THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Doctor of Philosophy (Genetics)

DEGREE

Genetics	Genetics
FIELD	**DEPARTMENT**

TITLE: Meiotic study, mitochondrial genome analysis and the *vitellogenin* gene cloning in giant water bug, *Lethocerus indicus*.

NAME: Miss Wijit Wisoram

THIS THESIS HAS BEEN ACCEPTED BY

THESIS ADVISOR

(Associate Professor Lertluk Ngernsiri, Ph.D.)

THESIS CO-ADVISOR

(Associate Professor Arinthip Thamchiapenet, Ph.D.)

DEPARTMENT HEAD

(Associate Professor Arinthip Thamchiapenet, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON _____

DEAN

(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

MEIOTIC STUDY, MITOCHONDRIAL GENOME ANALYSIS AND THE *vitellogenin* GENE CLONING IN GIANT WATER BUG, *Lethocerus indicus*

WIJIT WISORAM

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy (Genetics),
Graduate School, Kasetsart University
2014

Wijit Wisoram 2014: Meiotic Study, Mitochondrial Genome Analysis and the *vitellogenin* Gene Cloning in Giant Water Bug, *Lethocerus indicus.* Doctor of Philosophy (Genetics), Major Field: Genetics, Department of Genetics. Thesis Advisor: Associate Professor Lertluk Ngernsiri, Ph.D. 131 pages.

This study comprises three experiments. First, the meiotic chromosome of *Lethocerus indicus* was studied in insect samples collected from Thailand, Myanmar, Loas, and Cambodia. Testicular cells stained with lacto-aceto orcein, Giemsa, DAPI, and silver nitrate were analyzed. The results revealed that the chromosome complement of *L.indicus* was 2n = 22A + neo-XY + 2m, which differed from that of previous reports. Each individual male contained testicular cells with three univalent patterns. The frequency of cells containing neo-XY chromosome univalent (~5%) was a bit higher than that of cells with autosomal univalents (~3%). Some cells (~0.5%) had both sex chromosome univalents and a pair of autosomal univalents. Second, the first complete mitogenome of the *L. indicus* (Hemiptera: Belostomitidae) was sequenced using long PCR-based approach. This mitogenome was a 17,632 bp circular molecule with a total A+T content of 70.51% and 69.45% for coding regions. The gene content and order are consistent with common features found in mitogen-nome of hemipteran. The A+T rich region comprises of two types of extensive tandem repeat. First type included five regions, four copies of 174 bp and one copy of 175 bp. The other type was a shorter 114 bp sequence aligned in nine tandem repeats. Third, complete nucleotide and deduced amino acid sequences of giant water bug, *LiVg* cDNA were identified and characterized from fat body female. A 6,048 bp with 5,664 bp of open reading frame followed by a termination codon (TGA) and a poly-adenylation signal (AATAAA). The *LiVg* cDNA coded 1,888 amino acids and the calculated molecular weight was 212.40 kDa. The *LiVg* expression in fat body was detected in only adult female.

_____ _____ ____ /____ /____

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

I would like to gratefully thank Associate Professor Lertluk Ngernsiri for her valuable suggestions and invaluable guidance, encouragement and kindness during working in her laboratory. I would like sincerely to thank Assoc. Prof. Dr. Arinthip Thumchiapenet my committee for her valuable comments and suggestions.

This research could not be completed if there were no financial support provided by office of the higher education commission. I am extremely grateful this program to give me not only a grant, but also the great opportunity for the education in genetics study at Kasetsart University.

I am especially appreciated my parents, my sisters and brothers for their unconditional support and encouragements. Finally, I deeply appreciated my husband, Wasan Kuntum who always devoted time, support and will power my success as well as my little daughter who motivate to write this my thesis, this thesis and everything else I will ever do, will always be dedicated to them.

Wijit Wisoram

June, 2014

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

A	=	Adenine
aa	=	amino acid
°C	=	degrees Celsius
%	=	Percent
BLAST	=	Basic Local Alignment Search Tool
Bp	=	base pair
C	=	Cytosine
C_T	=	Baseline-subtracted Threshold
cDNA	=	complementary DNA
DEPC	=	diethylpyrocarbonate
DNA	=	deoxyribonucleic acid
dNTP	=	deoxyriboNucleotideTriPhosphate
EtBr	=	Ethidium Bromide
g	=	grams
G	=	Guanine
GI	=	Genbank ID
H	=	hour
IPTG	=	Isopropyl β-D-1-thiogalactopyranoside
kDa	=	Kilo Dalton
LB	=	Luria Bertani
M	=	Molar
mg	=	milligrams
ML	=	Maximum Likelihood method
ml	=	milliliters
mM	=	millimolar
mRNA	=	messenger RNA
µg	=	micrograms
µL	=	microliters
µM	=	micromolar

LIST OF ABBREVIATIONS (Continued)

NJ	=	Neighbor Joining method
ng	=	nanogram
nt	=	nucleotide
PCGs	=	protein coding genes
PCR	=	Polymerase Chain Reaction
*p*mol	=	picomole
RACE	=	Rapid Amplification of cDNA Ends
RNA	=	ribonucleic acid
rRNA	=	ribosomal RNA
tRNA	=	transfer RNA
rpm	=	revolutions per minute
RT	=	Reverse Transcription
T	=	Thymine
TAE	=	Tris Acetate EDTA
Tm	=	Melting Temperature
U	=	units

MEIOTIC STUDY, MITOCHONDRIAL GENOME ANALYSIS AND THE *vitellogenin* GENE CLONING IN GIANT WATER BUG, *Lethocerus indicus*

INTRODUCTION

Giant water bugs are the largest insects distributed worldwide. Their body length can be up to 12 centimeters. The insects belong to the Family Belostomatidae with three subfamilies, Belostomatinae, Horvathininia and Lethocerinae (Lauck and Menke 1961; Perez-Goodwyn, 2006). The last subfamily consists of three genera, *Lethocerus* Mayr, 1853, *Kirkaldyia* Montandon, 1909, and *Benacus* Stal, 1861. The giant water bugs belong to the genus *Lethocerus* Mayr. Until now, 22 species of giant water bug have been identified. Of these 22 species, 16 species are found in the American continent and 6 species are distributed in the remaining of the world (Perez-Goodwyn 2006). Giant water bugs, *Lethocerus indicus* is the native giant water bug of Southeast Asia. In Thailand, people have used the male bug as an aromatic ingredient in some native curry pastes; Nam Prik Malaeng Da (Maeng Da Na) that is a popular dish. The fragrance is indeed a sex pheromone produced by male bugs to attract the females (Butenandt and Tam, 1957; Devakul and Maarse, 1964) which the male secretes a fragrant liquid from two abdominal glands. The male bugs are found widely in the markets during the rainy season, when the price ranges from 10 to 15 baht. They are more expensive in the dry season, costing 15 baht or more. At present artificial water bug flavoring is now produced, but people still prefer to eat the real bugs. With the increasing demand, unsuccessful rearing and change of its environment, the number of *L. indicus* in the nature of Thailand has gradually decreased so the bugs have been imported from neighboring countries such as Cambodia, Myanmar and Laos.

In this study, the complete mitochondrial genome (mitogenome) of *L. indicus* that is the first species from the Belostomatidae had been successfully sequenced and analyzed. The *vitellogenin* (*Vg*) gene was identified from fat body of *L. indicus*. In

addition, the chromosome complement and the behavior of meiotic chromosomes were studied using lacto-aceto orcein squash technique, Giemsa, DAPI and silver staining, which presents clear evidence about the existence of micro-(m) chromosomes in Belostomatidae.

OBJECTIVES

1. To analyse of the meiotic chromosome complements of *L. indicus* from Thailand, Laos, Myanmar and Cambodia.

2. To identify the complete mitochondrial genome sequence of *L. indicus*.

3. To clone and sequence *vitellogenin* (*Vg*) genes of *L. indicus*.

LITERATURE REVIEW

Giant water bug (*Lethocerus indicus*)

Giant water bugs are the largest insects distributed worldwide. Their body length can be up to 12 centimeters. The insects belong to the Family Belostomatidae with three subfamilies, Belostomatinae, Horvathininia and Lethocerinae (Lauck and Menke 1961; Perez-Goodwyn, 2006). The last subfamily consists of three genera, *Lethocerus* Mayr, 1853, *Kirkaldyia* Montandon, 1909, and *Benacus* Stal, 1861. The giant water bugs belong to the genus *Lethocerus* Mayr. Until now, 22 species of giant water bug have been identified. Of these 22 species, 16 species are found in the American continent and 6 species are distributed in the remaining of the world (Perez-Goodwyn, 2006). *L. indicus* is the native giant water bug of Southeast Asia. In Thailand, people have used the male bug as an aromatic ingredient in some native curry pastes. The fragrance is indeed a sex pheromone produced by male bugs to attract the females (Butenandt and Tam, 1957; Devakul and Maarse, 1964).

Scientific classification
Kingdom Animalia
Phylum Arthropoda
Class Insecta
Order Hemipeta
Family Belostomatidae
Subfamily Lethocerinae
Genus *Lethocerus*
Species *indicus*

The life cycle of giant water bug

A gradual change in form through the stages of growth; there are three stages in incomplete metamorphosis: egg, nymph and adult (Figure 1). Wings develop

as the nymph molts and are functional when the insect reaches the adult stage. The life cycle from egg to adult may require 32 to 43 days, depending on the temperature and an adult to reproductive stage use 30-45 days (Pongsart, 1990). Their life span is one year or longer. Reproduction of giant water bug may occur as early as July-October during fall season. During the rainy months, giant water bug populations can increase very quickly. A female *L. indicus* lays her eggs above water on plants (rice, grass and aquatic plants) and other objects. Males taking care the eggs until they hatch, taking care to expose them to air periodically to prevent the growth of fungus. They also "brood pump": making water move over the eggs to increase oxygen diffusion.

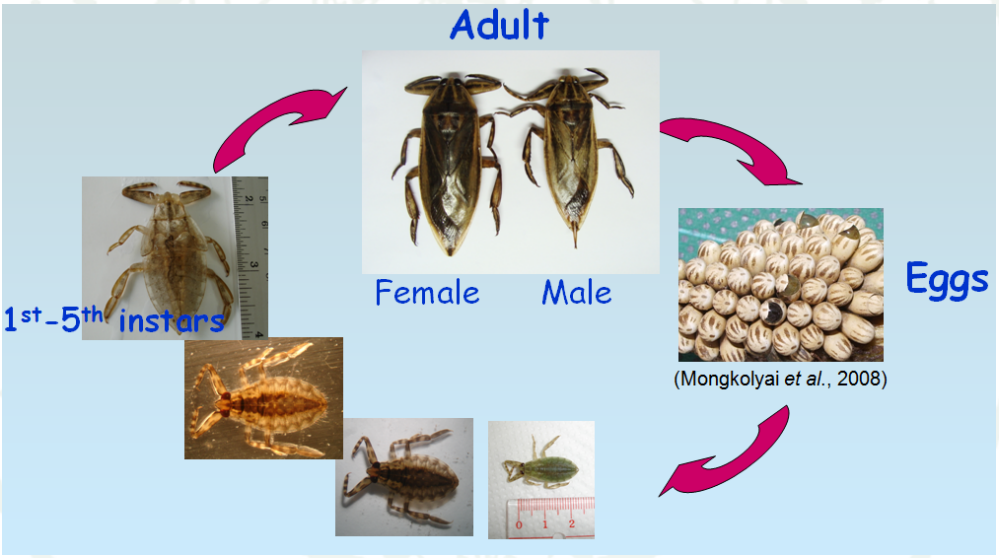


Figure 1 Life cycle of giant water bug (*Lethocerus indicus*)

Egg

An adult females lay her eggs above water anywhere such as on branches, grass and other objects, in clusters. Mucinous glue is secreted by the female prior to the deposition of each egg. Egg period is of 4-7 days. The eggs are whitish, oval in shape, 0.5 mm long (Pongsart, 1990; Sup-udom, 1992). Eggs hatch in 6 to 8 days (Mongkolvai *et al*., 2008). Twenty-four hours prior to hatching the free end of

the egg swells dorsal. Pressure from within eventually causes a rupture in the chorion around the cephalic cap (Figure 2). The cap is lifted by the head as the nymphal thorax emerges, but remains hinged to the egg on the side of the latter which faces the ventral surface of the emerging nymph. The nymph's head slips from under the operculum as peristaltic contractions free one haft of the body (Mongkolvai *et al*., 2008; Pongsart, 1990).

Figure 2 Egg cluster of giant water bug and the cephalic cap that is lifted bythe head.

Source: Mongkolvai *et al*., (2008)

Nymph

Giant water bug offspring are pale yellow in color for a few hours after birth and then assume their normal, darker color. Offspring spend much of their time near the water's surface, so they can occasionally stick their backsides out of the water to breathe. Nymphs have small tubes located on their hind end. These tubes act like snorkels and carry air throughout the animal's body. Metamorphosis of the giant water bug is incomplete, so nymphs look similar to their parents. Nymphs have five stages to change morphology by molting with the period of about 5-7 days. Total development time ranging from about 28-36 days before becoming adults that depending on temperature (Mongkolvai *et al*., 2008; Pongsart, 1990).

Adult

Adult giant water bugs molt from nymph and they develop to mature female and male for mating and lay egg immediately after emergence (Ayyar, 1934). Adults are light greyish-brown in color, 12 mm long and a wing span of about 15 mm, no any markings on the wings but veins are slightly darkened, a projected tuft of scales on head. An adult have long life around one year or longer. Sex pheromone is produced by adult males to attract the females. One female can lay eggs up to 19-320 within a few days after mating (Mongkolvai *et al*., 2008). Adult giant water bugs capture larger prey species by using their clawed front feet and chemicals which are injected into the body of the prey. The enzymes turn the prey's insides into liquid, which the giant water bug can suck up.

1.2 Physical Characteristics

Giant water bugs are approximately 5-7 cm in length and female size is normally bigger than male. The body is brown, flat and oval, giving them an appearance similar to that of a cockroach (Figure 3). The giant water bug's mouthparts are elongated into a beak-like structure designed for piercing and sucking (Figure 4). Instantly, the Giant water bug pierces its victim with its sharp beak and injects a powerful toxin. The toxin has two purposes. First, it paralyzes the prey. Then, it liquefies the internal parts of the prey's body so the hunter can suck up a liquid repast. The front legs are raptorial (grasping) to seize prey (Figure 5a). Their other two pairs of legs are flattened and fringed with hair to increase their surface area. These legs are used like paddles for propulsion (Figure 5b). Using its hind legs, the bug grabs hold of a plant growing close to the surface of a fresh water pond or stream and snatches passing prey with its powerful front legs. Adults have two pairs of wings (Figure 6), but they rarely fly unless forced to by unfavorable water conditions or lack of an adequate food supply. The posterior end of a giant water bug has two retractable, semi-cylindrical appendages which, when held together, form a breathing tube. This tube is used for underwater breathing. When in flight, air is exchanged through small openings of the respiratory system called spiracles (Figure 7).

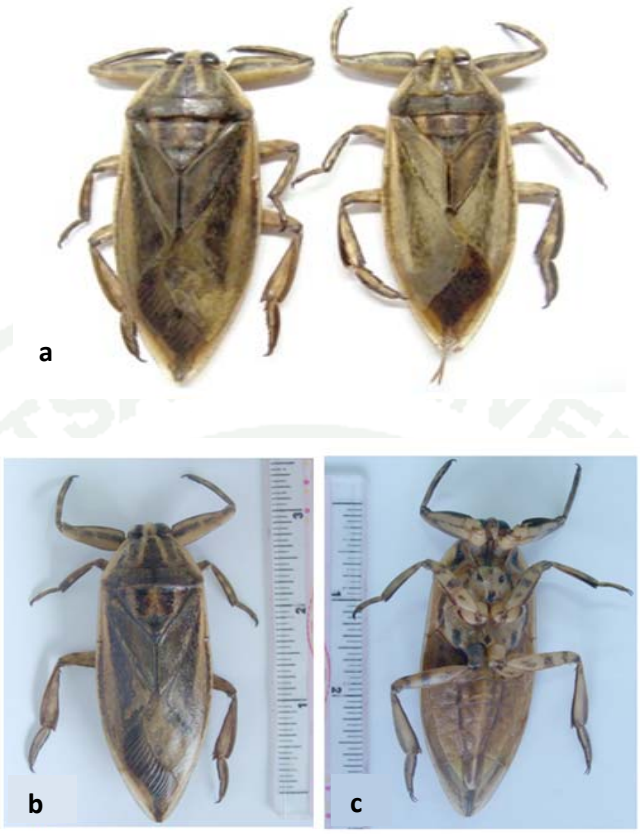


Figure 3 Dorsal view of an adult giant water bug. (a) Left: Female, Right: Male (b) Dorsal and (c) ventralview of an adult giant water bug. Photograph by Wijit Wisoram, Kasetsart University.

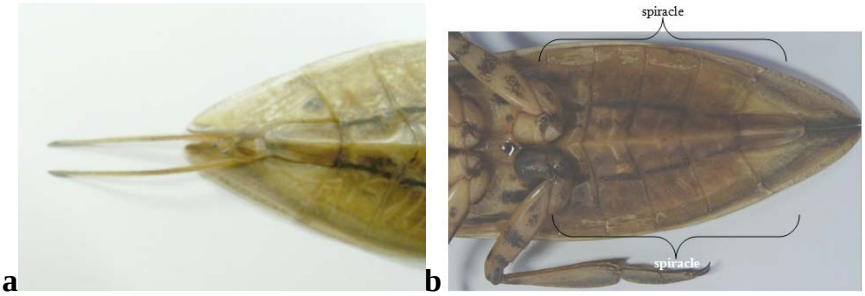
Figure 4 Side and ventral view of the head of an adult, a giant water bug showing the mouthparts is a beak-like structure. Photograph by Wijit Wisoram, Kasetsart University.

a

b

Figure 5 (a) Ventral anddorsalview of front legs (forelegs) and (b) two pairs of hind legs giant water bug. Photograph by Wijit Wisoram, Kasetsart University.

Figure 6 Showing the wing of an adult giant water bug, Photograph by Wijit Wisoram, Kasetsart University.



a b

Figure 7 Showing respiration system (a) the posterior end of a giant water bug has two retractable, semi-cylindrical appendages, and (b) the spiracle which behaves like grill in fish. Photograph by Wijit Wisoram, Kasetsart University.

1.3 Habitat

Giant water bugs live in freshwater streams and ponds, preferring those with aquatic vegetation. They like slowly moving water, especially where there is emergent vegetation such as cattails. They usually grab hold of a plant near the surface, and stick their short breathing tube out of the water to allow them to breathe while waiting for prey. With their powerful front legs they are able to grab other bugs and prey as big as small fish, frogs and other aquatic animals. They pierce their prey with their sharp beak and secrete enzymes that dissolve the body tissues, thus allowing them to suck up the resulting liquid (Figure 8)

Figure 8 The hunting of giant water bug. Photograph by Wijit Wisoram, Kasetsart University.

Chromosome analysis

The cytogenetics of Heteropteran insects are interesting primarily because they possess holocentric chromosomes. The chromosomes do not have a localized centromere, but the centromere is distributed along the length of the chromosome (Ueshima, 1979). Due to this characteristic, if a chromosome is broken, the fragments are not lost and still move to a pole at anaphase (Hughes-Schrader and Schrader 1961; LaChance *et al*., 1970). Moreover, the meiotic behavior of autosomes and sex chromosomes are different. As a rule, autosomes form bivalents with one chiasma per bivalent and divide pre-reductionally, while sex chromosomes are achiasmatic and form univalents at the first meiosis. The sex chromosoms divide equationally at anaphase I and segregate reductionally at anaphase II (Ueshima, 1979; Rebagliati *et al*., 2005). Moreover, some families also possess a pair of m-chromosomes, which are also achiasmatic. The m-chromosomes are unpaired and present as univalent chromosomes during early meiosis, but at metaphase I they form a pseudobivalent and divide reductionally at meiosis I and segregate equationally at meiosis II. Four sex chromosome systems have been described in 1,600 species of Heteroptera. The XX/XY system is the most commonly found (71.4%), the XO/XX and multiple system (XnY/XnXn, XnO/XnXn, XYn/XX) arefound in14.7% and 13.5%of species respectively and the rare system (0.5%) is the neo-sex chromosome system that hasbeen reported in seven species, including *L. indicus* (Grozeva and Nokkala, 1996; Bressa *et al*., 1999; Nokkala and Nokkala, 1999; Jacobs, 2004; Papeschi and Bressa, 2006).

The cytogenetics of Belostomatidae has been revealed by the studies on seventeen *Belostoma*species, three *Diplonychus* species, and seven *Lethocerus* species (Papeschi and Bressa, 2006). In these *Lethocerus* species, their chromosome numbers vary. Three species, *L. annulipes*, *L. griseus*, and *L. melloleitaoi* De Carlo, contain the same chromosome complement, 2n = 26A+ XY, while the chromosome complement of *L. indicus*is 2n = 24A+ neoX-neoY. It has been suggested that the neo-X and neo-Y chromosomes are established by the translocation of X and Y chromosomes to one pair of autosomes (Nokkala and Nokkala, 1999). The chromosome number is quite

reduced in *L. americanus* Leidy (2n = 6A + XY) and *Lethocerus* sp. (2n= 2A + neoX-neoY), while it is increased in *L. uhleri* (Montandon) (2n = ca. 30). However, the chromosomal behavior during spermatogenesis of the *Lethocerus* species has not been described, except the chromosome formula, because the cytogenetics of most species were studied during 1927 and 1959, and the original papers are difficult to access (Papeschi and Bidau, 1985; Papeschi and Bressa, 2006).

Mitochondrial genome

Mitochondria are key energy generators in most eukaryotic cells. Research on mitochondria has primarily focused on the process of ATP generation, phylogeny and evolutionary origins. Unique sequence signatures from mitochondrial DNA (mtDNA) have been used not only to categorize species, but also to study animal, bird, and human migration as well as in diagnostics and forensics. With the increase in the whole genome sequencing of eukaryotic genomes, mtDNAs are inevitably sequenced and this has facilitated comparative studies. Mitochondria are belived to have evoled in eukaryotes through a process called serial endosymbiosis from an unknown microbial ancestor (Chandra *et al.*, 2006).

Most animal mitogenome are about 16 Kb in size and mostly consisting of 37 genes: thirteen protein-coding genes (13 PCGs); (ATP synthase subunits 6 and 8 (*atp6, atp8*), cytochrome oxidase subunits 1–3 (*cox1–cox3*), cytochrome *b* (*cob*), and NADH dehydrogenase subunits 1–6 and 4L (*nad1–nad6, nad4L*), twenty-two transfer RNA genes (22 tRNA); (one for each amino acid except for leucine and serine, which have two genes: $trnL_1$ (uag), $trnL_2$ (uaa), $trnS_1$ (gcu), and $trnS_2$ (uga) and two ribosomal RNA (2 rRNA); large- and small-ribosomal subunit RNAs (*rrnL, rrnS*) (Wolstenholme, 1992; Boore, 1999). Additionally, there is a major non-coding region known as an A+T-rich region which contains essential regulatorly elements for the initiation of transcription and replication, and is therefore referred to the control region (Zhang and Hewitt, 1997). In recent years, mitogenomes are increasingly applied to comparative and evolutionary genomics, molecular evolution, phylogenetics, and population genetics (Wilson *et al.*, 2000; Salvato *et al.*, 2008),

because of their relatively simple structures and their mutation rates which are higher than in nuclear DNA (Avise *et al.*, 1987; Simon *et al.*, 1994) and its abundance in animal tissues, the small genome size, faster rate of evolution, low or absence of sequence recombination, and evolutionary conserved gene products (Lin *et al*., 2004; Simon *et al.*, 2006; Gissi *et al.*, 2008). Additionally, recent advancements in sequencing technology have lead to rapid growth of mitogenome data in Genbank. More than 3,000 complete mitochondrial sequences of metazoans have been deposited in the public databases (http://www.ncbi. nlm.nih.gov) last accessed January 7, 2014 and provide a foundation for large-scale comparative mitogenome studies. This number, however, is still far from enough, compared with the extreme species richness of metazoans, especially of arthropods. In addition, relatively few mitogenomes from closely related taxa are available to investigate mitogenome evolution over short time scales (Cameron *et al*., 2007).

The insect mitogenome is a circular, double-stranded molecule of 14–20 kb in length that comprises a set of 37 genes, including 22 tRNA genes, 2 ribosomal RNA genes and 13 protein-coding genes (Boore, 1999). In addition, it contains a major non-coding region, known as the A+T-rich region (or control region, CR) in insect mitochondrial DNA which has been widely studied (Saito *et al*., 2005; Tapper and Clayton, 1981; Wolstenholme, 1992). The A+T-rich region harbors initiation sites of mtDNA replication and transcription (Shadel and Clayton, 1993; Taanman, 1999). More than 270 insect mitogenomes have been sequenced, representing 26 of the 30 orders, with most from the mega-diverse orders such as Diptera, Lepidoptera, Hemiptera, Coleoptera and Hymenoptera (Zhao *et al*., 2013). Based on a series of analyses, mitogenomes may be superior to individual nuclear gene or partial mitochondrial genes to reconstruct phylogenetic relationships as they possess a number of evolutionarily interesting and potentially informative features such as length variation, altered tRNA anticodons or secondary structures, a typical start codons, base compositional bias, codon usage, and gene rearrangement (Sheffield *et al*., 2008; Boyce *et al*., 1989; Jia and Higgs, 2008).

***Vitellogenin* (*Vg*) gene**

Reproduction is the most important process of all living organisms for the existence of their species. In animals, the continuity of generations requires gametes, sperms and eggs, produced by males and females. Like most oviparous animals, almost all female insects produce eggs that will develop to be embryos after fertilized with sperms. Since most female insects lay eggs and then leave behind, the embryos within the eggs will develop to further stages using nutrients accumulated in the eggs no more supplementary food from their parent like mammalian embryos. The nutrients within insect eggs are yolk that is comprised of proteins, lipids, carbohydrates and other components (Dhadialla and Raikhel, 1990). The major yolk protein in all oviparous animals including insect is vitellin (Vn) derived from the precursor protein called vitellogenin (Vg). Vg, a major lipoprotein in many oviparous animals, is a precursor of major yolk protein vitellins. It has been shown that Vg is secreted by the liver in vertebrates and the intestine in nematodes. In vertebrates, Vg production is in control of estrogen receptor pathway. However, in invertebrates such as bivalves, vitellogenesis seems to be under the control of both estradiol and a neuropeptide (Osada *et al*., 2003,). In vertebrates, Vg levels can remain high in the plasma for some time as Vg proteins and then degrade slowly (Denslow *et al*., 1999). Therefore, measurement of plasma Vg levels, especially in immature females and males, is considered a useful biomarker of exposure to estrogenic compounds (Hansen *et al*., 1998; Okoumassoun *et al*., 2002; Martin and Matozzo, 2004; Matozzo *et al*., 2008).

In insects, the vitellogenesis generally proceeds from the biosynthesis of Vg, the processing in the fat body, the secretion of the processed Vg into the hemolymph, the selective uptake by competent oocytes and finally to the utilization of Vn by the developing embryo (Byrne *et al*., 1989; Wyatt, 1991; Raikhel and Dhadialla, 1992; Izumi *et al*., 1994; Hagedorn *et al*., 1998; Sappington and Raikhel, 1998; Giorgi *et al*., 1999). During these processes, Vg and Vn are modified through cleavage, glycosylation, lipidation and phosphorylation (Raikhel and Dhadialla, 1992; Sappington and Raikhel, 1998). *Vg* genes are generally expressed in sex-and tissue-specific manners in most animal species, however, expression analysis of the *U. major Vg* gene

revealed that it is expressed in both female and males (Kang *et al*., 2008), like adult sea urchin (Shyu *et al*.,1986), worker honey bee that presence of high Vg levels (Piulachs *et al*., 2003) although in males of some species expressed in smaller amounts (Engelmann, 1979; Trenczek & Engels, 1986; Valle, 1993; Piulachs *et al*., 2003).

Vg gene in insect

Vg genes code for the majoregg yolk protein precursor in insects (Wahli *et al*., 1981; Kunkel and Nordin, 1985; Sappington *et al*., 2002) and synthesized one or more large precursors (Della-Cioppa and Engelmann, 1987) in fat body and secreted into the hemolymph. After synthesized, the protein is released into haemolymph which brings the protein to developing oocytes. The Vg is taken up by the developing oocytes by receptor-mediated endocytosis (Dhadialla and Raikhel, 1990; Hagedorm *et al*., 1998; Tufail and Takeda 2009; Tufail *et al*., 2014). Vgs played important roles in promoting growth and differentiation of oocytes and transporting metallic ions, lipid, thyroxine, vitamin A, carotenoid and riboflavin into oocyte. In the oocytes, the Vgs are modified and stored in a crystalline form as vitellins, reserve food-source for the future embryo (Kunkel and Nordin 1985). The synthesis of Vg is regulated at the transcriptional level (Rhaikel *et al*., 2004). The hormones involved in *Vg* gene regulation include juvenile hormone (JH), ecdysone, and several neuro-peptides (Wyatt and Davey, 1996; Tufail and Takeda, 2008). In general, insects can be classified into three major groups based on hormonal regulation of *Vg* gene transcription. Group I includes insects that use only JH for *Vg* gene transcription. In most of the insect species investigated JH is the best-known gonadotropin (Engelmann, 1983; Wyatt and Davey, 1996; Belles, 1998; Raikhel *et al*., 2004). Group II includes the cases in which regulation of the *Vg* gene requires ecdysteroid, a product from ovaries (Hagedorn *et al*., 1975) in addition to the JH (such as Diptera). Group III includes lepidopterans and other insects that require JH, ecdysteroids and additional hormones to regulate their reproductive biology.

In most insects, Vgs are large molecules (~200 kDa), which are derived from a single *Vg* gene transcript of 6-7 kbp mRNA and the number of *Vg* genes vary (one to several) in different species (Tufail and Takeda, 2008).So far, Vgs have been sequenced from 25 insect species. Alignment of the Vgs sequences revealed that some conserved features such as cysteine residues, the GL/ICG and DGXR motifs, were found (Tufail and Takeda, 2008) at the carboxy terminal. Among the Lepidoptera insects, the Vgs have been identified from *Bombyx mori* (Yano *et al*., 1994a; Yano *et al*., 1994b), *Bombyx mandarina* (Meng *et al*., 2006b), *Actias selene* (Yin *et al*., 2007), *Antheraea pernyi* (Yokoyama *et al*., 1993; Liu *et al*., 2001, Zhu *et al*., 2010), *A. yamamai* (Liu *et al*., 2001; Meng and Liu, 2006a), *Saturnia japonica* (Meng *et al*., 2008), *Samia cynthia ricini*, *Lymantria dispar* (Hiremath and Lehtoma, 1997) and *Philosamia cynthia ricini* (Liu *et al*., 2003). Researches on insect Vgs provide great theoretical and practical significances for utilization of beneficial insects and prevention of harmful insects (Brownes, 1986). The primary structures of many insect Vgs (Tufail and Takeda, 2008) and some advanced structure are known (Sharrock *et al*., 1992).

Structure of Vitellogenins

Vitellogenins are large (200-700 kDa) homologous phosphoglyco-lipoproteins that are often oligomeric in their native state, with monomers consisting of one to four subunits. In most insect, Vg monomer are compose of two subunits (one of large subunit > 150 kDa and one of small subunit < 65 kDa) (Kunkel and Nordin, 1985; Raikhel and Dhadialla, 1992; Valle, 1993) derived from the cleavage of a single precursor in the fat body (Bose and Raikhel, 1988; Dhadialla and Raikhel, 1990; Heilmann *et al*., 1993; Kageyama *et al*., 1994; Yano *et al*., 1994; Hiremath and Lehtoma, 1997). Exceptions include the Vgs of higher Hymenoptera (suborder Apocrita) (Wheeler and Kawooya, 1990; Kageyama *et al*., 1994; Nose *et al*., 1997) and two species of whitefly (Homoptera) (Tu *et al*., 1997) which are not cleaved. There is evidence that some Orthopteran and sawfly Vgs consist of three or four subunits (Kunkel and Nordin, 1985; Della-Cioppa and Engelmann, 1987; Wyatt, 1988; Kim and Lee, 1994; Takadera *et al*., 1996). However, it is possible that some of

the observed multiple subunits are an artifact of cleavage by proteases during sample preparation, and more experiments are needed to confirm these observations.

Polyserine domains

The most striking feature of vertebrate and some insect Vg primary structures are the presence of one to three domains containing long runs of serine residues. Three of the six sequenced insect Vgs contain polyserine domains: the silk moth and sawfly contain two polyserine domains and the mosquito Vg habor three domains. There is a region in the wasp Vg rich in serines (corresponding to the second polyserine domain of mosquito, silk moth, and sawfly) but they are not arranged as a long run. The serine runs arose through serial amplification of serine codons, and are located in regions that are evolving at a faster rate than most of the rest of the gene (Byrne *et al*., 1989; LaFleur *et al*., 1995; Chen *et al*., 1997).

The role of the polyserine domains is unknown, and their presence cannot yet be correlated with peculiarities in the life histories of the insects harboring them. Their significance probably lies in being good substrates for kinases. Vertebrate phosvitin is indeed heavily phosphorylated (Byrne *et al*., 1989), as is mosquito Vg (Dhadialla and Raikhel, 1990), but the phosphorylation sites in the latter have not been identified. Phosphoserine tracts represent extreme concentrations of negative charge (Goulas *et al*., 1996), which may promote solubility of the Vg (Gerber-Huber *et al*., 1987) or provide a region for chelating essential metal ions such as Ca^{2+} and Fe^{3+} (Nardelli *et al*., 1987; Taborsky, 1991). Dephosphorylation of Vg reduces its uptake by oocytes (Miller *et al*., 1982; Dhadialla *et al*., 1992), suggesting a role of phosphorylated residues in VgR recognition or maintenance of tertiary structure. *Bombyx mori* Vg is further phosphorylated during embryogenesis, possibly as a signal to initiate its proteolysis (Takahashi *et al*., 1992; Izumi *et al*., 1994).

Cleavage site of the Vg precursor

Most of the insect Vgs for which the sequences are known are of the type that are cleaved once in the fat body to produce two subunits. The cleavage sites are immediately preceded by a motif, (R/K)X(R/K)R or RXXR, specifically recognized by the subtilisin-like proprotein endoproteases, convertases (Barr, 1991; Rouille*et al*., 1995). In the case of the wasp Vg which is not cleaved (Nose *et al*., 1997), the sequence at the presumed ancestral cleavage site has been mutated from RXRR to LYRR and is apparently no longer recognized by convertases. In addition to the recognition motif, each cleavage site is either immediately preceded or followed by a predicted β-turn, a requirement for optimal recognition by this family of endoproteases (Brakch *et al*., 1993). Indeed, a mosquito fat body-specific convertase has been cloned and sequenced and when coexpressed with mosquito Vg in an *in vitro* trancription/translation system, the convertase correctly cleaved the pro-Vg (Chen and Raikhel, 1996). The primary structure of this mosquito vitellogenin convertase (VC) is similar to that of human and *Drosophila* furins (Barr *et al*., 1991; Roebroek *et al*., 1991, 1992), and another *Drosophila* convertase, dKLIP-1 (Hayflick *et al*., 1992). Unlike the mosquito VC, however, the native substrates of the *Drosophila* convertases have not yet been identified. Interestingly, the putative cleavage site for mosquito VC auto-activation is a paired di-basic motif, RAKR211-RPKR215 (Chen and Raikhel, 1996).

The probable cleavage site of *Caenorhabditiselegans* Vg6 and *Dolichorab ditis* Vg (Winter *et al*., 1996) resemble those insects, but *C. elegans* Vg5 has an RSRR motif in the same relative position and is not cleaved. The reason for its not being cleaved is not readily apparent from its primary sequence, but unlike the other sites, the nematode Vg5 cleavage motif is immediately flanked on both sides by probable β-turns rather than having a turn on one side only. Although the cleavage sites in insect and nematode Vgs are structurally similar, their location in the precursor is not strictly conserved. In *Anthonomus*, *Aedes*, *Bombyx* and *Athalia*, the cleavage site is near the N-terminus of the precursor (Heilmann *et al*., 1993; Chen *et al*., 1994; Kageyama *et al*., 1994; Yano *et al*., 1994), but this site has been lost at least twice, once within the Lepidoptera e.g., *L. dispar* (Hiremath and Lehtoma, 1997), and once within the Hymenoptera e.g., *Pimpla nipponica* (Nose *et al*., 1997). In the case

of *Lymantria dispar*, the Vg is still cleaved, but the cleavage site is located near the C-terminus (Hiremath and Lehtoma, 1997).

In nematodes the site is nearer the center of the precursor (Winter *et al*., 1996). However, it is clear that the cleavage sites have not moved via exon shuffling, because multiple alignments indicates a one-to-one correspondence of domains along the length of the precursors (Chen *et al*., 1997). The common insect cleavage sites are in a region of relatively low conservation; an area not represented in vertebrate or nematode Vgs at all, and is flanked by polyserine domains. The putative nematode cleavage site (Winter *et al*., 1996) is also located in a region of low conservation (Chen *et al*., 1997). The *Lymantria* cleavage site is within a domain of comparatively high conservation, but its immediate vicinity appears to have undergone relatively rapid evolution (as suggested by low sequence conservation compared to other regions of the domain). Thus, the evolutionary loss or relocation of the Vg cleavage site is due to local modifications in the amino acid sequence, not to rearrangement of the gene.

METERIALS AND METHODS

Materials

1. Insects

Living adults of the giant water bug, *L. indicus* were collected from natural populations in three provinces of Thailand (Chiang Mai, Buri Ram and Sa Kaew). For samples from Myanmar, Laos and Cambodia, the living insects were bought from importing merchants at markets near the borders of Thailand.

Living productive period male and female were collected from Buri Ram province. All sampleswere dissected and stored at -80°C for further study.

2. Bacterial strain

Escherichia coli strain DH5α
Escherichia coli strain HT115 (Dolan DNA learning center, USA.)

3. Type of vector

pGEM-T® easy vectors (Promega: Fitchburg, United States)

Methods

1. Cytogenetic study

1.1 Slide preparations

The testes of *L. indicus* male were dissected and fixed in ethanol: acetic acid (3:1) overnight at 4°C. Then, they were stored in 70% alcohol at 4°C until used. For standard staining, the fixed testes stained with lacto-aceto orcein were squashed

on slides. This technique was used to examine chromosome types in all insect populations.

For Giemsa, DAPI, and silver nitrate stainings, the fixed testicular cells were dissociated in a few drops of 60% acetic acid for 5 min, spread on slides, placed on a slide warmer at 50° C, and kept at room temperature for further staining. The slides were stained with 4% Giemsa (Merck, www.merck.com) in phosphate buffer, pH 6.8, for 30 min, rinsed thoroughly with distilled water, and then air-dried.

