# Characterization of sex determination genes in lepidopterans

*Thesis submitted to the University of Hyderabad*
*for the degree of*

**Doctor of Philosophy**


*Department of Biochemistry*
*School of Life Sciences*
*University of Hyderabad*

*BY*

**Gajula Gopinath**

**(Regd. No. 10LBPH02)**

CDFD

*Laboratory of Molecular Genetics*
*Centre for DNA Fingerprinting and Diagnostics*
*Hyderabad*


*(June, 2016)*

# Characterization of sex determination genes in lepidopterans

**Thesis submitted to the University of Hyderabad**
**for the degree of**

**DOCTOR OF PHILOSOPHY**

**Department of Biochemistry**
**School of Life Sciences**
**University of Hyderabad**
**500 046**

BY

**Gajula Gopinath**

**(Regd. No. 10LBPH02)**

**Laboratory of Molecular Genetics**
**Centre for DNA Fingerprinting and Diagnostics**
**Hyderabad 500 001**

**(June, 2016)**

**Department of Biochemistry**
**School of Life Sciences**
**University of Hyderabad**
**Hyderabad 500 046**
**India**

---

# CERTIFICATE

This is to certify that the thesis entitled "**Characterization of sex determination genes in lepidopterans**" submitted by Mr. **Gajula Gopinath** to the University of Hyderabad for the award of the degree of Doctor of Philosophy is a bona fide record of the research work carried out by him under my supervision and guidance for a full period prescribed under the PhD ordinance of this University. The content of the thesis, in full or parts have not been submitted to any other Institute or University for the award of any other degree or diploma.

Thesis Supervisor(s)

**Dr. J. Nagaraju**
(demised)
Group Leader and Staff Scientist
Laboratory of Molecular Genetics
Centre for DNA Fingerprinting and Diagnostics
Hyderabad

**Dr. K. P. Arun Kumar**
(reassigned supervisor)
Group Leader and Staff Scientist
Laboratory of Molecular Genetics
Centre for DNA Fingerprinting and Diagnostics
Hyderabad

**Head**

Department of Biochemistry
University of Hyderabad

**Dean**

School of Life Sciences
University of Hyderabad

**Lab of Molecular Genetics**
**Centre for DNA Fingerprinting and Diagnostics**
**Hyderabad 500001**
**India**

---

# CERTIFICATE

This is to certify that the thesis entitled "**Characterization of sex determination genes in lepidopterans**" submitted by Mr. **Gajula Gopinath** to the University of Hyderabad for the award of the degree of Doctor of Philosophy is a bona fide record of the research work carried out by him under my supervision and guidance for a full period prescribed under the PhD ordinance of this University. The content of the thesis, in full or parts have not been submitted to any other Institute or University for the award of any other degree or diploma.

Thesis Supervisor(s)

**Dr. J. Nagaraju**
(demised)
Group Leader and Staff Scientist
Laboratory of Molecular Genetics
Centre for DNA Fingerprinting and Diagnostics
Hyderabad

**Dr. Sanjeev Khosla**
Dean
Centre for DNA Fingerprinting and Diagnostics
Hyderabad

**Dr. K. P. Arun Kumar**
(reassigned supervisor)
Group Leader and Staff Scientist
Laboratory of Molecular Genetics
Centre for DNA Fingerprinting and Diagnostics
Hyderabad

# DECLARATION

I **Gajula Gopinath,** hereby declare that the work presented in my thesis is entirely original and was carried out by me in the Lab of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, under the supervision of Dr**. K. P. Arun Kumar**/Late **Dr. J. Nagaraju.** I further declare that this work has not been submitted earlier, in part or in full, for the award of degree or diploma to University of Hyderabad or any other University or Institution.

Signature of the Supervisor(s):

Gajula Gopinath

Regd. No.: 10LBPH02

# DECLARATION

I **Gajula Gopinath** hereby declare that this thesis entitled "**Characterization of sex determination genes in lepidopterans**" submitted by me under the guidance and supervision of **Dr. K. P. Arun Kumar**/Late **Dr. J. Nagaraju** is a bonafide research work which is also free from plagiarism. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma. I hereby agree that my thesis can be deposited in Shodganga/INFLIBNET.

**A report on plagiarism statistics from the University Librarian is enclosed.**

Signature of the Supervisor(s):

Gajula Gopinath

Regd. No.: 10LBPH02

# Table of contents

# Acknowledgements

To begin with I am thankful to the almighty for blessings which gave me enough support and courage to accomplish my dream of becoming a researcher. My time as a researcher allowed great intellectual freedom and discovery. Research has allowed me to work on the cutting edge of my field. I believe research is about charting new territory, which is an exciting and challenging experience. This journey would not have been possible without the constant support and encouragement of many people. Here I take this opportunity to acknowledge all those people whose full-fledged co-operation and constant advises led me to achieve my goal.

At the foremost, I would like to express my heartfelt gratitude to my supervisor, Late Dr. J. NAGARAJU for giving me such a wonderful opportunity to join in his lab. The kind support, probing questions, careful criticism and remarkable patience he had, helped me in shaping out the outcome of my work.

I would like to express my deepest gratitude and sincere thanks to my reassigned supervisor, Dr. K. P. ARUN KUMAR. You have been a tremendous mentor for me. I would like to thank you for the unconditional support, utmost freedom and encouragement which helped me to grow as a researcher. Your advice on both research as well as on my career have been priceless.

My sincere thanks to former directors OF CDFD, Dr. J. Gowrishankar, Dr. G. R. Chandak and current director of CDFD, Dr. Ranjan Sen for their kind support during my stay in CDFD.

I would like to express my sincere thanks to Dr. Sanjeev Khosla, Dean of CDFD, my doctoral committee members Dr. Sanjeev Khosla, Dr. Gayathri Ramakrishan, Dr. Rashna Bhandari, and Dr. Rupinder Kaur for their passionate participation and input, which helped me complete this work faster.

My special thanks to all the teaching and non-teaching staff of Department of biochemistry, University Of Hyderabad for the timely help and cooperation.

I thank my seniors Dr. V. Sathish, Dr. Mrinal, Dr. Subbaiah, Dr. Jayendra Nath Shukla, Dr. Jyoti, Dr. Chandrapal Singh, Asha Minz, and Suresh Kumar Sawanth who are always there to help and support me with their valuable suggestions both scientific as well as personal. I really cherish the moments spent with them.

The members of Laboratory of Molecular Genetics, past and present form a very friendly and joyful atmosphere that is cherished always. A special thanks to my lab members, V. Satyavati, Annapurna, Lakshmi vaishna, Mutthu lakshmi, Sobhan Babu, Ramesh Babu, Sabahat, Murali, Manju, Swarna, Vandana, Sita Lakshmi, Akanksha, Praveen, Vijaysarathy, Adarsh, Babu, Santosh, Nagaraj, Rajendra, Kushal, Sasi Bhushan, Saikat, Sekhar, Pavani, Swetha, Prasanna,

*Sandhya, Vidya, Sunil, Rajeev, Chandrasekhar, Balaraj, Karunakar, Bhaskar, Balaraj and Wajida for their timely help.*

*I extend a heartfelt thanks to these colleagues; Dr. Keethana, Archana Tomar and Ch. Gangi reddy. They really helped me in completion of this work.*

*Finally I am really thankful to Council of Scientific and Industrial Research (CSIR) and CDFD for providing research fellowship. Special thanks to CDFD for funding my visit to attend conference held at Cold spring harbour laboratory (CSHL) in 2015, New York, USA.*

*I thank Department of Science & Technology (DST) and Japan Society for the Promotion of Science (JSPS) for funding my visit to University Of Tokyo, Japan in 2013 as a part of India-Japan Cooperative Science Programme (IJCSP).*

*Words are not enough to thank my wife, Archana Tomar for loving me unconditionally and standing beside me throughout my career and life. She has been a motivator and source of happiness and inspiration during my tough times. I wish to have the same support, love and care from her forever.*

*I would also like to thank all my friends who supported me in all possible ways, and helped me to strive towards my goal.*

*I am indebted to my parents who provided me unlimited love and support throughout my Ph.D. Their constant motivation and blessings kept me going in the most difficult situations. Their prayers for me are the reason I could sustain this far.*

Gopinath Gajula

# ABBREVIATIONS

| | |
|---|---|
| BAM | Binary alignment of SAM |
| BLAST | Basic local alignment search tool |
| 3' UTR | 3' Untranslated region |
| 5' UTR | 5' Untranslated region |
| BmIMP | Bombyx homolog of IGF-II mRNA binding protein |
| BmPSI | Bombyx homolog of P-element somatic inhibitor |
| CDC | Complete dosage compensation |
| CDD | Conserved domain database |
| cDNA | complementary DNA |
| CDS | Coding sequence |
| CE 1 | Conserved element 1 |
| CPM | Counts per million |
| Da | Daltons |
| DC | Dosage compensation |
| DCC | Dosage com |
| DDPCR | Differential display PCR |
| DGE | Differential gene expression |
| DNA | Deoxyribo nucleic acid |
| *Doa* | Darkener of apricot |
| dsRNA | double stranded RNA |
| DSX | Double sex |
| FBS | Fetal bovine serum albumin |
| FDR | False discovery rate |
| Fem | Feminizer |
| FISH | Fluorescent in situ hybridization |
| FPKM | Fragments per kilobase of exon per million fragments |
| *fru* | fruitless |
| GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |
| GSH | Genomic subtractive hybridization |

| | |
|---|---|
| IDC | Incomplete dosage compensation |
| IQR | Inter quartile range |
| IRES | Internal ribosome entry site |
| *ix* | intersex |
| JPSL | Japanese sex limited strain |
| kDa | Kilo Daltons |
| lnRNA | long non-coding RNA |
| logFC | log 2 of fold change |
| MA plot | M (log ratio) and A (mean average) plot |
| *Masc* | Masculinizer |
| MBE | Musashi binding element |
| MBNL3 | Muscleblind like protein 3 |
| mg | milligram |
| miRNA | micro RNA |
| mRNA | messenger RNA |
| MSL 2 | Male specific lethal 2 |
| MSL 3 | Male specific lethal 3 |
| MSL1 | Male specific lethal 1 |
| MWU | Mann Whitney U-test |
| NCBI | National centre for biotechnology information |
| NLS | Nuclear localization signal |
| NTC | negative control |
| PBS | phosphate buffer saline |
| PCR | Polymerase chain reaction |
| piRNA | Piwi-interacting RNA |
| QGSLO | Japanese sex marked strain QGSLO |
| qRT-PCR | quantitative Real time-PCR |
| RAPD | Random amplification of polymorphic DNA |
| RNA | Ribonucleic acid |
| RNAi | RNA interference |
| RNA-Seq | RNA-sequencing |

| | |
|---|---|
| SAM | Sequence alignment/map |
| siRNA | small interfering RNA |
| Sp1 | Sex specific storage protein 1 |
| *Sp25* | Silk protein 25 |
| *Sxl* | Sex-lethal |
| TE | Transposable element |
| *Tra* | Transformer |
| *Tra-2* | Transformer-2 |
| uORFs | Upstream open reading frames |
| *Vg* | Vitellogenin |
| XSE | X-linked signalling elements |
| µg | microgram |
| µl | microliter |

# SYNOPSIS

Sex determination is one of the most fundamental and astoundingly diverse biological processes in organisms. Evolution of sex determination should have started very long ago in living beings, which is evident from the appearance of two distinct sexes (male and female) in diverse organisms from corals to mammals (ARNOLD 2012; ARNOLD *et al.* 2013; CUTTING *et al.* 2013; GAMBLE and ZARKOWER 2012). A variety of sex determination mechanisms is observed in animals, most of which follow the chromosomal/genetic sex determination but in some reptiles the sex is determined by environmental factors like temperature at which eggs are incubated (crocodilians and turtles), settling of larval position (echiuroid worm *Bonellia*) and based on the opposite sex (slipper snail *Crepidula fornicata*). In the worm, *Caenorhabditis elegans* embryos with XX develop as females and hermaphrodites, the XO develop as males and the X/A ratio determines the sexual fate (MADL and HERMAN 1979). In mammals, the Y-chromosome confers maleness thus XY embryos develop into males and XX embryos develop into females (BERTA *et al.* 1990). In birds, insects and reptiles females are heterogametic (ZW) and males are homogametic (ZZ). In insects, the mechanism of sex determination is highly diverse and as there are no sex hormones the sex of each cell is maintained autonomously. Some species lack sex chromosomes where an autosomal gene controls the sex differentiation. In *Drosophila melanogaster*, the dose of the X-linked signalling elements (XSE) is the primary signal for sex determination (ERICKSON and QUINTERO 2007). In hymenopterans (*Apis mellifera*) the ploidy of the genome determines the sex, where the diploids develop as females and haploids develop as males (BEUKEBOOM 1995). In insects, *Megaselia scalaris*, *Ceratitis capitata, Lucilia cuprina,* and *Chironomous thummi* an epistatic male factor on Y chromosome confers maleness (BEDO and FOSTER 1985; HAGELE 1985; TRAUT 1994; WILLHOEFT and FRANZ 1996). In the order lepidoptera (moths and butterflies), females are heterogametic (ZW) and presence of W is believed to confer female development (ABE *et al.* 2008; HASHIMOTO 1933).

