3.3	791.36	2	-	1
		VTTLGNSDSNLASSR	scaffold00043:-:68442-68486			83	3.09	761.37	2	-	1
		TSHAVHAQTQPVVR	scaffold00043:-:68901-68942			85	5	510.94	3	-	2
8	ASTEI08928	GGPmVLGTDGTDFEHR	scaffold00079:160362-160409	Exon 3:scaffold0	Yes	102	7.87	852.89	2	M4(Oxidation)	7

			scaffold00079:-:160333-160353	0079:160333-160629						-	
9	ASTEI07748	VIGSGSSSSSSGSLK	scaffold00058:167079-167123	scaffold00058:167058-167228	No	120	4.36	670.30	2		7
		VQPPLGQQTTVYHGK	scaffold00058:167124-167168			56	3.07	826.94	2		2
		SKVQPPLGQQTTVYHGK	scaffold00058:167124-167174			69	7.03	623.33	3		1
10	ASTEI08405	KAQALTANDNR	scaffold00067:-:319292-319324	scaffold_00067:319286-319411	No	64	4.17	601.31	2		1
11	ASTEI07905	LQDSDADPVGSQR	scaffold00060:-:19451-19489	scaffold00060:19452-19493	Yes	81	3.72	694.32	2		2
12	ASTEI02092	VTVILAYYR	scaffold00008:692649-692675	scaffold00008:692585-692703	Yes	49	-	549.33	2		1
13	ASTEI02288	YVIDGGVMERPENcPDK	scaffold00009:-:14625-14675	scaffold00009:14517-14702	yes	24	-	660.30	3	C14(Carbamidomethyl)	1
14	ASTEI09661	VTKPGTEPTHER	scaffold00100:246621-246656	scaffold00100:246042-246662	Yes	58	4.09	676.35	2		12
15	ASTEI00924	TNLINELcK	scaffold00003:-:3973431-3973457	scaffold00003:3973393-3973494	Yes	71	3.31	552.78	2	C8(Carbamidomethyl)	7
		tNLINELck	scaffold00003:-:3973431-3973457			63	4.21	691.37	2	N-Term(iTRAQ4plex); C8(Methylthio); K9(iTRAQ4plex)	1
16	ASTEI02774	YGPAVTcDQcK	scaffold00011:+:826865-826897	Exon 5:scaffold00011:826,843-826,930	Yes	75	3.59	649.78	2	C7(Carbamidomethyl); C10(Carbamidomethyl)	4
17	ASTEI03067	WYYDPVQETcFAFR	scaffold00013:-:123064-123105	Exon 11:122,991-123,176	Yes	113	4.38	941.41	2	C10(Carbamidomethyl)	8

Appendix Table 4 E: A partial list of GSSPs mapping to exon-intron junctions of VectorBase genes leading to exon extension.

S.No.	Gene symbol	Peptide sequence	Genomic coordinates of identified peptide (Indian)	Category	Revised coordinates for the extended exon	Transcript evidence (Yes/No)	Mascot score	m/z [Da]	Charge	Modifications	PSMs
1	ASTEI09047	DFFTcSSGDYLEQNPEVR	scaffold00081:357330-357383	exon extension	scaffold00081:357331-358038	Yes	87	1082.47	2	C5(Carbamidomethyl)	1
2	ASTEI02109	VQVHEITIKPVGESM	scaffold00008:+:850585-850629	exon extension	scaffold00008:850459-850629	Yes	60	833.94	2		3
3	ASTEI00333	ALENYIGHTQNVLR	scaffold00001:5236043-5236084 (+)	exon extension	scaffold00001:5235647-5236432	No	79	814.43	2		15
		DVNVWSGIGGALDR	scaffold00001:5236085-5236126 (+)				99	729.87	2		6
4	ASTEI04362	EWGNPLPDMEGFQPEHSDH	scaffold00022:-:1352903-1352959	exon extension	scaffold00022:-:1352903-1,353,041	Yes	68	1111.47	2		51
5	ASTEI05220	ELNELQAIVGR	scaffold00029:-:1148259-1148291	exon extension	scaffold00029:1147833-1148576	Yes	76	621.346	2		17
6	ASTEI05228	NMQPQIDPSYNIIDTASGPQGPPSK	scaffold00029:-:1195682-1195756	exon extension	scaffold00029:1194893-1195813	Yes	121	1328.14	2		16
7	ASTEI01983	AINEDLEEVWEGTK	scaffold00007:2207151-2207192	exon extension	scaffold00007:2207151-2207330	No	97	816.892	2		1
8	ASTEI01906	HADNHIEATYDPR	scaffold00007:157370-157408	exon extension	scaffold00007:156925-157408	Yes	-	513.572	3		1
9	ASTEI01971	TASSSTAEVNEQK	scaffold00007:1920545-1920583	exon extension	scaffold00007:1920151-1920583	Yes	78	676.32	2		2
10	ASTEI01924	LRDEYARGEDQF	scaffold00007:444309-444344	exon extension	scaffold00007:444202-444588	Yes	-	749.854	2		2
11	ASTEI01928	QGYAVENQFLPQR	scaffold00007:506499-506537	exon extension	scaffold00007:506187-506888	Yes but not supporting extension	92	775.393	2		6
12	ASTEI02130	ESGELALENGSLGR	scaffold00008:1182036-1182077	exon extension	scaffold00008:1181909-1182292	Yes	62	716.355	2		2
13	ASTEI02132	TDDSGEADGFTIK	scaffold00008:1188860-1188898	Joining of exons	scaffold00008:1188830-1191943	Yes	93	678.301	2		7

		EPNAQPNVGTLAMNFR	scaffold00008:-:1189760-1189807				98	879.935	2		9
14	ASTEI02160	IHVQNLHTADKPIK	scaffold00008:1558549-1558590	exon extension	scaffold00008:1557331-1558658	Yes	25	404.234	4		1
15	ASTEI06327	LLEAAGVPDLR	scaffold00039:391951-391983	exon extension	scaffold00039:392735-391951	No	82	577.332	2		1
16	ASTEI06317	GSYSFTADDGQVYTVNYIADENGFQPEAPHLPK	scaffold00039:302010-302108	exon extension	scaffold00039:301739-302108	No	68	1210.9	3		2
17	ASTEI06611	LTDGGWASSQDTGLAmK	scaffold00042:1252490-1252540	exon extension	scaffold00042:1252289-1252681	No	89	877.405	2	M16(Oxidation)	3
18	ASTEI06828	NITQYDPVYPIVTNYLENVAK	scaffold00045:104662-104724	exon extension	scaffold00045:104045-104785	Yes	73	1227.63	2		4
19	ASTEI00212	fQQTFHQVMVDLk	scaffold00001:-:3540591-3540629	exon extension	scaffold00001:3540380-3541323	Yes	-	637.011	3	N-Term(iTRAQ4plex); K13(iTRAQ4plex)	1
20	ASTEI00096	VNLSANPR	scaffold00001:+:1433374-1433397	exon extension	scaffold00001:1432556-1437883	Yes	58	435.743	2		2
		GCLIGIVGPVGSGK	scaffold00001:+:1435151-1435187	exon extension	scaffold00001:1434318-1435311						
		SSILQVLLR	scaffold00001:+:1435151-1435188	exon extension	scaffold00001:1434318-1435311		73	514.828	2		10
		QNILFGDELDQYR	scaffold00001:+:1435151-1435189	exon extension	scaffold00001:1434318-1435311	Yes	98	805.894	2		5
		VcALQPDLAHLPAGDMTVIGER	scaffold00001:+:1435208-1435273	exon extension	scaffold00001:1434318-1435311	Yes	55	788.403	3	C2(Carbamidomethyl)	2
		FFETNPSGR	scaffold00001:+:1436291-1436317	exon extension	scaffold00001:1435673-1,437,025	No	-	527.751	2		2
		SPIFTHIAATLNGLPTIR	scaffold00001:+:1436618-1436671	exon extension	scaffold00001:1435673-1,437,025	No	51	641.365	3		2
		LPDASEAGGK	scaffold00001:+:1437221-1437247	exon extension	scaffold00001:1435673-1,437,025	No	26	472.736	2		2
		SSIIAAFLR	scaffold00001:+:1436996-1437026	exon extension	scaffold00001:1435673-1,437,025	No					
		SHISIIPQDPVLFSGTLR	scaffold00001:+:1437320-1437373	exon extension	scaffold00001:1435673-1,437,025	No	77	990.55	2		12
		NLDPFEAYPDAALWR	scaffold00001:+:1437377-1437421	exon extension	scaffold00001:1435673-1,437,025	No	84	889.43	2		11

		ALELVELR	scaffold00001:+:1437422-1437445	exon extension	scaffold00001:1435673-1,437,025	No	67	471.786	2		5
		ILVLDEATANVDPDTDR	scaffold00001:+:1437572-1437622	exon extension	scaffold00001:1435673-1,437,025	No	93	928.967	2		4
21	ASTEI00333	ALENYIGHTQNVLR	scaffold00001:5236085-5236126 (+)	exon extension	scaffold00001:5235647-5236432	No	79	814.43	2		15
		DVNVWSGIGGALDR					99	729.87	2		6
22	ASTEI08865	DMPPFLLVR	scaffold00077:117810-117836 (+)	exon extension	scaffold00077:117759-117871		48	544.304	2		3
		DmPPFLLVR				Yes	20	552.299	2	M2(Oxidation)	2
23	ASTEI08922	cGTAGVPGVYTNTAEYVDWIR	scaffold00079:41176-41238(+)	exon extension	scaffold00079:40661-41238	Yes	90	1165.05	2	C1(Carbamidomethyl)	2
24	ASTEI09057	GYTHccSVPAAQK	scaffold 00081:772559-772597	exon extension	scaffold00081:772182- 772596	Yes	-	739.829	2	C5(Carbamidomethyl); C6(Carbamidomethyl)	1
25	ASTEI09040	ILLNFSTLQLLEWATGTHT	scaffold00081:270623-270679	exon extension	scaffold00081:270623-270881	Yes	114	1079.59	2		10
26	ASTEI09042	QFVGFSDDDIFK	scaffold00081:282182-282217	exon extension	scaffold00081:282182- 282244	Yes	55	709.337	2		1
27	ASTEI08960	HVIFFEPGYAR	scaffold00079:680686-680718	exon extension	scaffold00079:680686-681054	Yes	55	668.351	2		10
		YSTAPSIVVR	scaffold00079:672839-672868	exon extension	Exon 8:672508-674005		69	546.805	2		15
		FFAYPITSYYSYR	scaffold00079:672800-672838	exon extension			83	839.402	2		30
		KLLISTLIR	scaffold00079:671685-671711	exon extension	scaffold00079:671252-671731		47	528.861	2		2
28	ASTEI09047	DFFTcSSGDYLEQNPEVR	scaffold00081:357330-357383	exon extension	scaffold00081:357331-358038	Yes	87	1082.47	2	C5(Carbamidomethyl)	1
29	ASTEI09103	TSPFPYDPMIIPR	scaffold00083:326911-326949	exon extension-5'	scaffold00083:326625-326960	Yes	79	767.392	2		3
30	ASTEI09302	GDSmTASFHPSDHANIEAFK	scaffold00088:448398-448457	exon extension	scaffold00088:448220-448458	Yes	86	1089.48	2	M4(Oxidation)	11

Appendix Table 4 F: A partial list of GSSPs mapping to the 5'and 3' boundary of existing genes leading to N or C- terminal extension of protein.