Silver nitrate staining was performed according to the protocol of Howell and Black (1980). Briefly, two solutions were prepared; one was colloidal developer solution (2 g powder gelatin, 1 ml formic acid in 100 ml distilled water) and the other was silver nitrate solution (4 g $AgNO_3$ dissolved in 8 ml distilled water). For staining, 200 µl of colloidal developer solution was mixed with 400 µl of silver nitrate solution and then dropped on the slides, covered with coverslip, and placed on a slide warmer at 70°C for 2 min. Then, distilled water was used to remove the coverslip and excess staining mixture. The stained slides were air-dried and examined. All slides stained with Giemsa and silver nitrate was observed under a BX41 Olympus microscope (www.olympus-global.com) using an immersion oil objective (100X) and all photo-graphs were taken with an Olympus DP72 digital camera.

For DAPI staining, DAPI (4-6-diamidino-2-phenylindole; Sigma-Aldrich, www.sigmaaldrich.com) was dissolved in distilled water to get 1 µg/ml. The DAPI solution was dropped on the slides and covered with coverslip. The slides were then observed immediately under a Zeiss fluorescence microscope (www.zeiss.com) with a 100X objective and the images were captured by the Isis FISH imaging system version 5.2 (www.isisimaging.com).

2. Mitochondrial genome analysis

2.1 Genomic DNA extraction

Total genomic DNA was extracted from foreleg tissue using DNeasy Tissue Kit (Qiagen: Hilden, Germany) following the manufacturer's protocol. Obtained DNA solution was stored at -20ºC.

2.2 PCR amplification and sequencing

A long PCR strategy was applied to amplify the whole mitochondrial (mt) genome. The 1,800 bp DNA fragment between the large and small ribosomal subunit was amplified using a primer set (Le16sF 5'-TAAAAGGTCGAACAGACC-3' and Le12sR 5'-AAACTAGGATTAGA TACCC-3'), which were designed by using a multiple alignment of several *rrnLs* and *rrnSs* sequences from hemipteran insects. The nucleotide sequence obtained from this segment was used to design the next primer for next overlapping fragments. Then, five (fragment 2-5 and16) overlapping fragments were amplified using long PCRs and eleven overlapping fragments were amplified using normal PCRs to get the complete mitogenome. Each obtained DNA fragment was sequenced directly with the same primers (sixteen primers set) (Table 1) and primer walking sequencing. PCR was performed in 50 µl reaction mixtures containing 30.5 µl sterile distilled water, 5 µl 10X High Fidelity PCR buffer (Fermentas Inc.: Burlington, Canada), 4 µl dNTPs (2.0 mM), 2 µl of each primer, 0.5 µl High Fidelity PCR Enzyme Mix (2.5 U) and 2.5 µl DNA template (10pg-1µg). Thirty-five PCR cycles for each PCR-primer set were performed using a personal Thermal Cycler (Biometra: Goettingen, Germany). For fragments of length less than 2 kb, PCR conditions were followed by 95°C for five min, thirty-five cycles of 94°C for 40 sec, 50-55°C (Annealing temperature was depending on primer combinations) for 40-50 sec, 72°C for 1-2.5 min (extension time was depending on putative length of the fragments) and a final extension step of 72°C for 5 min. For fragments of length longer than 2 kb, PCR conditions were followed by 95°C for five min, thirty-five cycles of 94°C for 50 sec, 50-55°C (Annealing temperature was depending on primer combinations) for 40-50 sec, 72°C for 3-7 min (extension time was depending on putative length of the fragments) and a final extension step of 72°C for 10 min. PCR products were purified using the FavorPrep^{TM}Gel/PCR Purification Kit (Favorgen: Ping-Tung, South Korea) and sent to Macrogen Inc. (Seoul, South Korea) for

sequencing. Then, internal primers were design to further sequence into the fragments.

As for longer fragment from the fifth primer set PCR product was cloned into pGEM®-T Easy vector and the resultant plasmid DNA was isolated using the FavorPrep™ Plasmid Extraction Maxi Kit. The plasmids DNA were sent to Macrogen for sequencing and using a primer-walking strategy initiated with vector-specific primer.

2.3 Sequence analysis

The obtained nucleotide sequences of all overlapping were characterized using BLASTN and assembled using CAP3 software (Huang and Madan, 1999). The nucleotide positions of 13 PCGs were located by comparison with *Diplonychus rusticus* and other insect mtDNA using BLASTN and clustalW. PCGs were translated using invertebrate mitochondrial genetic code. The 13 *PCGs* and *rRNA* genes were identified by sequence comparision with published insect mitogenome sequences. All location of 22 *tRNA* coding genes was identified using the same method for protein coding genes and was defined depending on their cloverleaf secondary structure using tRNA-scanSE online server (Lowe and Eddy, 1997) (http://www.genetics.wustl.edu/eddy/ tRNAscan-SE/) with the following setting: search mode: tRNA scan only; Source: Mito/Chloroplast; Cove score cutoff: 5.

2.4 Phylogenetic analysis

Phylogenetic analysis of mitogenome was performed using Neighbor-Joining and Maximum Likelihood analysis with 13 PCGs amino acid sequences of 42 mitogenome sequences from hemipteran insect species retrieved from the GenBank database (appendix table A1). Two species from Sternorrhyncha were selected as outgroups. All the sequences were aligned with ClustalX (version 1.83) by the multiple alignment method. Gaps and missing data were completely excluded from the analysis. Alignments of individual genes were then concatenated excluding the

stop codon. The generated data matrix was converted to mega format and analyzed with MEGA 5.03 program.

Table 1 The primers used in this study for mitogenome analysis

Primer name	Primer sequences (5'-3')	No. of fragment
Li16sF	TAA-AAG-GTC-GAA-CAG-ACC	1
Li12sR	AAA-CTA-GGA-TTA-GAT-ACC-C	1
Li_cox1_F2	GCA-GGG-GCA-ATT-ACT-AT	2[a]
Li_nad3_R2	GGT-CAA-ATC-CAC-ATT-CAA	2[a]
Li_cox3_F3	TTG-GAT-TTG-AAG-CA-GCA-GC	3[a]
Li_nadl4_R3	TGT-TTG-TGA-AGG-TGT-GTT-AGG	3[a]
Li_nad4_F4	TAG-GTG-CCT-CAA-CAT-GAG-C	4[a]
Li_16s_R4	TTA-TGG-GAC-GAG-AAG-ACC-C	4[a]
Li_12s_F5	ACT-CCT-ATA-TCC-TCT-GCT	5[a]
Li_cox1_R5	TAA-TTC-CAG-TTG-GTA-CAG-C	5[a]
Li_ND4L_F6	ATT-CGC-TCA-GGT-TGA-TAA-CC	6
Li_cytb_R6	AAG-AAT-CGT-GTT-AGT-GTG-GC	6
Li_cytb_F7	TCA-AGA-GTA-GCC-TCA-AAC-G	7
Li_cytb_R7	TTG-TAT-CTT-TTG-TCC-CAC-C	7
Li_ND5_F8	AAG-AAC-CAT-AGA-CAA-AGA-GC	8
Li_ND5_R8	GTG-ATT-GCT-GTT-GTT-TGT-CC	8
Li_inter_cox1F	AGT-CGT-AGC-GAT-GGT-AAT-GC	9
Li_inter_cox1R	CAG-CCT-ATA-ATT-CAG-CCA-CC	9
Li_link34_F	AGG-CTG-ATA-ACC-TCA-ACC	10
Li_link34_R	AGA-AGT-ATG-GGT-GGA-ATG-G	10
Li_ND1_F	ACG-ACC-AGC-AGA-AGA-ACC	11
Li_ND1_R	CCA-GGT-TGG-TTT-CTA-TCC	11
Li_cox2_F	ATC-CGA-CTA-CTA-GTA-ACA-GC	12
Li_atp6_R	TTC-CTA-CAG-GTA-TTA-GGT-GG	12
Li_nd1_F	CAA-CCA-GAA-AGT-AAA-ACA-CC	13
Li_l6s_R	ACC-GTG-CAA-AGG-TAG-CAT-A	13
Li_nd3_F	AGT-TGA-AGC-CAA-ATA-GAG-G	14
Li_nd4_R	ATT-CTG-CTA-GTC-AGC-ATG-G	14
Li_atp8_F	AAT-TCG-ACC-AGG-AGC-ACT-AG	15
Li_nd5_R	GGT-GAT-TCT-CAG-GAT-ATT-CG	15
Li_control_F	AAG-TTC-ACT-CAC-TCT-TAT-CG	16[a]
Li_control_R	TAT-GGT-CAG-GGT-ATG-AAC-C	16[a]

[a]Fragments were amplified using long PCR strategy

3. Cloning and sequencing of *Vg* gene in *L. indicus*

3.1 RNA isolation and synthesis the first strand cDNA

An adult specimen of *L. indicus* was collected in Buri Ram province, the North-East of Thailand. Fatbody was dissected from female *L. indicus* and was stored at -80ºC. Total RNA was extracted using TRIzol reagent (Gibco BRL: New York, United States). Fifty microgram of fatbody was homogenized by a pestle in 300 µl TRIzol reagent and incubated at room temperature for 5 min to dissociate nucleoprotein complex. The mixture was centrifuged at 12,000 rpm for 10 min at 4°C. Supernatant was removed into a new microcentrifuge tube and 200 µl of chloroform was added into that tube, shake the tube vigorously by hand for 15 sec and then incubated for 3 min at room temperature for precipitation of proteins. The mixture was centrifuged at 12,000 rpm at 4°C for 15 min. Following centrifugation, the mixture separated into a lower red phase, phenol-chloroform phase and a colorless upper phase. The upper aqueous phase was transferred to a new microcentrifuge tube. The RNA was precipitated by mixing with 150 µl of isopropanol. The sample was incubated at -20°C for 1 hr. Total RNA was precipitated by centrifugation at 12,000 rpm at 4°C for 10 min. After the supernatant was removed, the RNA pellet was washed twice with 300 µl of 75% (v/v) ethanol and centrifuged at 7,500 rpm at 4°C for 5 min. The washed RNA pellet was air-dried until ethanol evaporated completely and dissolved in 30 of µl RNase-free water (DEPC-treated water). Obtained total RNA solution was stored at -80ºC. These total RNA samples were used as templates for first-strand cDNA synthesis using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific: Massachusetts, United State).

3.2 Oligonucleotide primers design

Downloaded *Vg* nucleotide sequences of *L. deyrollei*, *Trigonotylus caelestialium* and *Nilaparvata lugens* from the database (www.ncbi.nlm. nih.gov) were aligned using ClustalW program (http://www.ebi.ac.uk/Tools/clustalw2 /index.html). The primers were designed on the basis of consensus nucleotide

sequences for three sets primers. The first set contains the forward (LiVgF1) and reverse primer (LiVgR1), second set contains the forward (LiVgF2) and reverse primer (LiVgR2) and the last set primer contains the forward (LiVgF3) and reverse primer (LiVgR3). All primer sequences are shown in the Table 2.

Table 2 List of primer for PCR amplification and 5' and 3' RACE of *Vg* gene

Primer name	**Sequence5'-3'**
LiVgF1	ATG-CTA-CAA-AAC-CAT-GGA-GG
LiVgR1	AAC-CTG-AAT-CTT-AGG-ACT-GTC
LiVgF2	AAA-GCC-CTC-AGG-TCA-TGA-AG
LiVgR2	TTC-CTC-CAT-CTG-TTG-TTC-GC
LiVgF3	ACC-GAC-TTC-CCA-GAT-GTA-AG
LiVgR3	TTT-TCT-CCG--TCG-AAG-GTG-C
GSP1-Vg-5R	TTC-CCA-GTT-GGT-GGG-ACC-GGT-AAT-TCC
GSP2-Vg-3F	GGG-AGA-CGA-CAT-AAT-CGA-GTT-ACA-GC

3.3 PCR amplification of *Vg* gene from cDNA

The *Vg* gene of *L. indicus* was amplified by polymerase chain reaction (PCR), using the first strand cDNA as a template and amplifiedthree fragments. The amplified reaction of fragments 1 using the *Taq* DNA Polymerase (recombinant) contained 1 µl of the first-stranded cDNA template, 5 µl of 10X *Taq* buffer, 5 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP, 2 µl of Forward primer (LiVgF1 10 pmol/ µl), 2 µl of reverse primer (LiVgR1 10 pmol/ µl), 0.5 µl of *Taq* DNA polymerase and add sterilized distilled water up to 50 µl. PCR procedure was as followed: preheated at 94ºC for 3 min, followed by 35 cycles of denaturing at 94 ºC for 30 sec, annealing at 55.5 ºC for 30 sec and extension at 72 ºC for 1 min. The final extension was carried out at 72 ºC for 10 min.

The amplified reaction of fragments 2 and 3 using the same amplified reaction solution of fragments 1, exception 2 µl of forward primer (LiVgF2 and 3,10 pmol/ µl), 2 µl of reverse primer (LiVgR2 and 3, 10 pmol/ µl). PCR procedure was as followed: preheated at 94 ºC for 3 min, followed by 35 cycles of denaturing at 94 ºC for 30 sec,

annealing at 57 and 55ºC for 30 sec and extension at 72 ºC for 2.5 and 1 min, respectively. The final extension was carried out at 72 ºC for 10 min.

Three fragments of PCR products of *Vg* gene was mixed with 6X loading buffer and the sample mixture was loaded into the wells of the submerged 1% agarose gel. The electrophoresis was carried out in 1X Tris Acetate EDTA (TAE) running buffer at 100V for 25 min. After finishing, the gel was stained in ethidium bromide (EtBr) solution for 10 min and destained by submerging in an excessive amount of water for 2 min. The nucleic acid bands were visualized under UV light.

PCR products were purified from agarose gel using Favorgen® Gel/PCR purification kit. The purified PCR products were sent to Macrogen Inc. for sequencing. Then, internal primers were design to further sequence into the fragments for fragment 2.

3.3 Cloning and sequencing cDNAs using 5' and 3' Rapid Amplification of cDNA Ends

5' and 3' RACEreactions was performed using the SMARTer™ RACE cDNA Amplification Kit User Manual (Clontech: California, United States) according to manufacturer's instructions, using one adapter-specific and the other gene-specific primers. The overview of full-length ds cDNA synthesis was shown in Figure 12

3.3.1 Designing PCR Primers for 5' and 3' RACE

The primers were designed from the sequence of *LiVg* nucleotide sequences. The gene-specific primers used for 5' and 3' RACE (GSP1_Vg_5R and GSP2_Vg_3F) are show in the Table 2 and the products are shown in the Figure 13.

3.3.2 Total RNA isolation from fat body of *L. indicus* female

Total RNA was extracted from fatbody of adult female using TRIzol reagent.5' and 3' RACE were performed to generate the full-length *Vg* cDNAs using the SMARTer^{TM} RACE cDNA Amplification Kit User Manual.

3.3.3 First-Strand cDNA Synthesis of 5' RACE

Prepare enough of the following Buffer Mix for 5'-RACE-Ready cDNA synthesis reactions by added 2.0 µl of 5X First-Strand Buffer, 1.0 µl of DTT (20 mM) and 1.0 µl dNTP Mix (10 mM), mix the reagents and spin briefly in a microcentrifuge, then set aside at room temperature. The reaction of 5'-RACE -Ready cDNA contained 2.75 µl of total RNA and 1 µl 5'-CDS Primer A. Mix contents and spin the tubes briefly in a microcentrifuge and incubate the tubes at 72°C for 3 min, then cool the tubes to 42°C for 2 min. After cooling, spin the tubes briefly for 10 seconds at 14,000 g to collect the contents at the bottom. Added 1 µl of the SMARTer IIA oligo per reaction, 4.0 µl of Buffer Mix, 0.25 µl of RNase Inhibitor (40 U/µl) and 1 µl of SMARTScribe Reverse Transcriptase (100 U). Mix the contents of the tubes by gently pipetting, and spin the tubes briefly to collect the contents at the bottom. Incubate the tubes at 42°C for 90 min in a hotlid thermal cycler and heat tubes at 70°C for 10 min and dilute the first-strand reaction product with Tricine-EDTA Buffer.

3.3.4 First-Strand cDNA Synthesis of 3' RACE

Prepare enough of the following Buffer Mix for 3'-RACE -Ready cDNA synthesis reactions by added 2.0 µl of 5X First-Strand Buffer, 1.0 µl of DTT (20 mM) and 1.0 µl dNTP Mix (10 mM), mix the reagents and spin briefly in a microcentrifuge, then set aside at room temperature. The reaction of 3'-RACE -Ready cDNA contained 3.75 µl of total RNA and 1 µl 5'-CDS Primer A. Mix contents and spin the tubes briefly in a microcentrifuge and incubate the tubes at 72°C for 3 min, then cool the tubes to 42°C for 2 min. After cooling, spin the tubes briefly for 10 seconds at 14,000 g to collect the contents at the bottom. Added 4.0 µl of Buffer Mix, 0.25 µl of RNase Inhibitor (40 U/µl) and 1 µl of SMARTScribe Reverse Transcriptase (100 U). Mix the contents of the tubes by gently pipetting and spin the tubes briefly to

collect the contents at the bottom. Incubate the tubes at 42°C for 90 min a hotlid thermal cycler and heat tubes at 70°C for 10 min and dilute the first-strand reaction product with Tricine-EDTA Buffer.

3.3.5 Rapid amplification of 5' and 3' ends of *LiVg*

5' RACE reactions was performed using the SMARTer™ RACE cDNA Amplification Kit User Manual according to manufacturer's instructions, using one adapter-specific and the other gene-specific primers. The gene-specific primers used for 5' RACE of *LiVg* was GSP1_Vg_5R. The amplified reaction used the Advantage 2 Polymerase Mix. 5' RACE-PCR procedure was as followed: preheated at 94 ºC for 3 min, followed by 5 cycles of denaturing at 94 ºC for 30 sec, annealing at 72 ºC for 3 min, 5 cycles of denaturing at 94 ºC for 30 sec, annealing at 70 ºC for 30 sec, extension at 72 ºC for 3 min and 20 cycles of denaturing at 94 ºC for 30 sec, annealing at 68 ºC for 30 sec, extension at 72 ºC for 3 min. The final extension was carried out at 72 ºC for 10 min. PCR products of 5' RACEfragments was detected by 1% agarose gel and purified from agarose gel using Favorgen® Gel/PCR purification kit cloned into pGEM-T®easy vector and transformed into *E coli* DH5α by heat shock transformation. The recombinants were identified through blue-white color selection in ampicillin-containing LB plates, extracted plasmid using FavorPrep™ Plasmid Extraction MiNi Kit and then sequenced on both strands.

3' RACEreactions was performed using the SMARTer™ RACE cDNA Amplification Kit User Manual according to manufacturer's instructions, using one adapter-specific and the other gene-specific primers. The gene-specific primers used for 3' RACEof *LiVg* was GSP2_Vg_3F. The amplified reaction used the Advantage 2 Polymerase Mix. 3' RACE-PCR procedure was as followed: preheated at 94 ºC for 3 min, followed by 5 cycles of denaturing at 94 ºC for 30 sec, annealing at 72 ºC for 3 min, 5 cycles of denaturing at 94 ºC for 30 sec, annealing at 70 ºC for 30 sec, extension at 72 ºC for 3 min and 20 cycles of denaturing at 94 ºC for 30 sec, annealing at 68 ºC for 30 sec, extension at 72 ºC for 3 min. The final extension was carried out at 72 ºC for 10 min.PCR products of 3' RACE fragment was detected by

1% agarose gel and purified from agarose gel using Favorgen® Gel/PCR purification kit cloned into pGEM-T®easy vector and transformed into *E coli* DH5α by heat shock transformation. The recombinants were identified through blue-white color selection in ampicillin-containing LB plates, extracted plasmid using FavorPrep™ Plasmid Extraction MiNi kit and then sequenced on both strands.

3.4 Nucleotide sequencing

PCR products were purified from agarose gel using Favorgen® Gel/PCR purification kit according to the manufacturer's protocol. The purified PCR products were sent to sequence at DNA sequencing unit, Macrogen Inc. All of the nucleotide sequences results from sequencing unit were checked with GenBank database by Blastn program (http://www.ncbi.nlm.nih.gov/) for confirmation that they were really *Vg* gene sequences. Then, internal primers were design to further sequence into the fragments for fragment 2. The sequence assembly program, CAP3 (http://pbil.univlyonl.fr/cap3.php), was used to assemble all of overlapping PCR fragments into each fragments of *Vg* gene.

3.5 Multiple alignment and Phylogenetic tree analysis of LiVg

Phylogenetic analysis of LiVg was performed using Neighbor-Joining and Maximum Likelihood analysis with amino acid sequences of 39 Vg sequences (excluding the signal peptide) from 38 arthropod and non-arthropod species retrieved from the GenBank database. This phylogenetic analysis consists of Hymenopteran, Lepidopteran, Colepteran, Dipteran, Dictyopteran and Hemipteran Vgs and one vertebrate, *Gall gallus* was the out group. All the sequences were aligned with ClustalX (version 1.83) by the multiple alignment method. Gaps and missing data were completely excluded from the analysis. The generated data matrix was converted to mega format and analyzed with MEGA 5.03 program. The Vg sequences used for comparison were show in appendix Table A2.

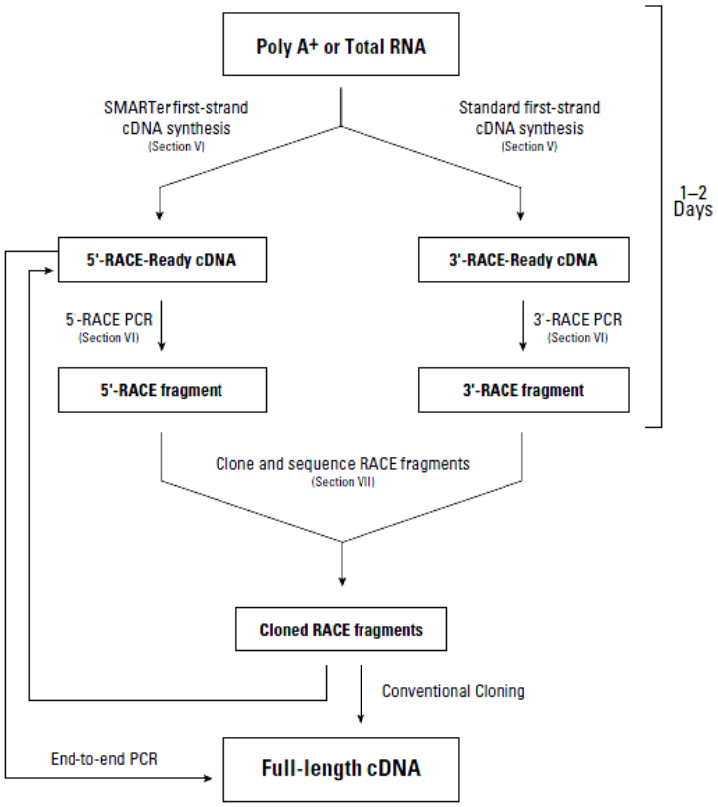


Figure 9 Overview of the SMARTer RACE procedure

Source: Clontech (2012)

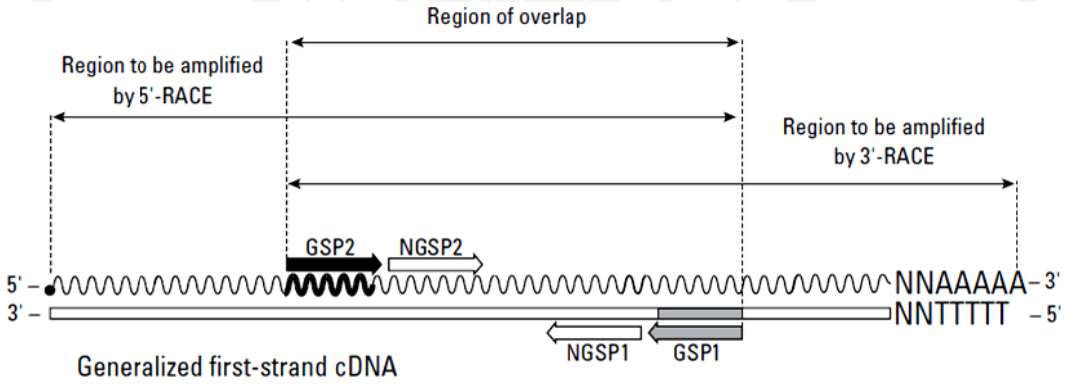


Figure 10 The relationship of gene-specific primers to the cDNA template.This diagram shows a generalized first-strand cDNA template.

Source: Clontech (2012)

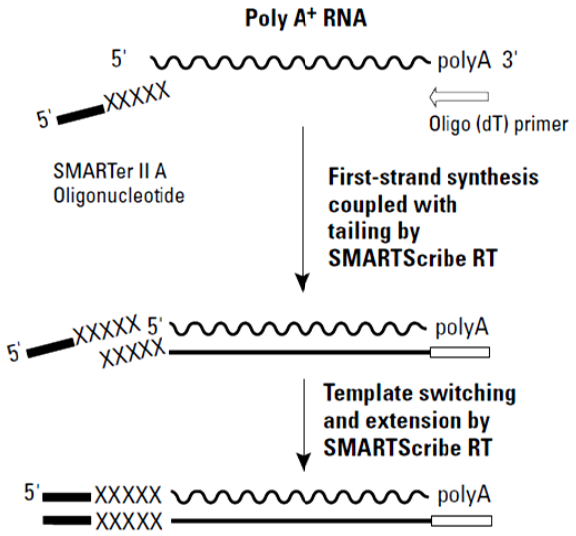


Figure 11 Mechanism of SMARTer cDNA synthesis

Source: Clontech (2012)

3.6 Detection the expression of *Vg* gene by RT-PCR

3.6.1 RNA Isolation and Synthesis the first strand cDNA

Total RNA were extracted from fat body of adult female and adult male of *L. indicus* using TRIzol reagent and synthesis the First strand cDNA using RevertAid First Strand cDNA Synthesis Kit.

3.6.2 Oligonucleotide primers design

The specific primers were designed on the sequence of *LiVg* nucleotide sequences. There are 2 primer sequences that contain the forward primer (Vg_RT_F1) and reverses primer (Vg_RT_R1), shown in the Table3. This primer set wasused for RT-PCR reaction that checked *Vg* gene expression in adult male and female.

Table 3 List of primer for RT-PCR amplification

Primer name	Sequence5-3
Vg_RT_F1	ATC-CTT-CAG-CGT-ATG-GCA-G
Vg_RT_R1	TTC-CAA-GAC-GTC-TTT-CTC-G

3.6.3 Expression of *LiVg* gene byRT-PCR amplification

The expression of *LiVg* gene was amplified by Reverse transcriptase polymerase chain reaction (RT-PCR), using the first strand cDNA as a template and amplied with two set of sequence primers.

The RT-PCR used Vg_RT_F1 and Vg_RT_R1 primer for *Vg* gene expression in adult male and female *L. indicus*. The amplified reaction of 50 µl using the *Taq* DNA Polymerase (recombinant) contained 1 µl of the first-stranded cDNA template, 5 µl of 10X *Taq* buffer, 5 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP, 2 µl of forward primer (Vg_RT_F1, 10 pmol/ µl), 2 µl of reverse primer (Vg_RT_R1, 10 pmol/ µl), 0.5 µl of *Taq* DNA polymerase and add sterilized distilled water up to 50 µl. PCR procedure was as followed: preheated at 94 ºC for 3 min, followed by 30 cycles of denaturing at 94 ºC for 30 sec, annealing at 55 ºC for 1.50 min and extension at 72 ºC for 30 sec. The final extension was carried out at 72 ºC for 10 min.

This fragment of RT-PCR products *Vg* gene was mixed with 6X loading buffer and the sample mixture was loaded into the wells of the submerged 1% agarose gel. The electrophoresis was carried out using 1X Tris Acetate EDTA (TAE) running buffer at 100V for 25 min. After finishing, the gel was stained in ethidium bromide (EtBr) solution for 10 min and destained by submerging in an excessive amount of water for 2 min. The nucleic acid bands were visualized under UV light.

RESULTS AND DISCUSSION

Results

1. Chromosome analysis

1.1 Chromosome complement

In this study, the behavior of chromosomes during spermatogenesis in *L. indicus* was observed. The mitotic metaphase of *L. indicus* spermatogonia was comprised of 26 chromosomes. The two large chromosomes and two m-chromosomes were obvious. It was difficult determine whether the remaining chromosomes were of middle or small size. Therefore, it was also quite difficult to perform a karyotype from mitotic chromosomes (Figure12a, b). The chromosome complement of *L. indicus* inferred from those in meiotic testicular cells at diakinesis and metaphase I stages was 2n = 22 + neo-XY + 2 m (Figure 12c). The meiotic karyotype was comprised of 13 bivalents: one large, nine medium, two small, and one m-chromosome (Figure 12d).

1.2 Meiotic chromosome behavior

The behavior of meiotic chromosomes was described in a typical cell. The polyploidy nuclei of nutritious cells at the seminiferous tubules walls had several heteropycnotic bodies with similar size contributing evenly (Figure 13a). In leptotene, the sex chromosomes appeared in one or two positively heteropycnotic bodies associated with the chromatin threads, while the nucleolus was less positively stained (Figure 13b). The chromosome threads were clearly seen in pachytene (Figure 13c). In diffuse stage, the heteropycnotic bodies were still obviously seen, while autosomes were partially decondensed and the less heteropycnotic positive nucleolus was round in shape (Figure 13d, 14a). At diplotene, 11 autosomal bivalents formed, and most bivalents or sometimes all bivalents appeared as rings. The ring bivalent indicated that a bivalent had two terminal chiasmata, while the neo-XY chromosomes

formed a bivalent with terminal association and were always associated with the nucleolus (Figure 13e, f, 14b). At diakinesis, the neo-XY chromosome could be seen conspicuously, with both ends staining darker than the remaining part, and the m-chromosomes could also be seen (Figure 13f, 14c). In some cells, two round nucleoli were observed. One was always associated with the neo-XY chromosome, and the other one could be located either close or far from the chromatin (Figure 13e). At late diakinesis, most bivalents appeared in short rod shapes, because one chiasma of the ring bivalent was released so that they became rod bivalents with terminal association, but some autosomal bivalents were still present in rings (Figure 12c). At metaphase I, nucleoli disappeared. The arrangement of bivalents did not form a particular pattern. However, in most cells, the neo-XY bivalent and one or two autosomal bivalents were surrounded by other autosomal bivalents (Figure 13g, 14d). The neo-XY bivalent was the biggest and was heteromorphic, comprising two chromosomes, one of which was a bit bigger than the other, associated together. The bigger one was probably the neo-X and the smaller might have been the neo-Y (Figure 12d, 13g, 14d). The autosomal bivalents were slightly different in size. From diplotene to telophase II, the m-chromosomes were always obviously seen as the smallest bivalent (Figure 13f-h). All bivalents showed axial orientation in the spindle. At anaphase I, all bivalents divided reductionally (Figure 13h, i, 14e, f). Therefore, at telophase I, two daughter nuclei had different sex chromosomes because one received the neo-X and the other obtained the neo-Y. However, it was impossible to identify which one received neo-X or neo-Y, since the size of both sex chromosomes was only slightly different (Figure 13i, 14e). Then, the chromosomes congregated together and the sex chromosome was obviously seen, with darker staining and the biggest size in the center (Figure 13j, 14f). In most meiosis II cells, metaphase II chromosomes had a radial configuration in which autosomal formed a ring with the sex chromosome as a part of the ring (Figure 13k, l, 14g). At anaphase II, all chromosomes divided equationally and all chromosomes congregated together with the sex chromosome, neo-X or neo-Y, obviously seen in the center at telophase II (Figure 13m, 14h). The spermiogenesis process began with a round spermatid containing a round, highly-stained, heteropycnotic body at the periphery or center (Figure 13n, 14j). Subsequently, the shape of spermatids elongated, and they could be seen as isosceles triangle-like with the sharpest corner

being heteropycnotic (Figure 13n, 14h). When the spermatid differentiated further, its shape became longer and the head of the spermatozoa was strongly stained (Figure 13p, 14l).

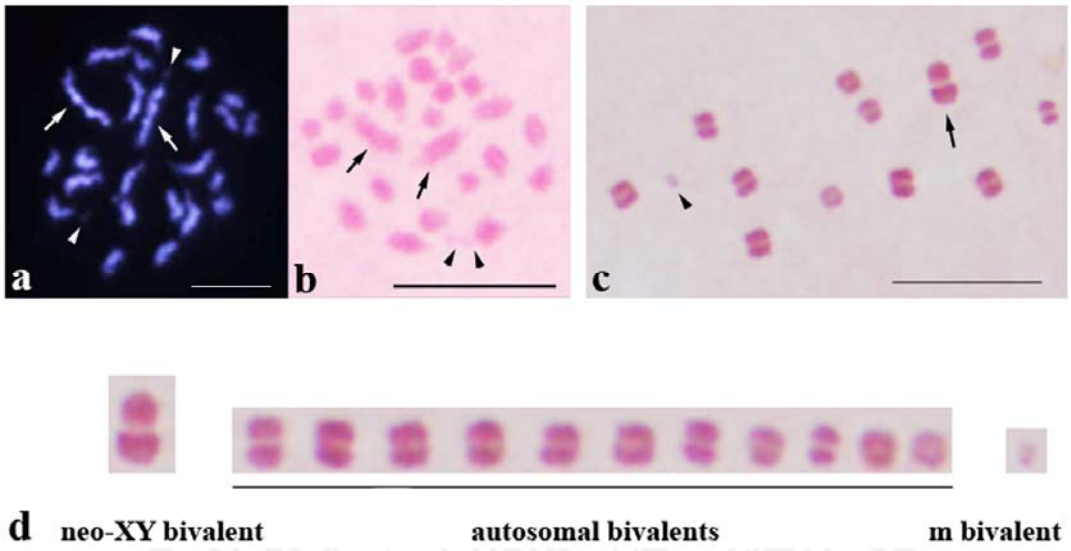


Figure 12 Male meiosis in *L. indicus* (2n = 22 A + neo-XY + 2m). (a) Mitotic prometaphase of spermatogonia stained with DAPI and (b) stained with lacto-aceto orcein. Arrows point to sex chromosomes and arrowheads point to m-chromosomes. (c) Metaphase I chromosomes stained with lacto-aceto orcein and (d) karyotype with 11 autosomal bivalents, a neo-XY bivalent, and an m-chromosome bivalent. Bar = 10 μm.

1.3 Univalent chromosomes

In the testis of an individual male, four cell types, designated A, B,C and D, with different patterns of chromosome univalents were observed. Type A cells with no univalent were the typical chromosome pattern of testicular cells of *L. indicus* in this study. Type A cells were present in majority (~92%), comprised 13 bivalents, including the neo-XY chromosome bivalent, which was the largest bivalent, 11 autosomal bivalents, which were similar in size, and an m-chromosome bivalent (Figure 15a, e). The other three cell types were characterized by the presence of chromosome univalents. Type B cells contained 12 bivalents (10 autosomal bivalents,

one sex chromosome bivalent, and an m-chromosome bivalent) and two autosomal univalents (Figure 15b, f). Type C cells also possessed 12 bivalents (11 autosomal bivalents and an m-chromosome bivalent) and two sex chromosome univalents, neo-X and neo-Y (Figure 15c, g). The frequency of type C cells (~5%) was a bit higher than that of type B cells (~3%). Some cells (~0.5%) were in type D, with 11 bivalents and four chromosomal univalents as the result of both the sex chromosomes and a pair of autosomes forming univalents (Figure 15d, h). Because autosomes were similar in size, it was impossible to identify with certainty which autosomal pair formed univalents, and whether autosomal univalents were presentin cells. The frequency of each cell type in each population is presented in Table 4.

1.4 DAPI staining

The testicular cells of *L. indicus* were stained with DAPI with the expectation that some fluorescent banding patterns would be observed; there was this expectation because DAPI stains AT-rich DNA in chromosomes. In mitotic chromosome (Figure 12a), all chromosomes contained DAPI-bright bodies. Conspicuously, the two longest chromosomes, neo-X and neo-Y, had three DAPI-bright bodies, two big bodies and a small one. The patterns of the DAPI-bright bodies in the two sex chromosomes were different. One sex chromosome contained a big bright, body at the end of chromosome, followed by another big one and then a small one near the middle of the chromosome. In the other sex chromosome, two big, bright bodies were located close to the middle, and the small one was next to one of the big bodies (Figure 12a). All autosomes contained one or two DAPI-bright bodies, while m-chromosomes showed relatively less brightly. In meiotic cells, one or two DAPI-bright bodies were found in a cell at the early stage of meiosis, leptotene (Figure 14a). At diplotene, a neo-XY chromosome was clearly distinguished from all autosomal bivalents and showed obviously DAPI-bright bodies at both chromosome ends. One end was bigger and brighter than the other, which displayed two DAPI-bright bodies (Figure 14b). From diakinesis onwards, all chromosomes showed almost the same level of DAPI-brightness except sex chromosomes, which displayed a bit brighter (Figure 14c-i). At the early spermiogenesis, each round spermatid contained a DAPI

bright body at one peripheral side (Figure 14j). Later, spermatids were elongated and a DAPI-positive signal had still been at their sharp corners (Figure 14k). In mature spermatozoa, the positive DAPI signals were present only in their heads (Figure 14l).

Table 4 Frequence of each cell types found in testes of *L. indicus* males collected from Thailand (Buri Ram, Sa Kaew and Chang Mai), Loas, Cambodia and Myanmar. Type A with no univalent, type B with autosomal univalent, type C with sex chromosome univalent and type D with both sex chromosome and autosome univalents.