Sex determination is followed by the process of sexual differentiation. In insects, the gene that governs this sexual differentiation is the *doublesex* (*dsx*), which codes for a transcription factor. Although the upstream factors of sex determination in several species are diverse, their bottom most gene of sexual differentiation i.e., *dsx* is highly conserved (Wilkins “bottom up” theory). This gene exhibits differential splicing in sexes by taking advantage of alternative splicing; i.e., in males it produces *dsxm* and in females it produces

*dsxf* types of splicing isoforms, resulting in different protein isoforms between the sexes. These proteins vary at their C terminal domain and are believed to be counteracting to each other at the level of gene expression via transcription, which leads to differences in the process of sexual differentiation. Thus the differential splicing of the *dsx* is a crucial factor that governs the sexual differentiation in the mechanism of insect sex determination pathway.

In domesticated silkworm *Bombyx mori*, W-chromosome ensures the development of zygote to female. Unlike *D. melanogaster* where triploids (XXY+3A) develop into intersexes and tetraploids (XXXY+4A) into normal females, in *B. mori* devoid of such ploidy differences, both the triploids (ZZW+3A) and tetraploids (ZZZW+4A) having W-chromosome develop into females, suggesting the presence of a strong epistatic factor on W-chromosome that directs the embryo to female development. Recent study on unravelling sex determination genes of *B. mori* has resulted in the identification of a W encoded piRNA called *fem* (KIUCHI *et al.* 2014). This piRNA is shown to be negatively regulating a CCCH zinc finger gene, *Masculinizer* (*Masc*), which is shown to be involved in differential splicing of *dsx* pre-mRNA (KIUCHI *et al.* 2014). In parallel to these findings, Satish and Azimura *et al.,* have also identified a group of CCCH and C2H2 zinc finger motif encoding genes on W-chromosome and their homologous copies on the 25$^{th}$ chromosome by employing differential display PCR (DDPCR) and genomic subtractive hybridization (GSH) techniques (SATISH *et al.* 2006). However, their existence on W-chromosome and role in sex determination remains to be validated.

I initiated my thesis work with the objective of characterization of the previously discovered CCCH zinc finger gene on 25$^{th}$ chromosome and its role in *B. mori* sex determination. At first (**Chapter I)**, sequence and expression analysis of the *Bmznf-1* and *Bmznf-2* transcripts was done using RNA isolated from 5$^{th}$ instar larval fat body of male and female larvae. The results suggest the presence of allelic variants for the autosomal copy, *Bmznf-2* gene. Further the relative overall expression of these CCCH zinc finger genes in various somatic tissues (fatbody, midgut, silkgland and head) showed a female biased expression and in gonads (ovary and testis) their expression is male biased. This expression trend suggests that these genes may be crucial for female somatic differentiation and they may be having a different function in gonads. Further, functional characterization of the autosomal copy of CCCH zinc finger motif encoding gene *Bmznf-2* was carried out (**Chapter II**) to study its role in *B. mori* sex determination. Several

analyses were conducted for functional characterisation of this gene in the ovary derived BmN cells that exhibit female mode of sexual differentiation by producing *Bmdsxf* type of splicing isoform. If *Bmznf-2* is a sex determination gene, a significant variation in its expression should affect the differential splicing of the *Bmdsx* gene. To test this, RNAi based knockdown of *Bmznf-2* was done in BmN cells and found no effect on the differential splicing of *Bmdsx*. But the over expression of *Bmznf-2* led to the production of *Bmdsxm*, male type of splicing in the female BmN cells suggesting that *Bmznf-2* may be a masculization factor. This splicing phenotype suggested the gain of function of *Bmznf-2* in BmN cells. Further, the already reported genes of sex determination cascade like *Bmpsi*, *Bmimp* and *Bmtra-2* were checked for their differential expression or splicing upon over expression of *Bmznf-2* in BmN cells. It was observed that *Bmtra-2* showed an enhanced expression and aberrant splicing leading to shorter splicing isoforms that code for TRA-2 proteins that lack RS2 domain.

The over expression of *Bmznf-2* CDS can also be viewed as over expression of the precursor for the ovarian small RNA-12564 as the CDS of *Bmznf-2* contains 100% sense to sense match with this RNA. If the protein product of *Bmznf-2* is responsible for the altered splicing phenotypes of *Bmdsx* and *Bmtra-2*, then are the CCCH tandem zinc finger motifs of BmZNF-2 are involved? To address this question, I did mutational study for the CCCH motif-1 and motif-2 individually and found that the altered splicing phenotypes of *Bmdsxm* and *Bmtra-2* are lost in BmN cells. This suggests that it is not the ovarian small RNA-12564 but the protein product of *Bmznf-2* and both the CCCH motifs are involved in the altered splicing of *Bmdsx* and *Bmtra-2* genes. As the process of splicing occurs in the nucleus of cells, I tested whether BmZNF-2 localizes to nucleus. For this immunostaining and protein tagging with m-cherry was performed and found the nuclear localisation of BmZNF-2, this suggests its activity in nucleus probably involved in alternative splicing. These experiments clearly suggest that *Bmznf-2* is a masculization factor that favours the male type of *Bmdsx* splicing. But the question arises why this gene does not masculinize the BmN cells normally and is there any endogenously driven negative regulation of this gene in BmN cells. To answer this I did 5'UTR analysis of *Bmznf-2* using dual luciferase assay and found translational repression of this gene in BmN cells. These experiments suggest that *Bmznf-2* is involved in sex determination and is a strong masculinization factor. It is also capable of inducing aberrant splicing of *Bmtra-2* resulting in shorter proteins that lack RS2 domain, whose function is unknown. A strong translational

repression operates on this gene, which could be the reason for its non-functionality in female cells (BmN).

Additionally in an attempt to identify possible new players of sex determination and W-encoded genes, RNA-sequencing was performed for early embryonic stages (**Chapter III**). The embryonic stages were selected based on the observation that *dsx* gene exhibits sex specific differential splicing at 96h. Hence, a stage before (78h) and a stage after (120h) 96h were selected for analysis. Analysis of these three stages suggested an early male biased expression at 78h and 96h stages, which gets normalized at 120h stage. The differential gene expression analysis has revealed a set of male biased and female biased genes at 78h, 96h and 120h stages. For the identification of W-derived fragments, the genome unmapped reads were subjected to de novo-assembly. This resulted in thousands of unmapped transcripts with ~200bp length (from male samples = 5726; female = 4667). BLAST analysis showed that nearly 50% of these transcripts (male = 2596; female = 2365) could be the precursor transcripts for the reported ovarian small RNAs in *B. mori*. These transcripts were further subjected to various levels of filtering, which resulted in 862 novel transcripts in which 225 were identified only in female samples and 423 were identified only in male samples. Out of the 225 female specific transcripts, 62 transcripts were predicted to be of W-origin based on the BLAST analysis against the W-chromosome derived BAC clones. Unfortunately no protein coding transcripts were identified among them and all the transcripts were non-coding in nature.

Further, dosage compensation (DC) and the sex determination are two intimately related processes. For example, in *Drosophila* the double dose of XSE decides the female sex (XX). Similarly in *C. elegans* the dose of "X-signal elements" determines the XX embryos to develop into females/hermaphrodites and in wild silk moths, where females are ZZ and males are ZO, the double dose of a subset of Z-linked loci is presumed to be determining the sex of the organism. A recent study has also revealed that a sex determination gene, *masc* also regulates dosage compensation in *B. mori* (KIUCHI *et al.* 2014). The sex chromosomal dose difference between sexes is often normalized by a gene regulatory mechanism called dosage compensation. DC mechanism operates in a fashion not only to reduce the sex bias in the expression of sex linked genes, but also to reduce the expression disparity of sex linked and autosomal genes (NGUYEN and DISTECHE 2006). DC in an organism is believed to be complete when the average expression of autosomes and sex chromosomes are equal or the gene expression difference is statistically insignificant in

both sexes. Generally, in XX/XY systems this pattern of DC is observed, whereas in case of ZZ/ZW systems this is not very common. The ZZ/ZW systems tested for DC were found to show a constant male biased expression of the Z-linked genes suggesting the incomplete DC. Hence it is believed that DC patterns are generally common in XX/XY systems than in ZZ/ZW systems. In *Bombyx mori* (lepidopteran), the initial reports for DC, based on expression profile of a set of a few Z-linked genes has suggested an incomplete DC, where males showed higher expression of Z-linked genes, (SUZUKI *et al.* 1998; SUZUKI *et al.* 1999). This was further confirmed by global, microarray analysis (ZHA *et al.* 2009). However, another study using reanalysis of the same microarray data draws a clue for the possibility of a globally operating DC mechanism in *B. mori* (WALTERS and HARDCASTLE 2011). In order to address this issue, the RNA-sequencing data of the embryonic stages (78h, 96h and 120h), was analysed to explore the real patterns of DC (**Chapter IV**). The RNA-sequencing data of male female larval head and BmN cells were used as reference samples for the developmentally more dynamic embryonic stages (78h, 96h and 120h). The results showed that the dosage of Z-linked genes is half to that of autosomes. In the early stages of embryos i.e., 78h and 96h, there is a male biased expression of Z-linked genes, which gets normalized at 120h stage. Based on this observation, I propose that after 96h stage, the embryos attain compensation of Z-linked gene dose. The second aspect of the study was how DC could have been achieved in *B. mori*. As the average Z-linked gene dose is half to that of autosomes and as it is the same even in the males with two Z-chromosomes, it suggests a probable suppression of Z-linked genes in males for equalizing the expression to that of females. Finally, we conclude that DC in *B. mori* to be complete as there is no statistically significant difference observed between Z-linked gene expressions between sexes at presumably dosage established samples like 120h embryos and larval head. But the patterns observed in *B. mori* suggest an unconventional and a unique type of DC mechanism.

"How sex is determination in species?" this puzzling aspect of biology had resulted in a pursuit, nearly a century ago to study the molecular mechanism behind this process. This has revealed an array of genetic cascades mostly determined by sex chromosomes. Studies on understanding the mechanism of sex determination in various taxa have led to the proposal of bottom-up theory by Adam Wilkins (WILKINS 1995), where the bottom most player of the cascade is highly conserved but the top players are diverse. In insects, sex is not influenced by hormones and every cell maintains its own sex, hence gynandromorphs are possible. The sex determination cascade involves a primary signal

mostly genetical, coming from sex chromosomes that activates a "key gene", which inturn takes control of sub-ordinate control genes - finally driving the double switch (*dsx* gene). The striking differences between male and female originate from the differential splicing of *dsx* pre-mRNA, producing sex-specific proteins that are antagonistic in the process of sexual differentiation and development. In most of the insects studied for sex determination, there is a conservation to some extent at the level of "key gene" (*tra*), whereas this gene is not found in *B. mori* by homology search. Additionally there seems to be many regulatory factors involved in the sex specific differential splicing of *B. mori dsx* pre-mRNA (*Bmdsx*) Eg., *Bmpsi*, *Bmimp, Masc* and currently identified *Bmznf-2*. These two observations make the cascade of sex determination in *B. mori*, remarkably different from that of other insects. The process of sex determination is undoubtedly linked to a vital process called "Dosage Compensation". The evolution of hetero-sex chromosomes (X and Y or Z and W) has created sex chromosome aneuploidy that resulted in sex-linked gene dose differences (X in males and XX in females or ZZ in males and Z in females), which has to be compensated for the sexual fitness and survival of the organism. It is interesting to check the mechanism of DC in *B. mori* where the sex determination cascade is remarkably different as it involves a bunch of novel genes or factors (*fem* piRNA, *Masc*, *BmPSI, BmIMP* and *Bmznf-2*). It is found that DC in *B. mori* is complete as ZZ=ZW as the double dose of ZZ expression in males is equal to that of single dose of Z in females, but it violates the evolutionary precedent of "Sex chromosomal expression should be equal to that of autosomal expression" ie., (male ZZ= female ZW≠ A) instead it is nearly half of autosomal (male ZZ=female ZW=A/2). In my thesis work, I have shed light on 1) the mechanism of *B. mori* sex determination cascade by revealing the sex determination function of *Bmznf-2* and 2) how DC is achieved, by RNA-seq analysis. Altogether, these two processes i.e., Sex determination and DC in *B. mori* should be recognised as remarkably different from other insects and further research is required for exploring even more exciting features in these areas of research.