S. No	Gene(s) involved (in Indian)	Category of protein correction (Indian)	Existing genomic coordinates of gene (Indian)	Revised genomic coordinates of the gene (Indian)	Peptide sequence	Genomic coordinates of identified peptide (Indian)	Transcript evidence	Mascot score	m/z	Charge	PSMs
1	ASTEI02407	N-terminal protein extension	scaffold00009:1890606-1892820	scaffold00009:1890606-1893095	TSSDTQLDKVQIK	scaffold00009:+:1890666-1890704	Yes	108	731.88	2	2
2	ASTEI06476	N-terminal protein extension	scaffold00041:65485-70221	scaffold00041:64991-70221	TELATDNMK	scaffold00041:65464-65490	Yes	56.0	511.74	2	1
3	ASTEI06523	C-terminal protein extension	scaffold00041:787057-787808	scaffold00041:787057-789812	ENLPTLTEESSMDDWWK	scaffold00041:787873-787923	Yes	81.0	1048.9	2	2
					VYDATVQHHQQFMGLK	scaffold00041:787924-787971		52.0	476.23	4	2
4	ASTEI06592	N-terminal protein extension	scaffold00042:629806-632031	scaffold00042:627792-632974	PTANQNLR	scaffold00042:632668-632691	Yes	53.0	457.24	2	3
					KLQQQDSSER	scaffold00042:-:631626-631653		62	609.80	2	3
					KLQQQDSSER	scaffold00042:-:632659-632660		62	609.80	2	3
5	ASTEI00095	C-terminal protein extension	scaffold00001:1422808-1429094	scaffold00001:1420450-1431890	DMVLATGPAESER	scaffold00001:-:1420929-1420967	Yes	62.0	688.33	2	3
					DMVLATGPAESER	scaffold00001:-:1420929-1420967		62.0	688.33	2	3
					LNTIMDSDR	scaffold00001:-:1421049-1421075		56.0	532.75	2	1
					LIQDTIR	scaffold00001:-:1421118-1421138		-	429.75	2	1
					VSIIPQDPVLFSGTLRR	scaffold00001:-:1421451-1421501		-	681.40	3	1
					SKVSIIPQDPVLFSGTLRR	scaffold00001:-:1421451-1421507		14.0	601.10	4	1

					VSIIPQDPVLFSGTLR	scaffold00001:-:1421454-1421501		99.0	871.49	2	17
					VSIIPQDPVLFSGTLR	scaffold00001:-:1421454-1421501		99.0	871.49	2	17
					SKVSIIPQDPVLFSGTLR	scaffold00001:-:1421454-1421507		45.0	653.04	3	2
					DLSFEIRPR	scaffold00001:-:1421646-1421672		34.0	378.20	3	3
					DLSFEIRPR	scaffold00001:-:1421646-1421672		34.0	378.20	3	3
					QSAEVANFMMSVER	scaffold00001:-:1421814-1421855		72.0	799.86	2	1
					SPVFSHLSASLAGLPTIR	scaffold00001:-:1422105-1422158		47.0	666.38	3	4
					SPVFSHLSASLAGLPTIR	scaffold00001:-:1422105-1422158		47.0	666.38	3	4
					DMGSVDELLPK	scaffold00001:-:1422336-1422368		56.0	602.30	2	6
					DMGSVDELLPK	scaffold00001:-:1422336-1422368		56.0	602.30	2	6
					DMGSVDELLPK	scaffold00001:-:1422336-1422368		56.0	602.30	2	6
					FYDTNPSGR	scaffold00001:-:1422390-1422416		37.0	528.74	2	11
					SISFYQTSVR	scaffold00001:-:1422474-1422503		57.0	594.30	2	4
					SISFYQTSVR	scaffold00001:-:1422474-1422503		57.0	594.30	2	4

6	ASTEI00276	N-terminal protein extension	scaffold00001:4341121-4345426	scaffold00001:4,341,121-4,345,549	AVAANGSPGLNNDGTAALR	scaffold00001:-:4345427-4345483	Yes	110.0	884.96	2	2
7	ASTEI01955	C-terminal protein extension	scaffold00007:1542277-1550545	scaffold00007:+:1542277-1553720	APESSQLTGITR	scaffold00007:+:1551116-1551151		41.0	630.33	2	1
8	ASTEI02033	N-terminal protein extension	scaffold00007:3683892-3684954	scaffold00007:3683718-3685400	QPYGVPLGQEAR	scaffold00007:-:3685150-3685185	Yes	51.0	657.84	2	1
9	ASTEI02146	N-terminal protein extension	scaffold00008:1423517-1428536	scaffold00008:1420696-1430235	SSGLQEPATQQEK	scaffold00008:+:1421475-1421513	Yes	39.0	701.84	2	1
					ETPAPEHATTPAR	scaffold00008:+:1422888-1422926		53.0	689.34	2	2
10	ASTEI06139	N-terminal protein extension	scaffold00036:1480584-1481423	scaffold00036:1480122-1481522	TVPAQAPFTVAPLPER	scaffold00036:+:1480437-1480484	Yes	90.0	847.46	2	6
					VQGPDSVSFLQGLMTNDMR	scaffold00036:+:1480497-1480553		38.0	1056.0	2	1
11	ASTEI06552	N-terminal protein extension	scaffold00041:1370188-1371360	scaffold00041:1369897-1371360	LVSNLLDNAPSDAVDHFEEYCR	scaffold00041:1369927-1369992	No	108.0	1282.6	2	16
					IDFLSPDYR	scaffold00041:1370002-1370028		70.0	563.28	2	1
					QTLSALAETENENVCYNSK	scaffold00041:1370098-1370154		108.0	1086.0	2	1

Appendix Table 4 G: List of peptides mapping to annotated exon but translated in a different frame than existing annotations

S. No.	Gene symbol	Peptide sequence	Genomic coordinates of identified peptide (Indian)	Category	Exon involved	Revised cooridantes of the exon	Mascot score	m/z [Da]	Charge	Modifications	PSMs
1	ASTEI04349	GGIDESEQFFDLK	scaffold00022:-:1085120-1085158	Insertion of T between 1085090 and 1085091	Exon 1 and 2	scaffold00022:1085090-1086205	93	742.848	2	-	6
		GGIDESEQFFDLK	scaffold00022:-:1085120-1085158				44	591.6362	3	N-Term(iTRAQ4plex); K13(iTRAQ4plex)	1
2	ASTEI06745	EETPAAVPAETVGDVEGEPK	scaffold00043:+:1293865-1293924	Exon split after peptide	Exon 2	scaffold00043:1293866-1293924	81	1157.089	2	N-Term(iTRAQ4plex); K20(iTRAQ4plex)	5
		EETPAAVPAETVGDVEGEPK	scaffold00043:+:1293865-1293924		Exon 3		79	1012.987	2	-	10
3	ASTEI04988	TAASSAVAATASVNAHGFASSDR	scaffold00027:+:467556-467624	Joining of exons : Deletion of C in scaffold00027:467,475	Exon 1 and 2	scaffold00027:467288-468208	69	717.0149	3	-	1
		CEQSATASVPR	scaffold00027:+:467670-467702				55	603.282	2	C1(Carbamidomethyl)	1
4	ASTEI01803	LFNLHERPAEEEDVEAE	scaffold00006:+:2913043-2913093	-	Exon 4	scaffold00006:+:2910491-2913093	47	724.3511	3	N-Term(iTRAQ4plex)	2
5	ASTEI03356	SYQLDGNTGSVALSELIQK	scaffold00014:+:1303232-1303288	-	Exon 3	scaffold00014:1303223-1303362	158	1012.022	2	-	32

Appendix Table 4 H: List of N-terminally acetylated peptides leading to identification of alternate start sites.

S. No.	Protein Group Accessions	Reference sequence	Peptide Sequence	Genomic coordinates (Indian strain)	Position of novel start codon (annotated start codon's position=0)	Transcript evidence	Ortholog evidence	Whether annotated as start site in ortholog	# PSMs	Modifications	Ion Score	Charge	m/z [Da]
1	ASTEI08087-PA	MAQNGFGHAATQLK	AQNGFGHAATQLK	scaffold00062:-:156856-156894	415	Yes	AGAP004793-PB	Yes	9	N-Term(Acetyl)	95.07	2	692.8
2	ASTEI08114-PA	MAANLRPLVPELAAIAR	AANLRPLVPELAAIAR	scaffold00062:-:555742-555789	884	Yes	AGAP004762-PA (is a case of gene splitting in stephensi)	Yes	50	N-Term(Acetyl)	91.45	2	859.0
3	ASTEI04621-PA	MADLGAPVR	ADLGAPVR	scaffold00024:-:1529447-1529470	12	Yes	AGAP006879-PA	Yes	33	N-Term(Acetyl)	69.78	2	420.7
4	ASTEI08702-PA	MDFYHLPLSAPCQAVR	MDFYHLPLSAPCQAVR	scaffold00073:+:859778-859825	20	Yes	AGAP004378-PA	Yes	33	N-Term(Acetyl); C12(Carbamidomethyl)	72.33	2	973.9
5	ASTEI05718-PA	MASSDNSYQVVQLPGPR	ASSDNSYQVVQLPGPR	scaffold00033:-:460414-460461	44	Yes	AGAP001501-PB	Yes	7	N-Term(Acetyl)	108.64	2	880.4
6	ASTEI08703-PA	MDLYYHIR	MDLYYHIR	scaffold00073:-:862950-862973	476	Yes	AGAP004380-PA	Yes	5	N-Term(Acetyl)	51.61	2	576.7
7	ASTEI07997-PA	MaSQVQSIHTPSVALSK	ASQVQSIHTPSVALSK	scaffold00060:-:1099900-1099947		Yes	AGAP002547-PA	Yes	29	N-Term(Acetyl)	89.46	2	847.9

8	ASTEI06800-PA	mDLVKPFMGLNVDR	MDLVKPFMGLNVDR	scaffold00044:+:715163-715204	-51	Yes	Hexokinase (Culex quinquefasciatus; Aedes aegypti; Anopheles arabiensis)	Yes	4	N-Term(Acetyl)	83.73	2	838.9
9	ASTEI01219-PA	MaKPAQLLNIDPANELK	AKPAQLLNIDPANELK	scaffold00004:+:4363817-4363864	-9	Yes	AGAP000626-PA	Yes	97	N-Term(Acetyl)	105.48	2	888.9
10	ASTEI06445-PA	mDATSIITQVSR	MDATSIITQVSR	scaffold00040:1315650-1315685	-227	Yes	CG5830, isoform A (Drosophila melanogaster)	Yes	1	N-Term(Acetyl); M1(Oxidation)	68.78	2	690.3

Appendix Table 4 I: RT-PCR validation and sequencing of novel coding regions identified against *An. stephensi* genome assembly version AsteI1.