Locations	individual	No. of cells in each type			
		Type A	Type B	Type C	Type D
Buri Ram	1	105	3	5	0
	2	130	1	4	0
	3	107	3	4	0
Sa Keaw	1	105	7	7	1
	2	105	2	7	0
	3	101	4	1	0
Chang Mai	1	130	6	10	2
	2	102	2	3	0
	3	110	4	6	0
Total	**9**	**995(92.1%)**	**32(3.2%)**	**47(4.4%)**	**3(0.3%)**
Loas	1	60	1	3	1
	2	169	2	7	0
	3	94	3	5	1
	4	140	4	5	0
Total	**4**	**463(93.5%)**	**10(2.0%)**	**20(4.0%)**	**2(0.2%)**
Cambodia	1	193	4	15	1
	2	160	1	4	0
	3	73	2	1	0
	4	106	5	6	0
	5	95	0	3	0
Total	**5**	**627(93.7%)**	**12(1.8%)**	**29(4.3%)**	**1(0.2%)**
Myanmar	1	100(78.7%)	11(8.7%)	13(10.2%)	3(2.4%)
Total	**19**	**2185(92.3%)**	**65(2.7%)**	**109(4.6%)**	**9(0.4%)**

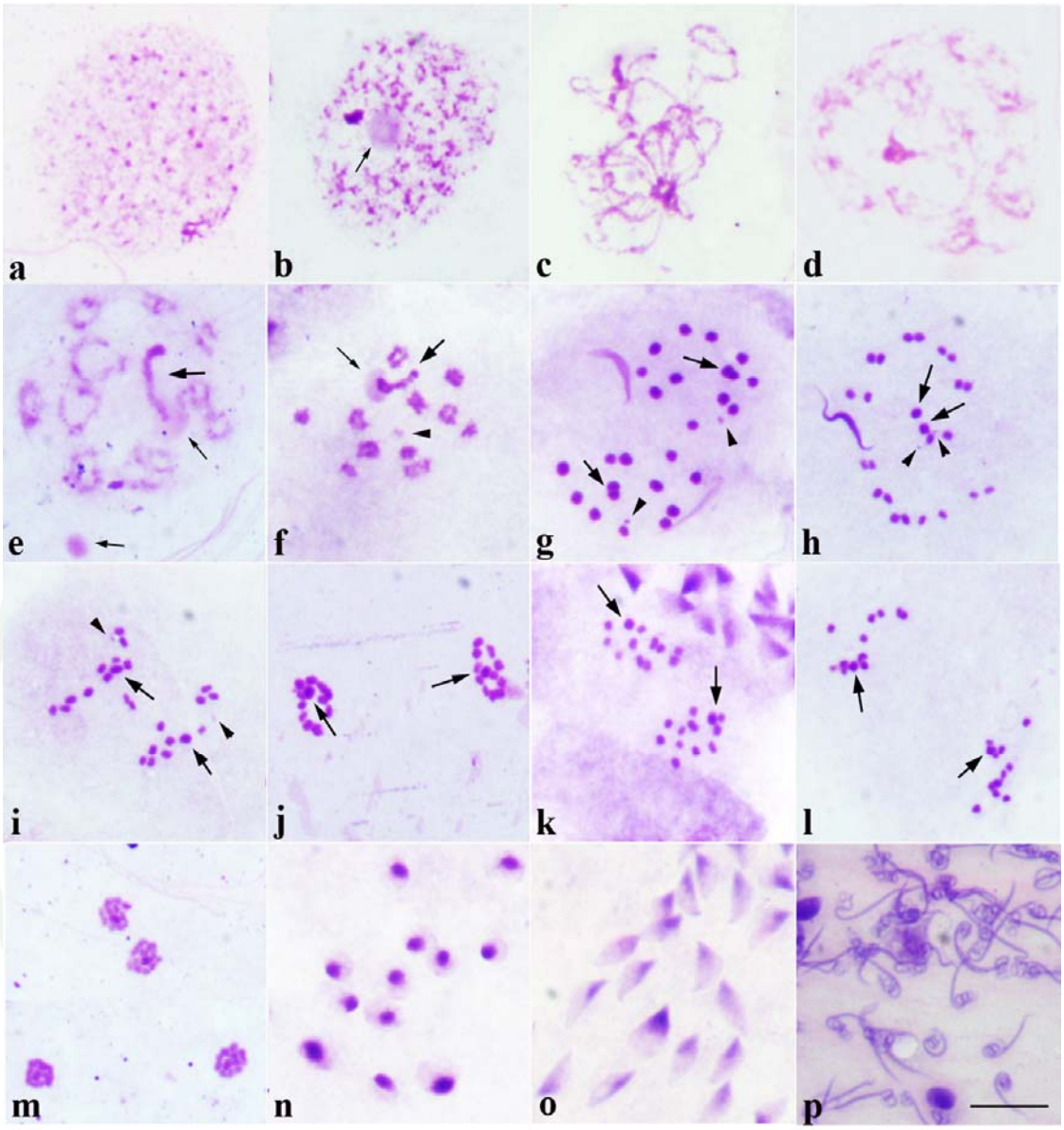


Figure 13 Meiotic chromosome behavior during *L. indicus* spermatogenesis stained with Giemsa. (a) polyploid nutritive cells with many heteropicnotic chromatin bodies; (b) leptotene-zygotene; (c) pachytene; (d) diffuse stage; (e) diplotene; (f) diakinesis; (g) metaphase I; (h, i) anaphase I; (j) telophase I;(k) metaphase II; (l) anaphase II;(m) telophase II; (n) round spermatids; (o) elongating spermatids; (p) head of spermatozoas. Big arrows indicate sex chromosomes, small arrows point to nucleoli and arrowheads indicate m-chromosomes. Bar = 10 μm

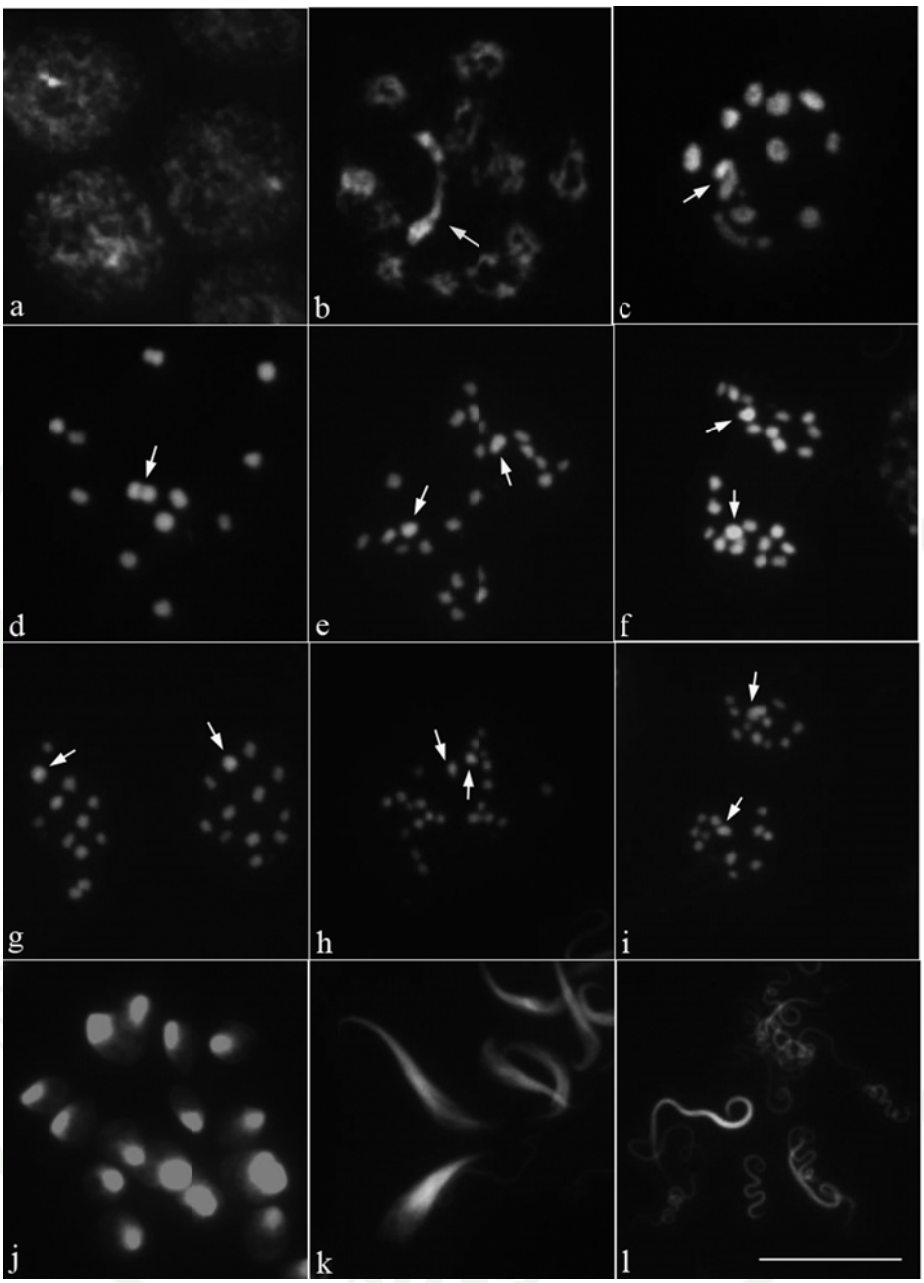


Figure 14 Meiotic chromosome behavior during *L. indicus* spermatogenesis stained with DAPI. (a) leptotene; (b) diplotene; (c) diakinesis; (d) metaphase I; (e) anaphase I; (f) telophase I; (g) metaphase II; (h) anaphase II; (i) telophase II; (j) early spermatids; (k) elongating spermatids; (l) head of spermatozoas. Arrows indicate sex chromosomes. Bar = 10 μm.

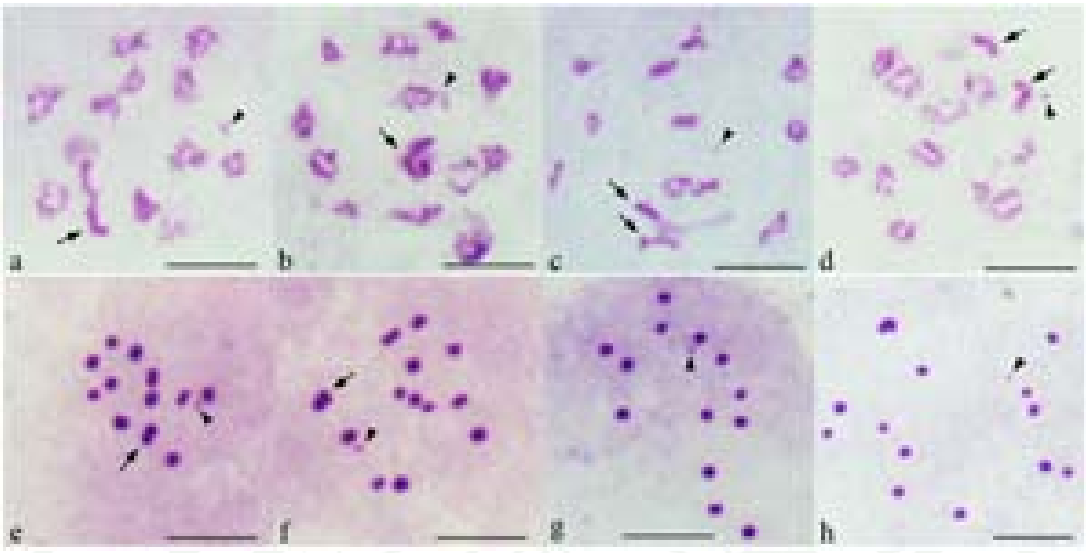


Figure 15 Four testicular cell types found in each testis of *L. indicus*.
(a, e) A typical cell type, type A cell have no univalent; (b, f) Type B cells with two autosomal univalents; (c, g) Type C cells with two sex chromosome univalents; (d, h) Type D cells with four univalent, two autosomal and sex chromosome univalents. Arrows indicate neo-XY chromsomes and arrowhead point to m-chromosomes. Bar = 10μm.

1.5 Silver staining

Silver nitrate has been used to reveal nucleolus and nucleolar organizing regions on chromosomes of many insect species, including Heteropteran insects. In this study, the nutritive polyploidy cells of testes stained with silver nitrate showed one or two impregnated regionsof nucleolar material (Figure 15a). Spermatogonial cells at early prophase I showed one or two nucleoli (Figure 16b, c). Most testicular cells at leptotene-pachytene contained two nucleolar bodies, one with strong staining associated with the chromatin, and another one with less staining located far from the chromatin (Figure 16d, e). In diplotene, the bodies were still stained and disappeared in late diakinesis (Figure 16f). From metaphase I to telophase I, all chromosome bivalents were positively stained with silver nitrate (Figure 16g, h). With this staining procedure, all obtained bivalents seemed smaller than those obtained with other staining procedures. Therefore, it was difficult to determine the exact locations of the

nucleolar organizing regions on the chromosome. At the beginning of spermiogenesis, the round spermatids had a round silver nitrate stained mass at the periphery (Figure 16i). The nucleolar mass still presented at the middle of elongated spermatids (Figure 16j). When spermatids developed to be spermatozoas, the sperm heads were more strongly stained than other parts of their bodies (Figure 16k, l).

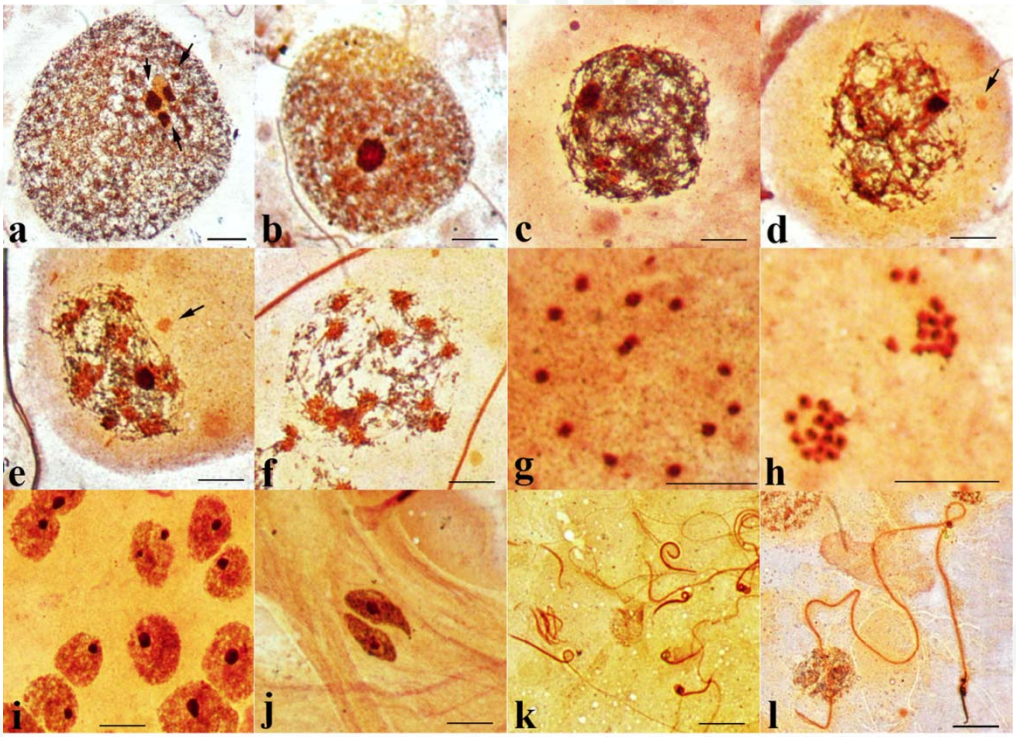


Figure 16 Silver nitrate staining the testicular cells of *L. indicus*. (a) polyploid nutritive cell showing two impregnated regions (arrows), (b) early leptotene with one nucleolar body; (c) zygotene; (d) pachytene; (d, e) arrows point to the nucleolus located out off chromatin region; (e, f) diplotene; (g) metaphase I; (h) telophase I; (i) round spermatids with nucleolar bodies at the periphery; (j) elongating spermatids with nucleolar bodies at the center of the heads; (k-l) mature spermatozoa showing the heads strongly impregnated with silver nitrate. Bar = 10 μm.

2. Mitochondrial genome analysis

2.1 Genome structure, organization and composition

The complete mitochondrial genome of *L.indicus*, (*Li*mt) (Hemiptera: Belostomitidae: Belostomatinae), was amplified by long polymerase chain reaction (PCR) in several overlapping fragments and sequenced directly. The entire sequence of the *Li*-mitogenome a 17,632-bp circular molecule, within containing 37 genes typically found in most insects is discovered: twenty-two transfer RNA (*tRNA*) genes, thirteen protein-coding genes (PCGs), two ribosomal (*rrnS* and *rrnL*) genes and a major non-coding region (the A+T-rich region or putative control region: CR) that are typically present in metazoan mitochondrial genome. A genes map of the genome and amplified regions are shown in Figure17 and Table 5 shows the positions of all major features of the genome.

The majority-coding strand (J-strand) contains 23 genes (9 PCGs: *cox1*, *cox2*, *cox3*, *cytb*, *atp6*, *atp8*, *nad2*, *nad3*, and *nad6*, and 14 *tRNA* genes) while the remaining 14 genes (4 PCGs: *nad1*, *nad4*, *nad4L*, and *nad5,* 8 *tRNA* genes and 2 *rRNA* genes) are locateded on the minority-coding strand (N-strand). The detailed positions and primary information of all genes are shown in Table 5. The nucleotide composition bias of the *L. indicus* mitogenome is 70.51% A+T and 70.33% for coding regions only. The A+T composition bias is 69.45% within protein-coding genes, and 72.78% for total tRNAs, 73.53% for *rRNA* genes, and 71.36% for the A + T-rich region (Table 6 and 7). The table 6 shows base composition of whole mitogenome was A>T>C>G and the strand bias is consistent with positive AT-skew and negative GC-skew for majority-coding strand (Table 7). To evaluate the degree of the base bias, the base-skew was measured and found that AT-skew and GC-skew in the complete genome of *L. indicus* were 0.2859 and -0.2323, respectively. Thereby indicating that more As and Cs are encoded in the whole mitogenome of *L. indicus*.

The overall organization of the *L. indicus* mitogenome is very compact, with only 58 nucleotides dispersed in 13 intergenic spacers and their ranged are

between 1-31 bp. The longest spacer sequence (31 bp) is located between *nad1* and $tRNA^{Leu(CUN)}$. The contiguous genes overlap in a total of 48 bp at 13 locations ranging from 1 to 8 bp, which the largest measuring 8 bp located between $tRNA^{Cys}$and $tRNA^{Tyr}$. In addition, the putative control region, the longest intergenic spacer is found between *rrnS* and $tRNA^{Ile}$ - $tRNA^{Gln}$-$tRNA^{Met}$ cluster gene.

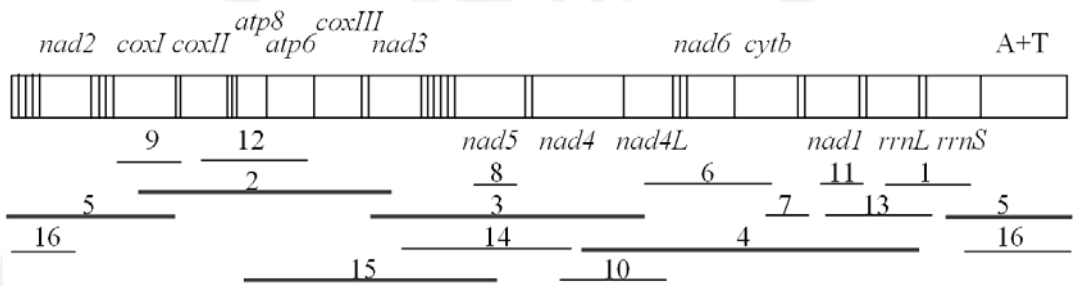


Figure 17 Schematic representation of amplification strategy employed for the mitochondrial genome of *L. indicus*. Lines below the genome map represent the amplification products. Numbers identify the primer pair (listed in Table 1). Normal lines represent amplification products obtained with standard PCR protocols, and bold lines (2, 3, 4, 5 and 15) represent those obtained through long PCR.

2.2 Protein-coding gene

The mtDNA of *L. indicus* contains the full set of PCGs usually present in animal mtDNA. A summary of the mitochondrial genes of *L. indicus* is given in Table 5. The genes are *NADH dehydrogenase subunit 1* (*nad1*), *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6* that encode forprotein complex I (NADH-ubiquinol oxido-reductase), *cytochrome b* (*cytb*) forprotein complex III (ubiquinone-cytochrome-c-oxidoreductase), *cytochrome c oxidase subunit 1* (*coxl*), *coxII*, and *coxII* for protein complex IV (cytochrome c oxidase), and *ATP synthase FO subunit 6* (*atp6*) and *atp8* for protein complex V (ATP synthase).

The total length of 13 PCGs was 11,094 bp in which 11,064 bp was translated 3,688 amino acids and the remaining 30 bp was stop codon. The base composition of all PCGs genes was T>A>C>G that shown in Table 6. Two different groups of overlapping PCGs genes were overlap 1 nucleotide composed of *atp6-coIII* and *nad6-Cytb* and seven nucleotides (7-bp) includes *atp8-atp6* and *nad4-nad4L* (Table 6). For *atp8-atp6* and *nad4-nad4L* gene pairs appear to overlap **ATGATAA** and **TTATCAT** sequences in different reading frames.

The PCGs of *L. indicus* initiated with ATN as the start codon, five with ATG (*nad2*, *atp6*, *coxIII*, *nad4*, and *cytb*), two with ATA (*coxII* and *nad6*), two ATT (*atp8, nad3*and *nad1*)and ATC (*nad5*) (Table 8). The exceptions were the *coxI* and *nad4L* gene, which used GTG and ACT as a start codon, respectively. Stop codon contained two complete termination codons TAA (*nad2*, *coxI*, *atp8*, *nad5*, *nad4L* and*nad6*) and TAG (*nad4* and *cytb*), whereas two incomplete termination codons T- (*coxII*, *atp6*, *coxIII* and *nad3*) and TA- (*nad1*) (Table 9), which could be completed as a result of posttranscriptional polyadenylation.

Codon usage in the mitochondrial proteins was calculated for all protein-coding genes (Table 9). Leucine, serine, phenylalanine, methionine and isoleucine are the five frequently used amino acids, accounting for 12.01%, 10.06%, 9.73%, 8.62% and 7.94%, respectively (Figure 18). The codons ATA (methionine), TTT (phenylalanine), TTA (leucine), ACA (threonine) and GTA (valine) were represented in the coding and only CGC was not represented (Table 9). The most widely used codons in *L. indicus* contain more A and T than their synonymous codon.The A and T biases of different codon positions are shown in Table 10. In general it is very similar among the first and second position (average 65.60 and 63.94%, respectively). However, the third position shows an extreme bias toward A and T (average 79.45%).

2.3 Transfer and ribosomal RNAs coding gene

The entire complement of 22 tRNAs typical of arthropod mitogenomes was found in *L.indicus* and schematic drawing of their respective secondary structures

were shown in Figure 19. All of the *tRNA* genes could be folded as classic clover-leaf structures that were identified using tRNAScan-SE. The size of *L. indicus tRNA* genes ranged frome 61-70 nucleotides. The *tRNA* genes were spread over the mitogenome. Thirteen *tRNA* genes were coded on J-strand and nine on N-strand.The base composition of all *tRNA* genes was A>T>C>G and AT content was 72.78% (Table 6).

Base on the secondary structure, the total of 5 unmatched base pairs were found in 4 secondary structures of *L. indicus* tRNAs. Four of them were U-U pairs (2 bp) of $tRNA^{Ala}$, G-A pairs (1 bp) of $tRNA^{Asn}$, and A-A pairs (2 bp) of $tRNA^{Aps}$and $tRNA^{Asn}$, which form a weak bond located in the AA stem. The remaining1 bp, G-G mismatches in the T stem of $tRNA^{Glu}$. The AC loop (7 nucleotides) and the size of anticodon stems of secondary structures of *L. indicus* tRNAs were conservative. The most aminoacyl (AA) stem were 7 bp, with exception of $tRNA^{Ala}$, $tRNA^{Asp}$ and $tRNA^{Gly}$ were 6 bp and $tRNA^{Ser}$was 5 bp. The most of the size variation was the DHU and TψC (T) arm, within which the loop size (3-10 bp) was more variable than the stem size (2-4 bp) (Figure 19).

The boundaries of *rRNA* genes were determined by sequence alignment with that of *Diplonychus resticus*, *Alloeorhynchus bakeri*, *Triatoma dimidiate* and *Enithares tibialis*. The small and large *ribosomal RNAs* genes in *L. indicus* were located between $tRNA^{Leu(CUN)}$ and $tRNA^{Val}$ and between $tRNA^{Val}$ and control region, respectively (Table 6). The *lrRNA* and *srRNA* genes were encoded in N-strand that had 1,269 and 790 bp in length, respectively, which were based on the location of neighboring tRNAs and comparison with other related sequences. The base composition of both *rRNA* genes was A>T>C>G and AT content was 74.55 % of *rrnL* and 71.90 % of *rrnS* as shown in the Table 6 and 7.

Table 5 Summary of the mitochondrial gene of *L. indicus*

Gene	Direction	Position	Size (bp)	Intergenic nucleotides	Anti-codon	Start codon	Stop codon
tRNA^{Ile}	F	65	65		GAT	-	-
tRNA^{Gln}	R	63-131	69	-3	TTG	-	-
tRNA^{Met}	F	131-199	69	-1	CAT	-	-
nad2	F	200-1210	1011	0	-	ATG	TAA
tRNA^{Trp}	R	1209-1274	66	-2	GCA	-	-
tRNA^{Cys}	R	1267-1331	65	-8	GTA	-	-
tRNA^{Tyr}	F	1334-1395	62	2	GTA	-	-
coxI	F	1397-2935	1539	1	-	GTG	TAA
tRNA$^{Leu(UUR)}$	F	2931-2995	65	-5	TAA	-	-
coxII	F	2996-3674	679	0	-	ATA	T--
tRNA^{Lys}	F	3675-3744	70	0	CTT	-	-
tRNA^{Asp}	F	3745-3807	63	0	GTC	-	-
atp8	F	3809-3967	159	1	-	ATT	TAA
atp6	F	3961-4624	664	-7	-	ATG	T--
cooxIII	F	4623-5412	790	-2	-	ATG	T--
tRNA^{Gly}	F	5413-5473	61	0	TCC	-	-
nad3	F	5475-5826	352	1	-	ATT	T--
tRNA^{Ala}	F	5828-5891	64	1	TGC	-	-
tRNA^{Arg}	F	5893-5953	61	1	TCG	-	-
tRNA^{Asn}	F	5955-6024	70	1	GTT	-	-
tRNA$^{Ser(AGN)}$	F	6019-6088	70	-6	GCT	-	-
tRNA^{Glu}	F	6088-6149	62	-1	TTC	-	-
tRNA^{Phe}	F	6149-6215	67	-1	GAA	-	-
nad5	R	6216-7925	1710	0	-	ATC	TAA
tRNA^{His}	R	7923-7986	64	-3	GTG	-	-
nad4	R	7987-9318	1332	0	-	ATG	TAG
nad4L	R	9312-9620	309	-7	-	ACT	TAA
tRNA^{Thr}	F	9629-9694	66	8	TGT	-	-
tRNA^{Pro}	R	9695-9762	68	0	TGG	-	-
nad6	F	9765-10265	492	2	-	ATA	TAA
cytb	F	10265-11392	1137	1	-	ATG	TAG
tRNA$^{Ser(UCN)}$	F	11391-11459	69	-2	TGA	-	
nad1	R	11491-12410	920	31	-	ATT	TA-
tRNA$^{Leu(CUN)}$	R	12417-12481	65	6	TAG	-	-
rrnL	R	12482-13750	1269	0	-	-	-
tRNA^{Val}	R	13753-13822	70	2	TAC	-	-
rrnS	R	13823-14612	790	0	-	-	-
Control region		14613-17632	3020	0	-	-	-

Table 6 Nucleotide composition of the *L. indicus* mitogenome

Genome feature	Number of nucleotides	Proportion of nucleotides				%AT
		A	C	G	T	
All nucleotides	17632	0.4534	0.1817	0.1132	0.2518	70.51
rrnS	790	0.4620	0.1772	0.1038	0.2570	71.90
rrnL	1269	0.4799	0.1552	0.0993	0.2656	74.55
tRNA	1451	0.4080	0.1496	0.1227	0.3198	72.78
Protein coding gene	11094	0.3159	0.1847	0.1209	0.3786	69.39
nad2	1011	0.4065	0.1840	0.1177	0.2918	69.74
coxI	1539	0.3496	0.2021	0.1592	0.2891	63.87
coxII	679	0.4109	0.2003	0.1340	0.2548	66.57
atp8	159	0.4969	0.1509	0.0566	0.2956	79.25
atp6	664	0.4157	0.1913	0.1145	0.2786	69.43
coxIII	790	0.3671	0.2025	0.1443	0.2860	65.32
nad3	352	0.4176	0.1847	0.1222	0.2756	69.32
nad5	1710	0.5281	0.1807	0.1076	0.1836	71.17
nad4	1332	0.5368	0.1599	0.1051	0.1967	73.35
nad4L	309	0.5928	0.1564	0.0814	0.1694	76.22
nad6	492	0.4878	0.1524	0.0915	0.2683	75.61
cytb	1137	0.4002	0.1996	0.1310	0.2691	66.93
nad1	920	0.5348	0.1851	0.1098	0.1739	70.87
Control region	3020	0.4679	0.1970	0.0894	0.2457	71.36

Table 7 Nucleotide strand asymmetry of the *L. indicus* mitogenome

Sequence	Size	%AT	AT Skew	GC skew
Whole genome	17632	70.51	0.2859	-0.0394
PCGs	11064*	69.39	-0.0907	-0.0394
rrnS	790	71.90	0.2852	-0.2613
rrnL	1269	74.55	0.2875	-0.2198
Control region	3020	71.36	0.3114	-0.3757

*Terminal codons were excluded in total codon count.

Table 8 Start and stop codon of the *L. indicus* mitogenome

Gene	Start codon	Stop codon
nad2	ATG	TAA
coxI	GTG	TAA
coxII	ATA	T-
atp8	ATT	TAA
atp6	ATG	T-
coxIII	ATG	T-
nad5	ATC	TAA
nad4	ATG	TAG
nad4L	ACT	TAA
nad6	ATA	TAA
cytb	ATG	TAG
nad1	ATT	TA-

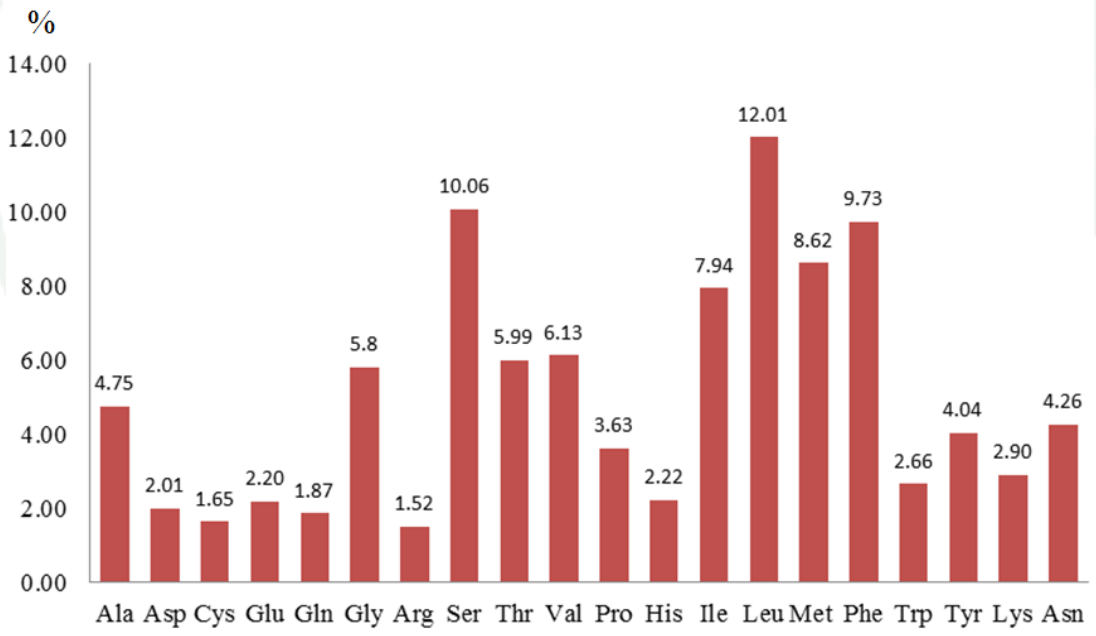


Figure 18 Amino acid content for 13 PCGs of *L. indicus* mitogenome. This analysis excludes stop codons (3,688 amino acid). The amino acids are shown by standard abbreviations.

Table 9 Relative codon usage table for *L. indicus* mitogenome

aa	Codon	Number	%	aa	Codon	Number	%
Ala	GCG	6	4.75	Asn	AAT	91	4.26
	GCA	86			AAC	66	
	GCT	52		Pro	CCG	2	3.63
	GCC	31			CCA	70	
Cys	TGT	39	1.65		CCT	54	
	TGC	22			CCC	8	
Asp	GAT	45	2.01	Gln	CAG	15	1.87
	GAC	29			CAA	54	
Glu	GAG	20	2.20	Arg	CGG	3	1.52
	GAA	61			CGA	37	
Phe	TTT	258	9.73		CGT	16	
	TTC	101			CGC	0	
Gly	GGG	22	5.80	Ser	AGG	11	10.06
	GGA	103			AGA	72	
	GGT	84			AGT	39	
	GGC	5			AGC	10	
His	CAT	38	2.22		TCG	11	
	CAC	44			TCA	105	
Ile	ATT	221	7.94		TCT	111	
	ATC	72			TCC	12	
Lys	AAG	34	2.90	Thr	ACG	7	5.99
	AAA	73			ACA	123	
Leu	TTG	73	12.01		ACT	76	
	TTA	187			ACC	15	
	CTG	9		Val	GTG	9	6.13
	CTA	108			GTA	122	
	CTT	61			GTT	92	
	CTC	5			GTC	3	
Met	ATG	51	8.62	Trp	TGG	23	2.66
	ATA	267			TGA	75	
Tyr	TAT	105	4.04				
	TAC	44					

Table 10 Nucleotide composition at different codon positions of 13 PCGs in *L. indicus* mitogenome

Base position	%A	%C	%G	%T	%AT	%GC	AT-skew	GC-skew
1st	33.64	13.32	21.1	31.95	65.60	34.42	0.0259	0.2261
2nd	19.50	20.85	15.21	44.44	63.94	36.06	-0.3902	-0.1564
3rd	41.95	12.67	7.87	37.50	79.45	20.54	0.0560	-0.2338

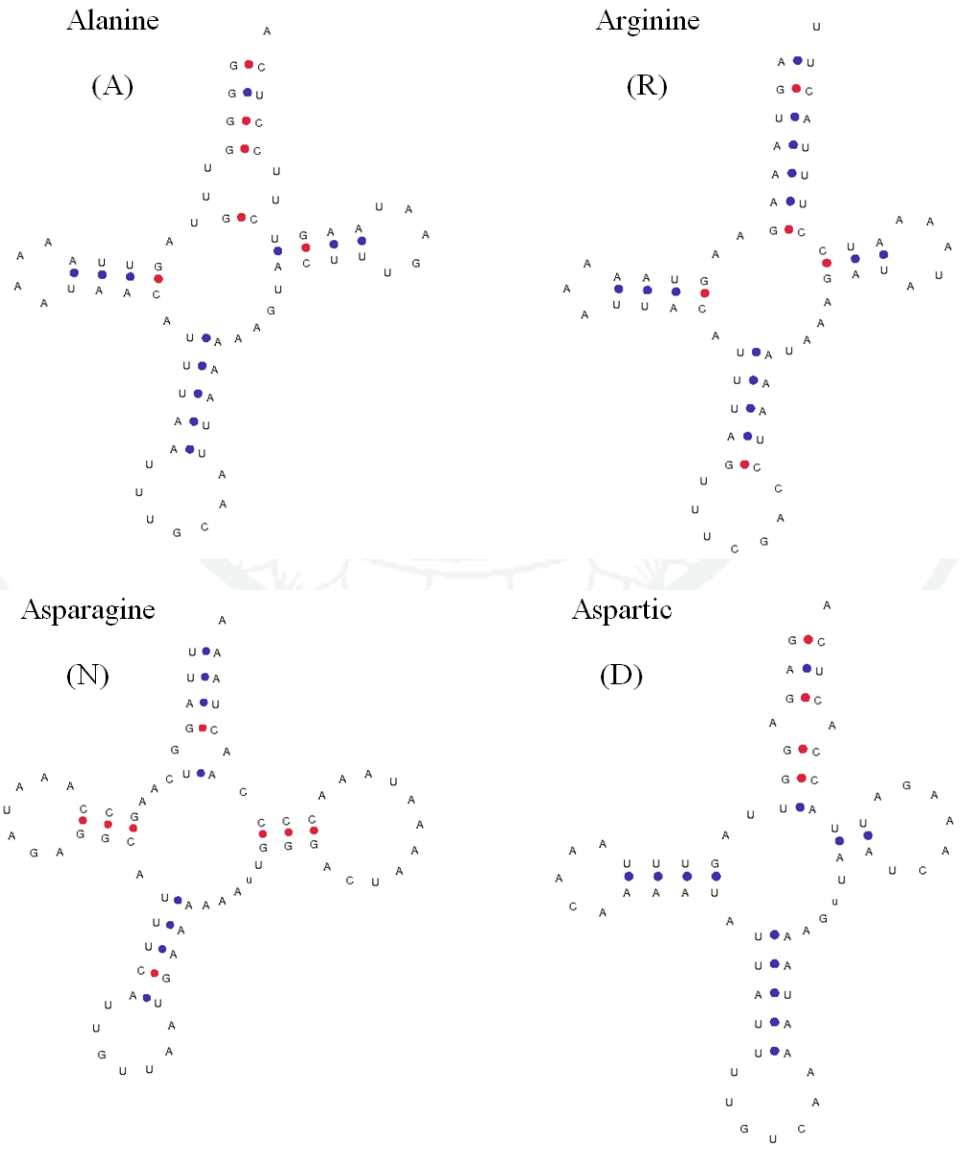


Figure 19 Putative secondary structures for 22 tRNAs of *L. indicus* mitogenome.