# Chapter-I

## Analysis of a novel group of CCCH zinc finger genes in *Bombyx mori*

## 1.1 Introduction

In female heterogametic systems, the ZZ embryos develop into males and ZW or ZO embryos develop in to females (TRAUT *et al.* 2007a). This is found in almost all the female heterogametic species studied so far, but the role of the sex chromosomes in the process of sex determination in not revealed. The possibilities include 1) 'dominant-W' mechanism where the female determining gene on W-chromosome drives the ZW-embryo to female development whereas the absence of W-chromosome lead to default male development. 2) 'Z counting' mechanism where the embryos with a single Z-chromosome develop in to females and with two ZZ-chromosomes develop in to males and even the 3) 'recessive Z' mechanisms operate in lepidopterans (TRAUT and MAREC 1996). The polyploidy and aneuploidy analysis has revealed the existence of "dominant W' in the domestic silkworm, *B. mori*. The region of W-chromosome that is capable of determining the femaleness called "Fem" is located by a couple of deletions and translocation studies in *B. mori* (ABE *et al.* 2008; ABE *et al.* 2005). Further a very recent report suggests that a piRNA called "*fem*" originating from W-chromosome is responsible for governing the femaleness (KIUCHI *et al.* 2014). This piRNA was also shown to be originating from the sex determining region of the W-chromosome (Fem) (KATSUMA *et al.* 2014) and down regulating a CCCH zinc finger gene, *Masculinizer* (*Masc*) that regulates the differential splicing of *Bombyx* double sex gene (*Bmdsx*). However, in parallel studies (from our laboratory) attempting to discover the genes involved in *B. mori* sex determination through differential display PCR (SATISH *et al.* 2006) and genomic DNA subtractive hybridization (AJIMURA *et al.* 2006) have resulted in identification of two novel groups of zinc finger genes (CCCH - *z1*, *z2*, *z3* and C2H2 - *z20*, *z21*, *z22*). Further analysis revealed the existence of both W-linked and autosomal ($25^{th}$ chromosome) copies of these genes (Figure 1). It was hypothesized that an ancestral autosomal pair (*z2* + *z21*) on $25^{th}$ chromosome has duplicated onto the W-chromosome (*z1*+ *z20*), thus acquiring unknown sex specific function and this copy has been further duplicated on to the W-chromosome as *z1-1b* + *z20-1b*, *z1-2* +*z20-2* and *z1-3* +*z20-3*. In addition, the translocation studies of the W-chromosomal fragments to autosomes have supported the existence of a strong putative epistatic female determining region called "Feminizer" (Fem) on the W-chromosome (HASHIMOTO 1933; TAZIMA 1954). A preliminary analysis using FISH has indicated that probably these zinc finger genes are linked to the "Fem" region of W-chromosome (AJIMURA *et al.* 2006; SATISH *et al.* 2006) and hence it is very appealing to discover the role of these novel zinc finger genes (CCCH and C2H2, Figure 1) in *B. mori* sex determination pathway.

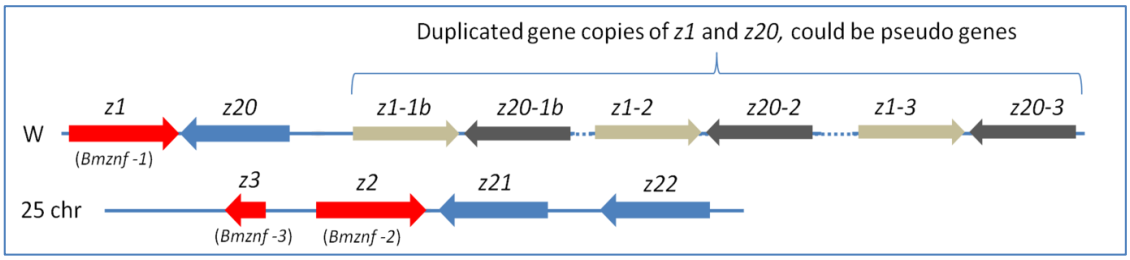


**Figure 1.** Genomic organisation of a novel group of CCCH and C2H2 zinc finger genes in *B. mori*. The genes *z1*, *z2*, and *z3* are referred to as *Bmznf-1*, *Bmznf-2* and *Bmznf-3* respectively.

Gene duplications have an exceptional role in the evolution of genes involved in sexual development or sex determination. To mention few examples, e.g. 1: the genes *sxl* and CG3056 are homologs generated by a duplication event in dipteran lineage. *Sxl* has adopted the function as a sex determiner gene whereas the parent gene CG3056 is believed to serve the same ancestral unknown functions (TRAUT *et al.* 2006). Hence in *D. melanogaster*, *D. erecta*, and *D. pseudoobscura* s*xl* and CG3056 are two paralogs (homologs with different function) which are orthologs (homologs with same functions) to the single s*xl* gene in *Anopheles*, *Apis*, *Bombyx*, and *Tribolium*. E.g. 2: In *Oryzias latipes* (medaka fish) the homolog *DMRT 1* on an autosome is clustered together with its paralogs, *DMRT 2* and *DMRT 3* (LUTFALLA *et al.* 2003; MATSUDA *et al.* 2003). A region on Y-chromosome contains a homolog of *DMRT1* gene, called *DMRT1Y*, which is expressed in male embryonic, larval and in sertoli cell of adult testes. E.g. 3: In *Danio rerio* (zebra fish), the gene *SOX9* gave rise to *SOX9a*/*SOX9b* gene pair by duplication (KLUVER *et al.* 2005). *SOX9a* is involved in testis development while SOX9b is considered to be involved in ovary development. Though both proteins contain similar DNA-binding and transcactivation properties, they show different expression profiles. E.g. 3: In *Xenopus laevis* (frog), *DM-W* a vital gene involved in ovary development, is considered to have evolved by the duplication of a testis formation gene *DMRT1*, hence both these genes are considered as paralogs (YOSHIMOTO *et al.* 2008). E.g. 4: The mammalian gene *Dmtr7* which plays a critical role in chromatin transition in meiosis is specific to males and is considered to have emerged by duplication of a *Dmrt* gene (KAWAMATA *et al.* 2007; KAWAMATA and NISHIMORI 2006). E.g. 5: In primates, the X-linked gene SCML1, with testis specific function, is thought to have arisen due to a duplication event from its ancient *SCML2* and *SCMH1* which have different functions (WU and SU 2008). It is apparent from all these examples that gene duplication is the root event in the evolution of genes involved in the process of sexual development or sex determination. In the present context, the

homologs of a novel CCCH and C2H2 zinc finger genes (Figure1) identified on W-chromosome and on 25$^{th}$ autosome in *B. mori* depict a similar scenario, giving an initiative for their function in the early sexual development or sex determination.

In general the zinc finger genes are involved in a variety of basic molecular functions like mRNA stability (CHOI *et al.* 2014; WELLS *et al.* 2015; YANG *et al.* 2015), transcription (OGO *et al.* 2015), translational repression (HUANG and HUNTER 2015), mRNA splicing (KONIECZNY *et al.* 2014), protein-protein interactions (MATTHEWS *et al.* 2000) and thus mediate myriad metabolic and developmental processes like DNA repair (PIERACCIOLI *et al.* 2015), insulin uptake (BUCHNER *et al.* 2015), gene silencing via epigenetic modulations (ICHIDA *et al.* 2015), follicular development (EFIMENKO *et al.* 2013), reproduction (KREBS and ROBINS 2010), salt resistance (HAN *et al.* 2014), drought tolerance (WANG *et al.* 2015), fertility and embryonic development (DE *et al.* 1999; LI and THOMAS 1998; RAMOS *et al.* 2004), etc. Therefore, *zinc finger* genes have evolved as crucial players in diverse functions, representing a group of key regulators of gene expression.

Zinc fingers (CCCH) are also reported to be involved in sexual differentiation and reproduction. In *Caenorhabditis elegans*, several C-x8-C-x5-C-x3-H type zinc finger genes like PIE-1, MEX-1, MEX-5/6 and POS-1, have been reported as maternal factors that are essential for the germ cell differentiation (DRAPER *et al.* 1996; GUEDES and PRIESS 1997; REESE *et al.* 2000; SCHUBERT *et al.* 2000; TENENHAUS *et al.* 2001). The RNA binding proteins, MEX-5 and MEX-3 are shown to be involved in translational repression by binding to 3'UTR of *pal-1* mRNA in the anterior embryo and thus contributing to asymmetric gene expression (HUANG *et al.* 2002), which is essential in development (HUANG and HUNTER 2015). Three zinc finger genes *moe-1*, *moe-2*, *moe-3* (SHIMADA *et al.* 2002) and recently discovered redundant OMA-1 and OMA-2 (KAYMAK and RYDER 2013) are involved in oocyte maturation. In the zebra fish, *Danio rerio*, the *zinc finger* gene *Zfcth1* codes for two putative CCCH zinc fingers that are essential for oocyte maturation (TE KRONNIE *et al.* 1999). In *Xenopus*, an ovary specific gene XC3H-4 with four zinc finger motifs is speculated to be involved in the regulation of mRNA stability in oocyte maturation and/or early embryogenesis (DE *et al.* 1999). In mice an mRNA-binding and destabilizing CCCH zinc finger protein, *Zfp36l2* functions in the control of female fertility physiologically (BALL *et al.* 2014; RAMOS *et al.* 2004). In *B. mori*, a CCCH zinc finger

motif encoding gene, *masc* is shown to be regulating the dosage compensation and also a regulator of *Bmdsx* splicing (KIUCHI *et al.* 2014). Thus the zinc finger genes have been evolved to be crucial factors that play a vital role in the physiology and development of organisms.

In the present work, I have studied the novel CCCH group of zinc finger genes, *Bmznf-1* and *Bmznf-2* for their allelic variants and overall expression in various somatic tissues and gonads. As this novel group CCCH zinc finger genes is associated with the W-chromosome (female determinant), they may be involved in the process sexual development of *B. mori*.

## 1.2 Materials and Methods

### 1.2.1 RNA isolation from larval fat body and RT-PCR

Total RNA from 5$^{th}$ instar larval fat body (~25 mg) was isolated by TriZol (Invitrogen) method. To remove DNA contamination, the total RNA was in solution DNase-I (Ambion) treated. The RNA concentration of samples was measured using Nanodrop 2000 (Thermo Scientific). cDNAs were synthesized using 1 µg of total RNA, using SuperScript III (Invitrogen). PCR was set using Emerald Amp GT PCR 2X master mix (Takara). The reaction conditions were 94˚C for 5 min, 35 cycles of 94˚C for 30 sec + 60˚C for 30 sec + 72˚C for 1 min and final elongation step of 10 min at 72˚C. The PCR samples were analysed on 1.5% agarose gels cast with ethidium bromide.

### 1.2.2 Cloning and sequencing of *Bmznf-1* and *Bmznf-2* transcripts from male and female fat body samples

The coding sequences (CDS) of *Bmznf-1* and *Bmznf-2* (1083 bp) were PCR amplified with the primers Zn2-F-SacI(5' GCGCGGGAGC TCGGGATGAAAAAATACTTAAAATAC3') and Zn2-R-NotI (5' TATTAGCGGCCGCCGCGATA CGTGCGTCTTATC 3') using cDNAs synthesized from the fat body of 5$^{th}$ instar male and female larvae independently to ensure isolation of transcripts sex specifically. The resultant amplicons were cloned into the pIZT-V5 insect glow vector (Invitrogen) using SacI and NotI restriction enzymes. The clones were selected on low salt Zeocin LB plates and the sequences were confirmed by OpIE2 F and OpIE2 R sequencing primers.

### 1.2.3 Relative quantification of the zinc finger genes *Bmznf-1, Bmznf-2* and *Bmznf-3* by quantitative RT-PCR

The relative expression level of *Bmznf-1, Bmznf-2 and Bmznf-3* in male and female somatic tissues and gonads was validated through quantitative real-time PCR (ABI 7500). The reaction was set using SYBR Premix Ex Taq (Tli RNaseH Plus, Takara Bio Inc) with cDNA sample of 3 µl (diluted to10 ng/µl) and 0.2 µM primers, in a final volume of 20 µl. Reaction conditions were: 95˚C for 30 sec, 95˚C for 5 sec and 60˚C for 34 sec. ABI SDS software version 1.2.3 was used for standard curve analysis. The relative expression was determined using ΔCt analysis and the reactions were carried out in triplicates. GAPDH was used as endogenous reference control.