Sl. No.	Primer Name	Primer Sequence	Novel/Modification	Type of gene	GenBank ID	GSSPs	Scaffold
1	seq28_FP	CAGTACTACGCACTGTTCGAC	Novel	UTR region	JZ152704 JZ152705	182.1 LQCDEDPPHAYSQTFVLKPLGGTFFCAHD IFR	scaffold02222
2	seq28_RP	GCCGAGTTGTGAATGTTGAG					
3	seq795_FP	GAGAGAGAAATAAACCGCATTC	Novel	UTR region	JZ152706 JZ152707	MGIG182.1 LQCDEDPPHAYSQTFVLKPLGGTFFCAHD IFR	
4	seq795_RP	GCACTCATACACACAGAACAG					
5	seq1110_FP	CTCGTTAATGTGCTTCGTC	Novel	Intergenic	JZ152708 JZ152709 JZ152710 JZ152711	MGIG564.1 KFVCHVVGITFQTIDR MGIG565.1 ALTNFPHTDFVLCK MGIG562.1 IIEAYVEEQAR	scaffold04663
6	seq1110_RP	CTGTACCAGTTCAATCCACAC					
7	seq916_FP	GATGCCGATTCAGGAAATG					
8	seq916_RP	CTCGTTAATGTGCTTCGTC					
9	seq660_FP	GAATCTAAGCTAGTTGTCG	Novel	UTR region	JZ152712 JZ152713	MGOG329 AIVGASLSEIMAK	scaffold01080
10	seq660_RP	CTTCGTCTTCTGCTGCTTAG					
11	seq659_FP	ATGAAAATCTGGAAGGATG	Novel	UTR region	JZ152714 JZ152715	seq440.1 LVDDVMYEVYGK seq489.1 VKLVDDVMYEVYGK seq733.1 TLGDVQLDGANPSAEEAEEGTESATETGV DIVLNHR seq928.1 LVDDVMYEVYGK seq659.1 KQFTTYLK seq153.1 LVETGFSDK seq74.1 SPGEVEVFK	scaffold03026
12	seq659_RP	AGAACTTCTCTTCCTCCAG					
13	seq_P6766302_FP	ATGAAAAACCTCACGGAAG	Novel	Different frame	JZ152716 JZ152717	NLTEVLHK seq856 EQFFGISPER	scaffold01705
14	seq_P6766302_RP	TACCCAATGAAGTTGAAC					
15	seq1052_FP	ATGACTGAAGAGGTTAAC	Novel	UTR region	JZ152718 JZ152719	MGIG818.1 ACGVSRPIVACSVTIDEGSQLK MGOG273.1 ACGVSRPIVACSVTIDEGSQLK	scaffold00511
16	seq1052_RP	TTAAACGAGCAGTCG					
17	seq146_FP	CCTCCGAAGAAGGATACCAAG	Novel	UTR region	JZ152720 JZ152721	MGIG745.1 LITPSVVSER MGOG159.1 LITPSVVSER	scaffold05487
18	seq146_RP	TTACGCCACCGGGTCGTC					
19	seq935_FP	GTACGCGTTGCGTCAGTaTTTC	Novel	Intergenic	JZ152722 JZ152723	seq935 GTLDVMPYLATTAK	scaffold01672

20	seq935_RP	TTACGCATTGTACTCTGGATCG				seq1160.1 DFLSIYQEK seq416.1 ASFFHTDYPAGR	
21	seq861_FP	GAACTGACCGATAGCGAC	Novel	UTR region	JZ152725 JZ152726 JZ152727	MGIG759.1 VNFAVSAKDDFQHELNEYGYDYTGDKPL VLAR	scaffold01363
22	seq861_RP	GGGAGAGAGGGTAGGAGATTAC					
23	seq245_FP	GTGCTCGATTTGGACCTG	Novel	UTR region	JZ152728 JZ152729	MGIG890.1 EVGNHLHESVPVSNDEDENRVER MGIG891.1 ELVSCSNCLDYQAR	scaffold03747
24	seq245_SP	CAGGAGGTGGCCCAGCTGTCC					
25	seq710_FP	CTACAAGAACGGTTTCAAGG	Novel	Intergenic	JZ152730 JZ152731	TLSFIPEPVLQGK TVAISSSELGAKDTLLAYVSNR QIKTETDIWDAMQWFHDK ELLPIYRDEIVPLADIITPNQYEVELLTGK LYTHLLTGYVGNPAFLR ELADVYAGLVDNDLHK VLSIQSHVVHGHVGNK	Orthology based
26	seq710_RP	GCTTTAGCTCTCGCTCTTGCG					
27	seq1833_FP	TAaGACCTGGCTGCACGATAAG	Novel	Intergenic	JZ152734 JZ152735	TPQELLSGSFSK KFADVLEYWQPIR	Orthology based
28	seq1833_RP	CTAACGGATGGGTTGCCAG					
29	seq679_FP	ATGGCCGGCGTTATCAAG	Novel	UTR region	JZ152737	SGIG509.1 QAPASQWAWPPANLPELK TstRP199.1 INCGQVEELIVQAENELTLAR LarIG197.1 INCGQVEELIVQAENELTLAR BrnRP178.1 INCGQVEELIVQAENELTLAR SGIG510.1 INCGQVEELIVQAENELTLAR FBIG326.1 INCGQVEELIVQAENELTLAR seq679.1 NPHHTLTALYNK seq18.1 ATGLTGLNVAK	scaffold01888
30	seq679_RP	CTTCAACTCGGGCAGATTGGC					
31	seqRP19_FP	ATGAATATACCCAATC	Novel	Intergenic	JZ152738 JZ152739	seq407.1 YVLPHVAR LarIG29.1 LCLIATFLEDGIR SGIG47.1 LCLIATFLEDGIR seq475.1 SLFAGVPSLGENKPK TstRP19.1 SLFAGVPSLGENKPK LarIG28.1 SLFAGVPSLGENKPK FBIG45.1 SLFAGVPSLGENKPK SGIG46.1 SLFAGVPSLGENKPK	scaffold02189
32	seqRP19_RP	TATGCTCATCCATCG					
33	seqRP76_FP	ATGAGTCTGCTCAAGTTTCAC	Novel	Intergenic	JZ152740 JZ152741	MGIG174.1 TLQTTDPHSVPVLIIGQLR MGIG175.1 AVSCLQPAPTDTCSLYLDMAR MGIG178.1 LIQSNTTGGSESIVIVCEK MGIG179.1 QDLLASACAVAR MGIG177.1 AALDPPALAVLSYEPK MGIG181.1 GATETVAWVGK MGIG180.1 ANVILDMATLTGAQGIATGK MGOG104.1 ANVILDMATLTGAQGIATGK MGIG173.1 YHGAILTNSDDWEAK MGOG103.1 YHGAILTNSDDWEAK	scaffold00913
34	seqRP76_RP	CACTTTATTCCGCTGCAC					

						102.1 TGDMLSPVPYCPELHFSEFNSAVADMK	
35	seq205_FP	ATGTCGAAGATTGGAATTAAC	Novel	UTR region	JZ152742 JZ152743	MGIG15.1 VPTPNVSVVDLTVRLSKPATYDQIK	scaffold01545
36	seq 205_RP	TTAATCCTTGGTCTGCATGTAC					
37	seq 146_SP	TAGCCGTGCTATCATCGGTG	Novel	UTR region	JZ152744 JZ152745	MGIG497.1 KDQDEFAINSYK MGOG136.1 QAGIGKEDVQEVYIGNVCAAGLGQAPAR MGOG137.1 HDVVIVAATR MGIG500.1 RHDVVIVAATR	scaffold02117
38	seq 146_RP	TTACAGCTTTTCGATCAGGATC					
39	seq 115_FP	ATGAGTCTGCCGCGTTG	Novel	intergenic	JZ152746 JZ152747	MGIG125.1 TAWLNDRHVVFGSVVEGMDVVR MGOG153.1 CFFDMSVDNQPVGR	scaffold02135
40	seq 115_RP	CAGCTGACCACAGTTCTC					
41	seq 243_FP	CCGTCGTTCGTTATCTTCCT	fusion of 2 exon	intronic	JZ152771 JZ152772	MGOG236.1 ILDDDSVITTLETLK	scaffold00922
42	seq 243_RP	TTCAGATGGATGCGACACAG					
43	seq 125_FP	TGGCACTGTCCCAGTTTATG	Novel	UTR region	JZ152748 JZ152749	MGOG6.1 TGEYPWAIIK	scaffold00062
44	seq 125_RP	CTGGATCTGCGACGTACC					
45	seq376_FP	ATGGTAAAACGGCAAAACTCG	Novel	intergenic	JZ152751 JZ152752	MGIG100 VQQLATFAVAHVNALLTGLRR	scaffold00629
46	seq376_FP	GCG CGA TTG CAG ATA CGT CT					
47	seq176_FP	CAATCCTAGcGCAATGG	Novel	UTR region	JZ152753 JZ152754	seq176 LVGLQDTDK seq52 IETIIGENEGK seq211 VQSPEEFNEK seq540 LVGLQDTDKLR	scaffold05822
48	seq176_FP	CTACTAGCTATTCCGAATC					
49	seq268_FP	CTAATGAAACGGCCATCG	Novel	Intergenic	JZ152755 JZ152756	MGIG539 YTLVEVSNCADTDASCVLKR MGOG265 YTLVEVSNCADTDASCVLK	scaffold00544
50	seq268_RP	CTTGGGATATTGCTTCAG					
51	seq 35_FP	AGCTCAAAATCTGTTCAAATAC	Novel	UTR region	JZ152757 JZ152758	seq151 AGIESALAK	scaffold03905
52	seq 35_RP	CTACTTCTCGCCCTCCTTCG					
53	seq 116_FP	ATGAGCAGCGCTGAGC	Novel	UTR region	JZ152761	MGIG414 LASTWEGIQAASVLEK MGIG416 NVGEIMALAGCDLLTISPK MGOG143 TYKPTDATTNPSLILSAAGMEQYQHLIDK	scaffold03609
54	seq 116_RP	TTCCAGCTTACGGCCATCG					
55	seq 156_FP	GCGGACGTGATGACACTC	Novel	Intergenic	JZ152762 JZ152763	seq204 YLLEVLEHSASR	scaffold03016
56	seq 156_RP	CTCGATAACGGTAATGCGTC					
57	seq 161_FP	TCGTTATGATTCCAGAACTAC	Novel	UTR region	JZ152764 JZ152765	MGIG365 HYGYQLYQSDPSGNYGGWK	scaffold03924
58	seq 161_RP	TATGATTTTTGACCCAGC					

59	seq 376_FP	GATTATGGAGCTGATCAACAAC	Novel	Intergenic	JZ152767 JZ152768	MGIG666 IAELGIYPAVDPLDSTSR	scaffold05388
60	seq 376_RP	CTGGAAGGGCTGCGACAG					
61	seq 3469_FP	TGAGCTGGAGAAGGATTGG	Gene extension	Intergenic	JZ152777	MGOG332 AQIHEIEQGLPQQNSR	scaffold04241
62	seq 3469_RP	TCATCAGTCGATAGCGGTTC					
63	seq 433_FP	AAGATCCTCAAGCCGGAATC	Gene extension	intronic	JZ152779 JZ152780	MGIG433 GCNEEQLTNAFTH	scaffold00405
64	seq 433_RP	TGTCGTTCATCTTCGTGAGG					
65	seq 567_FP	ATGGCATCGTTGGTGAATAG	Novel	Intergenic	JZ152769 JZ152770	MGIG567 LFRPADIVQYIADK MGIG319 VGNPANTNALVCSHYAPSIP	scaffold00160
66	seq 567_RP	TTCATAAATATCTTCCTTGTC					