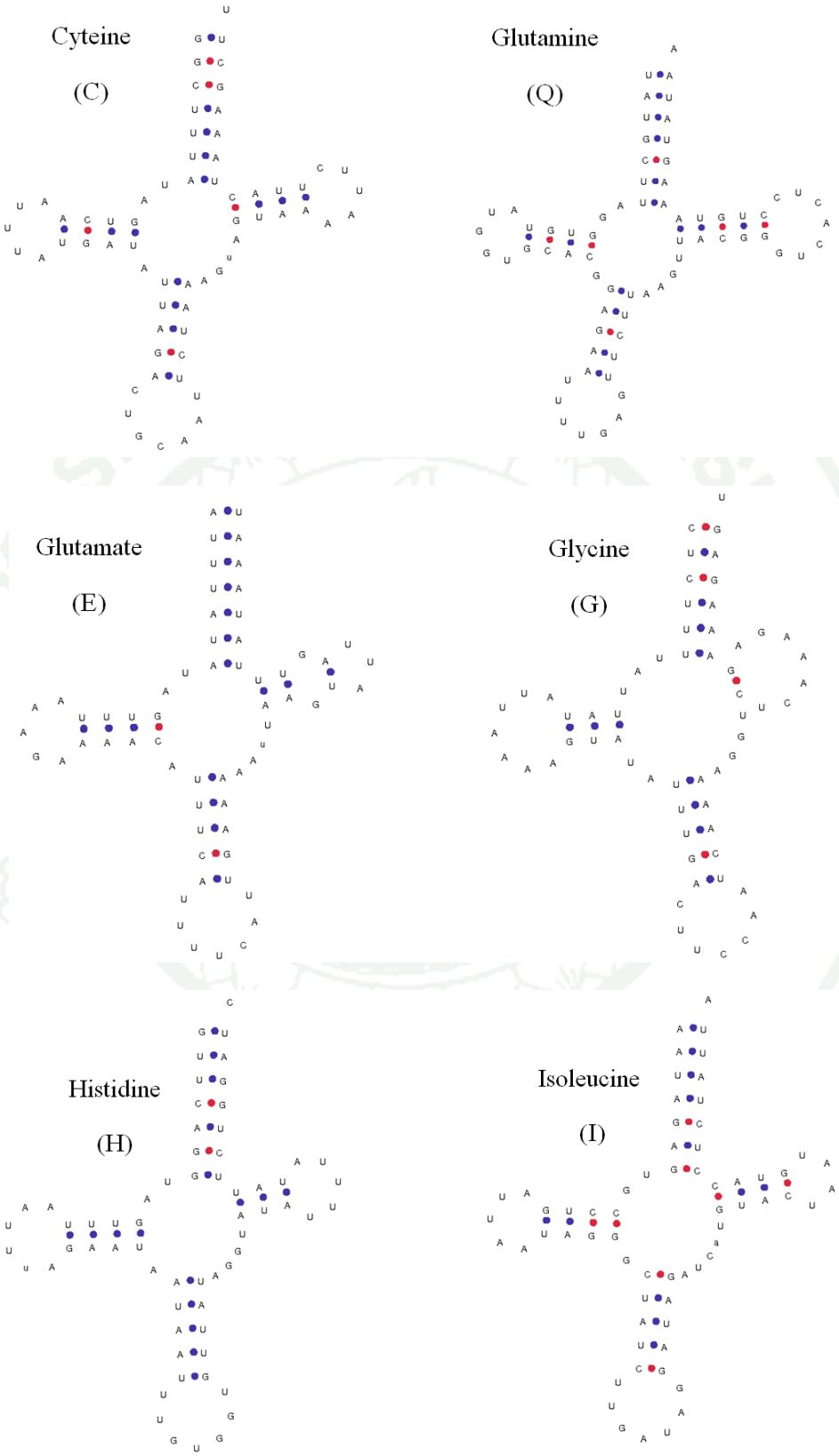


Figure 19 (Continued)

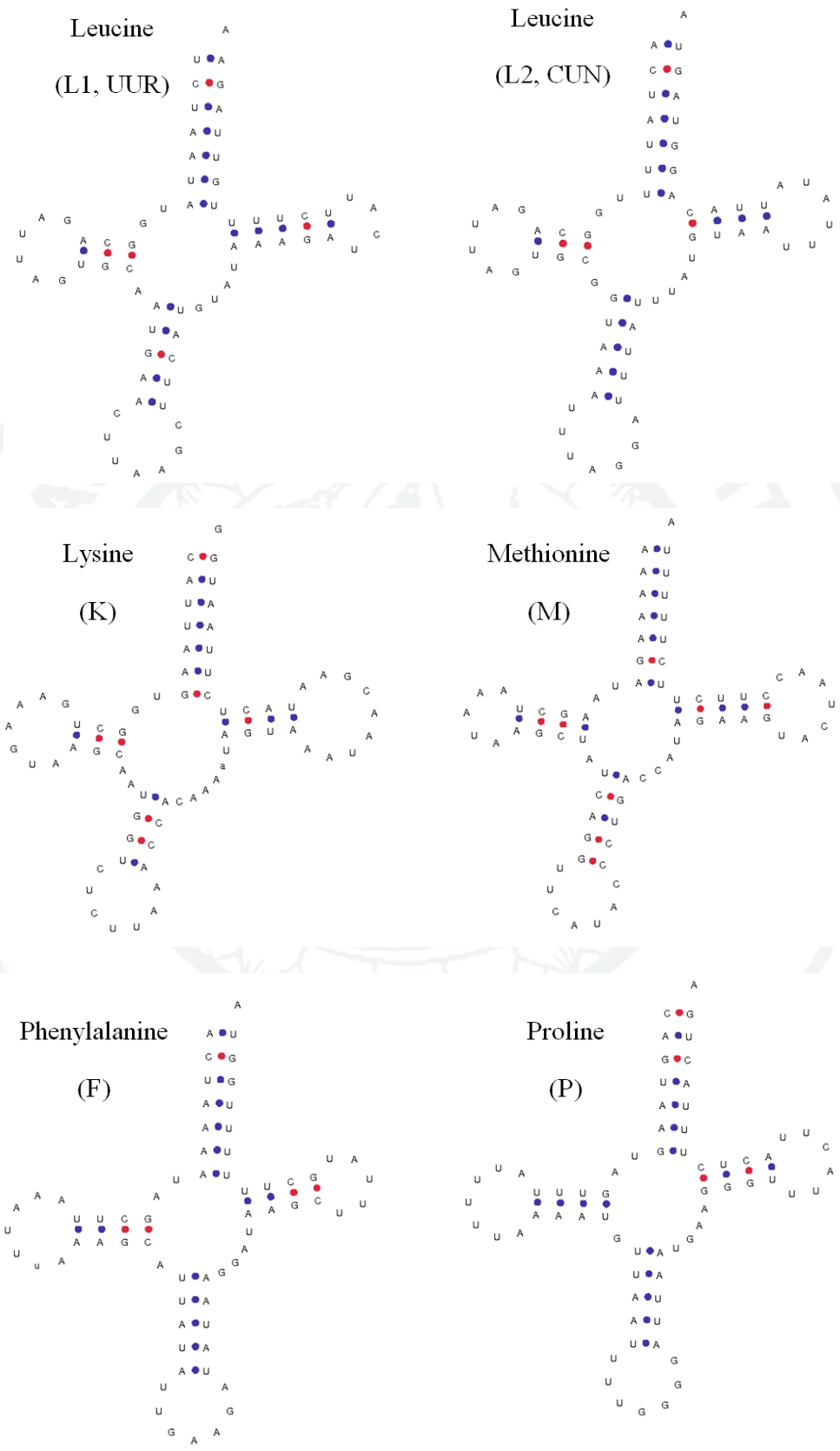


Figure 19 (Continued)

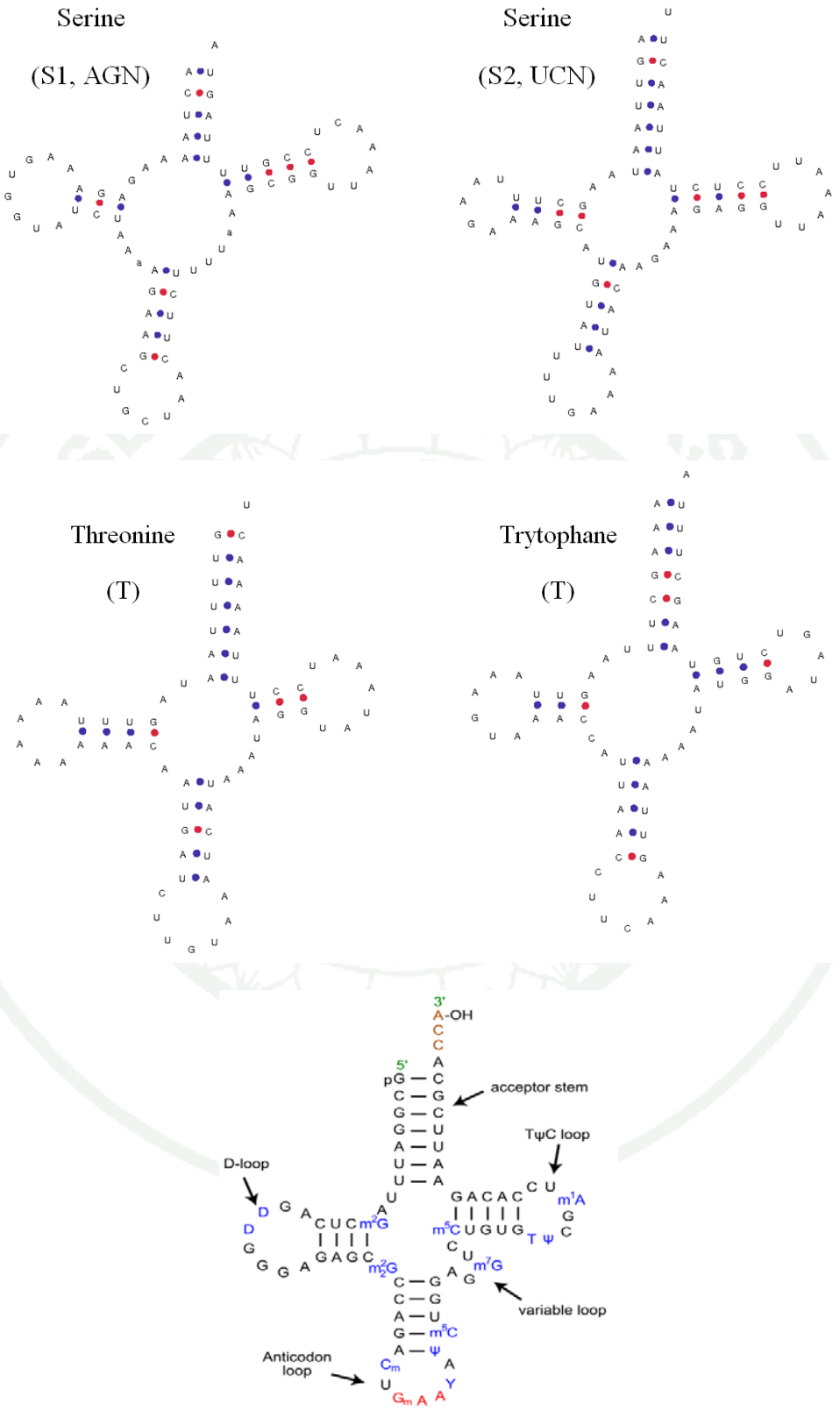


Figure 19 (Continued)

2.4 The control region or CR

The 3,020 bp long control region of *L. indicus* mitogenome was 17.13 % of whole genome. The CR was located at the conserved position between *rrnS* and $tRNA^{Ile}$- $tRNA^{Gln}$ - $tRNA^{Met}$ gene. CR was composed of 71.36% AT content (Table 6).

The control region can be divided into five parts (Figure 20) (1) a 740 bp region that was bordered by *rrnS*, of which the 36.9% GC content is higher than the whole genome; (2) a region composed of five tandem repeats, that each tandem repeat unit of 174 bp and one of 175 bp; (3) tandem repeats regions consist ofnine copies of 114-bp ; (4) a 210 bp in length near $tRNA^{Ile}$- $tRNA^{Gln}$ - $tRNA^{Met}$ gene, of which the 17.6% GC content is lower than the whole genome; (5) the 173 bp interspace sequence between first and second tandem repeats regions and this region consist of incomplete two tandem repeats sequence, that have 115 bp of first and 58 bp second tandem repeats region.

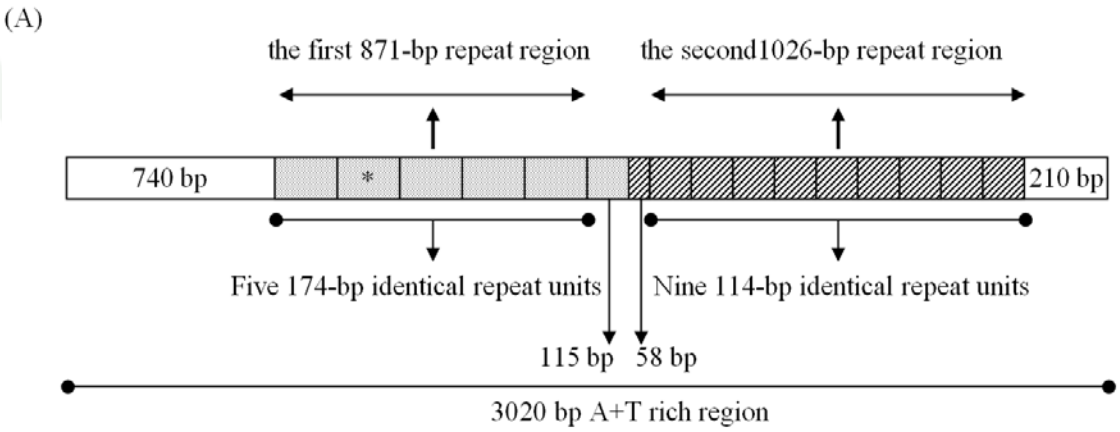


Figure 20 The A+T rich region of *L. indicus* mitogenome. (A) The structural organization of the A+T rich region of the *L. Indicus* mitogenome.(B) The sequence of the repeat unit in the the first region that have 174 and 175 bp. (C) The sequence of the repeat unit in the the second region, 114 bp.

(B)

```
ATAACAAATAAATAAAATTTATCTGATCTTTAATCAGAAATCAAAAAACA
CTGGTCCAAAACGTTTCCTATTTGGAAATTAATCATAAAATAAGAAATAA
CTGTTAAAAACCAAAGTCAAATTAATACCTGATATAAAATGTTAAGCAAA
CAAGCGCTTAACAAAGAAATTTCG-174 bp

ATAACAAATAAATAAAATTTATCTGATCTTTAATCAGAAATCAAAAAACA
CTGGTCCAAAACGTTTCCTATTTGGAAATTAATCATAAAATAAGAAATAG
CTGTTAATAAACCAAAGTCAAATTAATACCTGATATAAAATGTTAAGCAA
ACAAGCGCTTAACAAAGAAATTTCG-175 bp
```

(C)

```
CATAAACACAGAATAAATAAAATAATAAAAATAAATAAACGACACAA
TAAACCAAAACCAAATAGACCACAAAGTGTGAGCACCCAATAAATAA
GTAATTATCGCAGAGTCCTC-114 bp
```

Figure 20 (Continued)

2.5 Phylogenetic analysis

Phylogenetic tree was performed using the amino acid sequences of 13 PCGs genes of forty-two mitogenomes; including *L. indicus* represent six hemipteran higher superfamily (Sternorrhycha, Fulgoromorpha, Cicadomorpha, Orthptera, Psocoptera and Heteroptera). The amino acid sequences of 13 PCGs were aligned with Clustal X 1.83 and translated into amino acid sequences in the settings and implemented in the MEGA version 5.03. Neighbor Joining and Maximum Likehood analysis generated identical tree topologies (Figure 21 and 22). Using two Sternorrhycha species (*Schizaphis graminum* and *Bemisia tabaci*) as outgroups, the phylogenetic analysis presented: Sternorrhycha + (Fulgoromorpha + (Cicadomorpha + ((Orthoptera + Psocoptera) + (Heteroptera + (Pentatomomorpha + (Aradoidea + (Cimicomorpha+ (Leptopodomorpha + Nepomorpha)))))))) by ML analysis.The phylogenetic tree revealed five main clades. The species of the higher superfamilies Sternorrhycha, Pentatomomorpha, Cicadomorpha, Orthoptera +Psocoptera and Heteropteracluster monophyletic groups, respectively and are strongly supported (NJ analysis). In two sequenced Lethocerus species (belong Belostomatidae), the phylogenetic analyses support a close relationship between *L. indicus* and *D. rusticus* with 99% by NJ analyses and ML analyses.

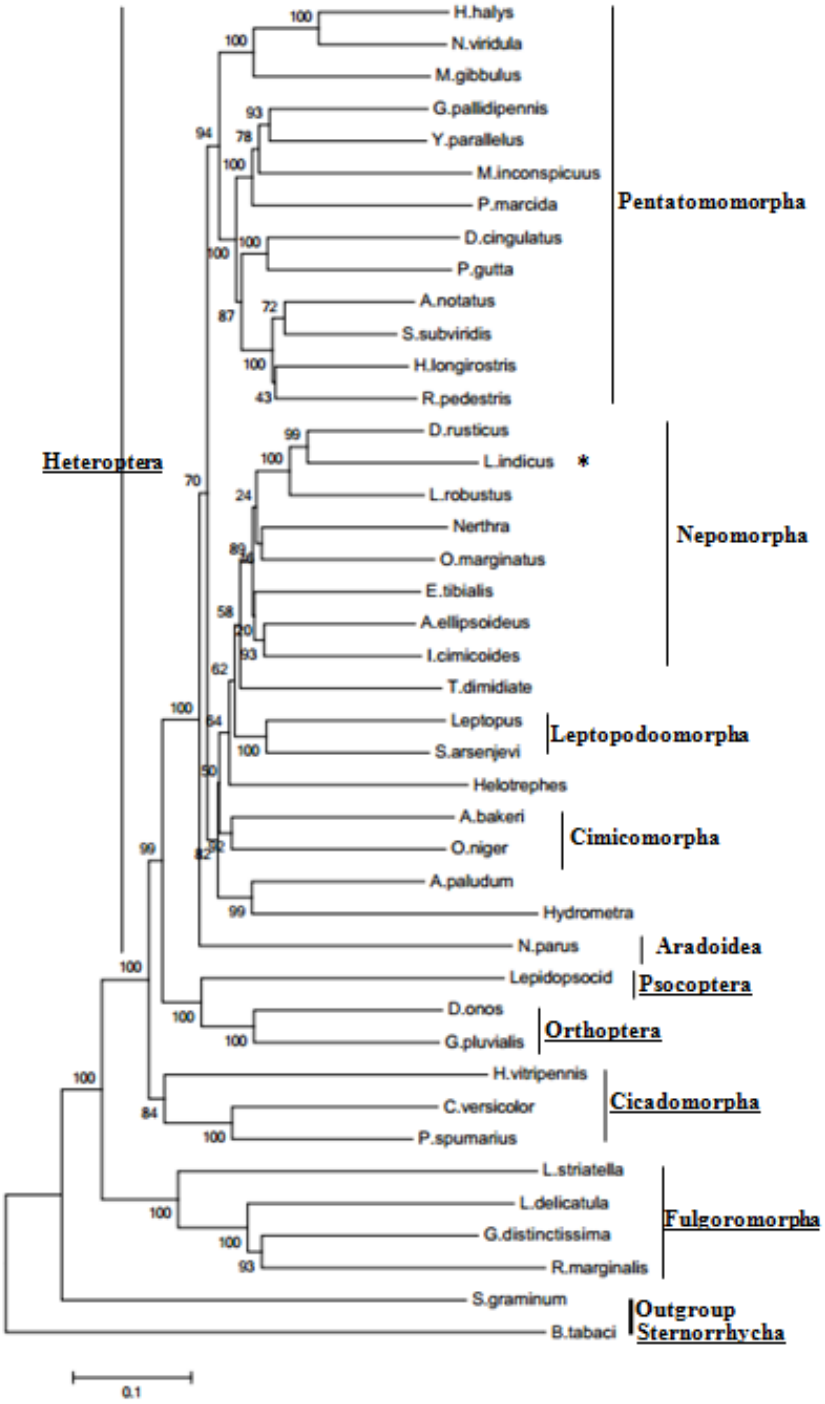


Figure 21 Phylogenetic analyses of hemipteran insects. Phylogenetic tree inferred from the amino acid sequences of 13 PCGs of the mitogenome using Neighbor Joining analysis. *B. tabaci* and *S. graminum* were used as outgroups. The numbers above the branches specify bootstrap percentages (1,000 replicates). (* in this study)

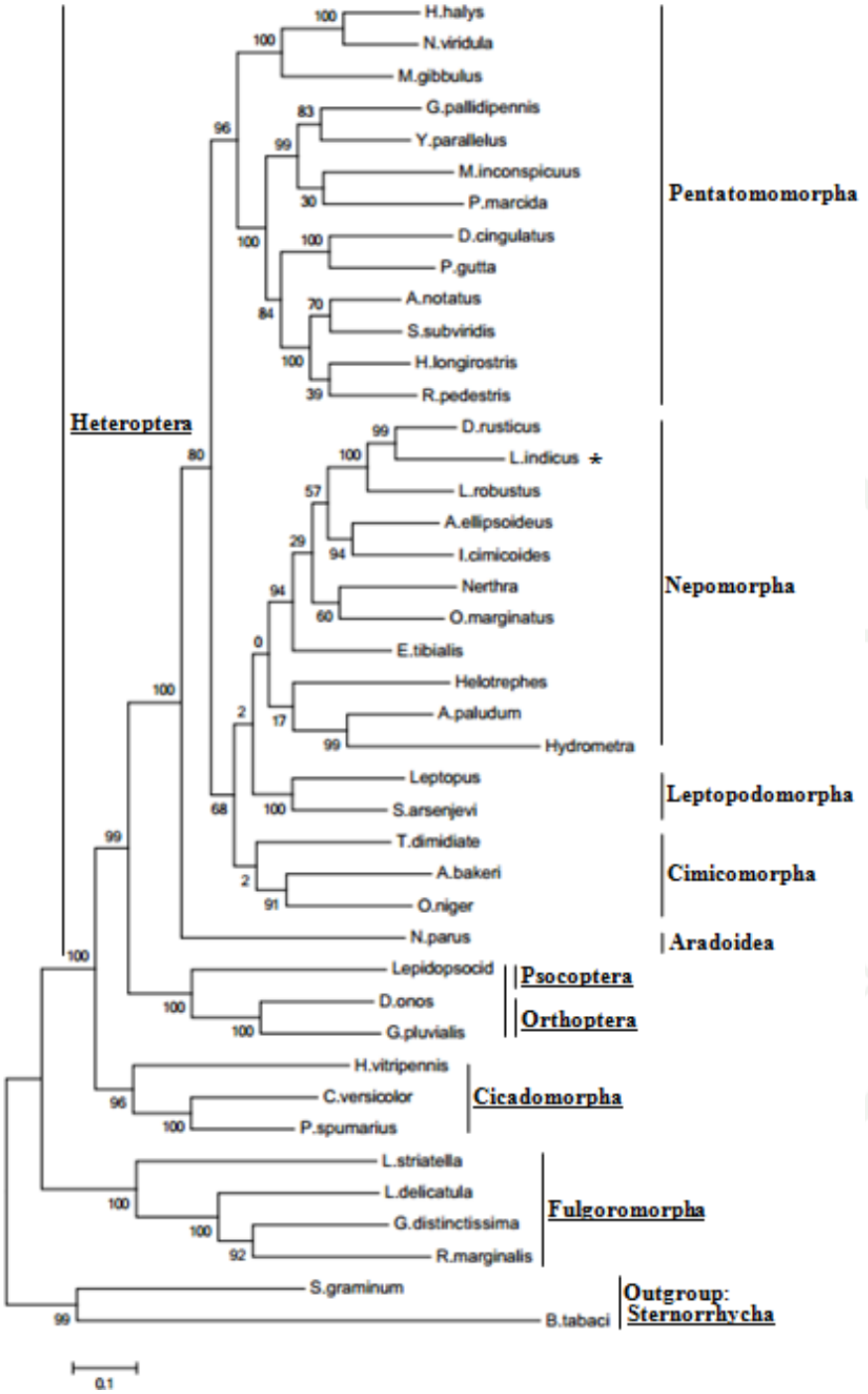


Figure 22 Phylogenetic analyses of hemipteran insects. Phylogenetic tree inferred from the amino acid sequences of 13 PCGs of the mitogenome using Maximum Likelihood analysis. *B. tabaci* and *S. graminum* were used as outgroups. The numbers above the branches specify bootstrap percentages (1,000 replicates). (* in this study)

3. Cloning and sequencing of *Vg* gene in *L. indicus*

3.1 PCR primer and amplification

The total RNA of adult female giant water bug was used as a template to synthesis the first strand cDNA using RevertAid First Strand cDNA Synthesis Kit. Then, the first strand cDNA were used as a template for amplification three fragments of *Vg* cDNAs by using three sets primer LiVgF1 and LiVgR1 for fragments 1 (Figure 23a), LiVgF2 and LiVgR2 for fragments 2 (Figure 23b) and LiVgF3 and LiVgR3 for fragments 3 (Figure 23c). Three sets of sequence primer for three fragment PCR products are show in the Table 2. The obtained three PCR products were about 1,000, 3,000 and 1,800 bp, respectively. Thes PCR products were purified using the FavorPrep™ Gel/PCR Purification Kit and sent to Macrogen for sequencing. Then, internal primers were design to further sequence (Fragment 2 and 3) into the fragments.

3.2 Rapid amplification of the 5'ends and 3' ends of *LiVg*

The 5' and 3' RACE reactions was performed using the SMARTer™ RACE cDNA Amplification Kit User Manual according to manufacturer's instructions, using one adapter-specific and the other gene-specific primers. The gene-specific primer used for 5' and 3' RACE of *LiVg* was GSP1_Vg_5R and GSP2_Vg_3F and its sequence was showed in Table 2.The size of PCR product were 969 and 1,901 bp long, respectively (Figure 24a and b) and purified PCR products from agarose gel and were ligated into pGEM-T® easy vectors. Then, the recombinant vectors were transformed into *E. coli* DH5α by heat shock transformation. Then, the products were purified using the FavorPrep™Gel/PCR Purification Kit and sent to Macrogen for sequencing.

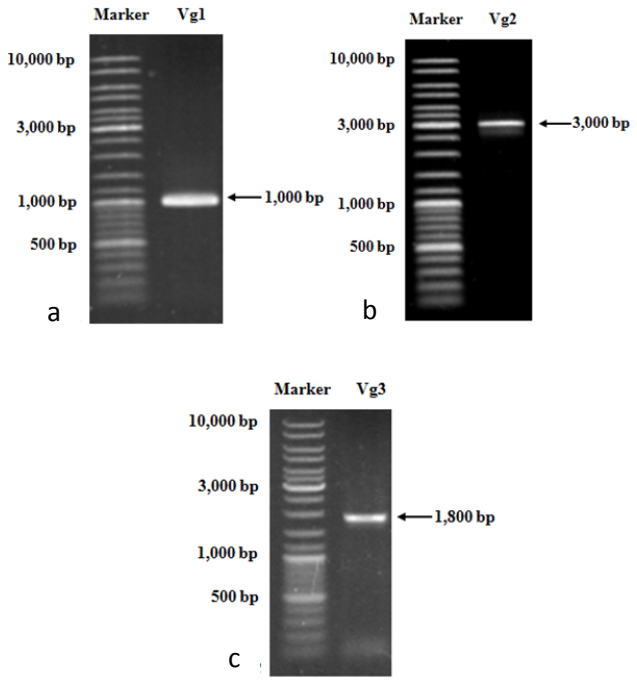


Figure 23 Agarose gel electrophoresis of the PCR product of *Vg* amplified from cDNA of fat body female. (a) PCR product of fragment 1, (b) PCR product of fragment 2 and (c) PCR product of fragment 3. Lane1: DNA ladder and Vg1-3: PCR products of *Vg* gene from the fat body female**.**

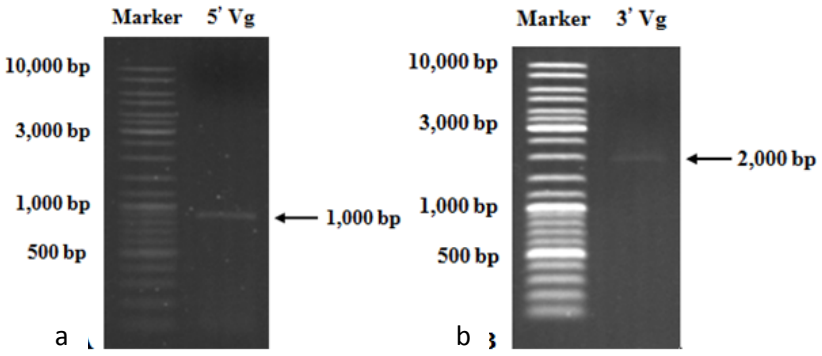


Figure 24 Agarose gel electrophoresis of the PCR product of *Vg* amplified from cDNA of fat body female. (a) 5'ends *Vg* cDNA Lane 1: DNA ladder, Lane 2: PCR products of 5' ends *Vg* gene; (b) 3' ends *Vg* cDNA. Lane 1: DNA ladder, Lane 2: PCR products of 3' ends *Vg* gene

3.3 Complementary DNA (cDNA) sequence of *L. indicus Vg* gene

The obtained nucleotide cDNA sequence of 5' RACE, 3' RACE and 3 fragments were assembly using by CAP3 program. The full-length cDNA of *Vg* is 6,048 bp long, comprising a 5'-untranslated region (5'-UTR) of 253 bp, and a 3'-UTR of 131bp long. The stop codon (TGA) was 5,665-5,667 bp position and a consensus polyadenylation signals sequence (AATAAA) was found at 76 nucleotides down-stream of the terminal codon (at 5,804-5,808 bp position) (Figure 25).

3.4 Amino acid sequence of *L. indicus Vg* gene

The open reading frame encoded 1,888 amino acids and the calculated molecular weight was 212.40 kDa and theoretical isoeletric point (*pI*) of 6.58 by ExPASy. The analysis of the primary protein revealed that the first 18 amino acids highly hydrophobic residues at the N-terminal correspond to a putative signal peptide [as analyzed with the SignalP-4.1 program (http://www.cbs.dtu.dk/services/SignalP/)] (Figure 25). The National Center for Biotechnology Information (NCBI) Conserved Domain Database (CDD) search predicted vitellogenin-N domains (also known as the lipoprotein N-terminal regions) at the amino terminus of the LiVg (amino acid position: 27-814), two domains of unknowns function (DUF1943) (amino acid position: 846-1140) and the von Willebrand factor type D (VWD) domain, was identified at the C-terminus (amino acid position: 1557–1727) (Figure 26). In addition, the DGXR motif, GL/ICG motif and the cysteine residues conserved at the C-terminus, which were identifiedcommon features found in Vg hemipteran. Interestingly, LiVg contains polyserine regions and four RXXR consensus cleavage sites (RNRR, RALR, RNER and RLTR) which were located at the N-terminus (Figure 25).

```
                                                        5'TCACACAGGAAAC -241
AGCTATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGC -181
ATCCAACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAG -121
TGATTCTAATACGAGTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACATGGGATT -61
CCAAAGTTTGGGGTTTTGGTGTTTTGACTGTTAACTGAAAGGTCGGCAGTTGAACACGAG -1

ATGCGCTCCTCGGTGATTCTATTCCTATTTGTGACAGTGACGGCGGCGGTGGCAGAGCCT 60
  1 M  R  S  S  V  I  L  F  L  F  V  T  V  T  A  A  V  A  E  P
TGGTCAAATGCAACCGCATGGAAGCCCAACACAGTGTACAAGTACACACTTCGTGGCCGG 120
 21 W  S  N  A  T  A  W  K  P  N  T  V  Y  K  Y  T  L  R  G  R
TCGATGACCGGACTCCATCAAGTATCCAATCAGTATAGCGGAATCCTCCTGAAGGCTCAA 180
 41 S  M  T  G  L  H  Q  V  S  N  Q  Y  S  G  I  L  L  K  A  Q
GTTGCAATCCAGCTTAAGACTCAGAACCTATTGTCGCTGAAGGTGAGCAAAGCGGAGTTC 240
 61 V  A  I  Q  L  K  T  Q  N  L  L  S  L  K  V  S  K  A  E  F
GCCGAATTGCACGCCAATTTGTCCAAGGGATGGAACACCGAAATCCCGGACCACAAATTG 300
 81 A  E  L  H  A  N  L  S  K  G  W  N  T  E  I  P  D  H  K  L
CACTGGCAACCGATGCCCCTGGCCGGCAAACCGTTCGAACTGAAGCTGAAGAATGGGACC 360
101 H  W  Q  P  M  P  L  A  G  K  P  F  E  L  K  L  K  N  G  T
GTGGACAGGATGTACGTGGACAAGAGCGTGCCGACGTGGGAGGAAAACTGGTACAAGAGT 420
121 V  D  R  M  Y  V  D  K  S  V  P  T  W  E  E  N  W  Y  K  S
ATCGCCTCCCAACTGCAAGTCGACACCGCCGCCAGGAACCTGCAGAAGTCCAGGATTAAC 480
141 I  A  S  Q  L  Q  V  D  T  A  A  R  N  L  Q  K  S  R  I  N
CAGCTGCCGGCGATGCACAAGCCGATGGGAGTCTACAAAACCATGGAGGACACAGTGACG 540
161 Q  L  P  A  M  H  K  P  M  G  V  Y  K  T  M  E  D  T  V  T
GGAGAATGCGAGACTGTGTACGACGTATCACCATTGCCCGACTACCTGCTTCAGTCCAAG 600
181 G  E  C  E  T  V  Y  D  V  S  P  L  P  D  Y  L  L  Q  S  K
CCAGAGCTGGCTCCTCTCCCGCACCTCAGGGCCGACGGAAGGATCATCGAAATCGTCAAG 660
201 P  E  L  A  P  L  P  H  L  R  A  D  G  R  I  I  E  I  V  K
ACCAGGAACTTCTCCGAATGCGACCAACGAGTCGCTTTCCATTTCGGAATTACCGGTCCC 720
221 T  R  N  F  S  E  C  D  Q  R  V  A  F  H  F  G  I  T  G  P
```

Figure 25 The full-length nucleotide sequence and the deduced amino acid sequence of *Vg* gene in *L. indicus*. Underlined indicate the start codon ATG and stop codon TGA, respectively. Double underlined indicate the signal for polyadenylation. Asterisk (*) indicates the position of the termination codon. Sequences with gray background are two polyserine regions. Bold underlines indicate the four RXXR consensus cleavage sites, the DGXR motif and the GL/ICG motif, respectively. C alphabets indicate cysteineat C-terminus**.**

```
    ACCAACTGGGAACCGGCCGGCAACCAGATGGGACCTTTCATGGCGCGCTCTTCGGTCAGC 780
241 T  N  W  E  P  A  G  N  Q  M  G  P  F  M  A  R  S  S  V  S
    CGTGTTATCATCTCGGGAAAAATAAACAAGTACACTATCCAATCGTCCGTAACGACAAAC 840
261 R  V  I  I  S  G  K  I  N  K  Y  T  I  Q  S  S  V  T  T  N
    AAGGTAATCCTGGCTCCTCACCTGTACAACAAGCAGAAAGGTATCGTGGCTAGTAGGATG 900
281 K  V  I  L  A  P  H  L  Y  N  K  Q  K  G  I  V  A  S  R  M
    AACCTAACTCTGGAGTCGGTGCAGCCGGCACAAGGCCAACCTCAAAATATTCCCGAGCCC 960
301 N  L  T  L  E  S  V  Q  P  A  Q  G  Q  P  Q  N  I  P  E  P
    AGGACCATTAAGGATCTCGTCTACGAGTACAACGCTGCCTCAGAAGACGCACACAAACAG 1020
321 R  T  I  K  D  L  V  Y  E  Y  N  A  A  S  E  D  A  H  K  Q
    GGACACGAAATCGGCCAAGACAGCAGCAGTTCCAGCAGCTCGGAGTCTAGCGAATCCTCC 1080
341 G  H  E  I  G  Q  D  S  S  S  S  S  S  S  E  S  S  E  S  S
    TCCTCCAGCAGCAGCAGCAGTAGCTCCTCTTCGAGCGAAAGTTCCCAGGAACAGCAAGCG 1140
361 S  S  S  S  S  S  S  S  S  S  S  S  E  S  S  Q  E  Q  Q  A
    CAAAAGCCGAAGCATCCCAGAAACAGGAGGGAGGCAGAAGATTCGTCATCATCGAGCAGC 1200
381 Q  K  P  K  H  P  R  N  R  R  E  A  E  D  S  S  S  S  S  S
    AGCAGCGAGAGCGATAGCTCTACTTCGAGCGACGACGACTACAGCTCATGTTCATCGTCC 1260
401 S  S  E  S  D  S  ST  S  S  D  D  D  Y  S  S  C  S  S  S
    TCGTCATCTTCCTCATCTTCCTCTTCGTCTACCTCTACCTCTTCGTCCTCGTCCTCTTCC 1320
421 S  S  S  S  S  S  S  S  S  S  T  S  T  S  S  S  S  S  S  S
    TCCAGTTCATCGAGCGACAGTGACTCCACTAGCCGCAGCAGTAGCGAAGAGTACTACCAG 1380
441 S  S  S  S  S  D  S  D  S  T  S  R  S  S  S  E  E  Y  Y  Q
    CCGAGACCCAAGCTCCGTATTCCTCCCAACTCGCCACTCCTGCCGTTCTTTATCGGTTAC 1440
461 P  R  P  K  L  R  I  P  P  N  S  P  L  L  P  F  F  I  G  Y
    CAAGGCAACTCGATCCAGGCGGCCCATAAAATTAATGTCGTACAACAGGCTCAGGAATTA 1500
481 Q  G  N  S  I  Q  A  A  H  K  I  N  V  V  Q  Q  A  Q  E  L
    GCGAAAGAAATCGCTCTCGACGTCCAGACGCCGAACGCTATCACCGGGGAAGACACCCTC 1560
501 A  K  E  I  A  L  D  V  Q  T  P  N  A  I  T  G  E  D  T  L
    ACCAAATTCACAGTTTTGATCCGAGCATTGCGCACGATGAGTCCCGAGCAGATCAAAGAG 1620
521 T  K  F  T  V  L  I  R  A  L  R  T  M  S  P  E  Q  I  K  E
    GCGGCTCAACAGCTCTACTTCCCGGCTCACAAAGCCTCCAACCACACAGTTACCGACGCC 1680
541 A  A  Q  Q  L  Y  F  P  A  H  K  A  S  N  H  T  V  T  D  A
    AAGAAATACCAAGCGTGGGCTGTCTACCGTGACGCAGTGGCGCAGGCCGGCACAGGTCCA 1740
561 K  K  Y  Q  A  W  A  V  Y  R  D  A  V  A  Q  A  G  T  G  P
    GCGTTGGTCGTTATTCAGGATTGGATCAGGAACGAGAGAGTGGTCCGGGAGAGAGCGGCC 1800
581 A  L  V  V  I  Q  D  W  I  R  N  E  R  V  V  R  E  R  A  A
    CAGTTAGTAGCCACACTCCAGGACAGTGCACGTCACCCGACCCTAGAGTACATGAACACA 1860
601 Q  L  V  A  T  L  Q  D  S  A  R  H  P  T  L  E  Y  M  N  T
    TTCTTCGAACTCGTAAAAAGCCCTCAGGTCATGAAGCAGCTGTACTTGAACACGAGTGCT 1920
621 F  F  E  L  V  K  S  P  Q  V  M  K  Q  L  Y  L  N  T  S  A
```