## 1.3 Results

### 1.3.1 Sequence and expression analysis of a novel group of CCCH zinc finger genes in *B. mori*

Among the CCCH zinc finger genes (*Bmznf-1* and *Bmznf-2*) the W-linked *z1* (*Bmznf-1*) and the autosomal *z2* (*Bmznf-2*) copies exhibit a very high degree of sequence conservation, 98% identity at mRNA (Figure 2) and 96% identity at protein levels (Figure 3), suggesting that these two genes are homologs (could be paralogs or orthologs) (AJIMURA *et al.* 2006; SATISH *et al.* 2006). The CDS of *Bmznf-1* and *Bmznf-2* is 1080 bases, which codes for 360 aa proteins of PI/MW = 8.67/41.7 kDa and 8.77/41.6 kDa respectively. But the autosomal gene *Bmznf-3* encodes for a shorter protein (PI/MW= 8.81/26.8 kDa) with an extra 30 a.a N-terminal sequence (Figure 4). The CDS regions of *Bmznf-1* and *Bmznf-2* transcripts vary at 25 positions (Figure 2) that resulted in only 14 a.a differences at their protein sequences, due to 11 synonymous substitutions.

The 5'UTR region of these zinc finger genes (~600 bp) is highly conserved between these two genes. UTRScan (itb tools.ba.itb.cnr.it/utrscan) revealed two upstream open reading frames (uORFs) in 5'UTR. It is believed that in eukaryotes, these uORFs are generally involved in the translational regulation of genes (BARBOSA *et al.* 2013). These transcripts also has two target sites for ovarian small RNAs, one on 5' UTR region [ovarian small RNA-24319 (GenBank accession: AB410509)] and the other spanning the junction of 5' UTR and start codon [ovarian small RNA-37041 (GenBank accession: AB423231)] (Figure 5). The CDS of these transcripts may be acting as precursors for the ovarian small RNA-12564 (GenBank accession: AB398754) based on 100% match in sense strand to sense strand orientation. The 3' UTR of these zinc finger genes are of ~400 bp and are highly conserved. The 3'UTR regions of these transcripts also have one internal ribosome entry site (IRES), two uORFs and three Musashi binding elements (MBE). Unlike the 5'UTR regions, the 3'UTR regions of these genes don't have target binding sites of any reported *B. mori* ovarian small RNAs.

```
Bmznf-1      ATGAAAAAATACTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGA 60
Bmznf-2      ATGAAAAAATACTTAAAATACAAGATCGGGGAGGAGAGTCACTCGCCCCAATATCATGGA 60
             ************************ ************* **********************

Bmznf-1      GTAGCAGAAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAG 120
Bmznf-2      GTAGCAGAAATCTCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAG 120
             ************ ************************************************

Bmznf-1      GTTCATCCACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAAT 180
Bmznf-2      GTTCATCCACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAAT 180
             ************************************************************

Bmznf-1      TTCAATAAGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAA 240
Bmznf-2      TTCAAAAAGGAAATCAATGAATGACGAGGAGACACCTTCAATAACCATAAATATCAAGAAA 240
             ***** *********** ** ******************** ******************

Bmznf-1      GATCCAACTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTT 300
Bmznf-2      GATCCAACTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTT 300
             ************************************************************

Bmznf-1      GTGCGTAACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCC 360
Bmznf-2      GTGCGTAACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCC 360
             ************************************************************

Bmznf-1      CAACTTAAAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCA 420
Bmznf-2      CAACTTAAAGGAGTTTACAAATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCA 420
             ****************** ****************************************

Bmznf-1      GAATGTTCATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTAT 480
Bmznf-2      GAATGTTCATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTAT 480
             ************************************************************

Bmznf-1      TTACCATCACATACATTGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACC 540
Bmznf-2      TTACCACCAAATACATTGAGCCATATTAAGAAAAAAACAGTCCTACCTCCAGCTAAAACC 540
             ****** ** ********************************* ****************

Bmznf-1      AAAGAAACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACC 600
Bmznf-2      AAAGAAACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACC 600
             ************************************************************

Bmznf-1      ACTTCTGTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAG 660
Bmznf-2      ACTTCTGTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCATACGAG 660
             ********************************************************** ******

Bmznf-1      AAGCAATGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCAT 720
Bmznf-2      AAGCAATGGCCGGAAATGGAAGACAAAACGTCGCCCGAGATGGCATATTATAGTGGGCAT 720
             *************************** ********************************

Bmznf-1      CCATCTTATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATA 780
Bmznf-2      CCATCTTATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATA 780
             ************************************************************

Bmznf-1      GTGCAAAATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAA 840
Bmznf-2      GTGCAAAATCCATACGTCACTGTAATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAA 840
             **************** ****** ************************************

Bmznf-1      GAACCCGTGAGTCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAG 900
Bmznf-2      GAACCCGTGAGTCAACCCACGCACGAGTACGCAGAGCTACGCTGAAACGGCTGGATCCAAG 900
             ******************************** ** ************************

Bmznf-1      AAATGCAGAAACTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATG 960
Bmznf-2      AAATGCAGAAACTGTGATGTCAACGAGTTCAGATTCCAGCAAAACAAAAACAAAATCATG 960
             ****************************************** ******************

Bmznf-1      AAAATGATAAAAGACACAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAAC 102(
Bmznf-2      AAAATGATAAAAGACACAGACGACTTGAATCATCGCGTGGGCCAAATAACTAAAAAGAAC 102(
             ****************************** ********** ******************

Bmznf-1      ACTAAATTGAATGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGC 108(
Bmznf-2      ACTAAATTGAATGAAGTACTTGTGCTGCTCGTCGAAGTTCTGATAAGACGCACTTTTCGC 108(
             ************** ******** ************** ************* * ****

Bmznf-1      TAG       1083
Bmznf-2      TAG       1083
             ***
```

**Figure 2.** The ClustalO alignment of *Bmznf-1* and *Bmznf-2* CDS regions to show a high degree of conservation, these two sequences vary at 25 positions (highlighted in red).

```
BmZNF-1         MKKYLKYKFGEESHSPQYHGVAEIPEAEAMKRKKHEQTEKVHPPLPNRPPLPPPEPLLAN 60
BmZNF-2         MKKYLKYKIGEESHSPQYHGVAEISEAEAMKRKKHEQTEKVHPPLPNRPPLPPPEPLLAN 60
                ********:*************** **********************************

BmZNF-1         FNKEINEDEETPSITINIKKDPTENYEKPNDVEETMCRNFVRNTCNRGASCRYLHKIIHS 120
BmZNF-2         FKKEINDDEETPSITINIKKDPTENYEKPNDVEETMCRNFVRNTCNRGASCRYLHKIIHS 120
                *:****:**************************************************

BmZNF-1         QLKGVYRFCIDFENKKCTRAECSYAHATVHEKEHFFRTGYLPSHTLSHIKKKTVQPPAKT 180
BmZNF-2         QLKGVYKFCIDFENKKCTRAECSYAHATVHEKEHFFRTGYLPPNTLSHIKKKTVLPPAKT 180
                ******:********************************** .********** *****

BmZNF-1         KETSSELSANYTGTTYPHSTTSVPMYTNASNISMMQNPYEKQWPEMEDKSSPEMAYYSGH 240
BmZNF-2         KETSSELSANYTGTTYPHSTTSVPMYTNASNISMMQNPYEKQWPEMEDKTSPEMAYYSGH 240
                ************************************************:**********

BmZNF-1         PSYPHSTSTGYPNAPPNTPIVQNPYATVMSPTKRQWTEMEEPVSQPTHEYAEYAETAGSK 300
BmZNF-2         PSYPHSTSTGYPNAPPNTPIVQNPYVTVMSPTKRQWTEMEEPVSQPTHEYAEYAETAGSK 300
                ************************.********************************

BmZNF-1         KCRNCDVNEFRFQHNKNKIMKMIKDTDDLNYRVGQITKKNTKLNEILVLLVEVLIRRTYR 360
BmZNF-2         KCRNCDVNEFRFQHNKNKIMKMIKDTDDLNHRVGQITKKNTKLNEVLVLLVEVVIRRTFR 360
                ****************************:**************:*******:****:*
```

**Figure 3.** The ClustalO alignment of BmZNF-1 and BmZNF-2 proteins to show a high degree of conservation at the level of protein sequences. These two protein sequences vary at 14 a.a positions.

```
BmZNF-3         MRSSDAGQTNFQSIGDEDSINFQPSPNFKHMKKYLKCKFGEESHSPQYHAVTEFPEAEAM 60
BmZNF-1         -----------------------------MKKYLKYKFGEESHSPQYHGVAEIPEAEAM 30
BmZNF-2         -----------------------------MKKYLKYKIGEESHSPQYHGVAEISEAEAM 30
                                             ****** *:**********.*:*: *****
BmZNF-3         EKKGNEQTEKVDPPLPNRPPLPPPDPLLANLKKEIVNEEYIDTSSITINIKTDPTENDGK 120
BmZNF-1         KRKKHEQTEKVHPPLPNRPPLPPPEPLLANFNKEINEDE--ETPSITINIKKDPTENYEK 88
BmZNF-2         KRKKHEQTEKVHPPLPNRPPLPPPEPLLANFKKEINDDE--ETPSITINIKKDPTENYEK 88
                ::* .******.***********:*****::*** ::* :* ********.*****  *
BmZNF-3         PNDVEETMCRNFVRNTCNRGASCRYLHKIIHSQLKGVYRFCIDFENKKCTRAECSYAHAT 180
BmZNF-1         PNDVEETMCRNFVRNTCNRGASCRYLHKIIHSQLKGVYRFCIDFENKKCTRAECSYAHAT 148
BmZNF-2         PNDVEETMCRNFVRNTCNRGASCRYLHKIIHSQLKGVYKFCIDFENKKCTRAECSYAHAT 148
                ************************************:********************
BmZNF-3         VHEKEHFFRTGYLPPNTLSHIKKKTVLPQAKTKGKKFIWKPYFILPK---------IYYS 231
BmZNF-1         VHEKEHFFRTGYLPSHTLSHIKKKTVQPPAKTKETSSELSANYTGTTYPHSTTSVPMYTN 208
BmZNF-2         VHEKEHFFRTGYLPPNTLSHIKKKTVLPPAKTKETSSELSANYTGTTYPHSTTSVPMYTN 208
                ************** .********** * **** .. . : . :* .
BmZNF-3         ----------Y---------------------------------------------------- 232
BmZNF-1         ASNISMMQNPYEKQWPEMEDKSSPEMAYYSGHPSYPHSTSTGYPNAPPNTPIVQNPYATV 268
BmZNF-2         ASNISMMQNPYEKQWPEMEDKTSPEMAYYSGHPSYPHSTSTGYPNAPPNTPIVQNPYVTV 268
                          *
BmZNF-3         ------------------------------------------------------------ 232
BmZNF-1         MSPTKRQWTEMEEPVSQPTHEYAEYAETAGSKKCRNCDVNEFRFQHNKNKIMKMIKDTDD 328
BmZNF-2         MSPTKRQWTEMEEPVSQPTHEYAEYAETAGSKKCRNCDVNEFRFQHNKNKIMKMIKDTDD 328

BmZNF-3         ------------------------------ 232
BmZNF-1         LNYRVGQITKKNTKLNEILVLLVEVLIRRTYR 360
BmZNF-2         LNHRVGQITKKNTKLNEVLVLLVEVVIRRTFR 360
```

**Figure 4.** The ClustalO alignment of BmZNF-1, BmZNF-2 and BmZNF-3 protein sequences. BmZNF-3 is a shorter protein with an extra 30 a.a N-terminal sequence.