Bibliography

1. Abai MR, Mehravaran A, Vatandoost H, Oshangi MA, Javadian E, Mashayekhi M, Mosleminia A, Piyazak N, Edallat H, Mohtarani F, Jabbari H, Raffi F (2008) Comparative performance of imagicide on *Anopheles stephensi*, main malaria vector in a malarious area, Southern Iran. J Vector Born Dis 45:307
2. Abraham EG, Pinto SB, Ghosh A, Vanlandingham DL, Budd A, Higgs S, Kafatos FC, Jacobs-Lorena M, Michel K (2005) An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. Proc Natl Acad Sci U S A 102:16327-16332.
3. Adak T, Singh OP, Das MK, Wattal S, Nanda N (2005) Comparitive Susceptibility of three important malaria vectors *Anopheles stephensi*, *Anopheles fluviatilis* and *Anopheles sundaicus to Plasmodium vivax.* J Parasitol 91:79-82
4. Amenya DA, Chou W, Li J, Yan G, Gershon PD, James AA, Marinotti O (2010) Proteomics reveals novel components of the *Anopheles gambiae* eggshell. J Insect Physiol 56:1414-1419
5. Angrisano F, Tan YH, Sturm A et al (2012) Malaria parasite colonisation of the mosquito midgut-placing the *Plasmodium* ookinete centre stage. International journal for parasitology 42:519-527
6. Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P et al (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science 330:86-88
7. Arrighi RB, Lycett G, Mahairaki V, Siden-Kiamos I, Louis C (2005) Laminin and the malaria parasite's journey through the mosquito midgut. J Exp Biol 208:2497-2502.
8. Bahia AC, Kubota MS, Tempone AJ, Araujo HR, Guedes BA, Orfano AS, Tadei WP, Rios-Velasquez CM, Han YS, Secundino NF et al (2011) The JAK-STAT pathway controls *Plasmodium vivax* load in early stages of *Anopheles aquasalis* infection. PLoS Negl Trop Dis 5:e1317.
9. Bamburg JR, Bernstein BW (2010) Roles of ADF/cofilin in actin polymerization and beyond. Biol Rep 2:62
10. Bansal S K, Singh KV (1996) Insecticide susceptibility status of some anophelines in district Bikaner, Rajasthan. Indian J Malariol 33

11. Basseri HR, Moosakazemi SH, Yosafi S, Mohebali M, Hajaran H, Jedari M (2005) Anthropophily of malaria vectors in Kahnouj District, south of Kerman Iran. Iranian J Publ Health 34:27-35
12. Baton L, Ranford-Cartwright LC (2005) Do malaria ookinete surface proteins P25 and P28 mediate parasite entry into mosquito midgut epithelial cells?. Malaria Journal 4
13. Batra CP, Mittal PK, Adak T, Subbarao SK (2006) Efficacy of Agnique MMF monomolecular surface film against *Anopheles stephensi* breeding in urban habitats in India. J Am Mosq Control Assoc 22:426-432
14. Batra CP, Reuben R (1979) Breeding of *Anopheles stephensi* (Liston) in wells and cisterns in Salem, Tamil Nadu. Indian J Med Res 70:114-122
15. Baudet M, Ortet P, Gaillard JC, Fernandez B, Guérin P, Enjalbal C, Subra G, Groot A, Barakat M, Dedieu A, Armengaud J (2010) Proteomics-based Refinement of *Deinococcus deserti* Genome Annotation Reveals an Unwonted Use of Non-canonical QTranslation Initiation Codons. Mol Cell Proteomics 9:415-426
16. Beerman RW, Jongens TA (2011) A non-canonical start codon in the *Drosophila* fragile X gene yields two functional isoforms. Neuroscience 181:48-66
17. Berois M, Romero-Severson J, Severson DW (2012) RNAi knock-downs support roles for the mucin-like (AeIMUC1) gene and short-chain dehydrogenase/reductase (SDR) gene in *Aedes aegypti* susceptibility to *Plasmodium gallinaceum*. Med Vet Entomol 26:112-115
18. Besansky NJ, Powell JR (1992) Re-association kinetics of *Anopheles gambiae* (Diptera: Culicidae) DNA. J Med Entomol 29:125-128
19. Bharti PK , Chand SK, Singh MP, Mishra S, Shukla MM, Singh R, Singh N (2013) Emergence of a new focus of *Plasmodium malariae* in forest villages of district Balaghat implications for the diagnosis of malaria and its control. Trop Med Int Health 18:7-12
20. 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 19:289-92

21. Biron DG, Agnew P, Marché L, Renault L, Sidobre C, Michalakis Y (2005) Proteome of *Aedes aegypti* larvae in response to infection by the intracellular parasite *Vavraiaculicis*. Int J Parasitol 35:1385-1397
22. Biswas D, Dutta RN, Ghosh SK, Chatterjee KK, Hati AK (1992) Breeding habits of *Anopheles stephensi* Liston in an area of Calcutta. Indian J Malariol 29:195-198
23. Blair A, Tomlinson A, Pham H, Gunsalus KC, Goldberg ML et al. (2006) Twinstar, the *Drosophila* homolog of cofilin/ADF, is required for planar cell polarity patterning. Development 133:1789-1797
24. 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.
25. Blitzer EJ, Vyazunova I, Lan Q (2005) Functional analysis of AeSCP-2 using gene expression knockdown in the yellow fever mosquito, *Aedes aegypti*. Insect Mol Biol 14:301-7
26. Bocchinfuso DG, Taylor P, Ross E et al (2012) Proteomic profiling of the planarian *Schmidteamediterranea* and its mucous reveals similarities with human secretions and those predicted for parasitic flatworms. Mol Cell Proteomics 11:681-691
27. Borchert N, Dieterich C, Krug K, Schutz W et al. (2010) Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. Genome Res 20:837-846
28. 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 180:1671-1678.
29. Bretscher A (1999) Regulation of cortical structure by the ezrin-radixin-moesin protein family. Curr Opin Cell Biol 11:109-16
30. Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, Deutsch EW, Panse C et al (2007) A high-quality catalog of the *Drosophila melanogaster* proteome. Nat Biotechnol 25:576-83
31. Busvine JR, Pal R (1969) The impact of insecticide resistance on control of vector and vector borne diseases. Bull World Hlth Org 40:731

32. Butler PR, Donald HD (2009) Proteomic analysis of the mosquito *Aedes aegypti* midgut brush border membrane vesicles. J Insect Physiol 55:264-272
33. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H et al (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature 455:75-763
34. Carter R (2001) Transmission blocking malaria vaccines. Vaccine 21:2309-14
35. Carter R, Mendis KN (2002) Evolutionary and Historical Aspects of the Burden of Malaria. Clinical Microbiology Reviews 15:564-594
36. Castellana NE, Shen Z, He Y, Walley JW, Cassidy CJ et al. (2014) An automated proteogenomic method uses mass spectrometry to reveal novel genes in *Zea mays*. Mol Cell Proteomics 13:157-167
37. Cázares-Raga FE, Chávez-Munguía B, González-Calixto C, Ochoa-Franco AP, Gawinowicz MA, Rodríguez MH, Hernández-Hernández FC (2014) Morphological and proteomic characterization of midgut of the malaria vector *Anopheles albimanus* at early time after a blood feeding. J Proteomics 5:111
38. Chaerkady R, Kelkar DS, Muthusamy B, Kandasamy K, Dwivedi SB, Sahasrabuddh NA, Kim MS, Renuse S, Pinto SM, Sharma R, Pawar H, Sekhar NR, Mohanty AK, Getnet D, Yang Y, Zhong J, Dash AP, MacCallum RM, Delanghe B, Mlambo G, Kumar A, Prasad TSK, Okulate M, Kumar N, Pandey A (2011) A proteogenomic analysis of *Anopheles gambiae* using high-resolution Fourier transform mass spectrometry. Genome Research 21:1872-81
39. Chakraborty S, Ray S, Tandon N (1998) Seasonal prevalence of *Anopheles stephensi* larvae and existence of two forms of the species in an urban garden in Calcutta City. Indian J Malariol 35:8-14
40. Chatterjee KK, Biswas D, Choudhuri DK, Mukherjee H, Hati AK (1993) Resting sites of *Anopheles stephensi* Liston in Calcutta. Indian J Malariol 30:109-112
41. Chen Y, Dong Y, Sandiford S, Dimopoulos G (2012) Transcriptional mediators Kto and Skd are involved in the regulation of the IMD pathway and *anti-plasmodium* defense in *Anopheles gambiae*. PLoS ONE 7:e45580.

42. Cheng J, Wang Y, Li F, Liu J, Sun Y, Wu J (2014) Cloning and characterization of a mannose binding C-type lectin gene from salivary gland of *Aedes albopictus.* Parasit Vectors 7:337
43. Cheon HM, Shin SW, Bian G, Park JH, Raikhel AS (2006) Regulation of lipid metabolism genes, lipid carrier protein lipophorin, and its receptor during immune challenge in the mosquito *Aedes aegypti*. J Biol Chem 281:8426-8435
44. Chertemps T, Mitri C, Perrot S, Sautereau J, Jacques JC, Thiery I, Bourgouin C, Rosinski-Chupin I (2010) *Anopheles gambiae* PRS1 modulates *Plasmodium* development at both midgut and salivary gland steps. PLoS ONE 5:e11538.
45. Cho WL, Tsao SM, Hays AR, Walter R, Chen JS, Snigirevskaya ES et al. (1999) Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extra ovarian precursor. J Biol Chem 274:13311-13321
46. Choumet V, Carmi-Leroy A, Laurent C, Lenormand P, Rousselle JC, Namane A, Roth C, Brey PT (2007) The salivary glands and saliva of *Anopheles gambiae* as an essential step in the *Plasmodium* life cycle: a global proteomic study. Proteomics 7: 3384-94
47. Cohuet A, Osta MA, Morlais I, Awono-Ambene PH, Michel K, Simard F, Christophides GK, Fontenille D, Kafatos FC (2006) *Anopheles* and *Plasmodium*: from laboratory models to natural systems in the field. EMBO Rep 7:1285-1289.
48. Corena M, VanEkeris L, Salazar M, Bowers D, Fiedler MM, Silverman D, Tu C, Linser PJ (2005) Carbonic anhydrase in the adult mosquito midgut. J Exp Biol 208:3263-3273
49. Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, Ratnam S, Rahman HA, Conway DJ, Singh B (2008) *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. Clin Infect Dis 46:165-71
50. Crabb JW, Carlson A, Chen Y, Goldflam S, Intres R, West KA, Hulmes JD, Kapron JT, Luck LA, Horwitz J, Bok D (1998) Structural and functional characterization of recombinant human cellular retinaldehyde-binding protein. Protein Sci 7:746-757
51. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20:1466-1467