Figure 25 (Continued)

```
    CTGATTTCTTTCACGCACCTCGTCAGGAAGGCTATGGTCAACAACGAAACCGTGCACCAC 1980
 641 L  I  S  F  T  H  L  V  R  K  A  M  V  N  N  E  T  V  H  H
    CGTTACCCCGTTCACGCCTACGGCCGGCTGACCCGCAAGAACGCGACTGTCGTCGTTCTG 2040
 661 R  Y  P  V  H  A  Y  G  R  L  T  R  K  N  A  T  V  V  V  L
    GAGAAATACATTCCTTACTTGGCCGAAAAATTGAGGAAGGCGGTCGAGGAGGAGGACAGT 2100
 681 E  K  Y  I  P  Y  L  A  E  K  L  R  K  A  V  E  E  E  D  S
    CCCAAGATTCAGGTTTACGCTCAGGCCCTCGGTAATATCGGCCACTACAAAATTTTGAGC 2160
 701 P  K  I  Q  V  Y  A  Q  A  L  G  N  I  G  H  Y  K  I  L  S
    GCCTTCGAGGACTACTTGGAAGGTAAGGTGAACGTGACCGACTATCAGCGTTTGGTGATG 2220
 721 A  F  E  D  Y  L  E  G  K  V  N  V  T  D  Y  Q  R  L  V  M
    GTGACCGCACTTCATAAGCTGACTGTGGTATACCCGAAGAGAGCTCTACCGGTTCTGTAC 2280
 741 V  T  A  L  H  K  L  T  V  V  Y  P  K  R  A  L  P  V  L  Y
    AAGATCTACCAGAATATAGGAGAAACGCCCGAGGTGAGAGTGGCGGCGGTCATGCTGATC 2360
 761 K  I  Y  Q  N  I  G  E  T  P  E  V  R  V  A  A  V  M  L  I
   ATGAGGACCAACCCACCGGCCCAGATCCTTCAGCGTATGGCAGAGTCGACCAGGTTCGAT 2420
 781 M  R  T  N  P  P  A  Q  I  L  Q  R  M  A  E  S  T  R  F  D
    CACAGTATCGACGTGAGGTCGGCCATTCAATCGGCCATTCAGAGTGCGGCGAAACTTACC 2480
 801 H  S  I  D  V  R  S  A  I  Q  S  A  I  Q  S  A  A  K  L  T
    GGACCCAAATACTACCAACTGTCCCAGAACGCCAAGGCCGCAGTTCACATGCTGAACCCG 2540
 821 G  P  K  Y  Y  Q  L  S  Q  N  A  K  A  A  V  H  M  L  N  P
    ACCCAATACGGATACCACTACTCGAAACACCATATGAGGTCGTACATCGTGGAGCAACAG 2600
 841 T  Q  Y  G  Y  H  Y  S  K  H  H  M  R  S  Y  I  V  E  Q  Q
    AACTTGGCGTACAAACACGACATATCCCTGATCCAGAGCAGCGACAGCATAATCCCCAGC 2560
 861 N  L  A  Y  K  H  D  I  S  L  I  Q  S  S  D  S  I  I  P  S
    ACTGTATACTACAACTTGAAACGAAAGCTCGGAGGATACAGGACTCAGGTGTTCAAGATG 2720
 881 T  V  Y  Y  N  L  K  R  K  L  G  G  Y  R  T  Q  V  F  K  M
    GCCTACATGACATCCAGCGTTGACGACCTGTTGGAATTGATTTTCGATCAGTTCGAAGAC 2780
 901 A  Y  M  T  S  S  V  D  D  L  L  E  L  I  F  D  Q  F  E  D
    GAAGAGCCGGAACATAAGTCGAAGCGTCCCTCCCACGGACCCGAAGAATGGACCCTGCAA 2840
 921 E  E  P  E  H  K  S  K  R  P  S  H  G  P  E  E  W  T  L  Q
    AAGATCAAGGAAATCCTGGATTTGGACGTCGATGTACAAGAGCAGGTCGAGGGCAATTTA 2900
 941 K  I  K  E  I  L  D  L  D  V  D  V  Q  E  Q  V  E  G  N  L
    CAAGCGACGTTGTTCGGCAGCAAGAGGTTCATGGCCTGGGACAACCATTCCATCGAGGCC 2960
 961 Q  A  T  L  F  G  S  K  R  F  M  A  W  D  N  H  S  I  E  A
    CTCCCTTACATCGTGAAGGAGGCGGCGAAATCACTGAAGGACGGCCAACCGTTCAACATG 3020
 981 L  P  Y  I  V  K  E  A  A  K  S  L  K  D  G  Q  P  F  N  M
    ACCAAGTGGTACAATCCTCTGAACATCGAGTTAGCTTTCCCGATGGCCACCGGTCTACCG 3080
1001 T  K  W  Y  N  P  L  N  I  E  L  A  F  P  M  A  T  G  L  P
    TTCGTGTACACCTACCGCACTCCCACTTACTTCTCTATCGGAGGCGAAGTGCGCGCCAAA 3140
1021 F  V  Y  T  Y  R  T  P  T  Y  F  S  I  G  G  E  V  R  A  K
```

Figure 25 (Continued)

```
     ACCACTCCGGATCTCGCTAAGGGAGACGACGATGAAATCAAAATTCCCGACACCGTGAAT 3200
1041 T  T  P  D  L  A  K  G  D  D  D  E  I  K  I  P  D  T  V  N
     GCCACCATTCAAGCCCGAGTGATATATTCGACCAAGACTCAGGCTACCATGGGATTCGTC 3260
1061 A  T  I  Q  A  R  V  I  Y  S  T  K  T  Q  A  T  M  G  F  V
     ACCCCGTTCAACCATCAACGTTACATCGCCGGTCTGGATAGGAATATCCACGTATACCTC 3320
1081 T  P  F  N  H  Q  R  Y  I  A  G  L  D  R  N  I  H  V  Y  L
     CCAATTAGCGGTAAGCTCGATGTGGACATCGAGAACACCAAGCTCTGGGCTACTCTGAAC 3380
1101 P  I  S  G  K  L  D  V  D  I  E  N  T  K  L  W  A  T  L  N
     CCGCTGAGGCCCAACAAGTATCAGAAATTGTTGGAGTACAGTACCGTCGCTTATACCACC 3440
1121 P  L  R  P  N  K  Y  Q  K  L  L  E  Y  S  T  V  A  Y  T  T
     AGGCACGAAATCCTCGACCTGAGCCCACCGGCACAAAAAAACAAAACGGAAGAAATCCAC 3500
1141 R  H  E  I  L  D  L  S  P  P  A  Q  K  N  K  T  E  E  I  H
     GTCCGACCACCCATGAGGGACTCACCGACTGTCGGACAGCACAGCACAGGATTCGCTTAC 3560
1161 V  R  P  P  M  R  D  S  P  T  V  G  Q  H  S  T  G  F  A  Y
     GATCTGAAGGCGGAAACCGAGAAAGACGTCTTGGAATTGTCTACCGTTTACAACGCACTT 3620
1181 D  L  K  A  E  T  E  K  D  V  L  E  L  S  T  V  Y  N  A  L
     CAGAGGCACGACATCGTGTCCGCCCTTCTCTACTCTACTCACGAGATGAGTATCACCAAC 3680
1201 Q  R  H  D  I  V  S  A  L  L  Y  S  T  H  E  M  S  I  T  N
     AGTAACTTCTCAGTTGCCTACAACCCCCAGAAGACCAGTGCCAAGGCTGCCAAAATAGTT 3740
1221 S  N  F  S  V  A  Y  N  P  Q  K  T  S  A  K  A  A  K  I  V
     TTGACTTACGACGATGACGACGATGATGAATCATCCTCTGGGAAGAAGCCCAGGAGTCAC 3800
1241 L  T  Y  D  D  D  D  D  D  E  S  S  S  G  K  K  P  R  S  H
     AACAAACATAGGGGTGCTAACATCGCGTACCCGATTTCCACCGATCCTGATAGCGCCGAA 3860
1261 N  K  H  R  G  A  N  I  A  Y  P  I  S  T  D  P  D  S  A  E
     CGCCAAGAGCAATACCTGGATGCGGTCGGAGAAGGAATTGAAGACAGTTTGGCACGCATG 3920
1281 R  Q  E  Q  Y  L  D  A  V  G  E  G  I  E  D  S  L  A  R  M
     ATCGATGTATCGGTTCAATTCGAGGGCGAATCCAAGGCCGAGTACGTGGCGACCGCTGCC 3980
1301 I  D  V  S  V  Q  F  E  G  E  S  K  A  E  Y  V  A  T  A  A
     TACGCGGACAGCCCCGTGTCCAATTATAGCAGGTGGTTGTTGTTCCTCTCGATGGACCCG 4040
1321 Y  A  D  S  P  V  S  N  Y  S  R  W  L  L  F  L  S  M  D  P
     GCCCAGAAGACACAGCAGACCAAACCATTCGAGCTGACATTGAACGTCACCACCGACTTC 4100
1341 A  Q  K  T  Q  Q  T  K  P  F  E  L  T  L  N  V  T  T  D  F
     CCAGATGTACCCGTCCTGAACTACAAGAAGGCACTCCGCGCAGATCCTACCTCCTCAGGG 4160
1361 P  D  V  P  V  L  N  Y  K  K  A  L  R  A  D  P  T  S  S  G
     TCTTCCGCACACCTCAACTTCGGTGAAAACGCCCAACACGGAGCATACGTCTACCTAGAG 4220
1381 S  S  A  H  L  N  F  G  E  N  A  Q  H  G  A  Y  V  Y  L  E
     GGCACACTCGAACAAACTAAACCGAGGATTCAGTACATCCAGAGGCATCCTCTTGCCAAG 4280
1401 G  T  L  E  Q  T  K  P  R  I  Q  Y  I  Q  R  H  P  L  A  K
     CTTTGCGAACAACAGATGGAGGAAGGACACAACATACTTCCTGCTTGCCGCAATATCACA 4340
```

Figure 25 (Continued)

```
1421 L  C  E  Q  Q  M  E  E  G  H  N  I  L  P  A  C  R  N  I  T
     ATGAGGGCCAATTTCCTTGATAGTTATAGCTTCGATGTTGAATACAAGGACGTTCCCGAG 4400
1441 M  R  A  N  F  L  D  S  Y  S  F  D  V  E  Y  K  D  V  P  E
     TACGCTAAGAACCTCAGTTACCATGCCTATGCCATGATTCGCCACATGTTCTACCCGTAC 4460
1461 Y  A  K  N  L  S  Y  H  A  Y  A  M  I  R  H  M  F  Y  P  Y
     GTGTCTGAGAACTTCATCGACCCCGAGAACGAGCACGGTAAACTGTCGGTGGATGTGGAC 4520
1481 V  S  E  N  F  I  D  P  E  N  E  H  G  K  L  S  V  D  V  D
     TTCGCTCCCAACTTGAGGTCGGTCAACGTGTCCATCGATACTCCTCTGTTTTCATCCGAG 4580
1501 F  A  P  N  L  R  S  V  N  V  S  I  D  T  P  L  F  S  S  E
     TTCAGGAACGTCAGTGTTCACCCTTGGGTCCGACCTATCGTCGTCTCTCATCCCGAATAC 4640
1521 F  R  N  V  S  V  H  P  W  V  R  P  I  V  V  S  H  P  E  Y
     GACGCTTACGACAGATTCGTTTATAAGGCATACAGAGCTCAGTACTTCCCTGTTTGCGCC 4700
1541 D  A  Y  D  R  F  V  Y  K  A  Y  R  A  Q  Y  F  P  V  C  A
     GTTGAAAAATCCCAGGCGACTACTTTCGACAACAAAACCTATCCGATCCGCCTCGGTAAC 4760
1561 V  E  K  S  Q  A  T  T  F  D  N  K  T  Y  P  I  R  L  G  N
     TGCTGGCACGTGATGGCCTACCACACCCCCGACGAGAGTCCCGAGTCTTCCGAGGAGGAG 4820
1581 C  W  H  V  M  A  Y  H  T  P  D  E  S  P  E  S  S  E  E  E
     GACGAAGACGAGGCTGAATTCGCCGTTCTCGTACGCGACGTTTCCTCCAACAGGAAGGAA 4880
1601 D  E  D  E  A  E  F  A  V  L  V  R  D  V  S  S  N  R  K  E
     GTGAGAGTTATTCTGGGAGACGACATAATCGAGTTACAGCCTAGTGGACAGAGGCCGAAC 4940
1621 V  R  V  I  L  G  D  D  I  I  E  L  Q  P  S  G  Q  R  P  N
     GTAATTGTGAACGGAGAGAAGATCAAATACGCTCGGGATAGTTTGGCCGAAGTGGAGGAT 5000
1641 V  I  V  N  G  E  K  I  K  Y  A  R  D  S  L  A  E  V  E  D
     GCGGAGGGCGAAACTCTGTTGGAGGTATACGCACTACCGGCGGGAGGCGTGAGGATGTAC 5060
1661 A  E  G  E  T  L  L  E  V  Y  A  L  P  A  G  G  V  R  M  Y
     GCCCCTCAACACGTGCTCGAGATCCTCTACGACGGGGAACGGGTGAAGCTGCACGCCAGC 5120
1681 A  P  Q  H  V  L  E  I  L  Y  D  G  E  R  V  K  L  H  A  S
     AACTGGTACAGGGACGAGATGCGAGGACTCTGCGGCACCTTCGACGGAGAAAAGGTCACC 5180
1701 N  W  Y  R  D  E  M  R  G  L  C  G  T  F  D  G  E  K  V  T
     GACTTCTTGGCCCCGAGAAACTGCATCCTGAAGAATCCGTCCCTATTCGTAGCCAGCTAC 5240
1721 D  F  L  A  P  R  N  C  I  L  K  N  P  S  L  F  V  A  S  Y
     TCCCTACCGGGGCACACCTGCCACCACCCCGCCGCCGAGGAACTGCGCCGCAAGGCACAG 5300
1741 S  L  P  G  H  T  C  H  H  P  A  A  E  E  L  R  R  K  A  Q
     AACGCCCCCTGCTACCAACCCGAAGTGATCCTCGGAGATGTGGTGAGCGAGCGAGATCTC 5360
1761 N  A  P  C  Y  Q  P  E  V  I  L  G  D  V  V  S  E  R  D  L
     GGTAAGACGAAGAAGCAGCCCAAGCCCCAGTGGAGCAACTCTCTGGAGAGGCCTGACAGC 5420
1781 G  K  T  K  K  Q  P  K  P  Q  W  S  N  S  L  E  R  P  D  S
     AAGAGGGCGTGCACCCACTACAAGGTGAAGGTGCTGGAGGTGGGCAAGAAGATGTGCTTC 5480
1801 K  R  A  C  T  H  Y  K  V  K  V  L  E  V  G  K  K  M  C  F
     AGCTTGAGGGCGCACATCGCCTGCGGTGCCGGATGCCAGCCGTCCGGCAGGATCGAGAAG 5520
1821 S  L  R  A  H  I  A  C  G  A  G  C  Q  P  S  G  R  I  E  K
```