The putative protein sequences of BmZNF-1 and BmZNF-2 have an N-terminal proline rich domain of 14 aa (PPLPNRPPLPPPEP) and a novel, non-canonical tandem CCCH zinc finger motif (C-x7-C-x5-C-x3-H-(x13)-C-x7-C-x4-C-x3-H; x=any aa) of 50 aa (Figure 6). The NCBI conserved domain database (CDD) search using the BmZNF-1 and BmZNF-2 protein sequences as query against the KOG v1.0-4825 PSSMS database resulted in the identification of two conserved domains namely, 1) KOG2494, C3H1-type Zn-finger protein, (E-value=$1.8e^{-3}$) and 2) KOG0850, Transcription factor DLX and related proteins with LIM Zn-binding and HOX domains (E-value = $9.43e^{-3}$). Further, a BLAST search using BmZNF-1 and BmZNF-2 protein sequences against the UniProt KB/Swiss-Prot gave hits to the CCCH tandem repeats of zinc finger genes from various organisms and specially to the Muscle blind like protein 3 (MBNL3) of *Mus musculus* (GenBank accession: Q8R003) and many splicing factors like U2AF small sub-unit A (GenBank accession: Q9ZQW8) and B (GenBank accession: Q6AUG0) sub-units and PFL2310w-pre-mRNA splicing factor from *Plasmodium falciparum* (GenBank accession: Q8I4V2). MBNL proteins are known to be mRNA splicing regulators and these genes regulate the alternative splicing of multiple genes (AMACK and MAHADEVAN 2004; CHENG *et al.* 2014; RANUM and COOPER 2006). The interaction of MBNL proteins with either exon or intron regions correlates their functional aspects in the alternative splicing. Their interaction with the 3' UTR sequences suggests their involvement in mRNA stability, localization and protein secretion associated functions (MASUDA *et al.* 2012; OSBORNE *et al.* 2009; POULOS *et al.* 2013; WANG *et al.* 2012a). Although we do not have any direct evidence for the interaction of BmZNF-1 or BmZNF-2 protein(s) with the mRNA(s), based on the homology of their CCCH motifs to MBNL group of proteins, these genes can be speculated to be splicing factors.

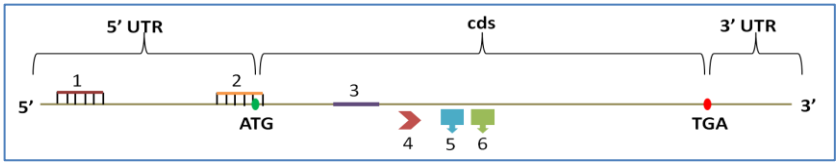


**Figure 5.** Schematic representation of the mRNA structure of zinc finger genes (*Bmznf-1* and *Bmznf-2*). The zinc finger genes (*Bmznf-1* and *Bmznf-2*) codes (CDS 1080 bp) for a putative protein of 41.6 kDa (360 amino acid). 5'UTR has target sites for two ovarian small RNAs-24319 (1) and 37041 (2). The CDS of these transcripts may be a precursor for the ovarian small RNA-12564 (3). The proteins contain one proline rich domain (4) and tandem CCCH zinc finger motifs (5 and 6). The green and red dots represent the start and the stop codons respectively.

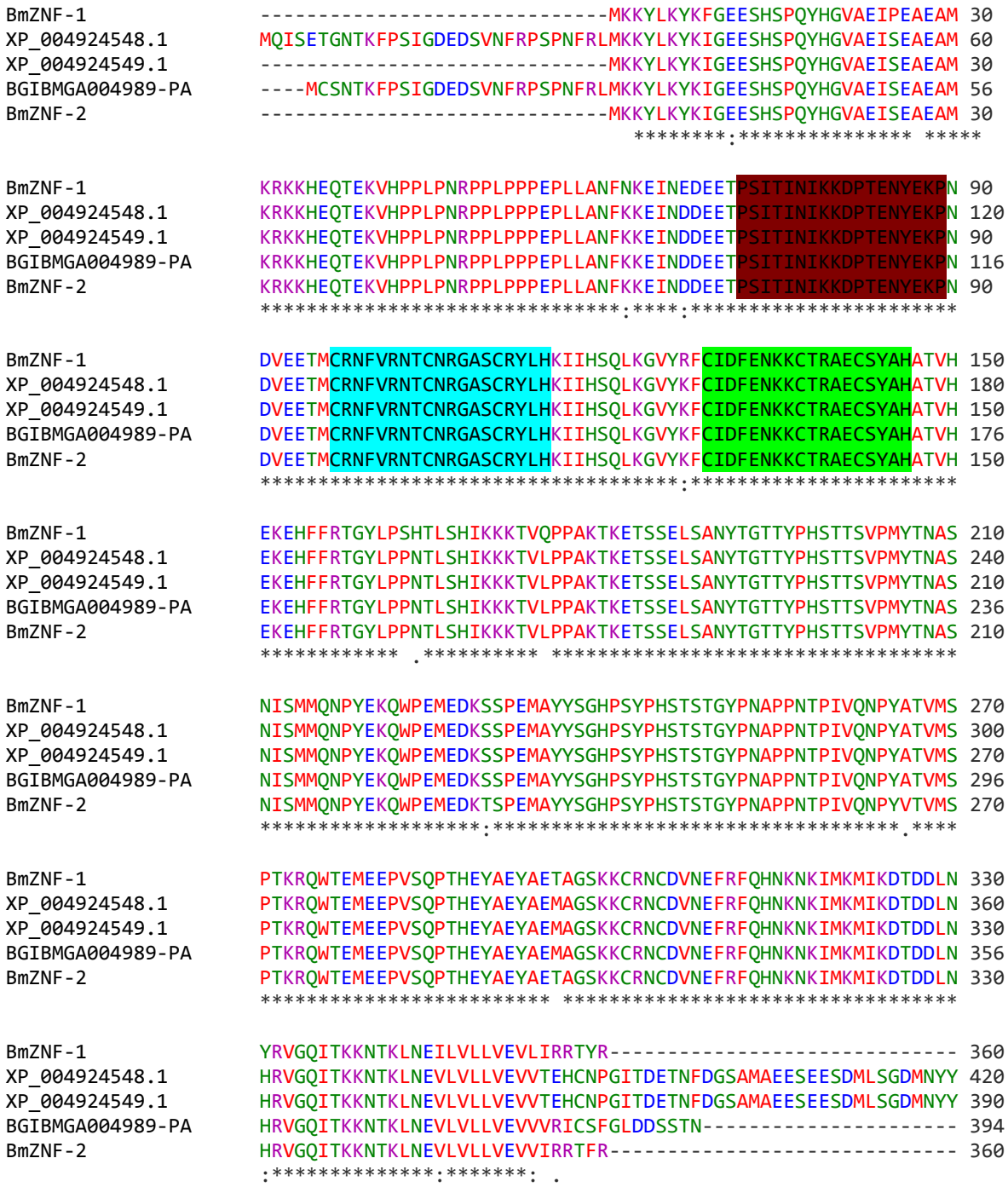


**Figure 6.** The ClustalO alignment of BmZNF-1 and BmZNF-2 with similar sequences found in NCBI (XP_004924548.1, XP_004924549.1) and SilkDB (BGIBMGA004989-PA) databases. The proline rich domain and the zinc finger motifs are highlighted with respective colors as depicted in Figure 5.

Further the overall expression level of these CCCH zinc finger genes (*Bmznf-1*, *Bmznf-2* and *Bmznf-3*) was analysed using qRT-PCR in various somatic tissues like fatbody, midgut, silkgland, head and in gonads (ovary and testis) of 5$^{th}$ instar larvae. These genes showed a female biased expression in somatic tissues and male biased expression in gonads (Figure 7). This suggests that the expression of these genes may be crucial for female somatic differentiation and in gonads these genes may have a different function.

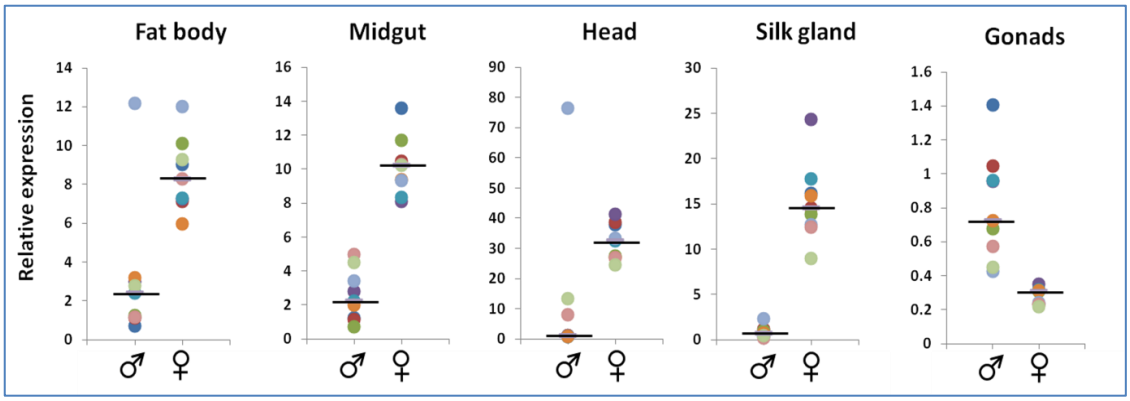


**Figure 7.** The relative expression of CCCH zinc finger genes (*Bmznf-1*, *2* and *3*) in tissues of 5$^{th}$ instar larvae, checked by qRT-PCR.

### 1.3.2 Identification of allelic variants of *Bmznf-2* zinc finger gene

The *Bmznf-1* and *Bmznf-2* transcripts from the male and female fat body samples were cloned in to pIZT vector using full length amplification primers with restriction sites *SacI* and *NotI*. The clones obtained were sequenced for the expression analysis. In females only the transcripts similar to *Bmznf-1* type were obtained (Annexure I), this could be due to the abundant expression of *Bmznf-1* from W-chromosome. As the *Bmznf-1* gene is W-specific, the males are not supposed to have the transcripts similar to this gene. But even in males, transcripts similar to *Bmznf-1* are obtained in abundance compared to transcripts similar to *Bmznf-2* type (Annexure II). The CDS regions of *Bmznf-1* and *Bmznf-2* transcripts exhibit variations at 25 nucleotides (Figure 2) and a group of clones obtained (Annexure III) from males showed similarity with *Bmznf-1* type of transcript at all these 25 nucleotide variations, suggesting the existence of allelic variants for the autosomal copy i.e., *Bmznf-2* gene in *B. mori*.

## 1.4 Discussion

*In silico* analysis of *Bmznf-1* and *Bmznf-2* transcripts in *B. mori* suggests that they encode for a novel, tandem CCCH zinc finger motif proteins with a proline rich domain. The zinc finger motifs are well known for the nucleic acid binding properties (HAGEN *et al.* 1994; MOLNAR *et al.* 1994; OGO *et al.* 2015; PIELER and BELLEFROID 1994) and the proline rich domains are known for the protein-protein interactions (KAY *et al.* 2000; WILLIAMSON 1994; YU *et al.* 1994). Especially the family of CCCH zinc finger genes may be considered as an emerging group of RNA interacting proteins, which are mainly involved in mRNA splicing and 3'UTR mediated regulation/suppression (LAI and BLACKSHEAR 2001; WANG *et al.* 2008). For example, the Muscleblind-like (MBNL) proteins in mammals are a group of CCCH zinc finger genes that are directly involved in the regulation of alternative splicing of various target gene pre-mRNAs. The aberrant splicing of these MBNL proteins leads to the disorder myotonic dystrophy (DM). A 35kDa U2AF, human splicing factor subunit (a CCCH gene) acts as critical factor in both constitutive and an enhancer induced splicing by promoting essential protein-protein interactions and protein-RNA interactions for the 3' splice site selection (PACHECO *et al.* 2006; SOARES *et al.* 2006; YOSHIDA *et al.* 2015). The CCCH zinc finger genes are also involved in regulating the mRNA metabolism by binding the AU rich sequence elements in the 3'UTR of the mRNA and directing them to degradation, thus regulating the gene expression (BLACKSHEAR 2002; HUDSON *et al.* 2004; LAI and BLACKSHEAR 2001). In humans, the CCCH zinc finger gene ZNF74 was shown to be a RNA binding protein and is involved in RNA metabolism (GRONDIN *et al.* 1996). Thus the role of zinc finger genes seems enormous as they exhibit a variety of functions and represent crucial factors in various metabolisms, especially in RNA metabolism.

The 5' and 3' UTR regions of *Bmznf-1* and *Bmznf-2* transcripts contain various UTR regulatory elements like Upstream open reading frames (uORFs), ovarian small RNA target sites, IRES and MBE in common. In eukaryotes the uORFs are involved in regulating the gene expression at the level of translation; there are different mechanisms by which uORFs mediate the translational regulation (BARBOSA *et al.* 2013). Studies have shown that 49% of the human transcriptome has uORFs, which are mostly conserved among species, indicating the evolutionary selection of uORFs that are functionally active (CALVO *et al.* 2009; IACONO *et al.* 2005; SUZUKI *et al.* 2000). The uORFs are common in certain classes of mRNAs, e.g., they are present in two third of oncogenes and genes

involved in important cellular processes like cell cycle regulation, differentiation and stress response (KOZAK 1987; KOZAK 1991; MORRIS 1995; MORRIS and GEBALLE 2000; SPRIGGS *et al.* 2010). About 100 eukaryotic transcripts including 30 of human were validated for their uORF mediated translational regulation (CALVO *et al.* 2009). Additionally recent studies have shown that changes in the 5'UTR sequences that create or disrupt a uORF are associated with development of human disease or disease susceptibility, suggesting their importance in gene regulation (CALVO *et al.* 2009).