52. Dana AN, Hillenmeyer ME, Lobo NF, Kern MK, Romans PA, Collins FH (2006) Differential gene expression in abdomens of the malaria vector mosquito, *Anopheles gambiae*, after sugar feeding, blood feeding and *Plasmodium berghei* infection. BMC Genomics 7:119
53. Daneshvar C, Davis TM, Cox-Singh J, Rafaee MZ, Zakaria SK, Divis PC, Singh B (2009) Clinical and Laboratory Features of Human *Plasmodium knowlesi* Infection. Clinical Infectious Diseases 49:852-860
54. Das A, Anvikar AR, Cator LJ, Dhiman RC, Eapen A, Mishra N, Nagpal BN, Nanda N, Raghavendra K, Read AF, Sharma SK, Singh OP (2012) Malaria in India: The Center for the Study of Complex Malaria in India. Acta Tropica 121:267-273
55. Das S, Malakar P, Saha GK, Dasgupta B, Hati AK (1996) A case of *Plasmodium malariae* infection in the Dooars region of West Bengal, India. Indian J Malariol 33:159-60
56. Dash AP, Valecha N, Anvikar AR, Kumar A (2008) Malaria in India: challenges and opportunities. J Biosci 33:583-592
57. David JP, Strode C, Vontas J, Nikou D, Vaughan A, Pignatelli PM, Louis C, Hemingway J, Ranson H (2005) The *Anopheles gambiae* detoxification chip: a highly specific microarray to study metabolic-based insecticide resistance in malaria vectors. Proc Natl Acad Sci 102:4080-4084
58. De Castro MC, Yamagata Y, Mtasiwa D, Tanner M, Utzinger J, Keiser J, Singer B (2004) Integrated urban malaria control: a case study in Dar es Salaam, Tanzania. Am J Trop Med Hyg 71:103-117
59. Desiere F, Deutsch EW, Nesvizhskii AI et al. (2005) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. Genome Biol 6:9
60. Dev V (2000) *Plasmodium malariae*- a case of quartan malaria in Assam. J Commun Dis 32:149-51
61. Dev V, Sharma VP (2013) The Dominant Mosquito Vectors of Human Malaria in India Edited by Sylvie Manguin, ISBN 978-953-51-1188-7, 828 pages, Publisher: Intech, Chapters published July 24, 2013 under CC BY 3.0 license

62. Dhingra N, Jha P, Sharma VP et al (2010) Adult and child malaria mortality in India: a nationally representative mortality survey. The Lancet 376:1768-74
63. Dimopoulos G, Muller H, Levashina EA, Kafatos FC (2001) Innate immune defense against malaria infection in the mosquito. Curr Opin Immunol 13:79-88
64. Dimopoulos G, Seeley D, Wolf A, Kafatos FC (1998) Malaria infection of the mosquito *Anopheles gambiae* activates immune responsive genes during critical transition stages of the parasite life cycle. EMBO J 17:6115-6123
65. Dinglasan RR, Devenport M, Florens L, Johnson JR, McHugh CA, Donnelly-Doman M, Carucci DJ, Yates JR, Jacobs-Lorena M (2009) The *Anopheles gambiae* adult midgut peritrophic matrix proteome. Insect Biochem Mol Biol 39:125-134
66. Dinglasan RR, Kalume DE, Kanzok SM, Ghosh AK, Muratova O, Pandey A, Lorena M J (2007) Disruption of *Plasmodium falciparum* development by antibodies against a conserved mosquito midgut antigen. Proc Natl Acad Sci USA 104:13461-13466
67. Djegbe I, Cornelie S, Rossignol M, Demettre E, Seveno M (2011) Differential Expression of Salivary Proteins between Susceptible and Insecticide- Resistant Mosquitoes of *Culex quinquefasciatus*. PLoS One 6(3): e17496
68. 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
69. Dong Y, Dimopoulos G (2009) *Anopheles* fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. J Biol Chem 284:9835-9844
70. Dwivedi SB, Muthusamy B, Kumar P, Kim MS, Nirujogi RS, Getnet D, Ahiakonu P, Dey G, Nair B, Gowda H, Prasad TSK, Kumar N, Pandey A, Okulate M (2014) Brain proteomics of *Anopheles gambiae*. OMICS: A Journal of Integrative Biology 18:421-37
71. Edwards FW (1932) Diptera, family Culicidae. In Genera Insectorum (ed. P. A. G. Wytsman), Fasc. 194. Desmet. Verteneuill, Brussels
72. Eede PV, Van HN, Overmeir CV et al (2009) Human *Plasmodium knowlesi* infections in young children in central Vietnam. Malaria Journal 8:249

73. Eng J, Mccormack AL, Yates JR (1994) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. J Am Soc Mass Spectrom 5:976-989
74. Eshghy N, Janbakhsh B (1976) Insecticide resistance of *Anopheles stephensi* mysorensis in the province of Fars, southern Iran, 1975. Mosq News 36:336-339
75. Farabaugh PJ (1996) Programmed translational frameshifting. Microbiol Rev 60:103-134
76. Felix RC, Silveira H (2011) The interplay between tubulins and P450 cytochromes during *Plasmodium berghei* invasion of *Anopheles gambiae* midgut. PLoS ONE 6:e24181.
77. Fermin D, Allen BB, Blackwell TW et al (2006) Novel gene and gene model detection using a whole genome open reading frame analysis in proteomics. Genome Biol 7:35
78. *Bacillus thuringiensis var israelensis* and *Bacillus sphaericus* formulations against Afrotropical anophelines in Western Kenya. Trop Med Int Health 8:37-47
79. Fontaine A, Fusaï T, Briolant S, Buffet S, Villard C, Baudelet E, Pophillat M, Granjeaud S, Rogier C, Almeras L (2012) *Anopheles* salivary gland proteomes from major malaria vectors. BMC Genomics 13:614
80. Fontaine MC, Pease JB2, Steele A3, Waterhouse RM, et al. (2015) Mosquito genomics. xtensive introgression in a malaria vector species complex revealed by phylogenomics. Science, 2; 347(6217):1258524
81. Francis C (2010) History of the discovery of the malaria parasites and their vectors. Parasites & Vectors 3:5
82. Freeman EG, Wisotsky Z, Dahanukar A (2014) Detection of sweet tastants by a conserved group of insect gustatory receptors. Proc Natl Acad Sci 111:1598-1603
83. Garcia GE, Wirtz RA, Rosenberg R (1997) Isolation of a substance from the mosquito that activates *Plasmodium* fertilization. Mol. Biochem Parasitol 88:127-135
84. Gardner MJ, Hall N, Fung E, White O, Berriman M et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419:498-511
85. 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

86. Garver LS, Dong Y, Dimopoulos G (2009) Caspar controls resistance to *Plasmodium falciparum* in diverse anopheline species. PLoS Pathog 5(3):e1000335
87. Gautam P, Nair SC, Gupta MK, Sharma R, Polisetty RV, Uppin MS et al. (2012) Proteins with Altered Levels in Plasma from Glioblastoma Patients as Revealed by iTRAQ-Based Quantitative Proteomic Analysis. PLos One 7(9):e46153
88. Geng YJ, Gao ST, Huang DN, Zhao YR, Liu J, Li XH, Zhang RL (2009) Differentially expressed genes between female and male adult *Anopheles anthropophagus*. Parasitol Res 105:843-851
89. Ghosh AK, Ribolla PE, Jacobs-Lorena M (2001) Targeting *Plasmodium* ligands on mosquito salivary glands and midgut with a phage display peptide library. Proc Natl Acad Sci USA 98:13278-81
90. Gillies MT, Coetzee M (1987) A supplement to the Anophelinae of Africa south of the Sahara (Afrotropical region). Johannesburg: Publication of the South African Institute for Medical Research, No.55
91. Goncalves RL, Oliveira JH, Oliveira GA, Andersen JF, Oliveira MF, Oliveira PL, Barillas-Mury C (2012) Mitochondrial reactive oxygen species modulate mosquito susceptibility to *Plasmodium* infection. PLoS ONE 7:e41083.
92. Gonzalez-Lazaro M, Dinglasan RR, Hernandez-Hernandez 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 39:395-402
93. Graf JF (1993) The role of insect growth regulators in arthropod control. Parasitology Today 9:471-474
94. Grewal SS (2009) Insulin/TOR signaling in growth and homeostasis: A view from the fly world. Int J Biochem Cell Biol 41:1006-1010
95. Gubler DJ, Reiter P, Ebi KL, Yap W, Nasci R, Patz JA (2001) Climate variability and change in the United States: potential impacts on vector and rodent borne diseases. Environ Health Perspect 109:223-233

96. Gupta A, Singh AK, Qayum A (2014) Evaluating Ancient Indian History and Culture towards Malaria Control Measures Based on Traditional Knowledge System. International Journal of Interdisciplinary and Multidisciplinary Studies 1:181-190
97. Gupta L, Noh JY, Jo YH, Oh SH, Kumar S, Noh MY, Lee YS, Cha SJ, Seo SJ, Kim I et al (2010) Apolipophorin-III mediates antiplasmodial epithelial responses in *Anopheles gambiae* (G3) mosquitoes. PLoS ONE 5:e15410.
98. Gupta S, Gunter JT, Novak RJ, Regens JL (2009) Patterns of *Plasmodium vivax* and *Plasmodium falciparum* malaria underscore importance of data collection from private health care facilities in India. Malaria Journal 8:227
99. Habtewold T, Povelones M, Blagborough AM, Christophides GK (2008) Transmission blocking immunity in the malaria non-vector mosquito *Anopheles quadriannulatus* species A. PLoS Pathog 4:e1000070.
100. Hahn MW, Lanzaro GC (2005) Female-biased gene expression in the malaria mosquito *Anopheles gambiae*. Curr Biol 15:192-193
101. Harbach RE (2004) The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. Bull Entomol Res 94:537-553
102. Harbach RE (2012) Mosquito taxonomic inventory, http://mosquito-Taxonomic-Inventory.info
103. Harris C, Lambrechts L, Rousset F, Abate L, Nsango SE, Fontenille D, Morlais I, Cohuet A (2010) Polymorphisms in *Anopheles gambiae* immune genes associated with natural resistance to *Plasmodium falciparum*. PLoS Pathog 6:e1001112.
104. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR et al. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298:129-149
105. Ito J, Ghosh A, Moreira LA, Wimmer EA, Jacobs-Lorena M (2002) Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. Nature 417:452-5
106. Jambulingam P, Mohapatra SS, Das LK, Das PK, Rajagopalan PK (1989) Detection of *Plasmodium ovale* in Koraput district, Orissa state. Indian Journal of Medical Research 89:115-116

107. Jaramillo-Gutierrez G, Molina-Cruz A, Kumar S, Barillas-Mury C (2010) The *Anopheles gambiae* oxidation resistance 1 (OXR1) gene regulates expression of enzymes that detoxify reactive oxygen species. PLoS ONE 5:e11168
108. Jaramillo-Gutierrez G, Molina-Cruz A, Kumar S, Barillas-Mury C (2010) The *Anopheles gambiae* oxidation resistance 1 (OXR1) gene regulates expression of enzymes that detoxify reactive oxygen species. PLoS ONE 5:e11168.
109. 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
110. 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.
111. Jariyapan N, Choochote W, Jitpakdi A, Harnnoi T, Siriyasatein P, Wilkinson MC, Junkum A, Bates PA (2007) Salivary gland proteins of the human malaria vector, *Anopheles dirus* B (Diptera: Culicidae). Rev. Inst. Med. trop. S. Paulo 49:5-10
112. Jariyapan N, Roytrakul S, Paemanee A, Junkum A, Saeung A, Thongsahuan S, Sor-suwan S, Phattanawiboon B, Poovorawan Y, Choochote W (2012) Proteomic analysis of salivary glands of female *Anopheles barbirostris* species A2 (Diptera: Culicidae) by two-dimensional gel electrophoresis and mass spectrometry. Parasitol Res 111:1239-49
113. Jiang X, Peery A, Hall AB, Sharma A, Chen XG, Waterhouse RM, Komissarov A, Riehle MM, Shouche Y, Sharakhova MV, Lawson D, Pakpour N, Arensburger P, Davidson VL, Eiglmeier K, Emrich S, George P, Kennedy RC, Mane SP, Maslen G, Oringanje C, Qi Y, Settlage R, Tojo M, Tubio JM, Unger MF, Wang B, Vernick KD, Ribeiro JM, James AA, Michel K, Riehle MA, Luckhart S, Sharakhov IV, Tu Z (2014) Genome analysis of a major urban malaria vector mosquito, *Anopheles stephensi*. Genome Biol 15:459
114. Kajla MK, Shi L, Li B, Luckhart S, Li J, Paskewitz SM (2011) A new role for an old antimicrobial: lysozyme c-1 can function to protect malaria parasites in *Anopheles* mosquitoes. PLoS ONE 6(5):e19649