Figure 25 (Continued)

```
      AAGGTGGAGTTCCACTGTGTGGAGAGGAGTGACGCCGCCCGCCACTGGGCCGACAGCATC 5580
1841  K  V  E  F  H  C  V  E  R  S  D  A  A  R  H  W  A  D  S  I
      GCCAAGGGACACAACCCCGACTTCAGCAAGAAGCAGCCCAACTACAGGGCCTCCATCAGG 5640
1861  A  K  G  H  N  P  D  F  S  K  K  Q  P  N  Y  R  A  S  I  R
      CTACCGGAGAAATGCATTCCTCATTGAAGAATTCACTGTGGAGCTATTCCTTAGGCATTT 5700
1881  L  P  E  K  C  I  P  H  *
      AATATAGTTTGTAAACTTGAAGCATCTCAAAAATCCCAAAAAAATAAATATTATGATTTT 5820
      GATATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 25 (Continued)

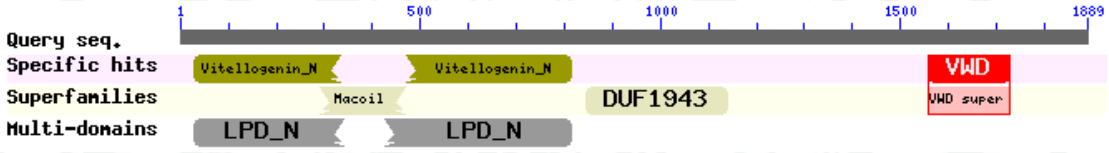


Figure 26 Diagram shows conserved domain structure of deduced amino acid sequence

3.5 Phylogenetic analysis

To clarify the evolutionary relationship of LiVg, we used a Neighbor Joining and Maximum Likelihood method. In this study, 39 Vg amino acid sequences represent 38 insect and one vertebrate species. The vertebrate Vg from *Gallus gallus* was used as an out group. All the sequences were aligned with ClustalW (ver. 1.83) by the multiple alignment method. Gaps and missing data were completely excluded from the analysis. The generated data matrix was converted to mega format and analyzed with MEGA 5.03 program. The phylogenetic relationships show that LiVg was grouped with most Hemiptera Vgs and closed to the Vg of *L. deyrollei*. The dendrogram obtained places the LiVgwith the other insects and particularly with Hemipterans as a distinct cluster. In addition, insect Vgs were separated into two major clades when analyzed by Neighbor Joining method. One clade consists of only all Lepidopteran Vgs, whereas the remaining insect Vgs were grouped together in the other large clade. This large clade consist of Hymenopteran, Lepidopteran, Colepteran, Dipteran, and most Hemipteran Vgs, except one Hemipteran Vg

(*Graptopsaltria nigrofuscata*) was separated in the small clade, that conclude with Dictyopteran Vgs.The topology tree obtained from Neighbor Joining method was different from Maximum Likelihood method. From this method, insect Vgs were separated into many clades. Dictyoptera and *G. nigrofuscata*, belongs Hemiptera were separated first group and major group consists of all remaining Vgs of Lepidopteran, Dipteran, Coleopteran, Hymenoptran and Hemipteran. Then, Dipteran and Hemipteran Vg were separated, respectively. The last Vg groups were grouped, are Coleopteran, Lepidopteran and Hymenopteran and Coleopteran Vg more similar to Lepidopteran than Hymenopteran Vgs.

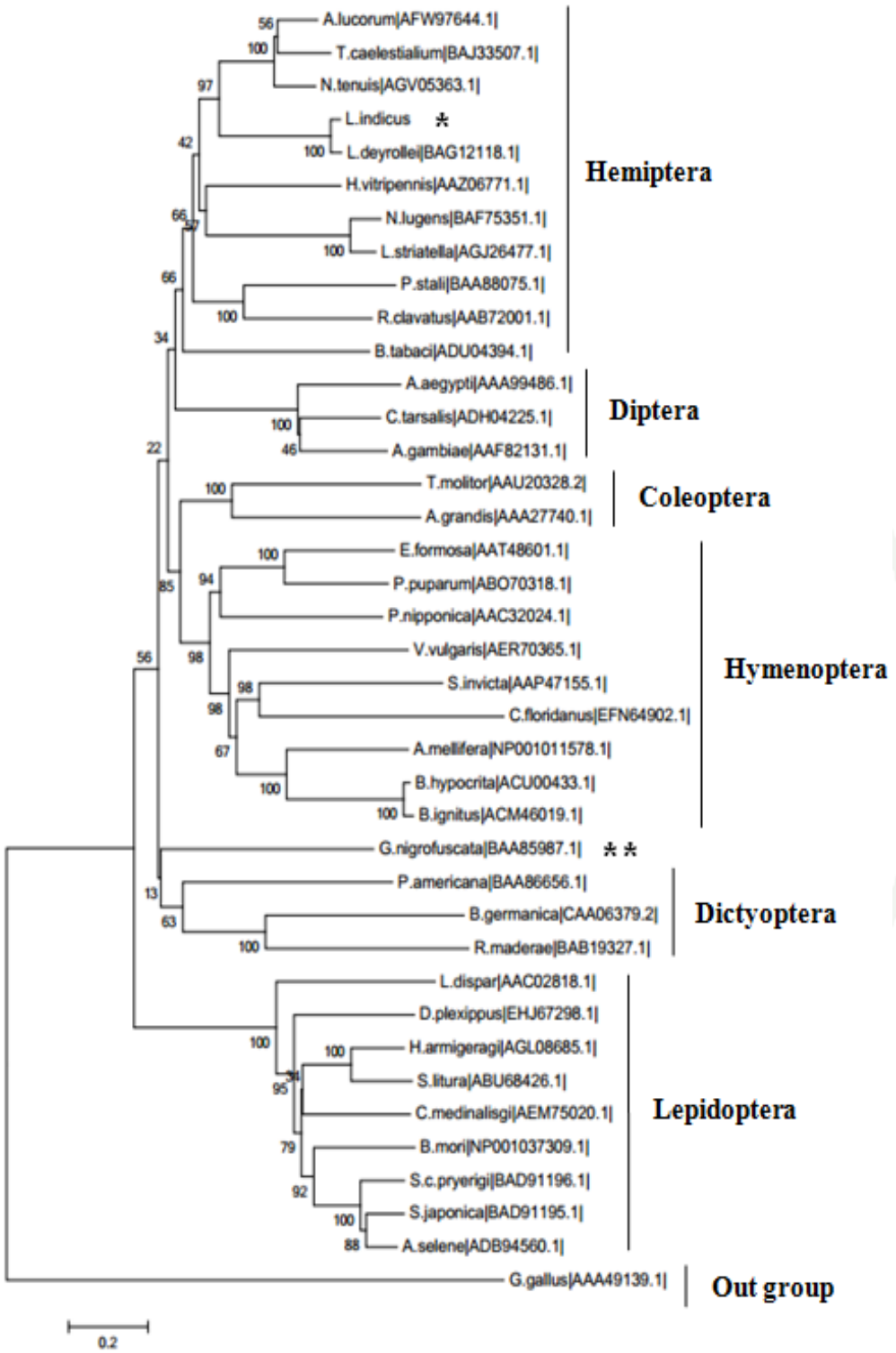


Figure 27 Phylogenetic analysis of the Vg amino acid sequences of insect using MEGA 5.03 program, Neighbor Joining method. *G. gallus* was used as outgroup. The numbers above the branches specify bootstrap percentages (1,000 replicates). (* in this study, ** Hemiptera species)

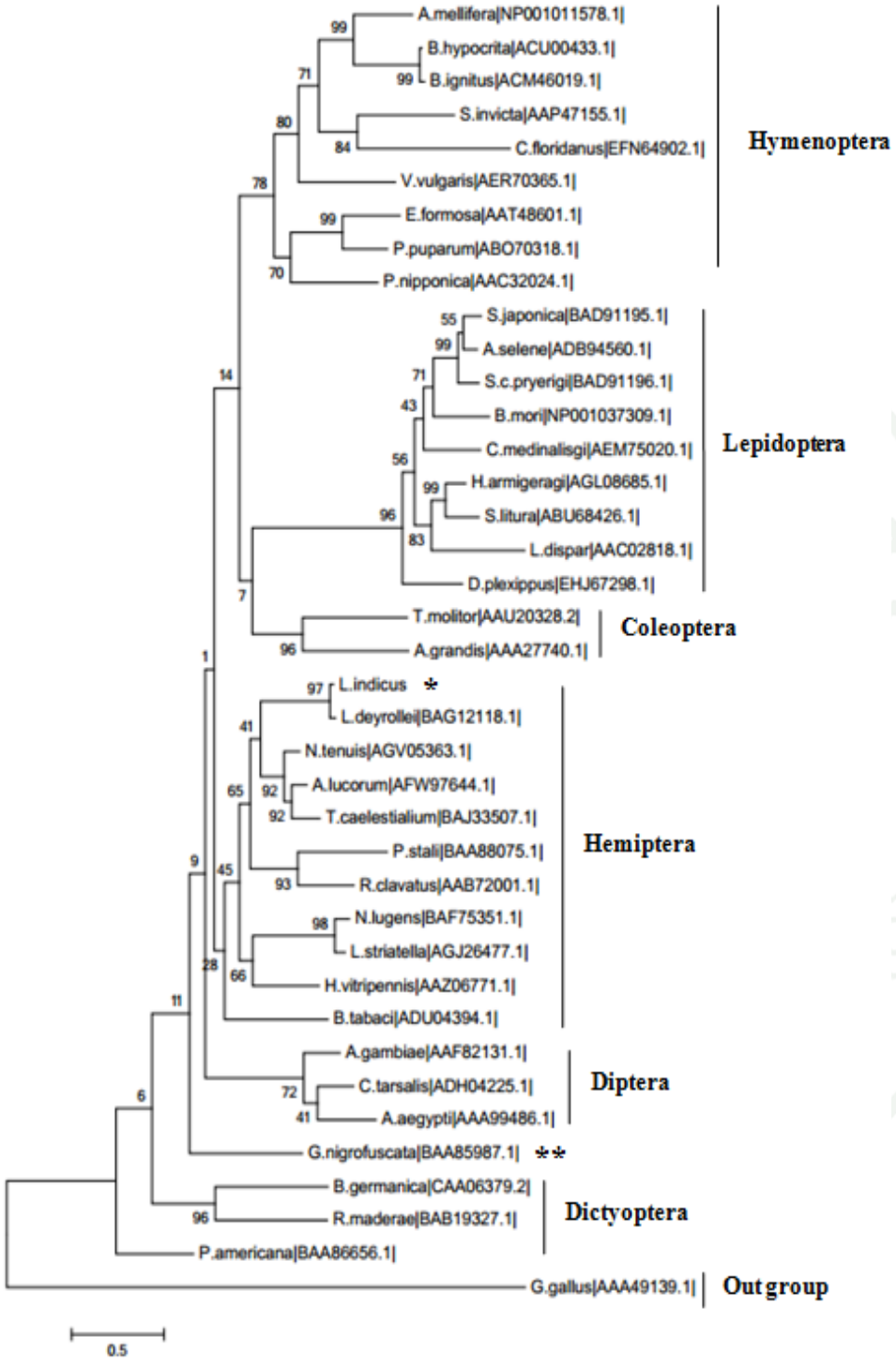


Figure 28 Phylogenetic analysis of the Vg amino acid sequences of insect using MEGA 5.03 program, Maximum Likelihood method. *G. gallus* was used as outgroup. The numbers above the branches specify bootstrap percentages (1,000 replicates). (* in this study, ** Hemiptera species)

3.6 *LiVg* gene expression analysis using RT-PCR

We used RT-PCR to determine the expression patterns of *LiVg* gene in adult male and female in spawning season.Total RNAexpression used specific primer: forward primer (Vg_RT_F1) and reverses primer sequences (Vg_RT_R1)to amplify a ~1,200 bp fragment from fat body tissue. The PCR products were clearly when analyzed by agarose gel electrophoresis. Result showed that the expression of *LiVg* gene was not detected in male (Figure 29).

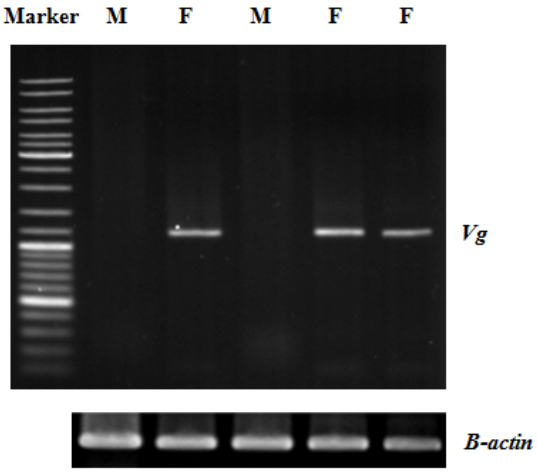


Figure 29 Reverse transcription-PCR was used to amplied *LiVg* gene expression. The upper panel shows amplification with *Vg*-specific primers using cDNA from fat body of adult male (M) and female (F). *β-actin* gene (lower panel) was used as a positive control.

Discussion

1. Chromosome analysis

In this study, the meiotic chromosomes from testicular cells of *L. indicus* were analyzed. Like other insects of Heteroptera, *L. indicus* possess holocentric chromosomes (chromosome without localized centromeres). Therefore, it is impossible to describe the morphology of chromosomes (Papeschi and Bressa, 2006). Two results

of major interest were obtained in the present study. One was the presence of sex chromosomes and autosomal univalents in some testicular cells and the other one was the clear presentation of m-chromosome existence in the family Belostomatidae.

1.1 Chromosome complement

The chromosome complement of *L. indicus* collected from Thailand, Loas, Myanmar and Cambodiawas 2n = 22A + neo-XY + 2m. In mitotic testicular cells, the two biggest chromosomes and the two smallest ones were obviously distinguished from the remaining autosomes. In mitotic cells stained with DAPI, the biggest chromosomes contained three DAPI-bright bodies, but their patterns were different. Prominently, one of the two biggest chromosomes had a DAPI-bright body at its end, while the other one had big DAPI-bright bodies in the middle region. The two biggest chromosomes were sex chromosomes, neo-Xand neo-Y. It was difficult to point out which one was neo-X or neo-Y because their sizes were almost the same. However, in meiotic cells at diplotene stage, the neo-XY bivalent was clearly seen to be composed of two chromosomes joined together, with one side bigger than the other. It is likely that the bigger one may have been the neo-X and the smaller the neo-Y. On the smaller side of the neo-XY bivalent, two DAPI-bright bodies were observed; one DAPI-bright body was located at the end. On the bigger side, two DAPI-bright bodies were located down the end. Which chromosome is neo-X and which is neo-Y can be determined by comparing the DAPI-bright body pattern on the sex chromosomes in mitotic cells and in meiotic cells.Hence, in Figure 12a, the chromosome with a DAPI-bright body located at its one end might be neo-Y and the other might be neo-X.

To determine the location of nucleolar organizing regions in the chromosome complement of *L. indicus*, silver nitrate staining was carried out. In metaphase I through telophase I, silver nitrate positive signal was found in all chromosome bivalents except the m-chromosome bivalent. The result was totally different from the previous reports of heteropteran insects. The location of nucleolar organizing regions in autosomal pairs was reported in some heteropteran insects such

as *Nezara viridula* (Camacho *et al*., 1985), *Pachylis argentines* (Papeschi *et al*., 2003), *Spartocera fusca* (Cattani and Papeschi, 2004) and *Limnogonus aduncus*, (Castanhole *et al*., 2008). Some species had nucleolar organizing regions located on sex chromosomes (Papeschi, 1995). Therefore, the location of nucleolar organizing regions in *L. indicus* should be investigated further using advent techniques.

1.2 Chromosome number change

The meiotic chromosome complement in testicular cells of *L. indicus* collected from Thailand, Loas, Myanmar and Cambodia was 2n = 22A + neo-XY + 2m. In previous reports, the chromosome complement of *L. Indicus* was 2n = 24A + neo-XY (Banerjee, 1958; Bagga, 1959; Jande, 1959). However, it is unknown whether the m-chromosome existed in the *L. indicus* chromosome complement reported previously, becausethe original papers could not be obtained and only the chromosome formula was presented (Papeschi and Bidau, 1985; Papeschi and Bressa, 2006). If the m-chromosomes existed and were defined as two small autosomes, the chromosome number of *L. indicus* obtained in this study and that of the previous study were the same. However, according to Papeschi and Bressa (2006), the m-chromosomes were absent in *L. indicus*. Therefore, there are two likely possibilities to explain the difference in chromosome number obtained from this study with the previous reports. First, the giant water bugs used in this study and the one studied previously might not be the same species. It is impossible to prove this possibility, becausethe previous one was studied long time ago. However, it has been reported that the giant water bugs distributed in Southeast Asia belong to the same species, *L. indicus* (Perez-Goodwyn, 2006). Genetic diversity studies of their mitochondrial genomes also support the report (unpublished data). Until now, the chromosome numbers of *Lethocerus* spp. have been studied in seven species (Papeschi and Bressa, 2006). Three species, *L. annulipes*, *L. griseus*, *L. melloleitaoi* De Carlo, have the same chromosome number, 2n = 28 = 26A + XY. In the previous report, the chromosome number of *L. indicus* was 2n = 26 = 24A + neoX-neoY. Together with the analysis of chromosome numbers, it was also suggested that the ancestral chromosome number of insects belonging to the family Belostomatidae was 2n = 26 + XY (Papeschi and

Bressa, 2006). Therefore, *L. indicus* should have had the same chromosome number as the above three *Lethocerus* species. The reduction of its chromosome number from 2n = 26 + XY to 2n = 24 + neo-XY occurred by the translocation of the X and Y chromosomes to one pair of autosomes, resulting in the neo-X and neo-Y chromosomes (Nokkala and Nokkala, 1999). The remaining three *Lethocerus* species, *L. americanus* Leidy (2n = 6+XY), *Lethocerus sp.* (2n = 2+ neoX-neoY) and *L. uhleri* (Montandon) (2n = ca. 30), had different chromosome numbers.

Second, the chromosome number difference might be due to the occurrence of chromosome polymorphism. The increase or decrease of chromosome numbers in insect species with holocentric chromosomes generally occurs by chromosome fragmentations and fusions (Ueshima, 1979; Nokkala *et al*., 2006). The results of the chromosome arrangements are presented as chromosome polymerphism in populations prior to the chromosome change being fixed in a population. In Heteroptera, chromosome polymorphism has been reported in some species. For example, in a species of *Belostoma*, two karyotypic types were found, asthe sex chromosomes showed to be polymorphic. In most individual males, the chromosome complement was 2n = 16 = 14 + XY,while some individuals had 2n = 17 = 14 +X1X2Y. It was proved that such chromosome polymorphism resulted from the breaking of the original X chromosome from the XY system into two unequal fragments, X1 and X2 (Papeschi, 1996). In *Belostoma plebejum*, the chromosome polymorphism was the result of autosomal fusion, so the decrease of the chromosome number from 2n = 14 + XY to 2n = 13 +XY was reported (Papeschi, 1994). In Homoptera, sex chromosome polymorphism was reported in *Cacapsylla* (Grozeva and Maryanska-Nadachowska, 1995). The karyotype of *Cacapsylla malimale* samples collected from populations in Finland and Russia was 2n = 22 + neo-XY, while the karyotype of samples collected from a population in Poland was 2n = 20 + neoX1X2Y. It is likely that chromosome polymorphism might occur in *L. indicus* populations from different areas because *L. indicus* is distributed from India to Southeastern Asia (Perez-Goodwyn, 2006). The *L. indicus* samples used in this study and the samples investigated in the previous report might have been collected from populations in different locations.

1.3 Origin of autosomal univalent chromosomes

In this study, the univalent chromosomes were found during the meiotic division in each testis, resulting in four cell types, A, B, C and D, with different univalent patterns. Type A cells with no univalent is the typical chromosome pattern of *L. indicus* used in this study. Type B and D cells contained two univalent autosomes since one autosomal pair did not form a bivalent. The presence of univalent autosomes was reported in some Heteroptera, such as *Myrmus miriformis* Fn. (Nokkala, 1985), *Calocoris quadripunctatus* (Nokkala, 1986), *Acanonicus hahni* (Papeshchi and Mola, 1990) and *Largus rufipennis* (Molaand Papeschi, 1993). The origin of univalent chromosomes is the result of homologous chromosomes asynapsis or desynapsis (White, 1973). The distinction between asynapsis and desynapsis is quite difficult to observe becauseboth processes occur at zygotene-pachytene, in which individual chromosomes are difficult to observe. In the case of asynapsis, pairing of homologous chromosomes is defective so the chromosomes fail to do synapsis at zygotene. In thecase of desynapsis, homologous chromosomes pair normally during zygotene, but chiasmata formation is defective, so the homologous chromosome undergoes desynaptic at late pachytene. John and Naylor (1961) suggested the occurrence of univalent chromosome formation might be caused by both genetics and environmental factors, but did not determine which one was the precise cause. For *A. hahni*, Papeshchi and Mola (1990) suggested univalent chromosomes did occur by desynapsis, becausethe univalent chromosomes were present in proximity and sometimes they were connected by a fine thread, which was the characteristic of desynapsis. In *L. rufipennis*, using the same characters as that of *A. hahni*, Mola and Papeschi (1993) suggested desynapsis to be the origin of autosomal univalents. With the above criteria, the origin of autosomal univalents found in type B and D cells may be desynapsis, because in some cells they were present close to each other. However, it was difficult to observe if autosomal univalents found in type B and D cells belongedto the same homologous couple, because most autosomes were similar in size.

1.4 Origin of neo-XY chromosome univalents

Four sex chromosome systems have been found in Heteroptera. There are XY/XX, XO/XX, different multiple,and neo-sex chromosome systems (Nokkala and Nokkala, 1999; Rebagliati *et al*., 2005; Papeschi and Bressa, 2006; Maryanska-Nadachowska *et al.*, 2008; Bressa *et al.*,2009; Grozeva *et al*., 2010). The first is the most commonly found system (71%), and the last is the most rare system, present in only seven species and subspecies of 1,600 species studied so far (0.4%), including *L. indicus* (Jande, 1959). The neo-XY system of *L. indicus* was the result of the translocation between the X-Y chromosomes and a pair of autosomes. The behavior of neo-XY chromosomes differed from that of the sex chromosome of other systems because the sex chromosomes, neo-X and neo-Y formed a bivalent at prophase I, while the sex chromosomes of other systems formed univalents. In this study, the neo-X and neo-Y chromosomes forming a bivalent were found in type A and B cells, whereas in type C and D cells, the neo-X and neo-Y chromosomes were present as univalents. The origin of the neo-X and neo-Y univalents may be inferred to be a result of desynapsis because of three reasons: (1) they werepresentin some type C cells at diplotene and diakinesis stages, (2) univalent neo-X and neo-Y chromosomes were always close to each other and (3) the fine thread linking the neo-X and neo-Y could be observed. The neo-X and neo-Y chromosomes have not been previously reported to form univalents.

1.5 Behavior of chromosomes

In this study, in type A cells with no univalents, all chromosomes, including autosomes, sex chromosomes and m-chromosomes, formed bivalents and divided prereductionally at anaphase I and segregated equaltionally at anaphase II. The meiotic behavior of autosomal bivalents did follow the rule of the order Heteroptera, in which the autosomal bivalents divide reductionally at anaphase I and segregate equaltionally at anaphase II (Ueshima, 1979; White, 1979). The general behavior of sex chromosomes during meiotic division of Heteropteran insects with XY/XX system instead followsthe rule that the sex chromosomes are achiasmatic and

form univalents. At anaphase I, the sex chromosomes segregate equationnally, in metaphase II they form a pseudobivalent and at anaphase II they divide reductionally (Ueshima, 1979; Suja *et al*., 2000). Since the sex chromosome system of *L. indicus* was the neo-XY system, the neo-X and neo-Y chromosomes did not follow this rule. On the contrary, in the neo-XY system, the sex chromosomes form a bivalent and divide exactly as autosomal bivalents, which divide prereductionally at meiosis I (White, 1973). In the case of autosomal univalents, they behave in the same way as sex chromosome univalents do. Therefore, the behavior of univalent chromosomes in type B, C and D cells were all the same, as they all divide equationally at anaphase I and segregated reductionally at anaphase II.

1.6 The existence of m-chromosome

The presence of an m-chromosome pair was claimed to be a characteristic of the family Belostomatidae (Ueshima 1979). However, cytogenetics studies in 27 species of this family revealed the absence of m-chromosomes in their chromosome complement (Papeschi and Bidau, 1985; Papeschi, 1996; Papeschi and Bressa, 2006). In this study, mitotic chromosomes of *L. indicus* showed two m-chromo-somes, so the existence of an m-chromosome pair in this familyis clear. The m-chromosomes were defined by their small size. The rule of m-chromosome behavior during meiosis I had been described as being usually chiasmatic chromosomes, not forming a bivalent, but in diakinesis they move closely to associate as ‘touch-and-go pairing in a form of pseudobivalent. In anaphase I, the m-chromosome pseudobivalent segregated reductionally and then divided equationally in anaphase II (Ueshima 1979; Gonzalez-Garcia *et al*., 1996; Suja *et al*., 2000; Papeschi and Bressa, 2006). However, the behavior of m-chromosomes in *L. indicus* represented another exception, because the m-chromosomes always appeared as a bivalent by associating distally during meiosis I in all cells. No univalent m-chromosome was observed. The presence of an m-chromosome bivalent behaving differently from the rule was reported in some species (Nokkala and Nokkala, 1983; Nokkala, 1986; Suja, 2000). In *Saldula orthochila* and *S. saltatoria*, their m-chromosomes were always present as bivalents, exactly like *L. indicus*. These m-chromosomes were not considered to be true m-chromosomes. The

appearance of an m-chromosome bivalent indicated that m-chromosomes behaved like other autosomes undergoing synapsis in prophase I to form a bivalent. Instead, the occurrence of m-chromosome univalents was the result of desynapsis, not of asynapsis (Nokkala and Nokkala, 1983).

2. Mitochondrial genome analysis

The sizes of most insect mitogenomes are ranging between 15 and 16 kbps. The differences are mainly depended on the non-coding regions (Beard *et al*., 1993). Within them, the length of the coding genes is relatively stable while the putative control region that is the largest non-coding region shows extremely significant variations both in length and in patterns (Liu and Liang, 2013). The 17,632 bp size of the whole *L. indicus* mitogenome is comparable to that reported for other hemipterans, which range from 14,496 bp in *Neomaskellia andropogonis* to 18,414 bp in *Trialeurodes vaporariorum* (Thao *et al*., 2004). Genome arrangement of the *L. indicus* mitogenome is as same as the ancestral insect (Boore, 1999) and is identical that of other published hemipteran mitogenome such as *Diplonychus rusticus* (Hua *et al*., 2009), *Triatoma dimidiata* (Dotson and Beaid, 2001), *Alloeorhynchus bakeri* (Li *et al*., 2012), *Agriosphodrus dohrni* (Li *et al*., 2011), *Philaenus spumarius* (Stewart and Beckenbach, 2005) and *Eusthenes cupreus* (Song *et al*., 2013).

The metazoan mitogenomes usually present a clear strand bias in nucleotide composition (Hassanin*, et al*., 2005, Hassanin, 2006), and the strand bias can be measured as AT- and GC-skews (Perna and Kocher, 1995). AT- and GC-skews of *L.indicus* mitogenomes are consistent to most of the strand biases of metazoan mtDNA (positive AT-skew and negative GC-skew for the J-strand). The average AT-skew of *L. indicus* mitogenome was 0.2859, out of ranging from 0.04 in *Hydaropsis longirostris* to 0.23 in *Leptopus* sp. The average GC-skew of true bug mt genomes was -0.18, ranging from -0.04 in *Yemmalysus parallelus* to -0.27 in *T. dimidiate* and the *Stenopirates* sp. mitogenome exhibited a marked GC-skew (-0.14) (*Li et al*., 2012). While species of three distantly related families of insects, i.e. Philopteridae (Phthiraptera), Aleyrodidae (Hemiptera) and Braconidae (Hymenoptera), have

positive GC-skew and negative AT-skew for the J-strand (Wei, *et al*., 2010). The underlying mechanism that leads to the strand bias has been generally related to replication, because this process has long been assumed to be asymmetric in the mtDNA and could therefore affect the occurrence of mutations between the two strands (Hassanin *et al*., 2005). It is possible that the overall genome A-bias is driven by mutational pressure on the N-strand and the GC-skew may be correlated with the asymmetric replication process of the mtDNA (Reyes *et al*., 1998).

Like other animal mitogenomes, the one of *L. indicus* mitogenome is compact and the overlapping genes reflect the high efficiency of nucleotides usages (Boore, 1999). In insects, two kinds of overlaps between mitochondrial genes can be observed (Kim *et al*., 2010): the first one occurs either between tRNAs or between tRNAs and rRNAs. These overlaps are processed through post-transcription processing (Boore, 1999). The second one occurs between PCGs. Three such overlapping regions are identified in the mitogenome of *L. indicus*: between *atp8* and *atp6*, *nad4* and *nad4L*, *nad6* and *cytb*, respectively. Interestingly, these overlappig are between a big (*atp6*, *nad4*, and *cytb*) and a small PCG (*atp8*, *nad4L*,and *nad6*). They also appear in the mitogenome of *D. plonychus* (Hua *et al.,* 2009) and other hemipteran species such as *Neomaskellia andropogonis* (Sternorrhyncha) (Thao *et al*., 2004) or *Geisha distinctissima* (Fulgoromorpha) (Song and Liang, 2009) or *P. spumarius* (Cercopidae) (Stewart and Beckenbach, 2005). The characteristic of these overlaps is the small gene open reading frames that are always located at the 5′ end of larger one. Additionally, these overlaps always locate at the 5′ end of the transcription and it is assumed that they might improve translation efficiency (Berthier *et al*., 1986). Similar to *L. indicus*, other hemipteran species also habour intergenic spacer sequences in the mitogenome. In most cases, the intergenic spacer sequences consist of only 1 or 2 bp. As for overlapping regions in the contiguous genes in other hemipterans, the total sizes range from 29 bp in *S. graminum* (Anstead *et al*., 2002) to 99 bp in *T. dimidiata* (Dotson and Beard, 2001, Thao *et al*., 2004). Another interesting aspect of *L. indicus* mitogenome is the 7-bp identical overlapping sequence at the junction of *atp8-atp6* genes and *nad4-nad4L* gene, which are often found in several hemipteran species. The reason for this phenomenon may be that these sequences are located at the same site

of the *atp8-atp6* gene cluster, and the *atp8* genes of these species have the similar transcriptional mechanism (Berthier *et al*., 1986). It has been proposed that the secondary structure of the transcribed polycistronic mRNA may facilitate cleavage between the proteins (Clary and Wolstenholme, 1985). For the *atp8-atp6* genes, because they are translated from a single mRNA, there is no cleavage of the mRNA for expression (Berthier *et al*., 1986). The same mechanism is applied to the *nad4L-nad4* genes (Berthier *et al*., 1986). The similar situations have been observed at the junctions of the same protein-coding genes in other hemipteran mitogenomes (e.g. in *P. spumarius* and *T. dimidiata*) (Stewart and Beckenbach, 2005, Dotson and Beard 2001) and the overlaps between *atp8-atp6* (7-bp) and *nad4-nad4L* gene (7-bp) are often been found across the Metazoa (Li *et al*., 2012, Stewart and Beckenbach, 2005, Carapelli *et al*., 2006).

The problematic translational start of the *cox1* locus has been extensively discussed in several arthropod species including insects (Caterino and Sperling, 1999; Wilson *et al*., 2000). It was postulated to be of tetranucleotide (ATAA, TTAA, and ATTA) or hexanucleotide (ATTTAA) nature.Whereas hemipteran species do not share this feature, for instance, *G. distinctissima*, *P. spumarius*, *T. dimidiata* and *Pachypsylla venusta* all use typical methionine (ATG) as the first amino acid for *coxI* (Song and Liang, 2009; Stewart and Beckenbach 2005; Dotson and Beard, 2001; Thao *et al*., 2004). This appears to suggest that such abnormalities are taxon-specific (Cha *et al*., 2007). For all PCGs of *L. indicus* are initiated by ATN codons, except for *coxI* and *nad4L* gene which is initiated by GTG and ACT, respectively.The *coxI* gene of *Dolycoris baccarum* (Zhang *et al*., 2013) and *E. cupreus* uses TTG as the initial codon but *nadI* gene uses GTG codon (Song *et al*., 2013), whereas, GTG uses for *nad4L* start codon of *Oncocephalus breviscutum* (Li *et al*., 2013). In *T. dimidiata*, the start codon for *nad1* is ATA, whereas, *A. dohrni* and *Valentia hoffmanni* have GTG as their start codon (Li *et al*., 2011). On the contrary, for *nad5*, *T. dimidiatesyarts* with GTG whereas, *A. dohrni* and *V. hoffmanni* use ATG and ATT as their respective start codons (Li *et al.,* 2011).The start codon for some *PCGs* genes of *B. afer* is GTG and TTG (Li *et al*., 2013), whereas, *Brachyrhynchus hsiaoi* is TTG (Li *et al*., 2014). For *coxI* of *A. bakeri* used TTG as start codon and had incomplete termination codons, TA

and T (Li *et al*., 2012). Incomplete termination codons (T or TA) are commonly found in many metazoan mitogenomes including that of insects (Bae *et al*., 2004; Cantatore *et al*., 1989). The common interpretation of this phenomenon is that TAA termini are created via posttranscriptional polyadenylation (Ojala *et al*., 1981; Gadaleta *et al*., 1989). Accordingly, partial stop codons have been observed in other hemipteran insects' mitogenomes. There seems to be a high degree of conservation of incomplete stop codons across the Hemiptera. For instance, *coxI*, *coxIII*, and *nad4* genes, they have incomplete termination codons (T) as stop signals, which are reported in *P. spumarius*, *T. dimidiata*, *P. venusta*, *Homalodisca coagulate*, *E. cupreus* and *Aradacanthia heissi* (Stewart and Beckenbach, 2005; Dotson and Beard, 2001; Nakabachi *et al*., 2010; Thao *et al*., 2004; Song *et al*., 2013; Shi *et al*., 2012).

In the *L. indicus* mitochondrial proteins, leucine, serine, phenylalanine, methionine, and isoleucine, the most frequently used amino acids are similarly, these amino acids are also frequently utilized in other hemipteran insects, for example, in *P. spumarius*, *T. dimidiate*, *E. cupreus* and *A. heissi* (Stewart and Beckenbach, 2005; Dotson and Beard, 2001;Song *et al*., 2013; Shi *et al*., 2012). The four most frequently used codons, ATA (methionine), TTT (phenylalanine), TTA (leucine), ATT (isoleucine) are all composed wholly of T and/or A. Like analysis of the codon usage reveals that codons with A or T in the third position are most prevalent in the mitogenome of *G. distinctissima*, *A. heissi* and *E. cupreus* (Song and Liang, 2009; Shi *et al*., 2012; Song *et al*., 2013). The four most frequently used codons, TTA (leucine), TTT (phenylalanine), ATT (isoleucine), and ATA (methionine), are all composed wholly of T and/or A and this implies that the amino acid composition was affected to a similar degree by the AT mutational bias (Song and Liang, 2009). Analysis of the codon usage reveals that codons with A or T in the third position are most prevalent in the mitogenome of *L. indicus*.

The putative control region is the most variable part of the mitogenome and shows complicated structures. No sequence similarity was found when blasting the putative control region sequences with the GenBank nucleotide records. For example, the small 310 bp long of the *Paphnutius ruficeps* mitogenome (Lui and Liang, 2013).

In the only other studied spittlebug *P. spumarius*, it is 1,835 bp (Stewart and Beckbach, 2005) in length while in the Sternorrhynchan *T. vaporariorum*, it is 3,725 bp (Thao *et al*., 2004). The shortest one 328 bp within Hemiptera was previously observed in *Neomaskellia andropogonis* (Thao *et al.*, 2004). In general, this region shows a high degree of variability among insects, from 4601 bp in *D. melanogaster* (Lewis *et at*., 1995) to only 70 bp in the Orthopteran *Ruspolia dubia* (Zhou *et al*., 2007). Boyce *et al*., (1989) reported that the bark weevil has a putative control region longer than 13 kbp, but so far no sequences are available in GenBank. Apparently, these changes occur on every systematic level, so no phylogenetic statement can be made. The 3,020 bp putative control region of *L. indicus* mitogenome is the main and largest non-coding region. There are five regions compost of two tandem repeat sequences (five of 174-bp and nine of 114-bp). The remaining region compose variable sequence that two large and one small. For large variable sequence consist of high and low GC content (36.9 % GC content of 740 bp and 17.6 % GC content of 210 bp) and small variable region consist of incomplete two tandem repeats sequence. Repeated sequences are common in the control region for most insects, and variable in length (Dotson and Beard, 2001). Consequently, analysis of the repeat units among individuals from different geographical location may shed light on the geographical structuring and phylogenetic relationships of species (Li *et al*., 2011).

Phylogenetic analysis, 42 mitochondrial genomes represent six higher superfamilies within the hemipteran insects and the concatenated amino acid sequences of the 13 PCGs genes. The topology tree obtains of taxa relationships of heteroptera is similar to previous report (Li *et al*., 2011; Li *et al*., 2012).

3. Cloning and sequencing of *Vg*gene in *L. indicus*

The *Vg* cDNA sequence of giant water bug, *LiVg* cDNA contained 6,048 bp with 5,664 bp of open reading frame which encodes a protein of 1,888 residues including an 18-residue putative signal peptide that similarity in the first investigation reporting a complete sequence for a Vg cDNA in a Lethocerus genus that first report is *Lethocerus deyrolli*, of 5,865 bp open reading frame which encodes a protein of

1,895 residues including an 19-residue putative signal peptide (Yoshiki *et al*., 2011). Several Vg mRNAs were isolated from insects, and these are mostly 6-7 kb long (Tufail *et al*., 2005; Tufail and Takeda, 2008). The molecular weight of primary LiVg protein is in the range of all insect Vgs that insects Vgs are large molecules (~200 kD) synthesized in the fat body in a process that involves substantial structural modifications (e.g., glycosylation, lipidation, phosphorylation and proteolytic cleavage, etc.) of the nascent protein prior to its secretion and transport to the ovaries (Tufail and Takeda, 2008). The exception is the yolk protein of higher Diptera that the molecular weight of their proteins are quite smaller, approximately 44-50 kDa (Tufail and Takeda, 2008).For example, the yolk protein precursor 1 (YP1) of *Drosophila melanogaster* was 48.7 kDa (Hovemann *et al*., 1981) and the vitellogenin-1 and vitellogenin-2 of *Ceratitis capitata* are 49 and 46 kDa, respectively (Rina and Mintzas, 1988). In most holometabolous insects, the molecular weight of their Vg primary precursors are approximately 200 kDa which are cleaved at the RXRR site into two polypeptides, one is small (49-60 kDa) and the other is large (140-190 kDa) (Tufail and Takeda, 2008). For example, the *Spodoptera litura* Vg precursor has 198.73 kDa which was cleaved into two subunits, 55.91 and 141.31 kDa (Shu *et al*., 2009) and the 210 kDa *A. yamamai* Vg precursor was cleaved into 184 and 42 kDa (Yokoyama *et al*., 1993). Its deduced amino acid sequence possessed two di-basic proteolytic processing sites R-X-K/R-R (Barr, 1991) that have also been found at subunit cleavage site of Vg's including *Anthonomus grandis* (Trewitt *et al*., 1992), *Aedes aegypti* (Chen *et al*., 1994), *B. mori* (Yano *et al*., 1994), *Lymantria dispar* (Hiremath *et al*., 1997) and *Riptortus clavatus* (Hirai *et al*., 1998). Even if the deduced LiVg contains four putative cleavage recognition sites (RNRR, RALR, RNER and RLTR), the 212.40 kDa LiVg precursor peptide should be cleaved at the RNRR site, yielding two subunits, small and large subunits. The small subunit with predicted 40.93 kDa derived from the amino acid sequences between the N-terminal excluded the signal peptide to the cleavage site, RNRR and the large subunit with 169.00 kDa was calculate from the first amino acid after the RNRR to the end of C-terminal sequences. The predicted molecular weights of two subunits were consistent with those of most insect Vgs. However, in *L. dispar*, the consensus RXXR sequence motif exists at the C-terminus. In hemimetabolous insects, where the Vg precursor is

cleaved into several subunits, RXXR motif was also determined near the C-terminus or at the center of the primary product (Hirai *et al*., 1998; Tufail *et al*., 2001, 2005, 2007; Tufail and Takeda, 2002). The recent molecular data showed that Vg molecules were cleaved in all insect species except Apocrita (higher Hymenoptera) where Vg gene product of~180 kDa remains uncleaved (Tufail & Takeda 2008). However, the Vgs of some hemimetabolous members (like cockroaches and the bean bug *Riptortus clavatus* Thunberg) are special, cleaved into more than two subunit polypeptides (Hirai*et al*. 1998; Tufail & Takeda 2002, 2008, 2009c). Such sites are recognized by subtilisin-like convertases (Rouille *et al*., 1995) which cleave the Vgsinto 2 or more subunits. Thompson *et al*. (2007) noticed that DvVn resolved on SDS-PAGE as seven subunits (210, 172,157, 111, 76.2, 58.7, and 50.8 kDa).

In insect, their Vgs contain a conserved GL/ICG domain near the C-terminus (Sappington and Raikhel, 1998; Tufail and Takeda, 2008) while Vgs of Crustacea contain a GLLG motif, and vertebrates have a longer TCGL/ICG motif (Mouchel *et al*., 1996). Comparison of the C-terminal part of all insect Vgs revealed a high conservation of cysteine residues at 9 positions following the GL/ICG motif. Asimilar but slightly longer motif (TCGLCG) was recognized in vertebrate Vgs (Mouchel et al., 1996), human von Willebrand factor (Baker, 1988), and the human intestinal mucine 2 glycoprotein (Gum *et al*., 1994). In addition, the DGXR motif isalso located 17–19 residues upstream of the GL/ICG motif in almost all (except *L.maderae*) insect Vg sequences. In *N. lugens*, *H. coagulate* and *S. invictaonly* arginine (R) is missing. The GL/ICG motif and cysteine residues are necessary for the oligomerization of vertebrate Vns (Mayadas and Wagner, 1992; Mouchel *et al*., 1996). One possible function of oligomerized insect Vns could be binding lipids (Hagedorn *et al*., 1998; Giorgi *et al*., 1999). It was proposed previously that the DG residues (of DGXR motif) together with GL/ICG motif and cysteine residues at conserved positions might form a structure necessary for Vns to function properly during embryogenesis (Tufail *et al*., 2001). For the DGXR motif, its amino acid sequence quite varied. The sequence of DGXR motif in LiVg was DGER which was the most often sequence found in insect Vgs. In addition, the DGXR motif is located 17 residues upstream of the GICG motif as present in all Lepidopteran Vgs, 18 residues in Hemipteran and

Hymenopteran Vgs, and 19 residues in Coleopteran Vgs. Another interesting conserved sequence at the C-terminal is the number of cystein residue. In this study, LiVg contained 18 residues, 10 cystein residues, two residues being at the N-terminal and the remaining eigtht located at C-terminal are found in all Lepidopteran Vgs. Moreover, some amino acid residues like D and E (acidic), P (hydrophobic), and Y (hydrophilic) were also highly conserved at the C-terminus of almost all Vgs. However, their specific role at this position remains unclear (Tufail and Yakeda, 2008).The number of conserved cystein residues counted from the DGXR motif until the C-terminus of insect Vg s could vary from 6-10 such as in 10 in all Hymenoptera (Li *et al*., 2010, Havukainen *et al*., 2011), 9 in giant water bug (Nagaba *et al*., 2011) recent study, 8 in parasitoid wasp (Donnell, 2004) and all Lepidopteran Vgs and 7 in *A. grandis* Vg. However, LiVg has 10 cystein residues. In vertebrate the GL/ICG motif and the conserve cysteins were suggested that they were necessary for the oligomerization Vitellins (Mayadas and Wagner, 1992, Mouchel *et al*., 1996), so these conserved sequences might do the same duty in insects.

Polyserine track is the prominentfeature of insect Vgs at the N-terminal (Tufail and Takeda, 2008). This property is considered to promote the solubility of Vg (Gerber-Huber et al., 1987) or chelation to essential metal ions (Taborsky 1991). Almost all Lepidopteran Vgs, two polyserine tracts are conserved at N-terminal (Tufail and Takeda, 2008), except *Lymantria dispar* Vg had no any polyserine tract (Hiremath and Lehtoma, 1997). The polyserine track at the C-terminal is often found in the Vgs of hemimetabolous insect such as mosquito, cockroach and brown planthopper (Romans *et al*., 1995, Tufail *et al*., 2010). In addition, in the between the two polyserine tracts, a conserved cleavage site, RXRR, was located except *Antherae apernyi* Vg and *Saturnia japonica* Vg had no the putative cleavage site flanked by two polyserine tracts (Liu *et al*., 2001, Meng *et al*., 2008). The polyserine regions were also suggested that they tend to be good sites for phosphorylation (Tufail and Takeda, 2008). Moreover, the role of the polyserine tracts was suggested that they might be phosphorylated and involved in the interaction between Vg and its receptor on the ovary membrane since the removal of phosphate caused the reduction in affinity between the Vg and its receptor (Dhadialla *et al*., 1992, Hirai *et al*., 1998). In

LiVg, there are 137 putative phosphorylation sites including all serine residues of the two polyserine tracks so this study could only support the suggaestion that the polyserine region is appropriate sites for phosphorylation.

Phylogenetic tree was performed using the deduced amino acid sequence of Vgs from *L. indicus* and some others insects available in the NCBI database. The vertebrate Vg from *Gallus gallus* was used as an out group. As expected, the result showed that LiVg was grouped with all Hemipteran Vgs and closed to the Vg of *L. deyrollei*. In addition, insect Vgs were separated into two major clades when analyzed by Neighbor-Joining method. One clade consists of only all Lepidopteran Vgs, whereas all the remaining insect Vgs were grouped together in the other clade. The topology of the tree obtained in this study was consistent with previous reports (Donnell, 2004, Meng *et al*., 2008, Tufail *et al*., 2010). Why the Lepidopteran Vgs differed from that of otherinsects? Multiple alignment of the C-terminal sequences of some insect Vgs available in the NCBI database showed some interesting regions that make them differ from other insect Vgs. One should be the conserved motifs GI/LCG in which all Lepidopteran Vgs possesses the GICG motif whereas all other insect harbor the GLCG motif. For all the remaining insect Vgs were grouped together in the other clade were divided into two clades, including only Dictyopteran Vgs and *G. nigrofuscata*, belongs Hemiptera and all remaining clade were separated into two groups; one group shown all Hemiperan closed to Dipteran Vgs and other group consist of Hymenopteran and Colepteran Vgs. These phylogenetic relations showed quite different result from the previous reports, Nose *et al*., (1997) constructed a phylogenetic tree using the complete amino acid sequences of Vgs from 6 insect species. The tree was in agreement with the accepted phylogenies based on morphological characteristics and ribosomal DNA sequences. Subsequently, Lee *et al*., (2000b) constructed a phylogenetic tree based on the complete amino acid sequences of 9 insects (3 hemimetabolous and 6 holometabolous); the results were similar to those reported by Nose *et al*. (1997). Tufail and Takeda (2002) reported a phylogenetic (Neighbour-joining) tree based on the entire sequences of 4 Vgs from three cockroach species (*P.americana*, *L. maderae* and *B. germanica*). The tree was in agreement with another molecular tree contructed based on the DNA sequences of 12

rRNA genes (Tufail and Takeda, 2002). The present phylogenetic tree (Figure 26 and 27), most Hemipteran Vgs were grouped in a lagre single clade. The exception was *G. nigrofuscata* seperated out of the clade. This result similar to the phylogenetic tree using 23 complete Vg sequences from 19 insect species by Tufail *et al.*, (2007) and the phylogenetic tree using 37 complete Vg sequences from 33 insect species by Tufail and Takeda (2014).

The polymerase chain reaction (RT-PCR) was used to determine the expression of *Livg* gene in adult male and female in spawning season. The *LiVg* mRNA was detected only in female as that of *L. deyrollei* (Nagaba *et al*., 2010), *G. nigrofuscata* (Lee *et al*., 2001) and *A*. *aegypti* (Martin *et al*., 2001). *Vg* gene is generally expressed in sex- and tissue-specific manners in most animal species, however, in some insects, *Vgs* express in both females and males such as ricemoth *Corcyra cephalonica* (Veerana *et al*., 2014), honey bee *Apis mellifera* (Piulachs *et al*., 2003) and *Rhodnius prolixus* (Valle *et al*., 1987).

CONCLUSIONS

1. Meiotic study

The giant water bug, *L. indicus* (Lepeletier and Serville) (Heteroptera: Belostomatidae), a native species of Southeast Asia, is one of the largest insects belonging to suborder Heteroptera. In this study, the meiotic chromosome of *L. indicus* was studied in insect samples collected from Thailand, Myanmar, Loas, and Cambodia. Testicular cells stained with lacto-aceto orcein, Giemsa, DAPI and silver nitrate were analyzed. The results revealed that the chromosome complement of *L.indicus* was 2n = 22A + neo-XY + 2m, which differed from that of previous reports. Each individual male contained testicular cells with three univalent patterns. The frequency of cells containing neo-XY chromosome univalent (~5%) was a bit higher than that of cells with autosomal univalents (~3%). Some cells (~0.5%) had both sex chromosome univalents and a pair of autosomal univalents. None of the m-chromosome univalents were observed during prophase I. In addition, this report presents clear evidence about the existence of m-chromosomes in Belostomatidae.

2. Mitochondrial genome

The complete mitogenome of *L. indicus*, (Hemiptera: Belostomitidae: Belostomatinae), was sequenced using long PCR-based approach. Like other animal mitogenomes, this mitogenome is a 17,632 bp circular molecular with a total A+T content of 70.51% and 69.39% for coding regions only. The entire genome contains 13 PCGs, 22 transfer RNA (*tRNA*) genes, 2 ribosomal RNA genes (*rrnS* and *rrnL*) and a putative control region. The major non coding region (the A+T rich region or a putative control region) was located between the small ribosomal subunit (*rrnS*) and $tRNA^{Ile}$- $tRNA^{Gln}$ - $tRNA^{Met}$ gene that comprises of two types of extensive tandem repeat. One type composses of five regions, four repeat regionsof 174 bp and one repeat region of 175 bp.The other type is a shorter 114 bp sequence aligned in nine tandem repeats. The gene content and order are consistent with *D. rustica* that is

common features found in hemipteran mitogenomes. The A+T rich region has similarity sequences with *Triatoma dimidiate.* The 3,020 bp putative control region, known as the A + T rich region in most insects, has an A+T bias of 71.36%. Of the 3,020 bp, the 2,070 bp tandem repeat sequence has 71.40% A +T content in that the A+T bias of the 114 bp repeat regions is lower (70.10%) than that of the 174 bp and 175 bp repeat regions (73.70%). Phylogenetic analysis shows that *L. indicus* is grouped with the insects in Belostomatidae.

3. *vitellogenin* gene

The full-length *LiVg* cDNA is 6,048 bp long, containing a 253 bp 5'-untranslated region (5'-UTR), a long 131 bp 3'-UTR, a stop codon (TGA) and a consensus polyadenylation signals sequence (AATAAA) at 76 nucleotides downstream of the terminal codon. The 5,664 bp open reading frame (ORF) encoded 1,880 amino acids with a predicted molecular mass of 212.40 kDa and theoretical isoeletric point of 6.58. The first 18 amino acids is a putative signal peptide. The LiVg contained vitellogenin-N domains, the domain of unknowns function (DUF1943) and the von Willebrand factor type D (VWD) domain. In addition, the protein contains two polyserine regions, four RXXR consensus cleavage sites (RNRR, RALR, RNER and RLTR), the DGXR motif, GL/ICG motif, and the cysteine residues conserved at the C-terminus. The *LiVg* expression was detected onlyin fat body females.

LITURATURE CITED

Anstead, J.A., J.D. Burd and K.A. Shufran. 2002. Mitochondrial DNA sequence divergence among *Schizaphis graminum* (Hemiptera: Aphididae) clones from cultivated and non-cultivated hosts: haplotypes and host associations. **Bull. Entomol. Res.** 92: 17-24.

Avise JC, J. Arnold, R.M. Ball, E. Bermingham, T. Lamb, J.E. Neigle, A.R. Carol and N.C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematic. **Annu. Rev. Ecol. Syst.** 18: 489-522.

Bae, J.S., I. Kim, H.D. Sohn and B.R. Jin. 2004. The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. **Mol. Phylogenet. Evol.** 32: 978-985.

Bagga, S. 1959. On the prereduction of the sex chromosomes during meiosis in a belostomatid *Lethocerus indicum*. Proceedings of the Zoological Society of Calcutta 12: 19-22. Cited in Papeschi AG, Bressa MJ. 2006. Evolutionary cytogenetics in Heteroptera. **J. Bio. Res.** 5: 3-21.

Baker,M.E. 1988. Invertrebrate vitellogenin is homologous to human von Willebrand factor. **J. Biochem.** 256: 1059-1061.

Banerjee, M.K. 1958. A study of the chromosomes during meiosis in twenty-eight species of Hemiptera (Heteroptera, Homoptera). Proceedings of the Zoological Society of Calcutta11: 9-37. Cited in Papeschi AG, Bressa MJ. 2006. Evolutionary cytogenetics in Heteroptera. **J. Bio. Res.** 5: 3-21.

Barr, P.J. 1991. Mammalian subtilisins: the long-sought dibasic processing endoproteases. **Cell** 66: 1-3.

Beard, C.B., D.M. Hamm and F.H. Collins. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. **Insect Mol. Biol.** 22:103-124.

Berthier, F., M. Renaud, S. Alziari and R. Durand. 1986. RNA mapping on Drosophila mitochondrial DNA: precursors and template strands. **Nucl. Acids Res.** 14: 4519-4533.

Boore, J.L. 1999. Animal mitochondrial genomes. **Nucl. Acids Res.** 27: 1767-1780.

Boyce, T.M., M.E. Zwick and C.F. Aquadro. 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. **Genetics** 123: 825-836.

Bressa, M.J, A.G. Papeschi, L.M. Mola and M.L. Larramendy.1999. Meiotic studies in Dysdercus Guérin Méneville 1831 (Heteroptera: Pyrrhocoridae). I. Neo-XY in *Dysdercus albofasciatus* Berg 1878, a new sex chromosome determination system in Heteroptera. **Chrom. Res.** 7: 503-508.

Bressa, M.J., A.G. Papeschi, M.Vitková, I. Ku-bicková-Fuková, M.I. Pigozzi and F. Marec. 2009. Sex chromosome evolution in cotton strainers of the genus Dysdercus (Heteroptera: Pyrrhocoridae). **Cytogenet. Genome Res.** 125: 292-305.

Butenandt, A. and N.D. Tam. 1957. Uber einen geschlechtsspezifischen Duftstoff der Wasserwanze Belostoma indica Vitalis (*Lethocercus indicus* Lep.). **Hoppe-Seyler's Zeitschrift für physiologische Chemie** 308: 277-283.

Byrne, B.M., M. Gruber and G. AB. 1989. The evolution of egg yolk proteins. **Prog. Biophys. Mol. Biol.** 53: 33-69.

Cantatore, P., M. Roberti, G. Rainaldi, M.N. Gadaleta and C. Saccone. 1989. The complete nucleotide sequence, gene organization, and genetic code ofmitochondrial genome of *Paracentrotus lividus*. **J. Biol. Chem.** 264: 10965-10975.

Camacho, J.P.M., J. Belda and J. Cabrero. 1985. Meiotic behavior of the holocentric chromosomes of *Nezava viridula* (Insecta, Heteroptera) analysed by C-banding and silver impregnation. **Can. J. Genet. Cytol.** 27: 490-497.

Cameron, S.L. and M.F. Whiting. 2007. Mitochondrial genomic comparisons of the subterranean termites from the Genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). **Genome** 50: 188-202.

Carapelli, A., L.Vannini, F. Nardi, J.L. Boore, L. Beani, R. Dallaiand and F. Frati. 2006. The mitochondrial genome of the entomophagous endoparasite *Xenos vesparum* (Insecta: Strepsiptera). **Gene** 376:248-259.

Castanhole, M.M.U., L.L.V. Pereira, H.V. Souza, H.E.M.C. Bicudo, L.A.A. Costa and M.M. Itoyama. 2008. Heteropicnotic chromatin and nucleolar activity in meiosis and spermiogenesis of *Limnogonus aduncus* (Herteroptera, Gerridae): a stained nucleolar organizing region that can serve as a model for studying chromosome behavior. **Gen. Mol. Res.** 7: 1398-1407.

Caterino, M.S. and F.A. Sperling. 1999. Papilio phylogeny based on mitochondrial *cytochrome oxidase I* and *II* genes. **Mol. Phylogenet. Evol.** 11: 122-137.

Cattani, M.V. and A.G. Papeschi. 2004. Nucleous organizing regions and semipersistent nucleolus during meiosis in *Spartocera fusca* (Thunberg) (Coreidae, Heteroptera). **Hereditus** 140: 105-111.

Cha, S.Y., H.J.Yoon, E.M. Lee, M.H. Yoon, J.S. Hwang, B.R. Jin and Y.S. Han. 2007. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). **Gene** 392: 206-220.

Chandra, A., P. S. Pathak and R. Kbhatt. 2006. Stylosanthes research in India: prospects and challenges ahead. **Curr. Sci.** 90: 915-921.

Chen, J.S., W.L. Cho and A.S. Raikhel.1994. Analysis of mosquito *vitellogenin* cDNA similarity with vertebrate phosvitins and arthropod serum proteins. **J. Mol. Biol.** 237: 641-647.

Clary, D.O. and D.R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. **J. Mol. Evol.** 22: 252-271.

Comas, D., M.D. Piulachs, and X. Belles. 2000. Vitellogenin of *Blattella germanica* (L.) (Dictyoptera, Blattellide): nucleotide sequence of the cDNA and analysis of the protein primary structure. **Arch. Insect Biochem. Physiol.** 45: 1-11.

Denslow, N.D., M.C. Chow, K.J. Kroll and L. Green. 1999. Vitellogenin as a biomarker of exposure for estrogen and estrogen mimics. **Ecotoxicology** 8:385-398

Dhadialla, T.S. and A.S. Raikhel. 1990. Biosynthesis of mosquito vitellogenin. **J. Biol. Chem.** 265: 9924-9933.

Donnell, D.M. 2004. Vitellogenin of the parasitoid wasp *Encarsia formosa* (Hymenoptera: Aphelinidae): gene organization and differential use by members of the genus. **Insect Biochem. Mol. Biol.** 34: 951-961.

Dotson, E.M. and C.B. Beard. 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. **Insect Mol. Biol.** 10: 205-215.

Engelman, F. 1979. Insect vitellogenin: identification, biosynthesis and role in vitellogenesis. **Adv. Insect Physiol.** 14: 49-109.

Gadaleta, G., G. Pepe, G. De Candia, C. Quagliariello, E. Sbisa and C. Saccone. 1989. The complete nucleotide sequence of the *Rattus norvergicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. **J. Mol. Evol.** 28: 497-516.

Gerber-Huber, S., D. Nardelli, J.A. Haefliger, D.N. Cooper, F. Givel and J.E. Germond. 1987. Precursor-product relationship between vitellogenin and the yolk proteins as derived from the complete sequence of a *Xenopus vitellogenin* gene. **Nucl. Acids Res.** 15: 4737-4760.

Giorgi, F., J.T. Bradley and J.H. Nordin. 1999. Differential vitellin polypeptide processing in insect embryos. **Micron** 30: 579-596.

Gissi, C., F. Iannelli, G. Pesole. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. **Heredity** 101: 301-320.

González-García, J.M., C. Antonío, J.A. Suja and J.S. Rufas. 1996. Meiosis in holocentric chromosomes: kinetic activity is randomly restricted to the chromatid ends of sex univalents in *Graphosoma italicum* (Heteroptera) **Chrom. Res.** 4: 124-132.

Grozena, S.M., and A. Maryanska-Nadachowska.1995. Meiosis of two species of Cacopsylla with polymorphic sex chromosomes in males (Homoptera, Psyllidae). **Folia Biology** 43: 93-98.

_____ and S. Nokkala. 1996. Chromosomes and their meiotic behavior in two families of the primitive infraorder Dipsocoromorpha (Heteroptera). **Hereditas** 125: 31-36.

Grozeva, S., V. Kuznetsova, and B. Anokhin. 2010. Bed bug cytogenetics: karyotype, sex chromosome system, FISH mapping of 18S rRNA, and male meiosis in *Cimex lectularius* Linnaeus, 1758 (Heteroptera: Cimicidae). **Comparative Cytogenetics** 4: 151-160.

Hagedorn, H.H., D.R. Maddison and Z. Tu. 1998. The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. **Adv. Insect Physiol.** 27: 335-384.

Hansen,P.D., H. Dizer, B. Hock, A. Marx, J. Shery, M. McMaster and Ch. Blaise. 1998. Vitellogenin a biomarker for endocrine disrupters. **Trends Anal. Chem.** 1: 1-4

Hassanin, A. 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: Strategies for limiting the misleading effects of multiple changes inpattern and rates of substitution. **Mol. Phylogenet. Evol.** 38: 100-116.

_____ , N. Leger and J. Deutsch. 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. **Syst. Biol.** 54: 277-298.

Havukainen, H., O. Halskau, L. Skjaerven, B. Smedal and G.V. Amdam. 2011. De constructing honey bee vitellogenin: novel 40 kDa fragment assigned to its N-terminus. **J. Exp. Biol.** 15: 528-592.

Hirai, M., D. Watanabe, A. Kiyota and Y. Chinzei.1998. Nucleotide sequence of *vitellogenin* mRNA in the bean bug, *Riptortus clavatus*: analysis of processing in the fat body and ovary. **Insect Biochem. Mol. Biol.** 28: 537-547.

Hiremath, S. and K. Lehtoma. 1997. Complete nucleotide sequence of the *vitellogenin* mRNA from the gypsy moth: Novel arrangement of the subunit encoding regions. **Insect Biochem. Mol. Biol.** 27: 27-35.

Howell, W.M. and D.A. Black. 1980. Controlled silver staining of nucleolus organizer regions with a protective colloidal developer: a 1-step method. **Experientia** 36: 1014-1015.

Hughes-Schrader, S. and F. Schrader. 1961. The kinetochore of the Hemitera. **Chromosoma** 12: 327-350.

Hua, J.M., M. Li, P.Z. Dong, Y. Cui, Q. Xie and W.J. Bu. 2009. Phylogenetic analysis of the true water bugs (Insecta: Hemiptera: Heteroptera: Nepomorpha): evidence from mitochondrial genomes. **BMC Evol. Bio.**9: 134.

Izumi, S., K. Yano, Y. Yamamoto and S.Y. Takahashi. 1994. Yolk proteins from insect eggs: structure, biosynthesis and programmed degradation during embryogenesis. **J. Insect Physiol.** 40:735-746.

Jacobs, D.H. 2004. The evolution of a neo-XY_1Y_2 sex chromosome system by autosome-sex chromosome fusion in *Dundocoris nodulicarinus* Jacobs (Heteroptera: Aradidae: Carventinae). **Chrom. Res.**12: 175-191.

Jande, S.S. 1959. An analysis of the chromosomes in the four species of the family Belostomatidae (Heteroptera, Cryptocerata). Res Bull (NS) Panjab Univ 10: 25-34, Cited in Papeschi AG, Bressa MJ. 2006. Evolutionary cytogenetics in Heteroptera. **J. Biol. Res.** 5: 3-21.

Jia, W., and P.G. Higgs. 2008. Codon usage in mitochondrial genomes: distinguishing context dependent mutation from translational selection. **Mol. Biol. Evol.** 25: 339-351.

John, B. and B. Naylor. 1961. Anomalous chromosome behavior in germ line of *Schistocerca gregaria*. **Heredity** 16: 187-198.

Kang, B.J., Nanri, T., Lee, J.M., Saito, H., Han, C.H. Hatakeyama, m., Saigusa, M., 2008. Vitellogenesis in both sexes of gonochoristic mud shrimp, Upogebia major (Crustacea): analyses of *vitellogenin* gene expression and vitellogenin processing.Comp. Biochem. **Physiol. Biochem. Mol. Biol**. 149: 589-598.

Kim, M.J., X.L.Wan, K.G. Kim, J.S. Hwang and I. Kim.2010. Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). **Afr. J. Biotechnol.** 9: 735-754.

Koyama, J., H. Kakinohana and T. Miyatake. 2004. Eradication of the melon fly, *Bactrocera cucurbitae*, in Japan: importance of behaviour, ecology, genetics, and evolution. **Annu. Rev. Entomol.** 49: 331-349.

LaChance L.E., M. Degrugillier and A.P. Leverich.1970. Cytogenetics of inherited partial sterility in three generation of the large milkweed bug as related to homokinetic chromosomes. **Chromosoma** 29: 24-41.

Lauck D.R. and A.S. Menke.1961. The higher classification of the Belostomatidae. **Ann. Entomol. Soc. Am.** 54: 644-657.

Lee, J.M., Y. Nishimori, M. Hatakeyama, T. W. Bae, and K. Oishi. 2000b. Vitellogenin of the cicada *Graptopsaltria nigrofuscata* (Homoptera): analysis of its primary structure. **Insect Biochem. Mol. Biol.** 30: 1-7.

Lewis, D.L., C.L. Farr and L.S. Kaguni. 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. **Insect Mol. Biol.** 4: 263-278.

Li, H., H.Y. Liu, L.M. Cao, A.M. Shi, H.L. Yang and W.Z. Cai. 2012b. The complete mitochondrial genome of the dam bug *Alloeorhynchus bakeri* (Hemiptera: Nabidae). **Int. J. Biol. Sci.** 8: 93-107.

_____, J. Huang, W. Cai, Z. Zhao, W. Peng and J. Wu. 2010. The vitellogenin of the bumblebee, *Bombus hypocrita*: studies on structural analysis of the cDNA and expression of the mRNA. **J. Comp. Physiol. B.**180: 161-170.

Lin, C.P. and B.N. Danforth. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. **Mol. Phylogenet. Evol.** 30: 686-702.

Liu, C.L., Z. Kajiura, K. Shiomi, R. Takei and M. Nakagaki. 2001. Purification and cDNA sequencing of *vitellogenin* of the wild silkworm, *Antheraea pernyi*. **J. Insect Biotecnol. Sericol.** 70: 95-104.

_____and A. Liang. 2013. The complete mitochondrial genome of spittlebug *Paphnutius ruficeps* (Insecta: Hemiptera: Cercopidae) with a fairly short putative control region.**Acta Biochim. Biophys. Sin.** 45: 309-319.

Lowe, T.M. and S.R. Eddy. 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. **Nucl. Acids Res.** 25: 955-964.

Lumsa-ard, J. 2001. Nutrition of edible insect of upper part in south of Thailand. **Keankakart** 29: 45-49.

Martin, M.G. and V. Matozzo. 2004. Vitellogenin induction as a biomarker of exposure to estrogenic compounds in aquatic environments. **Mar. Pollut. Bull.** 48: 835-839.

Maryanska-Nadachowska, A., V.G. Kuznetsova, H. Abdul-nour. 2008. A chromosomal study on a Lebanese spittlebug *Philaenus arslani* (Hemiptera: Auchenorrhyncha: Aphrophoridae). **Eur. J. Entomol.** 105: 205-210.

Matozzo, V., F. Gagne, M.G. Martin, F. Ricciardi and C. Blaise. 2008. Vitellogenin as a biomarker of exposure to estrogenic compounds in aquatic invertebrates: a review. **Environ. Int.** 34: 531-545.

Mola, L.M. and A.G. Papeschi. 1993. Meiotic studies in *Largus rufipennis* (Castelnau) (Largidae, Heteroptra): frequency and behavior of ring bivalents, univalents and B chromosomes. **Heredity** 71: 33-40.

Mayada, T.N. and D.D. Wagner. 1992. Vicinal cysteines in the prosequence play a role in von Willebrand factor multimet assembly. **Proceedings of the National academy of Science USA**. 89: 3531-3535.

Meng, Y., C.L. Liu, K. Shiomi, M. Nakagaki, Y. Banno and Z. Kajiura. 2008. Purification and cDNA cloning of *vitellogenin* of the wild silkworm, *Saturnia japonica* (Lepidoptera: Saturniidae). **J. Insect Biotechnol. Sericol.** 77: 35-44.

Mongkolvai, P., A. Wingsranoi, K. Sombum and J. Mongkolvai. 2008. Marketing and Culturing of Giant Water Bugs. Reports for faculty of Natural Resources, **Rajamangala University of Technology Isan: Sakonakhon Campus**.16 p.

Mouchel, N., V. Trichet, A. Betz, J.P. Le Pennec and J. Wolff.1996. Characteriza-tion of vitellogenin from rainbow trout (*Oncorhynchus mykiss*). **Gene** 174: 59-64.

Nagaba, Y., M. Tufail, H. Inui and M. Takeda. 2011. Hormonal regulation and effects of four environmental pollutants on *vitellogenin* gene transcription in the giant water bug, *Lethocerus deyrollei* (Hemiptera: Belostomatidae). **J. Insect Conserv.** 15: 421-431.

Nakabachi, A., S.Koshikawa, T. Miura, S. Miyagishima.2010. Genome size of *Pachypsylla venusta* (Hemiptera: Psyllidae) and the ploidy of its bacteriocyte, the symbiotic host cell that harbors intracellular mutualistic bacteria with the smallest cellular genome. **Bull. Entomol. Res.**100: 27-33.

Nokkala, S. 1985. Restriction of kinetic activity of holokinetic chromosomes in meiotic cells and its structural basis. **Hereditas** 102: 85-88.

_____. 1986. The mechanisms behind the regular segregation of autosomal univalents in *Calocoris quadripunctatus* (Vil.) (Miridae, Hemiptera). **Hereditas** 105: 199-204.

_____ and C. Nokkala. 1983. Achiasmatic male meiosis in two species of *Saldula* (Saldidae, Hemiptera). **Hereditas** 99: 131-134.

_____ and_____.1999. Chromosomes in two bug species of Hebrus (Hebridae, Heteroptera). The occurrence of neo-XY sex chromosome systemin Heteroptera.**Caryologia** 52: 27-30.

_____, V.G. Kuznetsova, A. Maryanska-Nadachowska and C. Nokkala. 2006. Holocentric chromosomes in meiosis. II. The mode of orientation and segregation of a trivalent. **Chrom. Res.**14: 559-565.

Nose, Y., J. M. Lee, T. Ueno, M. Hatakeyama and K. Oishi. 1997. Cloning of cDNA for vitellogenin of the parasitoid wasp, *Pimpla nipponica* (Hymenoptera: Apocrita: Ichneumonidae): vitellogenin primary structure and evolutionary considerations. **Insect Biochem. Mol. Biol.** 27(12): 1047-1056.

Ojala, D. J. Montoya and G. Attardi. 1981. tRNA punctuation model of RNA processing in human mitochondria. **Nature** 290: 470-474.

Okoumassoun, L. E., D. Averill-Bates, F. Gagne, M. Marion and F. Denizeau. 2002. Assessing the estrogenic potential of organochlorine pesticides in primary cultures of male rainbow trout (*Oncorhynchus mykiss*) hepatocytes using vitellogenin as a biomarker. **Toxicology** 178:193-207.

Osada, M. T. Takamura, H. Sato and K. Mori. 2003. Vitellogenin synthesis in the ovary of the scallop, Patinopecten yessoensis. Control by estradiol-17β and the central nervous system. **J. Exp. Zool.** 299:172-179.

Papeschi, A.G. 1994. Chromosome rearrangements in *Belostoma plebejum* (Stål) (Belostomatidae, Heteroptera). **Caryologia** 47: 223-231.

_____. 1995. Correspondance between C-banding and AgNOR in sex chromosome of *Belostoma oxyurum* (Belostomatidae, Heteroptera). **Cytologia** 60: 291-295.

_____ . 1996. Sex chromosome polymorphism in a species of *Belostoma* (Belostomatidae, Heteroptera). **Hereditas** 124: 269-274.

_____ and C. Bidua. 1985. Chromosome complement and male meiosis in four species of *Belostoma* Latreille (Heteroptera. Belostomatidae). **Braz. J. Genet.** 2: 249-261.

_____ and M.J. Bressa. 2006. Evolutionary cytogenetics in Heteroptera. **J. Biol. Res.** 5: 3-21.

_____ and L.M. Mola. 1990. Meiotic studies in *Acanonicus hahni* (Stål) (Coreidae, Heteroptera) I. Behavior of univalents in desynaptic individuals. **Genetica** 80: 31-38.

_____,_____, E.J. Bressa, V. Greizer-stein, V. Lia and L. Poggio. 2003. Behavior of ring bivalents in holokinetic systems: alternative sites of spindle attachment in *Pachylis argentines* and *Nezara viridula* (Heteroptera). **Chrom. Res.**11: 725-733.

Perez-Goodwyn, P.J. 2006. Taxonomic revision of the subfamily Lethocerinae Lauck & Menke (Heteroptera: Belostomatidae). **Stuttgarter Beiträge zur Naturkunde**, **Serie A (Biologie)** 695: 1-71.

Perna, N.T. and T.D. Kocher.1995. Patterns of nucleotide composition at four-fold degenerate sites of animal mitochondrial genomes. **J. Mol. Evol.** 41: 353-358.

Piulachs, M.D., K.R. Guidugli, A.R. Barchuk, J. Cruz, Z.L.P. Simoes and X. Belles. 2003. The vitellogenin of the honey bee: structural analysis of the cDNA and expression studies. **Insect Biochem. Mol. Biol.**33: 49-465.

Pongsart, C. 1990.Culturing of giant water bugs in laboratory.Spacial problem. **Srinakharinwirot University.Mahasarakham campus.**

Raikhel, A.S. and T.S. Dhadialla. 1992. Accumulation of yolk proteins in insect oocytes. **Annu. Rev. Entomal.** 37: 217-251.

Rebagliati, P., L.M. Mola, A.G. Papeschi and J. Grazia. 2005. Cytogenetic studies in Pentatomidae (Heteroptera): **Review. J. Zoo. Syst. Evol. Res.** 43: 199-213.

Reyes, A., C. Gissi, G. Pesole and C. Saccone.1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. **Mol. Biol. Evol.** 15: 957-966.

Rina, M.D. and A.C. Mintzas. 1988. Biosynthesis and regulation of two vitellogenins in the fat body and ovaries of *Ceratitis capitata* (Diptera). **Roux's Archiv. Dev. Biol.** 197:167-174.

Romans, P., Z. Tu, Z. Ke and H.H. Hagedorn. 1995. Analysis of a *vitellogenin* gene of the mosquito *Aedes aegypti* and comparisons to vitellogenins from other organisms. **Insect Biochem. Mol. Biol.** 25: 939-958.

Rouille, Y., S.J. Duguay, K. Lund, M. Furuta, Q. Gong and G. Lipkind. 1995. Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides: the subtilisin-like proprotein convertases. **Frontiers Neuroendocrinol.** 16: 322-361.

Saito, S., K.Tamuea, T. Aotsuka. 2005. Replicatiion origin of mitochondrial DNA in insects. **Genetics** 171: 433-448.

Salvato, P., M.Simonato, A. Battisti and E. Negrisolo. 2008. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). **BMC Genomics** 9: 331-345.

Sappington, T.W. and A.S. Raikhel. 1998. Molecular characteristics of insect vitellogenins and vitellogenin receptors**. Insect Biochem. Mol. Biol.** 28: 277-300.

Shadel, G.S. and D.A.Clayton. 1993. Mitochondrial transcription initiation: variation and conservation. **J. Biol. Chem.** 268: 16083-16086.

Sheffield, N.C., H. Song, S.L. Cameron and M.F. Whiting. 2008. A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. **Mol. Biol. Evol.** 25: 2499-2509.

Shi, A., H. Li, X. Bai, X. Dai, J. Chang, E. Guilbert and W. Cai. 2012. The complete mitochondrial genome of the flat bug *Aradacanthia heissi* (Hemiptera: Aradidae). Zootaxa **3238: 23-38.**

Shu, Y., J. Zhou, W. Tang, K. Lu, Q. Zhou and G. Zhang. 2009. Molecular characterization and expression pattern of*Spodoptera litura* (Lepidoptera: Noctuidae) vitellogenin, and its response to lead stress. **J. Insect Physiol.** 55: 608-616.

Simon, C., F. Frati, A.T. Beckenbach, B. Crespi, H. Lui and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. **Ann. Entomol. Soc. Am.** 87: 651-701.

Simon, C., T.R. Buckley, F. Frati, J.B. Stewart and A.T. Beckenbach. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. **Annu. Rev. Ecol. Evol. Syst.** 37:545-579.

Song, N., A.P. Liang and C. Ma. 2010. The complete mitochondrial genome sequence of the planthopper, *Sivaloka damnosus*. **J. Insect Sci.**10:76.

_____and_____. 2009. Complete mitochondrial genome of the small brown planthopper, *Laodelphax striatellus* (Delphacidae: Hemiptera), with a novel gene order. **Zool. Sci.** 26: 851-860.

_____, H. Li, F. Song, L. Liu, P. Wang, H. Xun, X. Dai, J. Chang and W. Cai. 2013. The complete mitochondrial genome of a tessaratomid bug, *Eusthenes cupreus* (Hemiptera: Heteroptera: Pentatomomorpha: Tessaratomidae). **Zootaxa** 3620: 260-272.

Stewart, J.B. and A.T. Beckenbach. 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumariuss* (Hemiptera: Auchenorrhyncha: Cercopoidae). **Genome** 48: 46-54.

Suja, J.A., A.L. del Cerro, J. Page, J.S. Rufas and J.L. Santos. 2000. Meiotic sister chromatid cohesion in holocentric sex chromosomes of three heteropteran species is maintained in absence of axial elements. **Chromosoma** 109: 35-43.

Sup-udom, W. 1992. Study of hating and nymph culturing of giant water bugs in laboratory. Spacial problem. **Srinakharinwirot University. Mahasarakham campus.**

Taanman, J.W. 1999.The mitochondrial genome: structure, transcription, translation and replication. **Biochim. Biophys. Acta**. 1410 (2): 103-123.

Taborsky, G. 1991. On the interaction of phosvitins with ferric ion: solubility of the Fe (III)-phosphoprotein complex under acid conditions is a function of the iron/phosphate ratio and the degree of phosvitin phosphorylation. **J. Inorg. Biochem.** 44: 65-77.

Tapper, D.A. and D.A. Clayton. 1981. Mechanism of replication of human mitochondrial DNA: localization of the 5′ends of nascent daughter strands. **J. Biol. Chem.** 256: 5109-5115.

Thao, M.L., L. Baumann and P. Baumann. 2004. Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). **BMC Evol. Biol.** 4: 25.

Thompson, D.M., S.M.S. Khalil, L.A. Jeffers, D.E. Sonenshine, R.D. Mitchell, C.J. Osgood and R.M. Roe. 2007. Sequence and the developmental and tissue-specific regulation of the first complete *vitellogenin* messenger RNA from ticks responsible for heme sequestration. **Insect Biochem. Mol. Biol.** 37: 363-374.

Trenzek, T.A. and W. Engels. 1986. Occurrence of vitellogenin in drone honeybees. **Int. J. Invert. Repr. Develop.** 10: 307-311.

Trewitt, P.M., L.J. Heilmann, S.S. Degrugillier and A.K. Kumaran. 1992. The boll weevil *vitellogenin* gene: nucleotide sequence, structure, and evolutionary relationship to nematode and vertebrate *vitellogenin* genes. **J. Mol. Evo.** 34: 478-492.

Tufail, M., J. Bembenek, A. M. Elgendy and M. Takeda. 2007. Evidence for two *vitellogenin-related* genes in *Leucophaea maderae*: the protein primary structure and its processind. **Arch. Insect Biochem. Physiol**. 66: 190-203.

Tufail, M., M. Hatakeyama, and M. Takeda. 2001. Molecular evidence for two *vitellogenin* genes and processing of *vitelloginins* in the American cockroach, *Periplaneta americana*. **Arch. Insect Biochem. Physiol**. 48:72-80.

Tufail, M. and M. Takeda. 2002. Vitelloginins of cockroach, Leucophaea maderae: nucleotide sequence, structure and analysis of processing in the fat body and oocytes. **Insect Biochem. Mol. Biol.** 32: 1469-1476.

_____ and _____. 2005. Molecular cloning, characterization and regulation of the cockroach vitellogenin receptor during oogenesis. **Insect Mol. Biol.** 14: 389-401.

_____ and._____ 2008. Molecular characteristics of insect vitellogenins. **J. Insect Physiol.** 54: 1447-1458.

_____ , M.Naeemullah, M. elmogy, P.N. Sharma, M. Takeda and G. Nakamura. 2010. Molecular cloning, transcriptional regulation, and differential expression profiling of vitellogenin in two wing-morphs of the brownplanthopper, *Nilaparvata lugens* Stal (Hemiptera: Delphacidae). **Insect Mol. Biol.** 19:787-798.

_____ , M., Y. Nagaba, M. Elgendy and M. Takeda.2014. Regulation of *vitellogenin* genes in insects. **Entomol. Sci.**http://onlinelibrary.wiley.com-/doi/10.1111/ens.12086/pdf.14 pages.

Ushima, N. 1979.**Animal Cytogenetics**, **Volume 3: Insecta 6, Hemiptera II: Heteroptera. Gebruder Borntraeger.**

Valle, D., J.E.P. Lima, S. Gomes, and E.S. Golddenberg. 1987. *Rhodnius prolixus* vitellogenesis: dependence upon the blood source. **J. Insect Physiol.** 33: 249-254.

Valle, D. 1993. Vitelogenesis in insects and other groups: a review. **Mem. Inst. Oswaldo Cruz.** 88: 1-26.

Veerana, M., Kubera, A. and L., Ngernsiri. Analysis of the *Vitellogenin* gene of rice moth, Corcyra cephalonica Stainton. **Archives of insect Biochemistry and Physilogy**.In press.

Wei, S.J., P. Tang, L.H. Zheng, M. Shi and X.X. Chen. 2010. The complete mitochondrial genome of *Evania appendigaster* (Hymenoptera: Evaniidae) has low A+T content and a long intergenic spacer between *atp8* and *atp6*. **Mol. Biol. Rep.** 37: 1931-1942.

White, M.J.D. 1973. **Animal Cytology and Evolution**. Cambridge University Press.

Wilson, K., V. Cahill, E. Ballment and J. Benzie. 2000. The complete sequence of the mitochondrial genome of the crustacean *Penaeus mondon*: are malacostracan crustaceans more closely related to insects than to branchiopods? **Mol. Biol. Evol.** 17: 863-874.

Wolstenholme, D.R. 1992. Genetic novelties in mitochondrial genomes of multi-cellular animals. **Curr. Opin. Genet. Dev.** 2: 918-925.

Wu, Y.P., J. Li, J.L. Zhao, T.J. Su, A.R. Luo, R.J. Fan, M.C. Chen, C.S. Wu, C.D. Zhu. 2012. The complete mitochondrial genome of the rice moth, *Corcyra cephalonica*. **J. Insect Sci.** 12:72.

Wyatt, G.R. 1991. Gene regulation in insect reproduction. **Invert. Reprod. Develop.** 20:1-35.

Yano, K., M.T. Sakurai, S. Izumi and S. Tomino. 1994a. Vitellogenin gene of the silkworm, *Bombyx mori*: structure and sex-dependent expression. **FEBS Lett.** 256: 207-211.

Yokoyama, M.N., Z. Kajiura, M. Nakagaki, R. Takei, M. Kobayashi, K. Tanaka. 1993. Storage proteins, vitellogenin and vitellin of wild silkworms, *Antheraea yamamai*, *Antheraea pernyi* and their hybrids. **Comp. Biochem. Physiol.** 106: 163-172.

Zhang, D.X. and G.M. Hewitt. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. **Biochem. Syst. Ecol.** 25: 99-120.

Zhao, J., L. Hu, S.L. Winterton and Z. Liu. 2013. Ancestral gene Organization in the mitochondrial genome of *Thyridosmylus langii* (Mclachaln, 1870) (Neuroptera: Osmylidae) and implications for lacewing evolution. **PLos ONE** 8: e62943.

_____, Z., Y. Huang and F. Shi. 2007. The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A+T-rich region of 70 bp in length. **Genome** 50: 855-866.

APPENDICES

Appendix A

Data for phylogenetic tree analysis

Appendix Table A1 List of taxa used in the phylogenetic analysis of mitogenome

Higher Taxon	Superfamily	Family	Species	Accession number
Fulgoromorpha	Fulgoroidea	Delphacidae	*Laodelphax striatellus*	FJ360695
Fulgoromorpha	Fulgoroidea	Fulgoridae	*Lycorma delicatula*	EU909203
Fulgoromorpha	Fulgoroidea	Ricaniidae	*Ricania marginalis*	JN242415
Fulgoromorpha	Fulgoroidea	Flatidae	*Geisha distinctissima*	FJ230961
Sternorrhyncha	Aphidoidea	Aphididae	*Schizaphis Graminum*	AY531391
Sternorrhyncha	Aleyrodoidea	Aleyrodidae	*Bemisia tabaci*	AY521259
Cicadomorpha	Membracoidea	Cicadellidae	*Homalodisca vitripennis*	NC_006899
Cicadomorpha	Cercopoidea	Aphrophoridae	*Philaenus spumarius*	AY630340
Cicadomorpha	Cercopoidea	Cercopidae	*Callitetix versicolor*	EU725832
Heteroptera	Gerroidea	Gerridae	*Aquarium paludum*	FJ456944
Heteroptera	Gerroidea	Helotrephidae	*Helotrephes sp.*	FJ456951
Heteroptera	Hydrometroidea	Hydrometridae	*Hydrometra* sp.	FJ456945
Heteroptera	Aradoidea	Aradidae	*Neuroctenus parus*	EU427340
Heteroptera	Cimicoidea	Anthocoridae	*Orius niger*	EU427341
Heteroptera	Leptopodoidea	Leptopodidae	*Leptopus* sp.	FJ456946
Heteroptera	Saldoidea	Saldidae	*Saldula arsenjevi*	EU427345
Heteroptera	Reduvioidea	Reduviidae	*Triatoma dimidiata*	AF301594
Heteroptera	Notonectoidea	Notonectidae	*Enithares tibialis*	FJ456949
Heteroptera	Naucoroidea	Naucoridae	*Ilyocoris cimicoides*	FJ456964
Heteroptera	Aphelocheiroidea	Aphelocheiridae	*Aphelocheirus ellipsoideus*	FJ456939
Heteroptera	Nepoidea	Nepidae	*Laccotrephes robustus*	FJ456948
Heteroptera	Nepoidea	Belostomatidae	*Diplonychus rusticus*	FJ456940
Heteroptera	**Nepoidea**	**Belostomatidae**	***Lethocerus indicus***	**in this study**
Heteroptera	Neboidea	Nabidae	*Alloeorhynchus bakeri*	HM235722

Appendix Table A1 (Continued)

Higher Taxon	Superfamily	Family	Species	Accession number
Heteroptera	Ochteroidea	Ochteridae	*Ochterus marginatus*	FJ456950
Heteroptera	Ochteroidea	Gelastocoridae	*Nerthra sp*	FJ456943
Heteroptera	Pentatomoidea	Cydnidae	*Macroscytus gibbulus*	EU427338
Heteroptera	Pentatomoidea	Pentatomidae	*Nezara viridula*	EF208087
Heteroptera	Pentatomoidea	Pentatomidae	*Halyomorpha halys*	FJ685650
Heteroptera	Coreoidea	Rhopalidae	*Stictopleurus subviridis*	EU826088
Heteroptera	Coreoidea	Alydidae	*Riptortus pedestris*	EU427344
Heteroptera	Coreoidea	Rhopalidae	*Aeschyntelus notatus*	EU427333
Heteroptera	Coreoidea	Coreidae	*Hydaropsis longirostris*	EU427337
Heteroptera	Pyrrhocoroidea	Pyrrhocoridae	*Dysdercus cingulatus*	EU427335
Heteroptera	Pyrrhocoroidea	Largidae	*Physopelta gutta*	EU427343
Heteroptera	Lygaeoidea	Berytidae	*Yemmalysus parallelus*	EU427346
Heteroptera	Lygaeoidea	Malcidae	*Malcus inconspicuus*	EU427339
Heteroptera	Lygaeoidea	Geocoridae	*Geocoris pallidipennis*	EU427336
Heteroptera	Lygaeoidea	Colobathristidae	*Phaenacantha marcida*	EU427342
Psocoptera	Atropetae	Lepidopsocidae	*Lepidopsocid sp.*	AF335994
Orthoptera	Tettigonioidea	Tettigoniidae	*Deracantha onos*	EU137664
Orthoptera	Grylloidea	Gryllotalpidae	*Gryllotalpa pluvialis*	EU938371

Appendix Table A2 The protein sequences of vitellogenin for phylogenetics analysis.

Order	Family	Species	Amino acid	GenBank Accession
Coleoptera	Curculionidae	*Anthonomus grandis*	1,790	AAA27740.1
	Tenebrionidae	*Tenebrio molitor*	1,821	AAU20328.2
Dictyoptera	Blaberidae	*Rhyparobia maderae*	1,913	BAB19327.1
	Cicadellidae	*Blattella germanica*	1,862	CAA06379.2
Diptera	Culicidae	*Aedes aegypti*	2,148	AAA99486.1
	Culicidae	*Anopheles gambiae*	2,051	AAF82131.1
	Culicidae	*Culex tarsalis*	2,119	ADH04225.1
Hymenoptera	Apidae	*Apis mellifera*	1,770	NP_001011578.1
	Apidae	*Bombus hypocrita*	1,772	ACU00433.1
	Apidae	*Bombus ignitus*	1,773	ACM46019.1
	Aphelinidae	*Encarsia formosa*	1,814	AAT48601.1
	Apocrita	*Pimpla nipponica*	1,807	AAC32024.1
	Formicidae	*Camponotus floridanus*	1,820	EFN64902.1
	Formicidae	*Solenopsis invicta*	1,641	AAP47155.1
	Leucospidae	*Pteromalus puparum*	1,803	ABO70318.1
	Vespidae	*Vespula vulgaris*	1,756	AER70365.1
Lepidoptera	Noctuidae	*Spodoptera litura*	1,748	ABU68426.1
	Nymphalidae	*Danaus plexippus*	1,763	EHJ67298.1
	Bombycidae	*Bombyx mori*	1,782	NP_001037309.1
	Crambidae	*Cnaphalocrocis medinalis*	1,789	AEM75020.1
	Erebidae	*Lymantria dispar*	1,747	AAC02818.1
	Noctuidae	*Helicoverpa armigera*	1,756	AGL08685.1
	Saturniidae	*Actias selene*	1,774	ADB94560.1
	Saturniidae	*Samia cynthia pryeri*	1,779	BAD91196.1
	Saturniidae	*Saturnia japonica*	1776	BAD91195.1
Galliformes	Phasianidae	*Gallus gallus*	1,852	AAA49139.1

Appendix Table A2 (Continued)

Order	Family	Species	Amino acid	GenBank Accession
Hemiptera	Alydidae	*Riptortus clavatus*	1,876	AAB72001.1
	Belostomatidae	*Lethocerus deyrollei*	1,895	BAG12118.1
		Lethocerus indicus	**1,888**	**in this study***
	Cicadellidae	*Homalodisca vitripennis*	1,890	AAZ06771.1
	Delphacidae	*Nilaparvata lugens*	2,063	BAF75351.1
	Aleyrodidae	*Bemisia tabaci*	2,184	ADU04394.1
	Pentatomidae	*Plautia stali*	1,907	ABB88075.1
	Alydidae	*Riptortus clavatus*	1,876	AAB72001.1
	Miridae	*Nesidiocori tenuis*	1,970	AGV05363.1
		Apolygus lucorum	1,995	AFW97644.1
		Trigonotylus caelestialium	1,976	BAJ33507.1
	Cicadadae	*Graptopsaltria Nigrofuscata*	1,987	BAAA85987.1

Appendix B

All amino acid sequence of *LiVg* gene

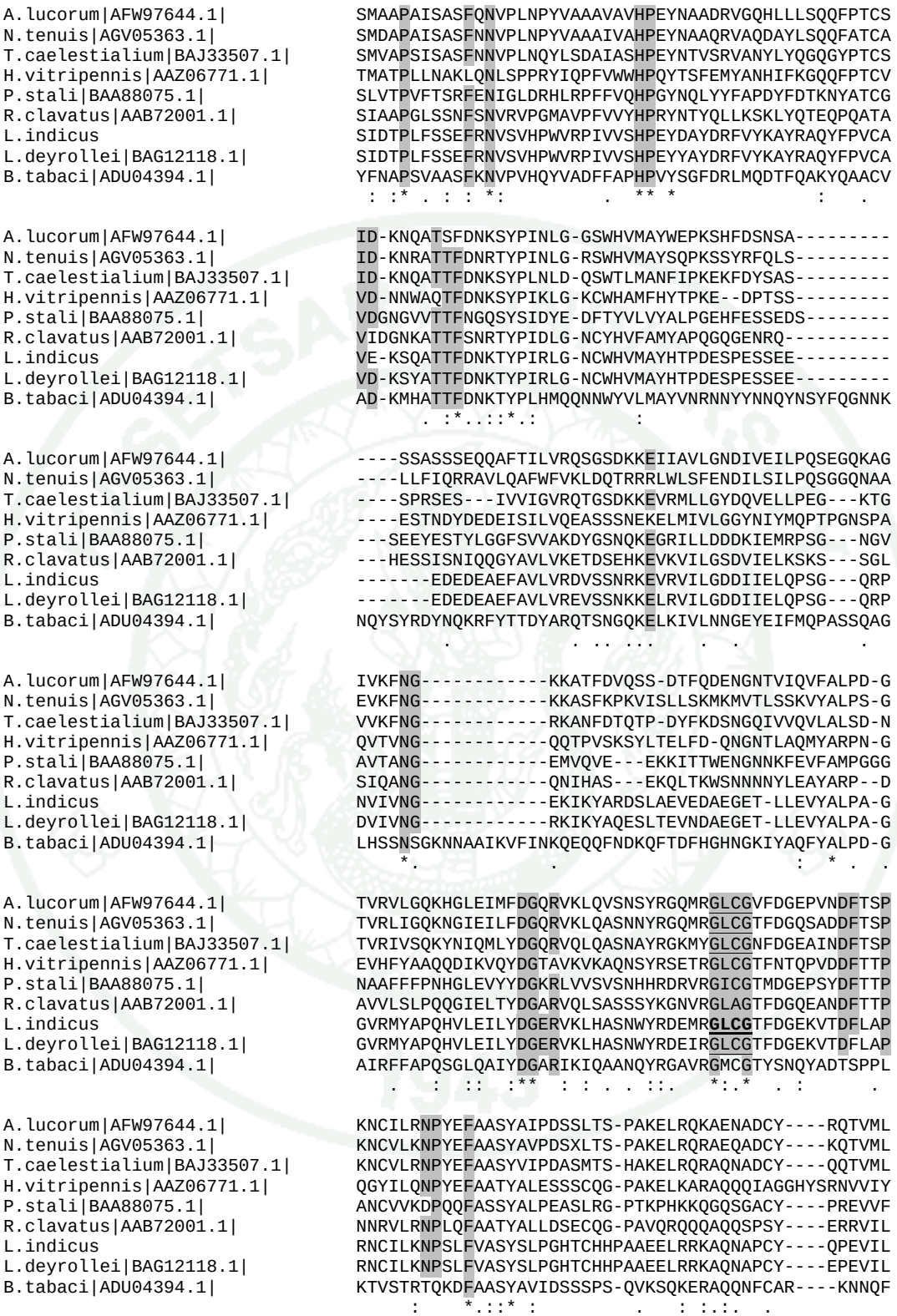
>*L.indicus vitellogenin* gene, 1888 aa

MRSSVILFLFVTVTAAVAEPWSNATAWKPNTVYKYTLRGRSMTGLHQVSNQYSGILLKAQVAIQLKTQNL
LSLKVSKAEFAELHANLSKGWNTEIPDHKLHWQPMPLAGKPFELKLKNGTVDRMYVDKSVPTWEENWYKS
IASQLQVDTAARNLQKSRINQLPAMHKPMGVYKTMEDTVTGECETVYDVSPLPDYLLQSKPELAPLPHLR
ADGRIIEIVKTRNFSECDQRVAFHFGITGPTNWEPAGNQMGPFMARSSVSRVIISGKINKYTIQSSVTTN
KVILAPHLYNKQKGIVASRMNLTLESVQPAQGQPQNIPEPRTIKDLVYEYNAASEDAHKQGHEIGQDSSS
SSSSESSESSSSSSSSSSSSSSESSQEQQAQKPKHPRNRREAEDSSSSSSSESDSSTSSDDDYSSCSSS
SSSSSSSSSSTSTSSSSSSSSSSSSDSDSTSRSSSEEYYQPRPKLRIPPNSPLLPFFIGYQGNSIQAAHK
INVVQQAQELAKEIALDVQTPNAITGEDTLTKFTVLIRALRTMSPEQIKEAAQQLYFPAHKASNHTVTDA
KKYQAWAVYRDAVAQAGTGPALVVIQDWIRNERVVRERAAQLVATLQDSARHPTLEYMNTFFELVKSPQV
MKQLYLNTSALISFTHLVRKAMVNNETVHHRYPVHAYGRLTRKNATVVVLEKYIPYLAEKLRKAVEEEDS
PKIQVYAQALGNIGHYKILSAFEDYLEGKVNVTDYQRLVMVTALHKLTVVYPKRALPVLYKIYQNIGETP
EVRVAAVMLIMRTNPPAQILQRMAESTRFDHSIDVRSAIQSAIQSAAKLTGPKYYQLSQNAKAAVHMLNP
TQYGYHYSKHHMRSYIVEQQNLAYKHDISLIQSSDSIIPSTVYYNLKRKLGGYRTQVFKMAYMTSSVDDL
LELIFDQFEDEEPEHKSKRPSHGPEEWTLQKIKEILDLDVDVQEQVEGNLQATLFGSKRFMAWDNHSIEA
LPYIVKEAAKSLKDGQPFNMTKWYNPLNIELAFPMATGLPFVYTYRTPTYFSIGGEVRAKTTPDLAKGDD
DEIKIPDTVNATIQARVIYSTKTQATMGFVTPFNHQRYIAGLDRNIHVYLPISGKLDVDIENTKLWATLN
PLRPNKYQKLLEYSTVAYTTRHEILDLSPPAQKNKTEEIHVRPPMRDSPTVGQHSTGFAYDLKAETEKDV
LELSTVYNALQRHDIVSALLYSTHEMSITNSNFSVAYNPQKTSAKAAKIVLTYDDDDDDESSSGKKPRSH
NKHRGANIAYPISTDPDSAERQEQYLDAVGEGIEDSLARMIDVSVQFEGESKAEYVATAAYADSPVSNYS
RWLLFLSMDPAQKTQQTKPFELTLNVTTDFPDVPVLNYKKALRADPTSSGSSAHLNFGENAQHGAYVYLE
GTLEQTKPRIQYIQRHPLAKLCEQQMEEGHNILPACRNITMRANFLDSYSFDVEYKDVPEYAKNLSYHAY
AMIRHMFYPYVSENFIDPENEHGKLSVDVDFAPNLRSVNVSIDTPLFSSEFRNVSVHPWVRPIVVSHPEY
DAYDRFVYKAYRAQYFPVCAVEKSQATTFDNKTYPIRLGNCWHVMAYHTPDESPESSEEEDEDEAEFAVL
VRDVSSNRKEVRVILGDDIIELQPSGQRPNVIVNGEKIKYARDSLAEVEDAEGETLLEVYALPAGGVRMY
APQHVLEILYDGERVKLHASNWYRDEMRGLCGTFDGEKVTDFLAPRNCILKNPSLFVASYSLPGHTCHHP
AAEELRRKAQNAPCYQPEVILGDVVSERDLGKTKKQPKPQWSNSLERPDSKRACTHYKVKVLEVGKKMCF
SLRAHIACGAGCQPSGRIEKKVEFHCVERSDAARHWADSIAKGHNPDFSKKQPNYRASIRLPEKCIPH

Appendix Figure B All amino acid sequence of *LiVg* gene

Appendix C

Structural comparison of the C-terminal part of *Vgs* gene



Appendix Figure C1 Structural comparison of the C-terminal part of *Vgs* gene

Appendix Figure C1 Structural comparison of the C-terminal part of *Vgs* gene