Nearly half of the *B. mori* genome is composed by transposable elements (TEs), also called selfish genetic elements (CONSORTIUM 2008; OSANAI-FUTAHASHI *et al.* 2008). These elements can insert themselves into new locations of host genomes and are capable of altering gene structure and expression. The expansion of these elements has to be regulated to conserve the genomic integrity especially in germline cells. Hence organisms have developed a special defense mechanism to silence these elements in germline. Among the three main classes of small RNAs (siRNAs, miRNAs and piRNAs classified based on their biogenesis), the piRNAs have been demonstrated to be actively involved in this type of genomic defense mechanism against transposons (ARAVIN *et al.* 2007; VAGIN *et al.* 2006). It has been found that a sub class of piRNAs called rasiRNAs is abundantly present in the *B. mori* ovary and ovary derived cells (CAI *et al.* 2012; KAWAOKA *et al.* 2008; KAWAOKA *et al.* 2009). In the present work, the transcripts of *Bmznf-1* and *Bmznf-2* contained target sites for two ovarian small RNAs and the CDS regions contain 100% conserved sequence of an ovarian small RNA, suggesting that these transcripts could be acting as precursors for this particular ovarian small RNA. This association between the transcripts and corresponding RNAs suggest that it might be very crucial for the cell to regulate the expression of these genes in order to achieve a desired phenotype of the cell.

Further three MBE elements are also present in the 3'UTRs of *Bmznf-1* and *Bmznf-2* transcripts. From literature it is known that the Musashi is an evolutionarily conserved neural RNA-binding protein that controls neural cell fate. In mammals, Musashi functions in maintenance of the stem cell state, differentiation and tumerogenesis by repressing the translation of a particular set of mRNAs (OKANO *et al.* 2002). Additionally an IRES is present in 3'UTR region of *Bmznf-1* and *Bmznf-2* transcripts. IRES is a nucleotide sequence that allows the translational initiation of mRNA and is present only in a subset of mRNAs. It generally occurs in the 5'UTR regions but the IRES in the 3'UTR of these

genes is unusual and its function is obscure. In the current context, the presence of uORFs and the MBEs in the 5'UTR and 3'UTR respectively indicates the possible involvement of these UTR elements in the translational regulation of the zinc finer genes *Bmznf-1* and *Bmznf-2* and this regulation may be tissue or sex specific or conditionally active.

Sequencing analysis of the CCCH zinc finger genes, *Bmznf-1* and *Bmznf-2* by cloning of *Bmznf-2* transcripts from male fat body samples has resulted in the identification of transcripts that are similar to the W-copy gene, *Bmznf-1*. Therefore these transcripts are considered to be the allelic variants of the autosomal *Bmznf-2* in *B. mori*. The reason why these zinc finger genes are highly homologous and why allelic variants are present for these genes is yet to be explored. This scenario reminds the existence of various natural sequence variants (14) for the CSD gene in *Apis* (BEYE *et al.* 2013), where the CSD genes are involved in absence/presence phenotypes of morphological development. In a similar fashion, the allelic variants of these zinc finger genes may also be involved in such phenotypes. Further, it is believed that the ancestral autosomal copy of *Bmznf-2* has duplicated onto the W-chromosome (*Bmznf-1*) in the course of evolution (SATISH *et al.* 2006), this duplication might be one of the major events in the process of acquiring a sex determination function. But if *Bmznf-2* is already a sex determining gene then the reason behind the duplication of this sex determining gene onto the W-chromosome is interesting to investigate. Further the necessity of this duplication probably suggests the association of these zinc finger genes in the absence/presence phenotypes.

## 1.5 Summary

The analysis of zinc finger genes, *Bmznf-1* and *Bmznf-2* suggests the presence of various regulatory elements of gene expression like uORFs, ovarian small RNA target sites and MBEs in their 5' and 3' UTR regions. The BLAST search of the CCCH zinc finger motifs of these genes indicates similarity with the CCCH zinc finger motifs of Muscleblind, a mRNA interacting protein. This suggests the probable involvement of the zinc finger genes, *Bmznf-1* and *Bmznf-2* in the mRNA metabolism. The expression analysis of *Bmznf-1* and *Bmznf-2* suggests that these genes are profoundly female biased in somatic tissues like fat body, mid gut, silk gland, and head. But in gonads, these genes showed a relative higher expression in testis. This expression pattern indicates that they may have different role in gonads. Further, the sequencing analysis of the clones obtained from male tissues has revealed the existence of allelic variants for the autosomal copy of zinc finger gene, *Bmznf-2* in *B. mori*. Further the role of these homologs in the mechanism of *B. mori* sex determination or sexual development is one exciting aspect to investigate.

# Chapter-II

## The role of *Bmznf-2*, an autosomal CCCH zinc finger gene in *Bombyx mori* sex determination

## 2.1 Introduction

Sex determination is a fundamental biological process that determines two distinct sexes. A variety of sex determination mechanisms is observed in animal species, most of which follow the chromosomal/genetic sex determination, except in few cases where the sex is determined by environmental factors like temperature (e.g. crocodiles, alligators and few lizards). Among insects, the mechanism of sex determination is well understood in *Drosophila* and serves as a reference for all insects. In *Drosophila*, (XX is female and XY is male) the sex is determined by the dose of X-linked signaling elements (XSE) (XSE are four transcription factors *Scute*, *SisA*, *Runt* and *Unpaired*) (ERICKSON and QUINTERO 2007), which in turn is determined by the number of X chromosomes. XSE, whose expression threshold can be reached only in female embryos, confines the production of Sex-lethal (SXL) protein to females. SXL produced in this way directs the female specific splicing of pre-mRNA of *transformer* (*tra*) gene resulting in functional TRA protein. The TRA interacts with non sex specific transformer2 (TRA2) protein and this complex binds to the *doublesex* repeat element (*dsx*RE) in the middle of fourth exon and forces the female specific splicing of *doublesex* (*dsx*) mRNA, producing the female DSX protein. These two proteins have been shown to exhibit antagonistic functions in the process of sexual differentiation (CHRISTIANSEN *et al.* 2002). In a few insect species like *Megaselia scalaris* (TRAUT 1994), *Ceratitis capitata* (WILLHOEFT and FRANZ 1996), *Bactotocera tryoni* (SHEARMAN and FROMMER 1998), *Lucilia cuprina* (BEDO and FOSTER 1985) and *Chironomus thummi* (HAGELE 1985), an epigenetic male factor from the Y chromosome decides the male development. *Culex tritaeniorhynchus* lacks the sex chromosomes and the maleness is conferred by an autosomal gene (BAKER and SAKAI 1976). The sex in *Aedes aegypti* is determined by *Nix* gene from M locus on Y chromosome like region (HALL *et al.* 2015). In hymenopteran species, the sex is maintained through haplodiploidy, where haploids develop as males and diploids develop as females. In *Nasonia vitripennis*, *transformer* (*Nvtra*) gene plays a crucial role in development of females, where it maintains its concentration by an autoregulatory loop through a maternally supplied TRA protein (VAN DE ZANDE and VERHULST 2014; VERHULST *et al.* 2010).

In lepidopterans (butterflies and moths) ZZ/ZW or ZZ/ZO chromosomal system of sex determination is observed. The heterogametic sex (ZW and ZO) is female and the homogametic sex (ZZ) is male. Sex in the domesticated silkworm, *Bombyx mori* is determined by a feminizing piRNA, *fem* on the W-chromosome (HASHIMOTO 1933;

KIUCHI *et al.* 2014). It has been reported that SXL is not regulated in a sex specific fashion in *B. mori* (NIIMI *et al.* 2006). *Tra* orthologue has not been identified so far in *B. mori*, probably owing to its rapid sequence divergence in the course of evolution (CONCHA and SCOTT 2009; HEDIGER *et al.* 2010; KULATHINAL *et al.* 2003; O'NEIL and BELOTE 1992; RUIZ *et al.* 2007). *Dsx* pre-mRNA has been shown to be lacking TRA/TRA-2 binding sites (SUZUKI *et al.* 2001). Though the orthologues of *tra2*, *intersex* (*ix*) and *fruitless (fru)* genes have been identified in *B. mori*, their functions remain elusive. Previous studies have resulted in the identification of two RNA binding splicing inhibitors: 1) *B. mori* homolog of IGF-II mRNA binding protein (BmIMP) (SUZUKI *et al.* 2010) and 2) *B. mori* homolog of P-element somatic inhibitor (BmPSI) (SUZUKI *et al.* 2008), which are involved in differential splicing of *Bmdsx* pre-mRNA. The involvement of *Bmpsi* and *Bmimp* renders this mechanism to be unique from any other group of insects.

Recently, the mechanism of *B. mori* sex determination was reported to be governed by a piRNA (*fem*) from the W-chromosome. The W-derived *fem* piRNA negatively regulates a Z-linked CCCH zinc finger gene, *Masculinizer* (*masc*). This has been shown to regulate the *Bmdsx* sex specific splicing by promoting the expression of male specific *Bmdsxm* type of splicing isoform and also dosage compensation by an unknown mechanism. Thus, this gene, *masc* is presumably non-functional in females, leading to female specific *Bmdsxf* type of splicing isoform (KIUCHI *et al.* 2014). Further studies have shown that the over expression of *masc* gene in BmN cells has enhanced the transcription of *Bmimp* gene and most probably through this the *masc* induces the expression of male specific *Bmdsxm* type of splicing isoform (SAKAI *et al.* 2015a). Thus, the reported studies have shown that the sex in *B. mori* is regulated by a W encoded *fem* piRNA that in turn negatively regulates the *masc* gene in females, which is a masculinization factor in males.

In *B. mori,* studies attempting to identify the genes involved in sex determination pathway have resulted in the identification of a female specific CCCH zinc finger motif encoding gene, termed as *z1* on W-chromosome and its homologous copies namely *z2* and *z3* on 25$^{th}$ chromosome (AJIMURA *et al.* 2006; SATISH *et al.* 2006) (please refer Chapter-I of this thesis for more information on these genes). Further, the studies of translocation of W-chromosomal fragments to autosomes have supported the existence of a strong putative epistatic female determining region called "Fem" on the W-chromosome (HASHIMOTO 1933; TAZIMA 1954). A preliminary analysis using FISH has indicated that probably these

zinc finger genes are linked to the "Fem" region of W-chromosome (AJIMURA *et al.* 2006; SATISH *et al.* 2006).

In the current study, my experimental evidences provide functional insights into the role of an autosomal CCCH zinc finger gene, *z2* (*Bmznf-2*) (GenBank accession: XP_004924549) in the *B. mori* sex determination. Through transient overexpression of *Bmznf-2* in BmN cells demonstrates that *Bmznf-2* affects the differential splicing of *Bmdsx* and *Bmtra-2* pre-mRNA. Further, I provide a critical evidence (Luciferase assays) for an endogenously existing 5'UTR mediated translational repression over *Bmznf-2* in BmN cells.

## 2.2 Materials and Methods

### 2.2.1 Cloning of *Bmznf-2* into insect expression vector, pIZT

The coding sequence (CDS) of *Bmznf-2* (1083 bp) was PCR amplified with the primers Zn2-F-SacI and Zn2-R-NotI (Annexure IV) using cDNAs synthesized from the fat body of $5^{th}$ instar male larvae to make sure the amplicon was derived from autosomal *Bmznf-2* gene. The resultant amplicon was cloned into the pIZT-V5 insect glow vector (Invitrogen) using *SacI* and *NotI* restriction enzymes, for pIZT-*Bmznf-2* plasmid. The clones were selected on low salt Zeocin LB plates and the sequences were confirmed by OpIE2 F and OpIE2 R sequencing primers. (ii) For the luciferase assay, a modified pmirGLO vector (Promega), which had OpIEI and OpIEII promoters from pIZT vector in places of human PGK and SV40 respectively were used. The 5'UTR regions (with two intact and scrambled putative target sites for the ovarian small RNAs) of *Bmznf-2* were cloned downstream of firefly luciferase CDS using restriction sites *XbaI* and *SacI*. The CDS of GFP (600 bp) was used as a negative control in this experiment. (iii) For the localization study of BmZNF-2 protein, the CDS of m-cherry was cloned inframe downstream of *Bmznf-2* in pIZT-*Bmznf-2* using the restriction sites *NotI* and *XbaI*, to get fused protein of *Bmznf*-2 with m-cherry at C-terminal end. The positive clones were confirmed through sequencing.