115. Kalra NL (1991) Forest malaria vectors in India: Ecological characteristics and epidemiological implications. In: Proceedings of an Informal Consultative Meeting WHO/MRC18-22 February (Eds. VP Sharma and AV Kondrashin, New Delhi), p93-114
116. Kalume D E, Peri S, Zhong J, Okulate M, Reddy R, Kumar N, Pandey A (2005) Genome annotation of *Anopheles gambiae* using mass spectrometry-derived data. BMC Genomics 6:128
117. Kamhawi S, Ramalho-Ortigao M, Pham VM, Kumar S, Lawyer PG, Turco SJ, Barillas-Mury C, Sacks DL, Valenzuela JG (2004) Role for Insect Galectins in Parasite Survival. Cell 119:329-341
118. Kar NP, Kumar A, Singh OP, Carlton JM, Nanda N (2014) A review of malaria transmission dynamics in forest ecosystems. Parasites & Vectors 7:265
119. Kelkar DS, Kumar D, Kumar P, Balakrishnan L, Muthusamy B, Yadav AK, Shrivastava P, Marimuthu A et al (2011) Proteogenomic analysis of *Mycobacterium tuberculosis* by high resolution mass spectrometry. *Molecular and Cellular Proteomics* 10:M111.011627
120. Kim M, Wessely V, Lan Q (2005) Identification of mosquito sterol carrier protein-2 inhibitors. Journal of Lipid Research 46:650-657
121. Kim MS, Pinto S, Getnet D, Nirujogi RS, Manda S, Chaerkady R, Madugundu A , Kelkar D et al (2014) A draft map of the human proteome. Nature 509:575-581
122. Koenraadt CJM, Githeko AK, Takken W (2004) The effects of rainfall and evapo-transpiration on the temporal dynamics of *Anopheles gambiae* s.s. and *Anopheles arabiensis* in a Kenyan village. Acta Trop 90:141-153
123. Koga H, Shimada K, Hara Y, Nagano M, Kohga H, Yokoyama R, Kimura Y, Yuasa S, Magae J, Inamoto S, Okazaki N, Ohara O (2004) A comprehensive approach for establishment of the platform to analyze functions of KIAA proteins: generation and evaluation of anti-mKIAA antibodies. Proteomics 4:1412-6
124. Kokoza V, Ahmed A, Cho WL, Jasinskiene N, James AA, Raikhel A (2000) Engineering blood meal activated systemic immunity in the yellow fever mosquito, *Aedes aegypti*. Proc Natl Acad Sci USA 97:9144-9

125. Korgaonkar NS, Kumar A, Yadav RS, Kabadi D, Dash AP (2012) Mosquito biting activity on humans and detection of *Plasmodium falciparum* infection in *Anopheles stephensi* in Goa, India. Indian Journal of Medical Research 135:120-6
126. Koutsos AC, Blass C, Meister S, Schmidt S, Maccallum RM, Soares MB, Collins FH, Benes V, Zdobnov E, Kafatos FC et al. (2007) Life cycle transcriptome of the malaria mosquito *Anopheles gambiae* and comparison with the fruitfly *Drosophila melanogaster*. Proc Natl Acad Sci 104:11304-11309
127. Krishnan KS. *A. stephensi* Liston, 1901 (1961) Vectors of malaria in India. 2nd ed. Delhi: National Society of India for Malaria and Other Mosquito-born Disease; p. 39-58
128. Kuang AA, Diehl GE, Zhang J, Winoto A (2000) FADD is required for DR4- and DR5-mediated apoptosis: Lack of TRAIL-induced apoptosis in FADD-deficient mouse embryonic fibroblasts. J Biol Chem 275:25065-25068
129. Kumar A (1997) Urban Malaria and its control in India. J Parasitic Dis 21:83-88
130. Kumar A, Chery L, Biswas C, Dubhashi N, Dutta P, Dua VK, Kacchap M, Kakati S et al (2012) Malaria in South Asia: Prevalence and Control . Acta Tropica 121:246-255
131. Kumar A, Dua VK, Rathod P (2011) Malaria –attributed death rates in India. The Lancet 337:991-992
132. Kumar A, Malhotra MS, DeSouza M, Valecha N (2014) Deciphering the recent malaria trends in India. Major tropical diseases: Public health perspective. Eds. Ashwani kumar, Savio Rodrigues and Amit Dias. Published by Broadway publishing house, p51-63
133. Kumar A, Sharma VP, Dash AP (2006) Biology and control of *Anopheles stephensi* Liston in Vector Biology Eds. VP Sharma and Jagbir Singh Kirti. Publ. The National Academy of Sciences of India, Allahabad, 55-69
134. Kumar A, Sumodan PK, Thavaselvam D (1996) Malaria Control with Biolarvicide *Bacillus thuringiensis* var. *israelensis* and Larvivorous fish *Aplocheilus blocki* in Candolim PHC, Goa, India. Global meet on Parasitic diseases. 18-22 March 1996, New Delhi p264

135. Kumar A, Thavaselvam D, Sharma VP, Sumodan PK, Kamat RH, Audi SS, Surve BN (1995) Control of *Culex quinquefasciatus* by *Bacillus sphaericus* B 101 serotype 5a 5b. *XII* National Congress of Parasitology 141:5
136. Kumar A, Valecha N, Jain T, Dash AP (2007) Burden of Malaria in India: Retrospective and Prospective View. Am J Trop Med Hyg 77:69-78
137. 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.
138. Kumari R, Thapar BR, Das Gupta RK, Kaul SM, Lal S (1998) Susceptibility status of malaria vectors to insecticides in India. J Commun Dis 30:179-185
139. Lal AA, Patterson PS, Sacci JB, Vaughan JA, Paul C, Collins WE, Wirtz RA, Azad AF (2001) Anti-mosquito midgut antibodies block development of *Plasmodium falciparum* and *Plasmodium vivax* in multiple species of *Anopheles mosquitoes* and reduce vector fecundity and survivorship. Proc Natl Acad Sci USA 98:5228-33
140. Lal S, Sonal GS, Phukan PK (2000) Status of Malaria in India. Journal of Indian Academy of Clinical Medicine 5:19-23
141. Lapcharoen P, Komalamisra N, Rongsriyam Y, Wangsuphachart V, Dekumyoy P, Prachumsri J, Kajla MK, Paskewitz SM (2012) Investigations on the role of a lysozyme from the malaria vector *Anopheles dirus* during malaria parasite development. Dev Comp Immunol 36:104-111
142. Laplante M, Sabatini DM (2012) mTOR signaling in growth control and disease. Cell 149:274-293
143. Lasonder E, Ishihama Y, Andersen JS, Vermunt AM et al (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. Nature 419: 537-542
144. Lefevre T, Thomas F, Schwartz A, Levashina E, Blandin S, Brizard JP, Bourligu L, Demettre E, Renaud F, Biron DG (2007) Malaria *Plasmodium* agent induces alteration in the head proteome of their *Anopheles* mosquito host. Proteomics 7:1908-1915

145. Li RYJ (2003) Towards global analysis of mosquito chorion proteins through sequential extraction, two-dimensional electrophoresis and mass spectrometry. Proteomics. 3:2036-2043
146. Liston (1901) *Anopheles stephensi* The Indian Medical Gazette 36:361-366, 441-443
147. Lund AA, Rhoads DM, Lund AL, Cerny RL, Elthon TE (2001) In vivo modifications of the maize mitochondrial small heat stress protein, HSP22. J Biol Chem 276:29924-29929
148. Manguin S, Garros C, Dusfour I, Harbach RE, Coosemans M (2008) Bionomics, taxonomy and distribution of the major malaria vector taxa of *Anopheles* subgenus *Cellia* in Southeast Asia: An updated review. Infect Genet Evol 8:489-503
149. Mann M, Pandey A (2001) Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. Trends Biochem Sci 26:54-61
150. Manouchehri AV, Javadian E, Eshighy N, Motabar M (1976) Ecology of *Anopheles stephensi* Liston in southern Iran. Trop Geogr Med 28:228-32
151. Marathe A, Date V, Shah HN, Tripathi JR (2006) *Plasmodium ovale* a case report from Gujarat. Journal of Vector Borne Disease 43:206-208
152. Marinotti O, Calvo E, Nguyen QK, Dissanayake S, Ribeiro JMC et al. (2006) Genome-wide analysis of gene expression in adult *Anopheles gambiae*. Insect Mol Biol 15:1-12
153. Marinotti O, Cerqueira GC, de Almeida LGP, Ferro MIT, da Silva Loreto EL (2013) The Genome of *Anopheles darlingi*, the main neotropical malaria vector. Nucleic Acids Research 15:7387-7400
154. Marinotti O, Ngo T, Kojin B, Chou SP, Nguyen B, Juhn J, Carballar-Lejarazú R, Marinotti B, Jiang X, Walter M, Tu Z, Gershon PD, James AA (2014) Integrated proteomic and transcriptomic analysis of the Aedes aegypti eggshell. BMC Developmental Biology 14:15
155. Marsick BM, Miguel-Ruiz JES, Letourneau PC (2012) Activation of Ezrin/radixin/moesin mediates attractive growth cone guidance through regulation of growth cone actin and adhesion receptors. J Neurosci 32:282-296

156. Megy K, Emrich SJ, Lawson D, Campbell D, Dialynas E, Hughes D, Koscielny G, Louis C et al. (2012) VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res 40:729-734
157. Meister S, Agianian B, Turlure F, Relogio A, Morlais I, Kafatos FC, Christophides GK (2009) *Anopheles gambiae* PGRPLC-mediated defense against bacteria modulates infections with malaria parasites. *PLoS Pathog* 5:e1000542.
158. 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 102**:**11420-11425
159. Mendes AM, Awono-Ambene PH, Nsango SE, Cohuet A, Fontenille D, Kafatos FC, Christophides GK, Morlais I, Vlachou D (2011) Infection intensity-dependent responses of *Anopheles gambiae* to the African malaria parasite *Plasmodium falciparum*. *Infect Immun* 79:4708-4715.
160. Mendes AM, Schlegelmilch T, Cohuet A, Awono-Ambene P, De Iorio M, Fontenille D, Morlais I, Christophides GK, Kafatos FC, Vlachou D (2008) Conserved mosquito/parasite interactions affect development of *Plasmodium falciparum* in Africa. PLoS Pathog 4(5):e1000069
161. Menon R, Zhang Q, Zhang Y, Fermin D, Bardeesy N, De Pinho RA, Lu C, Hanash SM, Omenn GS, States DJ (2009) Identification of novel alternative splice isoforms of circulating proteins in a mouse model of human pancreatic cancer. Cancer Res 69:300-309
162. Michel K, Budd A, Pinto S, Gibson TJ, Kafatos FC (2005) *Anopheles gambiae* SRPN2 facilitates midgut invasion by the malaria parasite *Plasmodium berghei*. EMBO Rep 6:891-897.
163. Michel K, Suwanchaichinda C, Morlais I, Lambrechts L, Cohuet A, Awono-Ambene PH, Simard F, Fontenille D, Kanost MR, Kafatos FC (2006) Increased melanizing activity in *Anopheles gambiae* does not affect development of *Plasmodium falciparum*. Proc Natl Acad Sci U S A 103:16858-16863.
164. Mikolajczak SA, Silva-Rivera H, Peng X, Tarun AS, Camargo N, Jacobs-Lorena V, Daly TM, Bergman LW, de la Vega P, Williams J (2008) Distinct malaria parasite

sporozoites reveal transcriptional changes that cause differential tissue infection competence in the mosquito vector and mammalian host. Mol Cell Biol 28:6196-6207

165. Mishra B, Mirdha BR, Samantray JC (1999) *Plasmodium ovale* malaria in Delhi. Indian Journal of Pediatrics 66:143-144

166. Mitri C, Jacques JC, Thiery I, Riehle MM, Xu J, Bischoff E, Morlais I, Nsango SE, Vernick KD, Bourgouin C (2009) Fine pathogen discrimination within the APL1 gene family protects *Anopheles gambiae* against human and rodent malaria species. PLoS Pathog 5:e1000576.