```
A.lucorum|AFW97644.1|           GDVISENEAGRS-------------------------------------
N.tenuis|AGV05363.1|            GDVINENEAGRS-------------------------------------
T.caelestialium|BAJ33507.1|     GDVISENEAGRQ-------------------------------------
H.vitripennis|AAZ06771.1|       GNVVTDADAYHY-------------------------------------
P.stali|BAA88075.1|             ADVVSDTDAGR-------------------------------------
R.clavatus|AAB72001.1|          GDVVNELEAGRQQQQN-----------------------------------
L.indicus                       GDVVSERDLGKTK-------------------------------------
L.deyrollei|BAG12118.1|         GEVVSERDLGKVR-------------------------------------
B.tabaci|ADU04394.1|            GNYVSRSDAGYGYKYNNNDKYYESAYKNTKYYDPSNSQYNKYYKNNKYAK
                                .: :.  :

A.lucorum|AFW97644.1|           ------------------------------------------------
N.tenuis|AGV05363.1|            ------------------------------------------------
T.caelestialium|BAJ33507.1|     ------------------------------------------------
H.vitripennis|AAZ06771.1|       ------------------------------------------------
P.stali|BAA88075.1|             ------------------------------------------------
R.clavatus|AAB72001.1|          ------------------------------------------------
L.indicus                       ------------------------------------------------
L.deyrollei|BAG12118.1|         ------------------------------------------------
B.tabaci|ADU04394.1|            DQDSSYSSSSSSSSDSSSSSSSDSSSSSSSSSPSSSSSSSDSSSSSSSSQ


A.lucorum|AFW97644.1|           --------------------------TKRSSSKSSNSIGNRNSKMSSPA
N.tenuis|AGV05363.1|            --------------------------SKKLWSKSSSNSNGRSQASTSPS
T.caelestialium|BAJ33507.1|     --------------------------PNKPKSGNSPHNRN---AAYTGS
H.vitripennis|AAZ06771.1|       --------------------------SRSNSNKYGRLRILNKEKYVSCS
P.stali|BAA88075.1|             --------------------------GRRLQTMPGNDIEGKLPRMPSCS
R.clavatus|AAB72001.1|          --------------------------NKSKKNKANGSEERMKVQGSSCS
L.indicus                       ----------------------------KQPKPQWSNSLERPDSKRACT
L.deyrollei|BAG12118.1|         ----------------------------KQPKPQWSNSLERSDSKRACT
B.tabaci|ADU04394.1|            SNDNNRNNNNRNNSNNNNNRNNNRNNNNNNNSSSSQEASYEQRNQNGPSI
                                                           .           .