### 2.2.2 Transfection and overexpression of *Bmznf-2* in BmN cells

BmN cells were cultured and passaged using TC-100 insect medium (SIGMA-ALDRICH) with 10% FBS (Thermo Scientific-Gibco). Transfections were carried out in 12 well plates for ectopic expression and RNAi experiments, and 24 well plates for luciferase reporter assays. Cells were seeded at a concentration of $1x10^5$ per well (for 12 well) and $5x10^4$ (for 24 well) followed by transfection after 24h. Transfection mixtures were prepared by mixing plasmid DNA or dsRNA with (200 ng of plasmid for ectopic expression, luciferase assays and one µg of dsRNA for RNAi) 3 µl of transfection reagent TransIT2020 (Mirus Bio) in a final volume of 100 µl using serum free medium per well. This mixture was incubated at room temperature for 20 minutes for complex formation and distributed dropwise to each well containing cells in a volume of 1 ml (for 12 well plate) and 0.5 ml (for 24 well plate) medium, following manufacturer's instructions.

### 2.2.3 Isolation of total RNA from BmN cells and RT-PCR

Total RNA from tissues (~25 mg) was isolated following TriZol (Invitrogen) method. DirectZol kit (Zymo Research) was used for isolation of total RNA from BmN cells. To

remove DNA contamination, the total RNA was DNase-I (Zymo Research) treated as per manufacturer's instructions. The RNA concentration of samples was measured using Nanodrop 2000 (Thermo Scientific). cDNAs were synthesized using 1 µg of total RNA, using SuperScript III (Invitrogen). PCR was set using Emerald Amp GT PCR 2X master mix (Takara). The reaction conditions were 94˚C for 5 min, 35 cycles of 94˚C for 30 sec + 60˚C for 30 sec + 72˚C for 1 min and final elongation step of 10 min at 72˚C. The PCR samples were analysed on 1.5% agarose gels cast with ethidium bromide.

### 2.2.4 Quantitative RT-PCR

The relative expression level of *Bmdsxm* isoform was validated through quantitative real-time PCR (ABI 7500) (for primers see Annexure IV).The reaction was set using SYBR Premix Ex Taq (Tli RNaseH Plus, Takara Bio Inc) with cDNA sample of 3 µl (diluted to 10 ng/µl) and 0.2 µM primers, in a final volume of 20 µl. Reaction conditions were: 95˚C for 30 sec, 95˚C for 5 sec and 60˚C for 34 sec. ABI SDS software version 1.2.3 was used for standard curve analysis. The relative expression was determined using ΔCt analysis and the reactions were carried out in triplicates. GAPDH was used as endogenous reference control.

### 2.2.5 Mutational PCR

To construct the mutant clones, point mutations were introduced into the two CCCH tandem repeats of *Bmznf-2* clones by using two mutant primers (for primers see Annexure IV), changes two cysteines to serines and one hystidine to leucine in the resultant protein products. Amplicons of Zmut-pair1F + Zn-R-NotI and Zn1-F-SacI + Zmut-pair 2R were used as templates for overlapping PCR and the resultant final fragment was cloned in pIZT vector. The clones obtained were screened for the desired mutations through sequencing and the selected mutant clones were used for transfection experiments.

### 2.2.6 Stem loop RT-PCR

cDNA for the ovarian small RNAs was synthesized by following the previously described protocol (SINGH *et al.* 2010). The reverse transcription reaction was carried out using SuperScript™ III (Invitrogen) by following the manufacturer's instructions. In brief, the reaction mixture of 20 µl included 1 µg RNA, 1 µl 10 mM dNTPs (Fermentas) and 0.5 µl 100 nM stem-loop primer (for primers see Annexure IV).

### 2.2.7 Luciferase reporter assay

Dual luciferase reporter assays were performed in duplicates (3-biological x 2-technical replicates). BmN cells, two days post transfection with pmirGLO constructs were slogged in 160 µl of PBS per well. 75 µl of these slogged cells were distributed into two wells of 96 well plate for the assay. Dual luciferase reporter assays were performed in the GloMax multi plus machine (Promega), using the dual-luciferase reporter assay system (Promega), following the manufacturer's protocol. Briefly, to 75 µl of slogged cells, 75 µl of Dual-Glo Luciferase Reagent was added and were set for agitation for 10 minutes (for cell lysis) followed by taking the readings for firefly luciferase activity. The plate was removed from the machine and 75 µl of Dual-Glo Stop and Glow solution was added to the wells followed by agitation for 10 minutes and renilla luciferase activity was recorded. For all the samples, the ratio of firefly and renilla luciferase activities was estimated and the percentage of relative luciferase activity was plotted.

### 2.2.8 Immunofluorescence

A day before transfection with pIZT-*Bmznf-2* construct, the BmN cells were seeded on 18mm cover slips placed in 24 well plate,. For immunofluorescence imaging, on 2$^{nd}$ day post transfection with desired pIZT constructs cells were fixed with 4% paraformaldehyde for 10 min, permeabilized with 0.2% Triton X-100 in PBS for 5 min and then blocked with 1% bovine serum albumin at room temperature (RT) for 1 h. The cells were subsequently incubated overnight with primary antibody- Anti His (Invitrogen) at 1:100 dilution, at 4˚C. After washing thrice (10 min each) with PBS, the cells were incubated with Alexa 594 conjugated anti- anti-mouse antibody (Invitrogen-Molecular Probes) at a dilution of 1:500, and incubated at RT for 1 h followed by three washes each for 10 min with PBS. The nuclei were counterstained with 4, 6-diamidino- 2-phenylindole (DAPI) and cells were mounted using VECTASHIELD Mounting Medium (Vector Laboratories). The fluorescent images were captured on Zeiss LSM 510 META inverted confocal microscope and analyzed with LSM software (Zeiss).

### 2.2.9 RNA interference

RNAi was performed in BmN cells using short double stranded RNAs. A short dsRNA was designed to target two regions of a transcript. The CDS of *Bmtra-2* and *Bmznf-2* were scanned for the potential siRNAs using siRNA wizard (HTTP://WWW.SIRNAWIZARD.COM/DESIGN.PHP 2016). A set of two long primers (Annexure

IV) were designed per gene by fusing two predicted siRNAs based on not less than 50% GC content. These primers contained a tag sequence which is complementary to the T7 primer of mirVANA probe construction kit (Ambion) that facilitated the construction of two DNA templates of ~75 bp length. These two templates were set for in vitro transcription for producing the sense and antisense transcripts which were mixed at equimolar concentration and annealed by heating at 95˚C for 2 min on a thermomixer. This mixture was then cooled slowly to room temperature and stored in -20˚C until use.

### 2.2.10 Sequence confirmation

All clones and the differential bands obtained as a result of altered splicing were either agarose gel extracted or cloned into TOPO vector (Invitrogen). These amplicons or clones were then confirmed by sequencing through Sanger method.

## 2.3 Results

### 2.3.1 Masculinization of BmN cells upon over expression of *Bmznf-2*

In this study, the role of zinc finger gene *Bmznf-2* is revealed in the sex specific differential splicing of the *Bmdsx* pre mRNA in the ovary derived BmN cells. These cells produce the female type of *Bmdsx* (*Bmdsxf*) splicing isoform, representing their female mode of sexual differentiation. To decipher the role of *Bmznf-2* in promoting differential splicing of *Bmdsx* pre-mRNA, RNAi based knockdown of *Bmznf-2* was performed in BmN cells using short dsRNA. The knockdown achieved for *Bmznf-2* gene was 75 to 90%, which is considerably high and presumably enough for generally interfering the gene activity in *Bombyx* (MON *et al.* 2012; SUZUKI *et al.* 2010; SUZUKI *et al.* 2008). But it did not affect the innately expressing *Bmdsxf* splicing isoform level (Figure 8A, 8B), which indicates the lack of activity of *Bmznf-2* in achieving *Bmdsxf* splicing isoform in BmN cells (female). Further, the overexpression of *Bmznf-2* in BmN cells promoted male specific splicing isoform (*Bmdsxm*) and this correspondingly decreased *Bmdsxf* (Figure 9A and 9B). This shift of splicing phenotype is referred as "masculinization" (KIUCHI *et al.* 2014). The masculinization induced by *Bmznf-2* overexpression denotes the "gain of function" of *Bmznf-2* in BmN cells (female cells), which indirectly suggests that *Bmznf-2* may be generally inactive in female cells.

As mentioned previously, the CDS region of *Bmznf-2* mRNA sequence could be a putative precursor of the ovarian small RNA-12564. In such a case, the overexpression experiments of *Bmznf-2* may also be treated as the over expression of the ovarian small RNA and possibly the observed masculinization could be either by the putative BmZNF-2 protein or by some kind of gene regulation induced by the ovarian small RNA-12564. Therefore, to test which of the above two factors (BmZNF-2 protein or ovarian small RNA-12564) is actually associated in inducing masculinization of BmN cells, I have performed site directed mutagenesis of the two CCCH motifs of putative BmZNF-2 protein, in order to to unravel their role in masculinization. By keeping the region of the ovarian small RNA-12564 intact, two mutant pIZT constructs were generated and over expressed in BmN cells, each expressing the mutated BmZNF-2 proteins at its a) CCCH motif 1 and b) CCCH motif 2 respectively.

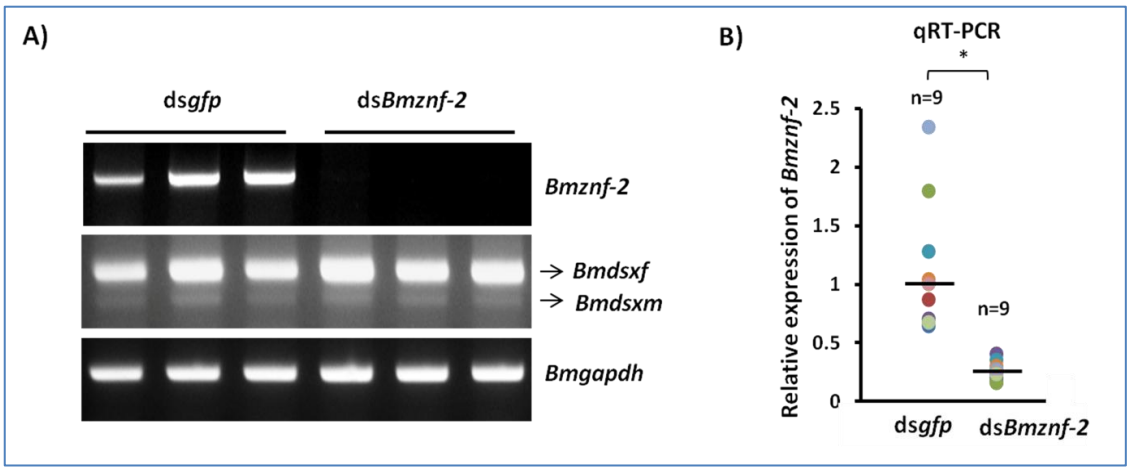


**Figure 8.** RNAi of *Bmznf-2* did not affect the differential splicing of *Bmdsx* in BmN cells. A) RNAi shown by RT-PCR and B) qRT-PCR to show their RNAi efficiency (75% to 90%) in BmN cells. The results suggest no role of *Bmznf-2* in female type of *Bmdsx* splicing (*Bmdsxf*).

The mutations resulted in the replacement of the 2nd and 3rd cysteines to serines and the hystidine to leucine amino acids in the CCCH motifs, which has previously been demonstrated to affect the structure of the zinc finger motif, and would seriously compromise the function of CCCH zinc finger protein (ADDEPALLI and HUNT 2007; GUO *et al.* 2000). The point mutations in either the CCCH motif 1 or CCCH motif 2 has abolished the phenotype of masculinization (Figure 9C and 9D), indicating the involvement of putative BmZNF-2 protein and the essentiality of zinc finger motifs in inducing masculinization of BmN cells. These experiments in BmN cells revealed the association of *Bmznf*-2 encoded protein in regulating the sex specific differential splicing of *Bmdsx*, and thus signifying its activity in controlling the processes of sex determination and differentiation (Figs. 9A and 9B).