167. Mohapatra PK, Prakash A, Bhattacharyya DR (2008) Detection & molecular confirmation of a focus of *Plasmodium malariae* in Arunachal Pradesh, India. Indian Journal of Medical Research 128:52-56

168. Molina H, Bunkenborg J, Reddy GH, Muthusamy B, Scheel PJ, Pandey A (2005) A proteomic analysis of human hemodialysis fluid. Mol Cell Proteomics. 4:637-650

169. 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 109:1957-1962

170. Mongin E, Louis C, Holt RA, Birney E, Collins FH (2004) The *Anopheles gambiae* genome: an update. Trends Parasitol 20:49-52

171. Moreno SNJ, Docampo R (2003) Calcium regulation in protozoan parasites. Curr Opin Microbiol 6:359-364

172. Motabar M, Tabibzadeh I, Manouchehri AV (1975) Malaria and its control in Iran. Trop Geogr Med 27:71-78

173. Mukunda L, Miazzi F, Kaltofen S, Hansson BS, Wicher D (2014) Calmodulin modulates insect odorant receptor function. Cell Calcium 55:191-199

174. Murray CJL, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, Fullman N, Naghavi M, Lozano R, Lopez AD (2012) Global malaria mortality between 1980 and 2010: a systematic analysis. The Lancet 379:413-431

175. Murray CJL, Lopez AD (1996) Evidence-based health policy-lessons from the Global Burden of Disease Study. Science 274:740-743

176. Murray CJL, Lopez AD (1997) The Global Burden of Disease 1990-2020: alternative projections of mortality and disability by cause for eight regions. The Lancet 349:1498-1504
177. Nagpal BN, Sharma VP (1995) Indian Anophelines. Oxford & IBH Publishing Co. Pvt. LTD. New Delhi (ISBN 81-204-0929-9), p416
178. National Vector Borne Disease Control Programme. Government of India. http:// www.nvbdcp.gov.in (accessed 30 June 2015)
179. Neafsey DE, Christophides GK, Collins FH, Emrich SJ, Fontaine MC, Gelbart W, Hahn MW, Howell PI, Kafatos FC, Lawson D, Muskavitch MAT, Waterhouse RM, Williams LJ, Besansky NJ (2013) The Evolution of the *Anopheles* 16 Genomes Project G3 (Bethesda):1191-1194
180. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arcà B, Arensburger P, Artemov G et al (2015) Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. Science 347: 6217
181. Neira OM, Vanekeris L, Corena-McLeod MD, Linser PJ (2008) A microarray-based analysis of transcriptional compartmentalization in the alimentary canal of *Anopheles gambiae* (Diptera: Culicidae) larvae. Insect Mol Biol 17:61-27
182. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ et al. (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316:1718-1723
183. Nesvizhskii A, Vitek O, Aebersold R (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. Nature methods 4:787-797
184. Nirujogi RS, Pawar H, Renuse S et al. (2013) Moving from unsequenced to sequenced genome: Reanalysis of the proteome of *Leishmania donovani*. J Proteomics 93:48-61
185. Nolan T, Petris E, Müller HM, Cronin A, Catteruccia F et al. (2011) Analysis of two novel midgut-specific promoters driving transgene expression in *Anopheles stephensi* mosquitoes. PLoS One6: e16471
186. Oliveira JH, Goncalves RL, Oliveira GA, Oliveira PL, Oliveira MF, Barillas-Mury C (2011) Energy metabolism affects susceptibility of *Anopheles gambiae* mosquitoes to *Plasmodium* infection. Insect Biochem Mol Biol 41:349-355

187. Olsen JV (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 4:2010-2021
188. Osta MA, Christophides GK, Kafatos FC (2004) Effects of mosquito genes on *Plasmodium* development. Science 303:2030-2032.
189. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, Kimura K, Makita H, Sekine M, Obayashi M, Nishi T, Shibahara T, Tanaka T, Ishii S et al. (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet 36:40-5
190. Panda R, Verma KV, Rahman SJ (1990) Present status of *Plasmodium malariae* infection in Bastar district (M.P). J Commun Dis 22:185-90
191. Pant CP (1988) Malaria vector control: imagociding. In: Malaria: Principles and practicie of malariology. Wernsdorfer WH, McGregor I,eds. Churchill Livingstone, Edinburgh, UK. 1173-1212
192. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S et al. (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. Proc Natl Acad Sci 103:18928-18933
193. Paskewitz SM, Shi L (2005) The hemolymph proteome of *Anopheles gambiae*. Insect Biochem Mol Biol 35:815-824
194. Payne SH, Huang ST, Pieper R (2010) A proteogenomic update to *Yersinia*: Enhancing genome annotation. BMC Genomics 11:460
195. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551-67
196. Pervez SD, Shah IH (1989) Role of *Anopheles stephensi* as malaria vector in rural areas of Pakistan. Pak J Health 26:73-84
197. Pinto SB, Lombardo F, Koutsos AC, Waterhouse RM, McKay K, An C, Ramakrishnan C, Kafatos FC, Michel K (2009) Discovery of *Plasmodium* modulators by genome-wide analysis of circulating hemocytes in *Anopheles gambiae*. Proc Natl Acad Sci U S A 106:21270-21275.

198. 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.
199. Prakash A, Mohapatra PK, Bhattacharya DR, Goswami BK, Mahanta J (2003) *Plasmodium ovale*: First case report from Assam. Curr Sci 84:1187-8
200. Prasad TS, Harsha HC, Keerthikumar S et al (2012) Proteogenomic analysis of *Candida glabrata* using high resolution mass spectrometry. J Proteome Res 11:247-260
201. Prévot GI, Laurent-Winter C, Feldmann AM, Rodhain F, Bourgouin C (1998) Two-dimensional gel analysis of midgut proteins of *Anopheles stephensi* lines with different susceptibility to *Plasmodium falciparum* infection. Insect Mol Biol 7:375-383
202. Prevot GI, Laurent-Winter C, Rodhain F, Bourgouin C (2003) Sex-specific and blood meal-induced proteins of *Anopheles gambiae* midguts: analysis by two-dimensional gel electrophoresis. Malaria Journal 2:1-7
203. Puttamallesh VN, Sreenivasamurthy SK, Singh PK, Harsha HC, Ganjiwale A, Broor S, Pandey A, Narayana J, Prasad TSK (2013) Proteomic profiling of serum samples from chikungunya-infected patients provides insights into host response. Clinical Proteomics 10:14
204. Raikhel AS, Kokoza VA, Zhu J, Martin D, Wang S-F, Li C, Sun G, Ahmed A, Dittmer N, Attardo G (2002) Molecular biology of mosquito vitellogenesis: from basic studies to genetic engineering of anti-pathogen immunity. Insect Biochem Mol Biol 32:1275-1286
205. 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. PNAS 112:1273-1280
206. Rao BA, Sweet WC, Rao AMS (1938) Ova measurements of *An. stephensi* type and *An. stephensi var. mysorensis.* J Malar Inst Ind 1:261-266
207. Rao TR (1984) The anophelines of India. Malaria Research Centre, Indian Council of Medical Research, New Delhi, India

208. Reinert JF, Harbach RE, Kitching IJ (2009) Phylogeny and classification of tribe *Aedini* (Diptera: Culicidae). Zool J Linnean Soc 157:700-799
209. Reisen WK (1986) Population dynamics of some Pakistan mosquitoes: the impact of residual organophosphate insecticide spray on anopheline relative abundance. Ann Trop Med Parasitol 80:69-75
210. Remm M, Storm CE, Sonnhammer ELL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314:1041-1052
211. Renuse S, Chaerkady R, Pandey A (2011) Proteogenomics. Proteomics 11:620-30
212. Rich SM, Ayala FJ (2006) Evolutionary Origins of Human Malaria Parasites. In Krishna R. Dronamraju, Paolo Arese (Ed). Emerging Infectious Diseases of the 21st Century: Malaria - Genetic and Evolutionary Aspects. Springer US. pp.125-146
213. Riehle MM, Markianos K, Niare O, Xu J, Li J, Toure AM, Podiougou B, Oduol F, Diawara S, Diallo M et al (2006) Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. Science 312:577-579.
214. 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.
215. Rodrigues J, Oliveira GA, Kotsyfakis M, Dixit R, Molina-Cruz A, Jochim R, Barillas-Mury C (2012) An epithelial serine protease, AgESP, is required for *Plasmodium* invasion in the mosquito *Anopheles gambiae*. PLoS ONE 7:e35210.
216. Rono MK, Whitten MM, Oulad-Abdelghani M, Levashina EA, Marois E (2010) The major yolk protein vitellogenin interferes with the anti-plasmodium response in the malaria mosquito *Anopheles gambiae*. PLoS Biol 20(7):e1000434
217. Roy RG, Chakrapani KP, Zacharias J (1977) *Plasmodium malariae* in Kerala and Tamil Nadu States of India (1969-75). Trans R Soc Trop Med Hyg 71:447
218. Roy RG, Shanmugham CAK, Chakrapani KP, Madesayya NM, Ghosh RB (1979) Distribution of human *plasmodia* in Karnataka and Tamil Nadu States in the reappearing phase of malaria in 1970 and 1975. Indian J Med Res 69:53-9
219. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-386