A.lucorum|AFW97644.1|           RLRVLVIHANGQTCFSTRPQLSCQSQSQEANTAQKNVDFHCVSDATAAK-
N.tenuis|AGV05363.1|            SLRVKVIEHNGQTCFSLRPYVTCASRSQEANKIEKSVDFHCVSDAQAAR-
T.caelestialium|BAJ33507.1|     ALRVKVIEHNGQTCFSKRPQLACSSQAQESNQTSKNLEFHCVSDATAAK-
H.vitripennis|AAZ06771.1|       QRRLMTMHKDGKQCFSVQPQLHCTDQCSPQGYVNKEVQFMCVLTSSVSE-
P.stali|BAA88075.1|             TYKIQIVKEGGRSCFSLRPQVSCNPNCKPTKNIQKHIEFHCRPDNDNLTG
R.clavatus|AAB72001.1|          KLMTQIVEENDKSCFSTTPQPACASHCRPTGKIQKIVDFHCVQASSSSR-
L.indicus                       HYKVKVLEVGKKMCFSLRAHIACGAGCQPSGRIEKKVEFHCVERSDAAR-
L.deyrollei|BAG12118.1|         HYRVKVLEVGKKMCFSLRAHIACGAGCQPASKLEKKVDFHCIENSAAAQ-
B.tabaci|ADU04394.1|            RKLYRAINQGDDMCFTINAIPTCRYPAKPVGSAKKMVDFYCAPKSSSEAQ
                                    :. .   **:  .   *   .     .* ::* *

A.lucorum|AFW97644.1|           RWEDQIKRGASPDFSKKGANYKHS-MKLPSRCNA--COOH
N.tenuis|AGV05363.1|            RWVEQIKKGANPDFSQKGANHRAT-IRLPGKCNA--COOH
T.caelestialium|BAJ33507.1|     RWEEQIKKGANPDFSRKPTNHTAN-VRVPTSCKA--COOH
H.vitripennis|AAZ06771.1|       HWKKLVNRGINPDFSDKHAYFQTYNVQVPQSCQAN-COOH
P.stali|BAA88075.1|             HWLRMVQKGANPDLSQKSSNSKMS-VMLPEGCTSIHCOOH
R.clavatus|AAB72001.1|          NWAQMIKNGANPDFSAKEKHRQIL-LQLPTGCVPKACOOH
L.indicus                       HWADSIAKGHNPDFSKKQPNYRAS-IRLPEKCIPH-COOH
L.deyrollei|BAG12118.1|         HWADSIAKGHNPDFSKKQPNYRAT-VRLPENA----COOH
B.tabaci|ADU04394.1|            HFSKLIAKGAAPSQLSLKKPNQKFEVNIPEYCVA--COOH
                                .:   : .*  *.          : :*  .
```

Appendix Figure C1 (Continued)

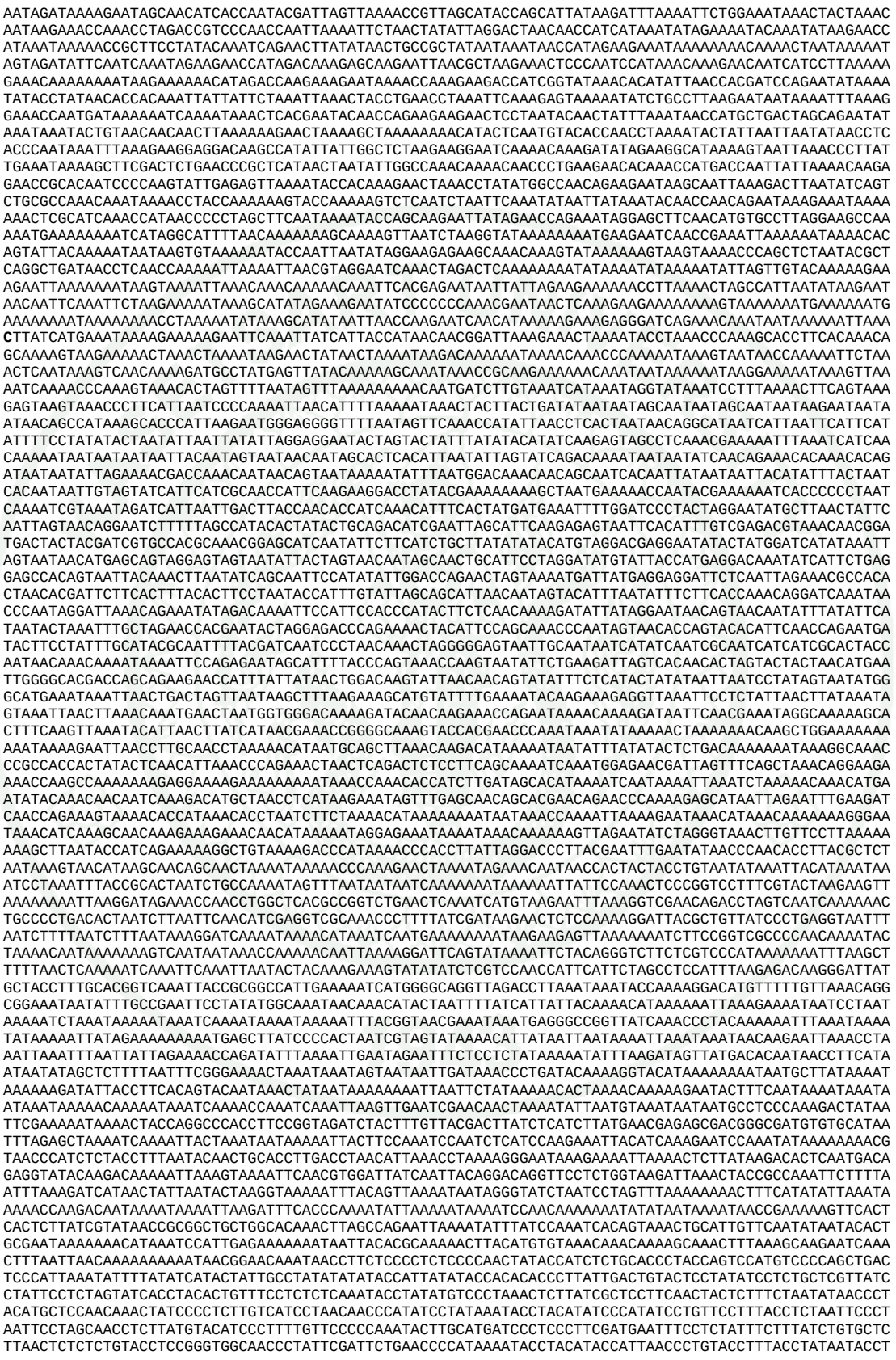
Appendix D

Nucleotide sequence of *L. indicus* mitogenome

>*Lethocerus indicus*, mitochondrial genome sequence, 17,632 bp

AATAGAGTGCCTGATTAATAGGGCTATCTTGATAGGATAGATCATGTACTAATGTACCTCTATTATACTTTACAGGAGTGACCCGTAACTTAAGAACT
CAAAATTCTCCGTGCACCATACACCTAAGCATAAAAAGATAAGCTAAATAAGCTATCAGGTTCATACCCTGACCATAGAAGTACTAACCTTCTTCTTT
TTAATGGCTGGCAATTCAAGCAAGCTAATATTTATCGCTATTATAATTTTAGGAACATTAACTACTATCAACGCATCAAATTGACTTGGAATTTGAAT
AGGACTCGAAATTAACTTGATTGCGTTTACCCCTTTAATCTTCAAACCTAAGCAAGCATCAATATCAGAAGCATCAATAATTTATTTCTTAGTGCAAT
CATTAGGGTCAGCCATAATATTAATAATTATTGTAACAAACAACTTACTAATAGAACCTATTTCATCAATCAATAACTCAACAGCCACAGCACTAGCT
GCTAGATTAATACTAAAACTAGGGATAGCACCAATACACATATGATTTCCAGAAGTAACAAGAAAAATATCCTGAAGTAGAGCGCTACTATTATTAAC
ATGACAAAAAATTGCCCCCCTTGTAGTACTATCATATGTAGTAGAAAAGTCATTCGTAATTTACATTCTAGTAATAACAAGAACATTCATTGGTGCAG
TAGGAGGATTAAACCAAACATCACTACGAAAAATTATAGCTTACTCATCAATTAGACACATAGGATGAATAATTGCTTGCATAAAAATAGAAAATAAC
CTATGAATAGCTTACTTAGCAATTTACATTGCCATAGTAACAATAGCAATCATTCTGTTCGAAAAAACATCAATATACCATATTAATCAAGTAAACAT
AAACTCATTAACACCCACTAGAAAAATTTATATACATATCATTATTTATAAGACTAGGAGGGTTGCCACCATTCCTAGGGTTTTTGCCAAAATGAATA
GTTATCCAATCAATAATAGGACACAGATTATACATAGTACTAACTGTAATAGTGCTAACAACCCTAATCACTCTACTATACTATTTACGAATAATAAC
AATAATATCAACACTTACACAAGAACAAACAAGTGAATTACAGAATTTAAACCACCACACAAAAATTTAGTAATAATAGTATTATTAGTAAATCTGTC
AACCCCAATCGTAACTGTATTCATATTCATATAAAGCTTTAAGTTAAAGTAAACCATTAACCTTCAAAGTTAAAAATATGGATAGTCTGTAAGCTTTA
GTAAGAATTTTACTACTTTAGAATTGCAGTCTAATATCATAAATTGACTATAAAGCCTGGCAGGAGGTAATTAACCATATATAAATTTACAATTTACA
GCCTATAATTCAGCCACCCTACTCGTGAATAAATGAATATTTTCAACAAACCACAAGGATATTGGAACTTTGTACTTTTTATTCGGGATCTGGTCGGG
AATAATTGGTACAGCCCTAAGATGATTAATCCGAATCGAACTAGGGCAACCAGGATCCTTTATTGGAGATGACCAAATCTATAACGTAATTGTAACAG
CCCACGCATTCATTATAATTTTCTTTATAGTTATACCAATCATAATTGGTGGATTTGGAAACTGACTTCTACCTTTAATAATCGGAGCACCAGATATA
GCATTCCCGCGAATAAATAACATGAGATTCTGACTACTACCTCCATCACTAACACTTCTTCTATCGAGAAGAATTGTTGAAAAGGGAGCAGGTACTGG
TTGAACAGTGTACCCACCACTATCAACTAATATCTCACACAGAGGAGCATCAGTAGATCTTGCAATCTTCTCCCTCCACCTAGCAGGAGTATCGTCAA
TTCTAGGAGCAGTAAACTTTATCTCAACCGTAATCAACATACGAGCTACAGGAATAACACCAGATCGAATACCTTTATTTGTTTGATCAGTATCAATT
ACAGCCCTATTACTATTATTATCACTACCAGTACTAGCCGGAGCAATCACTATATTATTAACAGATCGAAACTTCAACACATCATTCTTTGACCCTGC
AGGAGGGGGTGACCCAATCCTATATCAACACTTATTTTGATTCTTCGGGCATCCAGAAGTATACATTTTAATTTTACCCGGATTCGGATTAATTTCCC
ACATTATTAGACAAGAAAGAGGAAAAAATGAAGCATTCGGATCACTAGGAATAATCTACGCAATAATAGCAATTGGACTATTACGATTTGTAGTATGA
GCCCACCATATATTTACTGTAGGAATAGACGTAGACACGCGAGCATACTTTACATCAGCTACAATAATTATTGCAGTACCAACAGGAATTAAAATTTT
CAGATGAATAGCAACAATACATGGAACAGTTATTAATTATAGACCATCAACTTTATGAGCACTAGGATTCGTATTTCTATTCACCTTAGGTGGACTAA
CAGGTGTAGTATTAGCAAACTCATCAATTGATATTGTACTGCACGACACATACTATGTAGTAGCACATTTCCACTACGTACTATCAATAGGAGCCGTA
TTTGCCATTATCGGAAGATTTATTCAATGATACCCTCTATTTACTGGAATTACCATAAATCCCAAATGACTTAAAATCCATTTTATAGTAATATTCAT
CGGAGTAAACGTAACATTCTTTCCACAGCACTTCCTAGGATTAAGAGGAATGCCACGACGATATTCAGATTACCAGACAGTTCACAACATGAAATG
TAGTATCGAGAATAGGATCAACAATATCAGTAATAGGAGTAATATTATTTATCTTCATCATATGGGAAAGAATAGTAGCTAAACGAATAATAATATTT
TCATCAAACATGCACTCAAGAATCGAATGACTACAAAAATATCCGCCAGCAGAACACTCCTATTCAGAGCTGCCAATCATAACAGAATTCTAATATGG
CAGATTAGTGCAATGAACTTAAGCTTCATGTATAAAGATCATTCTTTTGTTAGAAATAGCAACATGATCTAATCTTACCTTACAAGACGCCAATTCCC
CAATTATAGAACAAATAATCTTCTTCCACGACCATGCTATAATAATCCTAGCAATAATTACTACTATAGTAGGATACTTAATAGCAAGAATACTAAAA
AACAAACTAATTAACCGATACTTGCTAGAAGGACAAACAATTGAACTAGTATGAACAATTATGCCCGCGGTAACATTAGTATTTATTGCATTACCATC
GCTACGACTACTATATTTAATGGACGAAATATGAAATCCTATTATAACAGTAAAAGCAATTGGGCATCAATGATACTGATCATATGAATATACAGACT
TCTCTAACGTAGAATTTGACTCATATATATAAAACCATCGTCAGAAATAGAGATAATAGAATTTCGATTACTAGACGTAGATAACCGAATAATAATACCA
ATAAACACACAAATCCGACTACTAGTAACAGCAGCAGATGTATTACATTCATGAGCAATACCATCACTGGGAATCAAAATTGACGCAACCCCAGGACG
ACTAAACCAAGGAAGAATCTCAATTATACGACCAGGACTAATATTCGGACAATGCTCAGAAATCTGCGGAGCAAACCACAGATTCATACCAATTGTAG
TAGAAAGAACACCAATATCTAACTTTATTAATTGACTAAAATCAAACTCATTAAGTGGCTGAAAGTAAGCAATGGTCTCTTAAACCACAAAATAGTAA
ATAACGAATACTCTTAATGGGAGAGGTTAGTTTAAACAAAATATTATTTTGTCAAAATAAAGTTAATCAAAGATTACCACTCACATTCCACAAATATC
ACCAATATGATGATCAATATTATTCATTATATTCATTACAGCATTTATAGTAATATGCACAATAACATACTTCATAGTATTTTACACTAAAGTAAAA
CAACACATAAAAAAATACAAAAAACATCAATAAATTGAAAATGATAACAGATTTATTCTCAACATTTGACCCAGCAACATCAGCATCATCATCAATAA
ACTGAACAAGAACATTCATTGGAATAATTTTAATACCATCAATATATTGAATAATACCAAATCGATATACTATAGCAGTAAAAATAATAATAATAACA
CTACACAACGAATTCAAAACTTTACTAGGAAAAACAAAGGGAACAACATTAATAATGGTATCATTATTCATATTTATTCTATTCAACAACATCATAGG
GTTACTTCCATACGTATTTACAAGATCAAGACATCTAGTATTTACAATATCAATAGCCCTTCCAATATGAATATCATTAATGCTATTTGGATGGGTAA
ATAAAACCCAAGAAATATTTGCCCACCTAATACCTGTAGGAACTCCTCCAGCTCTAATACCATTTATAGTAATCATCGAAACAATTAGAAACTTAATT
CGACCAGGAGCACTAGCCGTACGACTAACAGCTAATAAATTGCAGGACACCTATTATAAGACTACTAGGAAACAACATAGAAAAAAATTCAATGATTA
TTGTAACATCACTAATCACGATTCAAATAATACTAATACTATTCGAAACAACCGTAGCCAGAATTCAAGCTTATGTATTTTCAGTACGGTCAACCCTT
TACAATGAGAAGTAGA*ATG*GCAATAAATAAAAATCACCCATTCCACTTAGTAGATGCAAGACCATGGCCTCTCACAGGATCAATTGGAGCAATAACTA
TAGTATCAGGACTAGTAATGTGATTTCATCAATACAAAACAGGACTACTAATAATAGGAATATTAATTACTCTATTAACTATATTCCAATGATGACGA
GACGTATCACGAGAAGGAACATTTCAAGGGAAACACACAATTGCCGTAACAAATGGACTAAAACTAGGAATAATTTTATTCATTATTTCCGAAGTATT
TTTCTTCATTTCATTCTTCTGAAGATTTTTCCACAGAAGACTAGCACCAACAATTGAAATTGGAGCAATATGACCACCTGTTGGAATCAAAACATTCA
ACCCTATACAAATCCCTCTACTAAATACAATAATCTTACTATGCTCAGGAATTACTGTAACGTGAGCACATCACAGATTAATAGAAAGAAACCACTCA
CAAGCCAGACAAAGATTATTCTTTACCGTAATATTAGGAATATACTTCCACCATGCTACAAGCATTCTGAATACTATGAATCAAGATTTACAATCAGA
GACTCAATTCTACGGATCAACATTCTTCATAGCAACGGGATTCCACGGACTACACGTAATTATCGGAACAACATTCCTATTAGTTTGCTTAATACGAC
ACATAATATGCCACTTCTCTAGAAAACATCACTTCGGATTCGAAGCAGCAGCATGATACTGACACTTTGTGGACGTAGTATGATTATTCCTATATATT
TCTATCTATTGATGAGGTAG*ATC*TCTTTTATTATATTAAAGTATATTTGACTTCCAATCAAAAGGTCTCAAAGAGAAAGAGTAATTTATTTACTAAT
AATAGCAGCATCAATAGCAATAATAATCAGAATAGGACTTATAATACTGTGCACAGTAATTTCAAAAAAAAGAATTATAGACAAAGAAAAAATATCCC
CATTTGAATGCGGATTTGACCCAAATAGATCAGCACGAATACCATTCTCAATCCAATTCTTCCTAATCGCCGTACTATTCCTAATCTTCGATATCGAA
ATTGTAATTATCATGCCCATAATCATTACAGTAAAAATAAGAATAATAGCACAATGATTCACAACTGTAACAATATTCGTAATAATCATCTTAGCAGG
ACTATTCCACGAATGAAAAAACGGAGTACTAGAATGAGCCAAATGGGGGTTGTAGTTAAAAATAACATTTAATTTGCAATTAAAAAGTACTTTGAATA
AGTCTTCCTCAAAGTAAAGAAGTAAAAATTACATTTAGTTTCGACCTAAATAAAGATATAAAATCCTTTACTTATTAGGTCAAGCCAAATAGAGGCAT
TTCATTGTTAATGAAAAAATTGGGACTAAAATAAACCCCAACTAAAAGAGAAAGTGGTATCTAAAAGAAGCTGCTAACTTCTTTTAAAGCGGTTAAAC
TCCGTTTTAGTATTTATATAGTTTAAAGAAAACATTTCATTTTCATTGAAAAATTAAGTATTAGTTTATAAATACCAAAAAAGCATAAAGCTTATCCT
TATATCTTCAATATAATGCTTTAAATTTAAGCTATTTTAGTTTAAATTAAAATAAAGAAAAGAACAAAAAAGAAAGAAGCCATAAGCAACTTAAATGT
GTTACTTTGACAATAAGTATTAAATTGACTAAGTTTAGAAAAAAAAATAAGAGAATAATTAGAAACAACCCTCTCACCTCAACCATATCTATAACTC
TAAACAAACCACTAGAAAGTTTAAAAGCAGGAGAATAAACCATATAAGTACTAAACCTAGGCATAAACCACATAGAACCCATAAATACCGAAATAGAA
TAAAAACGTAAAGAAAACAAATCAAACCCCCTGTAATAAACAGACAACTCGTAACCTAAAAAAGAACCCAGAATAATAAAAAATAAAGAAAGAATCTT
AGTACCAAAAGGAATAACAACCAAATGAGGATAAGGAAAAATAAGCCAAGACAAAATTCTACCTCTAAAAATAGATATAAAAGTTAAAAAAATTATAG
GAAAAATCATAGTAAAATCCTCCACATAAGATTGACAAACATAACAACCTGTTCCATCACTAACGCAATAGTAAACCAAACGCATTCTATAAGAAACA
GTAAGACCAATAGAAACATAAAAAATTAAATAAAAAAACAAATTACCACCAGATATTCTATAAGACTCAACAATAAAATCCTTAGAATAAAAACCAGA
TAAAAAAGGCAAACCACAAAGAGATAAAGAAGAAATACAAAAACAAGCTCTAGTAAAAGGCAAATGAACAGAGACACCACCTATATGACGAATATCCT
GAGAATCACCTATACAATGAATAATTAAACCCGCACACAAAAACAACAAGGCCTTAAAAAAAGCATGAGTTAACAAATGAAAAAAAGACAAAATAGGA
AACCCTAAAAATAAAATACCCATTATAAGACCCAATTGACTTAATGTAGACAAAGCAATAATCTTCTTTAAATCATACTCAAAACTAGCCCCTAAACC
AGCTATAAACATAGTAACCAAACCCAAAAGAGCAAAAAACCAACAATCAAAATGAACAAATATAGAACTAAAACGAATTAATAAGTAAACACCAGCAG
TAACTAAAGTAGAAGAATGAACTAAAGCAGAAACAGGAGTAGGGGCAGCCATAGCCGCAGGTAACCAAGAAGAAAAAGGAATCTGAGCACTCTTAGTA
AAACCAGCCAAAACAACTAAAATAAATAAATAAAAACTCCAAGATCTTAAATCAGAATATATAAAAAAATTTCAACCACCAAAATTTAATATCCAAGA

Appendix Figure D1 Nucleotide sequence of *L. indicus* mitogenome



Appendix Figure D1 (Continued)

ATGTTCCCCCCTTATATACATTATCCAGTGGATAGAATAAATAATTATATGTCAGACACTCCTGCCCCCCCCCCTTAATCACATAAATTATATATAAA
TATAACCTATAGATTAAAAGAAGAATTGTATATGTAATAATAAATCTAGTAATAAGTAATTTTAATAACAAATAAATAAAATTTATCTGATCTTTAAT
CAGAAATCAAAAAACACTGGTCCAAAACGTTTCCTATTTGGAAATTAATCATAAAATAAGAAATAACTGTTAAAAACCAAAGTCAAATTAATACCTGA
TATAAAATGTTAAGCAAACAAGCGCTTAACAAAGAAACTTCGATAACAAATAAATAAAATTTATCTGATCTTTAATCAGAAATCAAAAAACACTGGTC
CAAAACGTTTCCTATTTGGAAATTAATCATAAAATAAGAAATAGCTGTTAATAAACCAAAGTCAAATTAATACCTGATATAAAATGTTAAGCAAACAA
GCGCTTAACAAAGAAATTTCGATAACAAATAAATAAAATTTATCTGATCTTTAATCAGAAATCAAAAAACACTGGTCCAAAACGTTTCCTATTTGGAA
ATTAATCATAAAATAAGAAATAACTGTTAAAAACCAAAGTCAAATTAATACCTGATATAAAATGTTAAGCAAACAAGCGCTTAACAAAGAAATTTCGA
TAACAAATAAATAAAATTTATCTGATCTTTAATCAGAAATCAAAAAACACTGGTCCAAAACGTTTCCTATTTGGAAATTAATCATAAAATAAGAAATA
ACTGTTAAAAACCAAAGTCAAATTAATACCTGATATAAAATGTTAAGCAAACAAGCGCTTAACAAAGAAATTTCGATAACAAATAAATAAAATTTATC
TGATCTTTAATCAGAAATCAAAAAACAATGGTCCAAAACGTTTCCTATTTGGAAATTAATCATAAAATAAGAAATAGCTGTTAAAAACCAAAGTCAAA
TTAATACCTGATATAAAATGTTAAGCAAACAAGCGCTTAACAAAGAAATTTCGATAACAAATAAATAAAATTTATCTGATCTTTAATCAGAAATCAAA
AAACAATGGTCCAAAACATTTCCTATTTGGAAATTAATCATAAAATAAGAAATAGCTGTTAAAAACCAAAACCAAATAGACCACATAGTGTGAGCACA
CAATAAATAAGTAATTATCGCAGAGTCCTCCATAAACACAGAATAAATAAAATAATAAAAATAAATAAACGACACAATAAACCAAAACCAAATAGACC
ACAAAGTGTGAGCACCCAATAAATAAGTAATTATCGCAGAGTCCTCCATAAACACAGAATAAATAAAATAATAAAAATAAATAAACGACACAATAAAC
CAAAACCAAATAGACCACAAAGTGTGAGCACCCAATAAATAAGTAATTATCGCAGAGTCCTCCATAAACACAGAATAAATAAAATAATAAAAATAAAT
AAACGACACAATAAACCAAAACCAAATAGACCACAAAGTGTGAGCACCCAATAAATAAGTAATTATCGCAGAGTCCTCCATAAACACAGAATAAATAA
AATAATAAAAATAAATAAACGACACAATAAACCAAAACCAAATAGACCACAAAGTGTGAGCACCCAATAAATAAGTAATTATCGCAGAGTCCTCCATA
AACACAGAATAAATAAAATAATAAAAATAAATAAACGACACAATAAACCAAAACCAAATAGACCACAAAGTGTGAGCACCCAATAAATAAGTAATTAT
CGCAGAGTCCTCCATAAACACAGAATAAATAAAATAATAAAAATAAATAAACGACACAATAAACCAAAACCAAATAGACCACAAAGTGTGAGCACCCA
ATAAATAAGTAATTATCGCAGAGTCCTCCATAAACACAGAATAAATAAAATAATAAAAATAAATAAACGACACAATAAACCAAAACCAAATAGACCAC
AAAGTGTGAGCACCCAATAAATAAGTAATTATCGCAGAGTCCTCCATAAACACAGAATAAATAAAATAATAAAAATAAATAAACGACACAATAAACCA
AAACCAAATAGACCACAAAGTGTGAGCACCCAATAAATAAGTAATTATCGCAGAGTCCTCCATAAACACAGAATAAATAAAATAATAAAAATAAATAA
ACGACACAATAAACCAAAACCAAATAGACCACAAAGTGTGAGCACCCAATAAATAAGTAATTATCGCAGAGTCCTCTACAAATAAAAAGCAACTAAAA
CAATAAAAATAAAAATAAAGGAACACAAAATAAAGTTAATAAAAAAAAAAATAAAAAAAAAAATTAAAAAAATTAAAAAAATAAATAATTCACAAAAA
TAAAGGATTATGGATCCAATACTTTACAACCCTCTATTTAAATACTAAAATTTCCCAAGAGGAAAAAAAATGGGAAATTTAATTTATCCC

Appendix Figure D1 (Continued)

Appendix E

Amino acid sequence of 13 PCGs of *L. indicus* mitogenome

>Lethocerus indicus
MAGNSSKLMFIAIMILGTLTTINASNWLGIWMGLEINLIAFTPLIFKPKQASMSEASMIY
FLVQSLGSAMMLMIIVTNNLLMEPISSINNSTATALAASLMLKLGMAPMHMWFPEVTSKM
SWSSALLLLTWQKIAPLVVLSYVVEKSFVIYILVMTSTFIGAVGGLNQTSLRKIMAYSSI
SHMGWMIACMKMENNLWMAYLAIYIAMVTMAIILFEKTSMYHINQVNMNSLTPTSKIYMH
IIIYKTSSVATIPSVFAKMNSYPINNSTQIMHSTNCNSANNPNHSTMLFTNNNNNINTYT
STNKWITEFKPPHKNLVMMVLLVNLSTPIVTVFMFM- 336 aa-ND2

>Lethocerus indicus
MNKWMFSTNHKDIGTLYFLFGIWSGMIGTALSWLIRIELGQPGSFIGDDQIYNVIVTAHA
FIMIFFMVMPIMIGGFGNWLLPLMIGAPDMAFPRMNNMSFWLLPPSLTLLLSSSIVEKGA
GTGWTVYPPLSTNISHSGASVDLAIFSLHLAGVSSILGAVNFISTVINMRATGMTPDRMP
LFVWSVSITALLLLLSLPVLAGAITMLLTDRNFNTSFFDPAGGGDPILYQHLFWFFGHPE
VYILILPGFGLISHIISQESGKNEAFGSLGMIYAMMAIGLLRFVVWAHHMFTVGMDVDTR
AYFTSATMIIAVPTGIKIFSWMATMHGTVINYSPSTLWALGFVFLFTLGGLTGVVLANSS
IDIVLHDTYYVVAHFHYVLSMGAVFAIIGSFIQWYPLFTGITMNPKWLKIHFMVMFIGVN
VTFFPQHFLGLSGMPRRYSDYPDSFTTWNVVSSMGSTMSVMGVMLFIFIMWESMVAKRMM
MFSSNMHSSIEWLQKYPPAEHSYSELPIMTEF- 512 aa-CoI

>Lethocerus indicus
MATWSNLTLQDANSPIMEQMIFFHDHAMMILAMITTMVGYLMASMLKNKLINRYLLEGQT
IELVWTIMPAVTLVFIALPSLRLLYLMDEMWNPIMTVKAIGHQWYWSYEYTDFSNVEFDS
YMKPSSEMEMMEFRLLDVDNRMMMPMNTQIRLLVTAADVLHSWAMPSLGIKIDATPGRLN
QGSISIMRPGLMFGQCSEICGANHSFMPIVVESTPMSNFINWLKSN- 226 aa-CoII

>Lethocerus indicus
MPQMSPMWWSMLFIMFITAFMVMCTMTYFMVFYTKSKTTHKKMQKTSMNWKW- 52 aa ATP8

>Lethocerus indicus
MMTDLFSTFDPATSASSSMNWTSTFIGMILMPSMYWMMPNRYTMAVKMMMMTLHNEFKTL
LGKTKGTTLMMVSLFMFILFNNIMGLLPYVFTSSSHLVFTMSMALPMWMSLMLFGWVNKT
QEMFAHLMPVGTPPALMPFMVIIETISNLIRPGALAVRLTANKLQDTYYKTTSKQHSKKF
NDYCNITNHDSNNTNTIRNNRSQNSSLCIFSTVNPLQWEVE- 221 aa-ATP6

>Lethocerus indicus
MAMNKNHPFHLVDASPWPLTGSIGAMTMVSGLVMWFHQYKTGLLMMGMLITLLTMFQWWR
DVSREGTFQGKHTIAVTNGLKLGMILFIISEVFFFISFFWSFFHSSLAPTIEIGAMWPPV
GIKTFNPMQIPLLNTMILLCSGITVTWAHHSLMESNHSQASQSLFFTVMLGMYFHHATSI
LNTMNQDLQSETQFYGSTFFMATGFHGLHVIIGTTFLLVCLMRHMMCHFSSKHHFGFEAA
AWYWHFVDVVWLFLYISIYWWGS- 263 aa-CoxIII

>Lethocerus indicus
MYLLMMAASMAMMISMGLMMLCTVISKKSIMDKEKMSPFECGFDPNSSARMPFSIQFFLI
AVLFLIFDIEIVIIMPMIITVKMSMMAQWFTTVTMFVMIILAGLFHEWKNGVLEWAK- 117aa-ND3

>Lethocerus indicus
MVVNMCLYRWSSLVLFFLGLCFFLFFLFLFKDDCSLFMDWEFLSVNSCSLSMVLLFDWMS
TIFISFVFFISSMVIYYSGSYMSSDLYSKRFLFMVLMFVFSMFMMVVSPNMVSILIGWDG
LGLVSYCLVVYFQNFKSYNAGMLTVLTNRIGDVAILLSISWMLNFGGWNFFMYSDLSSWS
FYLFILVVLAGFTKSAQIPFSSWLPAAMAAPTPVSALVHSSTLVTAGVYLLIRFSSMFVH
FDCWFFALLGLVTMFMAGLGASFEYDLKKIIALSTLSQLGLMMGILFLGFPILSFFHLLT
HAFFKALLFLCAGLIIHCMGDSQDIRHMGGVSVHLPFTSACFCISSLSLCGLPFLSGFYS
KDFIVESYSMSGGNLFFYLIFYVSIGLTVSYSMRLVYYCVSDGTGCYVCQSYVEDFTMIF
PMIFLTFMSIFSGSILSWLIFPYPHLVVIPFGTKILSLFFIILGSFLGYELSVYYSGFDL
FSLRFYSISVFMGSMWFMPSFSTYMVYSPAFKLSSGLFSVMDMGWGESVVSNYSLIFFSK
LSQFNTYCQSNTFKLLMASFFFVLFFILI- 569 aa-ND5

Appedix Figure E1 Amino acid sequence 13 PCGs of *L. indicus* mitogenome

>Lethocerus indicus
MMSLIFYYLFLIPLSFYVDSWLIMCFMFLGFFYFFSFFHFFTFFSSLSYSFGGDILSMCF
IFLSIWIVILMLMASFKVFSSNNYSREFVFVCLILLIFLILSFCTTNIFMFYIFFESSLI
PTLILIFGWGYQPERISAGFYLLFYTLFASLPMLIGIFYTYYFCNTVFYFLISVDSSFFL
YLSLTFAFFVKMPMIFFHFWLPKAHVEAPISGSMILAGILLKLGGYGLMRVFLFLYSVGC
IYNYIWISLSLFGTFLVGFICLAQTDIKSLIAYSSVGHMGLVLCGILTLNTWGLCGSLVL
MIGHGLCSSGLFCLANISYERVQSRSFYFNKGLITFMPSMSLFWFLLSANNMACPPSLNL
LGEVMLINSILGWCTLSMFFLAFSSFLSCCYSIYLYSASQHGYLNSCISSSSSGCIREFI
LIFYHWFPLNFIILKADIFTLWI- 443 aa-ND4

>Lethocerus indicus
MLGFDFNFIFPYFLLFVFSCGLFAFCMTHSHLLLTLLSLEFLVITLFLGLFYFLSYFSYS
SYFSLVFLTFAVCEGALGLGILVSLIRCYGNDNLNSFSFISW- 102aa-ND4L

>Lethocerus indicus
MMMAMMAMMSMMMTAMKHPLSMGGVLMVQTMLTSLMTGMIINSFMFSYMLMLIMLGGMLV
LFMYMSSVASNEKFKSSTKMMMMITMVMTMALTLMLVSDKMMMSTETQTQMMMLENDQTM
TVMKMFNGQTTAITIMMITYLLITMIVVSFIATIQEGPMRKKS-163aa-ND6

>Lethocerus indicus
MKNQYEKITPLIKIVNSSLIDLPTPSNISLWWNFGSLLGMCLTIQLVTGIFLAMHYTADI
ELAFKSVIHICRDVNNGWLLRSCHANGASMFFICLYMHVGRGMYYGSYKLVMTWAVGVVM
LLVTMATAFLGYVLPWGQMSFWGATVITNLMSAIPYIGPELVKWLWGGFSISNATLTRFF
TLHFLMPFVLAALTMVHLMFLHQTGSNNPMGLNSNMDKIPFHPYFSTKDIMGMTVTMFMF
MMLNLLEPRMLGDPENYIPANPMVTPVHIQPEWYFLFAYAILRSIPNKLGGVIAMIMSIA
IIIALPMTNKMKFQSMAFYPVNQVMFWSLVTTLVLLTWIGARPAEEPFIMTGQVLTTVYF
SYYMINPMVMWAWNKLTD- 378 aa-cytB

>Lethocerus indicus
MVSILVLWVFILVAVAYVTLLERKVLGYIQIRKGPNKVGFMGLLQPFSDGIKLFFKEQVY
PSYSNFFVYFISPIFMLFLSLLLWCLFPFFVYVYSFNFGLLFFLCFSSLGVYGVLLSGWS
SNSNYALLGSVRAVAQTISYEVSMSLIVVCMFMFVFSFNFIDFMCYQDGVWFIFFSFPLF
LAWFSSCLAETNRSPFDFAEGESELVSGFNVEYSGGGFAFIFLSEYMNIIFMSCLSCIMF
LGCKVNSFIFFFQLVFLVFMFIWVRGTLPRFRYDKLMYLTWKCFLPISLNYLLFYSGFLL
YLLSHH- 306 aa-ND1

Appedix Figure E1 (Continued)

Appendix F

Reagent preparation

General reagent

0.5 M EDTA, pH 8.0

EDTA	18.61 g

The chemical was dissolved in distilled water, adjusted to pH 8.0 with 1 M NaOH. Distilled water was added to the final volume 100 ml. The solution was sterilized by autoclaving for 15 minutes at 121 ºC, 15 lb/square inches.

TE, pH 8.0

1 M Tris-HCl, pH 8.0	1.0 ml
1 mM EDTA, pH 8.0	0.2 ml

The mixture was mixed thoroughly and adjusted to the final volume of 100 ml with distilled water.

3.0 M Sodium acetate, pH 5.2

Sodium acetate	24.2 g

The chemical was dissolved in distilled water, adjusted to pH 5.2 with glacial acetic acid. The final volume was adjusted to 100 ml with distilled water.

10 % SDS

Sodium dodecyl sulfate	10 g
Distilled water was added to final volume	90 ml

The chemical was dissolved at 68 ºC and adjusted to pH 7.2 by adding 2-3 drops of concentrated HCl. The solution was adjusted to the volume 100 ml

Reagent for molecular cloning

50X TAE buffer (stock)

Tris	240.2 g
Glacial acetic acid	57.0 ml
0.5 M EDTA (pH 8.0)	100.0 ml

The chemicals were dissolved in distilled water, and adjust the final volume to 1 liter with distilled water. (Working solution is 1 X)

Chloroform: Isoamyl alcohol (24: 1, v/v)

Chloroform (Merck)	24.0 ml
Isoamyl alcohol	1.0 ml

Both reagents were mixed together and stored in a dark bottle at room temperature.

1.0 M IPTG

Isopropylthio-ß-D-galactoside	2.38 g
Distilled water	100 ml

The solution was sterilized by filtration through a 0.2 µm filter and dispensed the solution into 1 ml aliquot tube and stored at –20 ºC.

X-gal

5-bormo-4-chloro-3-indolyl-ß-D-galactoside	100 mg

The chemical was dissolved in 2 ml of dimethyl-formamide. The solution was stored in a tube covered with aluminum foil and stored at -20ºC.

Ethidium bromide (10 mg/ml)

Ethidium bromide	1.0 g
Distilled water	10.0 ml

The solution was stored in a dark bottle at room temperature.

6X gel-loading dye buffer

0.25 % bromophenol blue
0.25 % xylene cyanol FF
30 % glycerol in water

The chemicals were dissolved and adjusted the final volume with distilled.

Media for bacterial culture

Luria-Bertani medium (LB medium per liter)

Tryptone (Difco)	10 g
Yeast extract (Difco)	5.0 g
NaCl	10 g

Adjust pH to 7.0 with NaOH. Then the solution was adjusted to the final volume of 1000 ml with distilled water and sterilized by autoclaving.

LA plates with ampicillin

Fifteen grams of agar was added to 1 l of LB medium then the media was sterilized by autoclaving. The medium was allowed to cool to 50ºC before adding ampicillin to a final concentration of 100 μg/ml. The medium (30-35 ml) was poured into 85 mm petri dishes. The agar was allowed to harden. Agar plates were stored at 4ºC for up to 1 month or room temperature for up to 1 week.

LA plates with ampicillin / IPTG / X-gal

The LB with ampicillin was prepared then supplement with 0.5 mM IPTG and 80 µg/ml X-Gal and pour the plates. Alternatively, 100 µl of 100 mM IPTG and 20 µl of 50 mg/ml X-gal may be spread over the surface of an LA ampicillin plate and allowed to absorb for 30 minutes at 37ºC prior to use

CURRICULUM VITAE

NAME : Miss Wijit Wisoram

BIRHT DATE : November 18, 1975

BIRTH PLACE : Buri Ram, Thailand

EDUCATION :

YEAR	INSTITUTE	DEGREE/DIPLOMA
1998	Suranaree Univ.Tech.	B.Sc. (Animal production Tech.)
2004	Kasetsart Univ.	M.Sc. (Genetics)

POSITION/TITLE : -

WORK PLACE : Rajamangkala University of TechnologyTawan-ok, Bangpra

SCHORASHIP/AWARDS: Grant for study and this work from the Office of the Higher Education Commission, Thailand (year 2008) and this work was granted from Kasetsart University Development Institut