220. Saboia-Vahia L, Borges-Veloso A, Cuervo A, Junqueira M, Mesquita-Rodrigues C, Britto C, Domont GB, Jesus JBD (2012) Protein expression in the midgut of sugar-fed *Aedes albopictus* females. Parasite Vectors 5:290
221. Sehrawat N, Gakhar SK (2014) Mosquito Proteomics: Present and Future Prospective. Research in Biotechnology 5:25-33
222. Serrano-Pinto V, Acosta-Perez M, Luviano-Bazan D, Hurtado-Sil G, Batista C, Martinez-Barnetche J, Lanz-Mendoza H (2010) Differential expression of proteins in the midgut of *Anopheles* infected with *Plasmodium berghei*. Insect Biochemistry and Molecular Biology 40:752-758
223. Shahabuddin M, Kaslow DC (1994) *Plasmodium*: Parasite chitinase and its role in malaria transmission. Exp Parasitol 79:85-88
224. Shahabuddin M, Lemos FJA, Kaslow DC, Jacobs-Lorena M (1996) Antibody-mediated inhibition of *Aedes aegypti* midgut trypsins blocks sporogonic development of *Plasmodium gallinaceum*. Infect Immun 64:739-743
225. Shahabuddin M, Toyoshima T, Aikawa M, Kaslow DC (1993) Transmission-blocking activity of a chitinase inhibitor and activation of malarial parasite chitinase by mosquito protease. Proc Natl Acad Sci 90:4266-70
226. Sharma V P (1998) Fighting malaria in India. Curr Sci 75:1127-1140
227. Sharma VP (1993) Ecosystem approach to malaria control. Proc Natl Acad Sci India 63:47-55
228. Sharma VP (1999) Current scenario of Malaria in India. Parassitologia 41:349-53
229. Sharma VP (2012) Continuing challenge of malaria in India. Current science 102:678-682
230. Sharma VP, Ansari MA, Mittal PK, Razdan RK, Batra CP (1986) Impact of malathion thermal fogging on mosquito population in Delhi and its place in malaria control. Indian J Malariol 23:65
231. Shih KM, Fallon AM (2001) Two dimensional electrophoretic analysis of *Aedes aegypti* mosquito fat body proteins during a gonotropic cycle. Am J Trop Med Hyg 65:42-46
232. Shin SW, Zou Z, Raikhel AS (2011) A new factor in the *Aedes aegypti* immune response: CLSP2 modulates melanization. EMBO Rep 12:938-943.

233. Silva-Neto MA, Atella GC, Shahabuddin M (2002) Inhibition of Ca^{2+} calmodulin-dependent protein kinase blocks morphological differentiation of *Plasmodium gallinaceum* zygotes to ookinetes. J Biol Chem 277:14085-14091
234. Singh B, Sung LK, Radhakrishnan A (2004) A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. The Lancet 363:1017-1024
235. Singh OP (2002) Bionomics of malaria vectors in India. Malaria Research Centre 19-31
236. Singh R, Jain V, Singh P P, Bharti P K, Thomas T, Basak S, Singh N (2013) First report of detection and molecular confirmation of *Plasmodium ovale* from severe malaria cases in central India. Tropical Medicine and International Health 18:1416-1420
237. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, Gething PW, Elyazar IR, Kabaria CW, Harbach RE, Hay SI (2011) The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. Parasit Vectors 4:89
238. 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.
239. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. Nature 434:214-17
240. Sor-Suwan S, Jariyapan N, Roytrakul S, Paemanee A, Saeung A, Thongsahuan S, Phattanawiboon B, Bates PA, Poovorawan Y, Choochote W (2013) Salivary gland proteome of the human malaria vector, Anopheles campestris-like (Diptera: Culicidae). Parasitol Res 112:1065-75
241. Sreenivasamurthy SK, Dey G, Ramu M, Kumar M, Gupta M K, Mohanty A K, Harsha HC, Sharma P, Kumar N, Pandey A, Kumar A, Prasad TSK (2013) A compendium of molecules involved in vector-pathogen interactions pertaining to malaria. Malaria Journal 12:216
242. Strüchler D (1989) How much malaria is there worldwide?. Parasitol Today 5:39-40
243. Subbarao SK (1998) Anopheline Species Complexes in South-East Asia. World Health Organization Technical Publication 18:82

244. Subbarao SK, Sharma VP, Vasantha K, Adak T (1994) Effect of malaria spraying on four anopheline species and the development of resistance in *Anopheles stephensi* in Mandora, Haryana. Indian J Malariol 21:109
245. Subbarao SK, Vasantha K, Adak T, Sharma VP, Curtis CF (1987) Egg-float ridge number in *Anopheles stephensi*: ecological variation and genetic analysis. Med Vet Entomol 1:265-71
246. Sumodan PK, Kumar A, Yadav RS (2004) Resting Behaviour and Incrimination of *Anopheles stephensi* in Goa, India. Journal of American Mosquito Control Association 20:317-318
247. Sweet WC, Rao B (1937) Races of *Anopheles stephensi* Liston, 1901. Ind Med Gaz 72:665-674
248. Tahar R, Boudin C, Thiery I, Bourgouin C (2002) Immune response of *Anopheles gambiae* to the early sporogonic stages of the human malaria parasite. *Plasmodium falciparum*. EMBO J 21:6673-6680
249. Tchankouo-Nguetcheu S, Khun H, Pincet L, Roux P, Bahut M, Huerre M, Guette C, Choumet V (2010) Differential Protein Modulation in Midguts of *Aedes aegypti* Infected with Chikungunya and Dengue 2 Viruses. PLoS ONE 5(10):e13149
250. Thammasittirong A, Dechklar M, Leetachewa S, Pootanakit K, Angsuthanasombat C (2011) *Aedes aegypti* Membrane-Bound Alkaline Phosphatase Expressed in *Escherichia coli* Retains High-Affinity Binding for *Bacillus thuringiensis* Cry4Ba Toxin. Applied And Environmental Microbiology 77:6836-6840
251. Thavaselvam D, Kumar A, Sumodan PK (1993) Insecticide susceptibility status of *Anopheles stephensi, Culex quinquefasciatus and Aedes aegypti* in Panaji, Goa. Indian J Malariol 30:75-79
252. Tikar SN, Mendki MJ, Sharma AK, Sukumaran D, Veer V, Prakash S (2011) Resistance status of the malaria vector mosquitoes, *Anopheles stephensi* and *Anopheles subpictus* towards adulticides and larvicides in arid and semi-arid areas of India. J Insec Sc 11:85
253. Tiwari S, Ghosh SK, Ojha VP, Dash AP, Raghavendra K (2010) Reduced susceptibility to selected synthetic pyrethroids in urban malaria vector *Anopheles stephensi*: a case study in Mangalore city, South India. Malar J 9:179

254. Valenzuela JG, Francischetti IMB, Pham VM, Garfield MK, Ribeiro JMC (2003) Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. Insect Biochemistry and Molecular Biology 33:717-732
255. Vannini L, Reed TW, Willis JH (2014) Temporal and spatial expression of cuticular proteins of *Anopheles gambiae* implicated in insecticide resistance or differentiation of M/S incipient species. Parasit Vectors 7:24
256. Vatandoost H, Oshaghi MA, Abaie MR, Shahi M, Yaaghoobi F, Baghaii M, Hanafi-Bojd AA, Zamani G, Townson H (2006) Bionomics of *Anopheles stephensi* Liston in the malarious area of Hormozgan province, Southern Iran, 2002. Acta Tropica 97:196-203
257. Vijay S, Rawal R, Kadian K, Raghavendra K, Sharma A (2015) Annotated Differentially Expressed Salivary Proteins of Susceptible and Insecticide-Resistant Mosquitoes of *Anopheles stephensi.* PLoS One 10(3):e0119666
258. 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 up-regulation. PLoS One6(4):e18400
259. Vijay S, Rawat M, Sharma A (2014) Mass Spectrometry Based Proteomic Analysis of Salivary Glands of Urban Malaria Vector *Anopheles stephensi*. Article ID 686319, 12 pages
260. Vlachou D, Schlegelmilch T, Christophides GK, Kafatos FC (2005) Functional genomic analysis of midgut epithelial responses in *Anopheles* during *Plasmodium* invasion. Curr Biol 15:1185-1195
261. Volz J, Muller HM, Zdanowicz A, Kafatos FC, Osta MA (2006) A genetic module regulates the melanization response of *Anopheles* to *Plasmodium*. Cell Microbiol 8:1392-1405.
262. Volz J, Osta MA, Kafatos FC, Muller HM (2005) The roles of two clip domain serine proteases in innate immune responses of the malaria vector *Anopheles gambiae*. J Biol Chem 280:40161-40168.

263. Wang B, Hajano JUD, Ren Y, Lu C, Wang X (2015) iTRAQ-based quantitative proteomics analysis of rice leaves infected by Rice stripe virus reveals several proteins involved in symptom formation. Virology Journal 12:99
264. Warr E, Aguilar R, Dong Y, Mahairaki V, Dimopoulos G (2007) Spatial and sexspecific dissection of the *Anopheles gambiae* midgut transcriptome. BMC Genomics 8:37
265. Warr E, Das S, Dong Y, Dimopoulos G (2008) The Gram-negative bacteria-binding protein gene family: its role in the innate immune system of *Anopheles gambiae* and in anti-*Plasmodium* defence. Insect Mol Biol 17:39-51
266. WHO. The world health report (1999) making a difference. http://www.who.int/whr/1999/en/index.html
267. WHO. World Malaria Report (2010) Geneva: World Health Organization, 2010. http://www.who.int/malaria/world
268. WHO. World Malaria Report (2014) WHO, Geneva. 2014. Available at http://apps.who.int/iris/bitstream/10665/144852/2/9789241564830_eng.pdf
269. WHO. World malaria situation in (1994) http://www.who.int/malaria/publications/atoz/wer7236-37-38/en/index.html
270. Wilke AB, Nimmo DD, St John O, Kojin BB, Capurro ML, Marrelli MT (2009) Mini-review: genetic enhancements to the sterile insect technique to control mosquito populations. Asia Pac J Mol Biol Biotechnol 17:65-74
271. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery I, Hochstrasser DF, Williams KL (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnology and Genetics Engineering Reviews 13:19-50
272. www.VectorBase.org, Bioinformatics resource for invertebrate vectors of human pathogens. accessed on 30th April and August 2015
273. Xia D, Sanderson SJ, Jones AR, Prieto JH et al. (2008) The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. Genome Biol 9:116
274. Yap HH (1985) Biological control of mosquitoes, especially malaria vectors, *Anopheles* species. Southeast Asian J Trop Med Pub Health 16:163-172

275. Yocupicio-Monroy M, Padmanabhan R, Medina F, Del Angel RM (2007) Mosquito La protein binds to the 3' untranslated region of the positive and negative polarity dengue virus RNAs and relocates to the cytoplasm of infected cells.Virology 357:29-40
276. Zahar AR (1974) Review of the ecology of malaria vectors in WHO Eastern Mediterranean Region. Bull Org Mond Sante 50:427
277. Zaim M, Ershadi MR, Manouchehri AV, Hamdi MR (1986) The use of CDC light traps and other procedures for sampling malaria vectors in southern Iran. J Am Mosq Control Assoc 2:511-515
278. Zhang Y, Cai Y, Yu H, Li H (2014) iTRAQ-Based Quantitative Proteomic Analysis Identified HSC71 as a Novel Serum Biomarker for Renal Cell Carcinoma. Bio Med Research International P1-6
279. Zhou D, Zhang D, Linna Shi GD, Hou Q (2014) Genome sequence of *Anopheles sinensis* provides insight into genetics basis of mosquito competence for malaria parasites. BMC Genomics 15:42
280. Zou Z, Souza-Neto J, Xi Z, Kokoza V, Shin SW, Dimopoulos G, Raikhel A (2011) Transcriptome analysis of *Aedes aegypti* transgenic mosquitoes with altered immunity. PLoS Pathog 7(11):e1002394