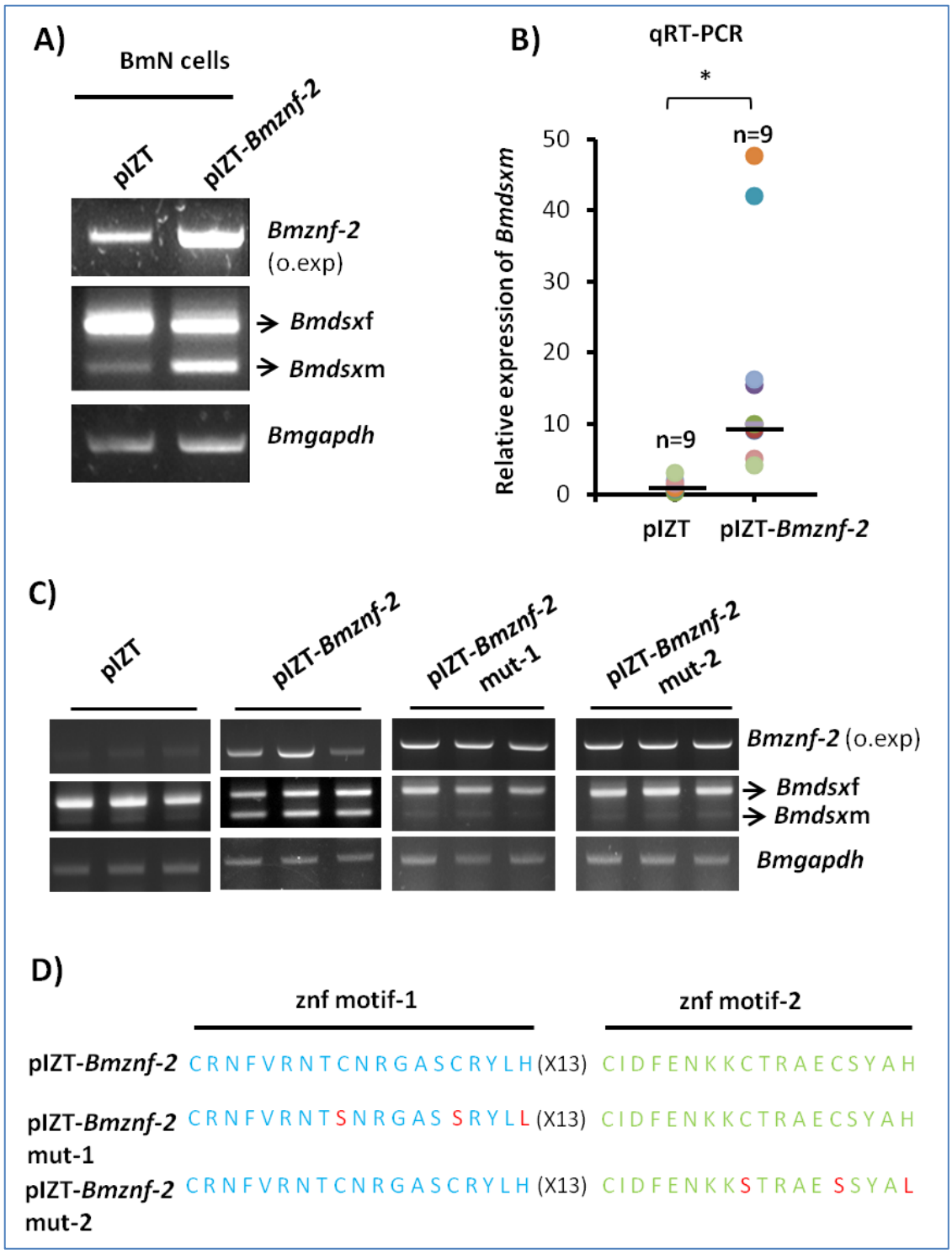


**Figure 9.** *Bmznf-2* promotes the male specific splicing of *Bmdsx* in BmN cells upon overexpression. A) Increased expression of *Bmdsxm* splicing isoform (masculinization) upon *Bmznf-2* over expression. B) Relative quantification of the *Bmdsxm* splicing isoform using real-time qPCR (* indicates a significant difference, t-test - $p<0.05$) between control-pIZT and BmZNF-2 induced-pIZT-*Bmznf-2* samples. The dark lines represent the median values of the data points. C) Point mutations (three cysteines to serines and one histidine to leucine) in at least one CCCH motif out of two resulted in the loss of masculinization phenotype. D) The sequences of the wild type and the mutated zinc finger motifs of the three clones used for transfection.

The above experiment suggests the role of BmZNF-2 protein in the alternative splicing of *Bmdsx* and as the mechanism of alternative splicing operates only in the nucleus of cells, I further checked the localization of *Bmznf*-2 protein in BmN cells. For this, the *Bmznf-2* CDS in pIZT construct was fused with m-cherry at its C-terminal end and over-expressed in BmN cells. The fluorescent imaging clearly indicated the nuclear localization of *Bmznf*-2 and m-cherry fused protein (Figure 10), which was further supported by immunostaining of BmN cells transiently over-expressed with the His-tagged (C-terminal) *Bmznf*-2 protein (Figure 11). This study implies its functional activity in the nucleus and its probable involvement (either direct or indirect) in the nuclear processes like mRNA splicing.

### 2.3.2 Differential splicing of *Bmtra-2* pre-mRNA induced by *Bmznf-2* over expression.

Apart from testing the effect of overexpression of *Bmznf-2* on *Bmdsx* splicing its effect on differential expression of genes reported to be involved in *Bombyx* sex determination (*Masc*, *Bmtra-2*, *Bmpsi* and *Bmimp*) was also checked. A unaltered expressions of these genes was observed (data not shown) except for *Bmtra-2*, which showed an enhanced expression coupled with aberrant splicing.

*Bmtra-2* splicing (from masculinised BmN cells) was analysed by using a series of forward and reverse primers as indicated in Figure 12A. The sequencing analysis (Sanger method) of full-length amplicons (Figure 12B and 12C) using primer sets $F_1$-$R_5$ and $F_1$-$R_6$, revealed the aberrant splicing of *Bmtra-2*. All the major and the minor (based on relative band intensity as shown in Figure 12C) aberrant splicing isoforms are depicted in Figure 12D. The major aberrant isoforms are resultant of two splicing events 1) Due to the skipping of $7^{th}$ exon alone and 2) Skipping of $6^{th}$ and $7^{th}$ exons together. Both types of transcripts presumably produce truncated versions of BmTRA-2 proteins that lack RS2 domain. The minor aberrant isoforms include two types of transcripts, formed by 1) Skipping of $4^{th}$ to $7^{th}$ exons and 2) Skipping of $5^{th}$ to $7^{th}$ exons (Figure 12C). These minor isoforms presumably result in much truncated versions of *Bmtra-2* proteins that lack a major portion of protein (even the RRM motifs), whose function is unknown.

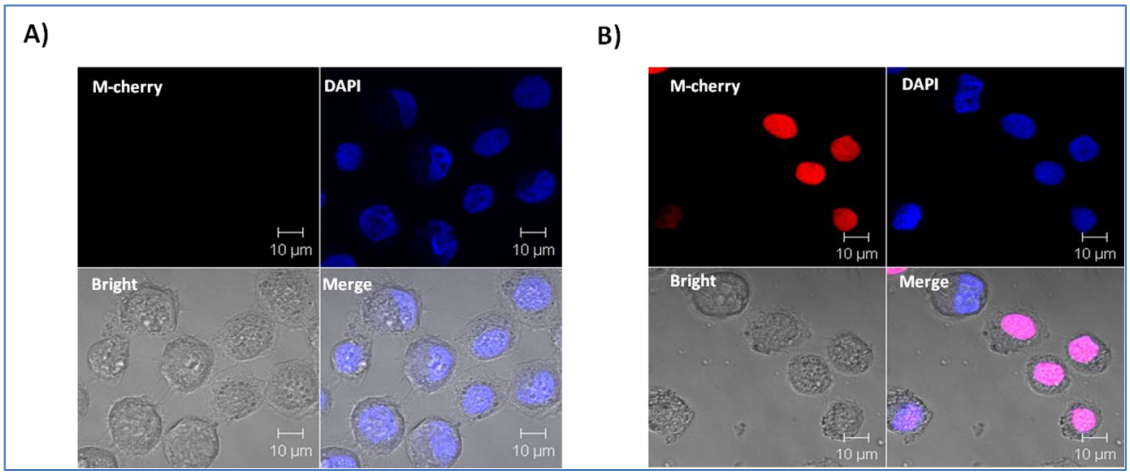


**Figure 10.** Nuclear localisation of BmZNF-2 protein. A) Un-transfected control BmN cells and B) BmN cells transfected with pIZT-Bmznf-2_m-cherry plasmid that expresses fused protein.

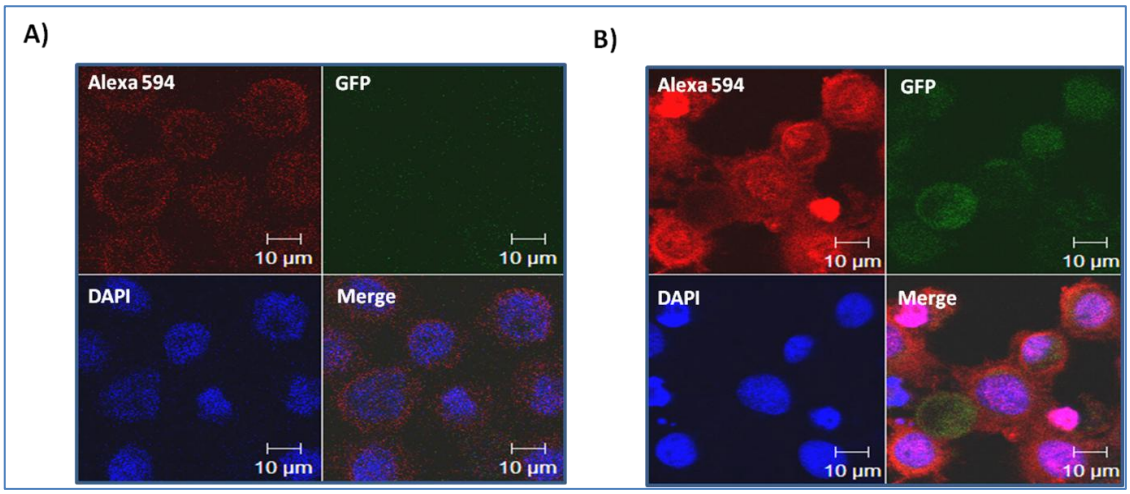


**Figure 11.** Confocal imaging of the nuclear localised BmZNF protein. A) BmN cells un-transfected, used as control for immunostaining. B) BmN cells transfected with pIZT-*Bmznf-2* for over expression of c-terminal His tagged BmZNF protein. Cells were immuno-stained using mouse Anti His primary and Alexa 594 anti-anti mouse secondary antibodies.

An enhanced expression of *Bmtra-2* was observed upon *Bmznf-2* over expression as measured by qRT-PCR, selecting two regions of the transcript, proximal (Figure 12E, exon 2-3) and distal ends (Figure 12F, exon 6-8). Though a significantly (t-test, $p<0.05$) enhanced expression was observed in both the cases (Figure 12E and 12F), the enhanced expression was found to be considerably higher from the distal end of the transcript (Figure 12F). This difference was also reflected from full-length amplicons using semi-quantitative RT-PCR (Figure 12B and 12C). These findings, suggest an inherent differential selection/exemption of $8^{th}$ and $9^{th}$ exons (containing stop codons), making their expression weak, and thus in general a major proportion of *Bmtra-2* transcripts may be directed towards an unknown event of splicing.

Further, in order to study the possible relationship between the observed differential splicing of *Bmtra-2* and *Bmdsxm* type of splicing, RNAi based knockdown was performed for the gene *Bmtra-2* in BmN cells (efficiency achieved was 75% to 90%). The knockdown of *Bmtra-2* did not affect the sex specific splicing of *Bmdsx* (Figure 13), indicating that it has no role in achieving default splicing isoform (*Bmdsxf*) of *Bmdsx* in BmN cells. Previous studies conducted for the functional analysis of *Bmtra-2* in *B. mori* embryos revealed no role of *Bmtra-2* in the *Bmdsx* sex specific splicing (SUZUKI *et al.* 2012). Though the RNAi results are concordant with the previous RNAi study in embryos (SUZUKI *et al.* 2012), the unusual splicing events observed in *Bmtra-2* upon overexpression of *Bmznf-2* presents an elusive role of *Bmtra-2* in the mechanism of *Bombyx* sex determination.

To evaluate the functional role of the observed aberrantly spliced shorter transcripts of *Bmtra-2* upon overexpression of *Bmznf-2* in masculinization, the shorter isoforms of *Bmtra-2* were cloned and over-expressed in BmN cells, this did not affect the sex specific splicing of *Bmdsx* (data not shown). Hence the reason for the occurrence of shorter *Bmtra-2* transcripts which presumably generate putative proteins, lacking RS2 domain upon *Bmznf-2* overexpression, is not clear and needs to be investigated.

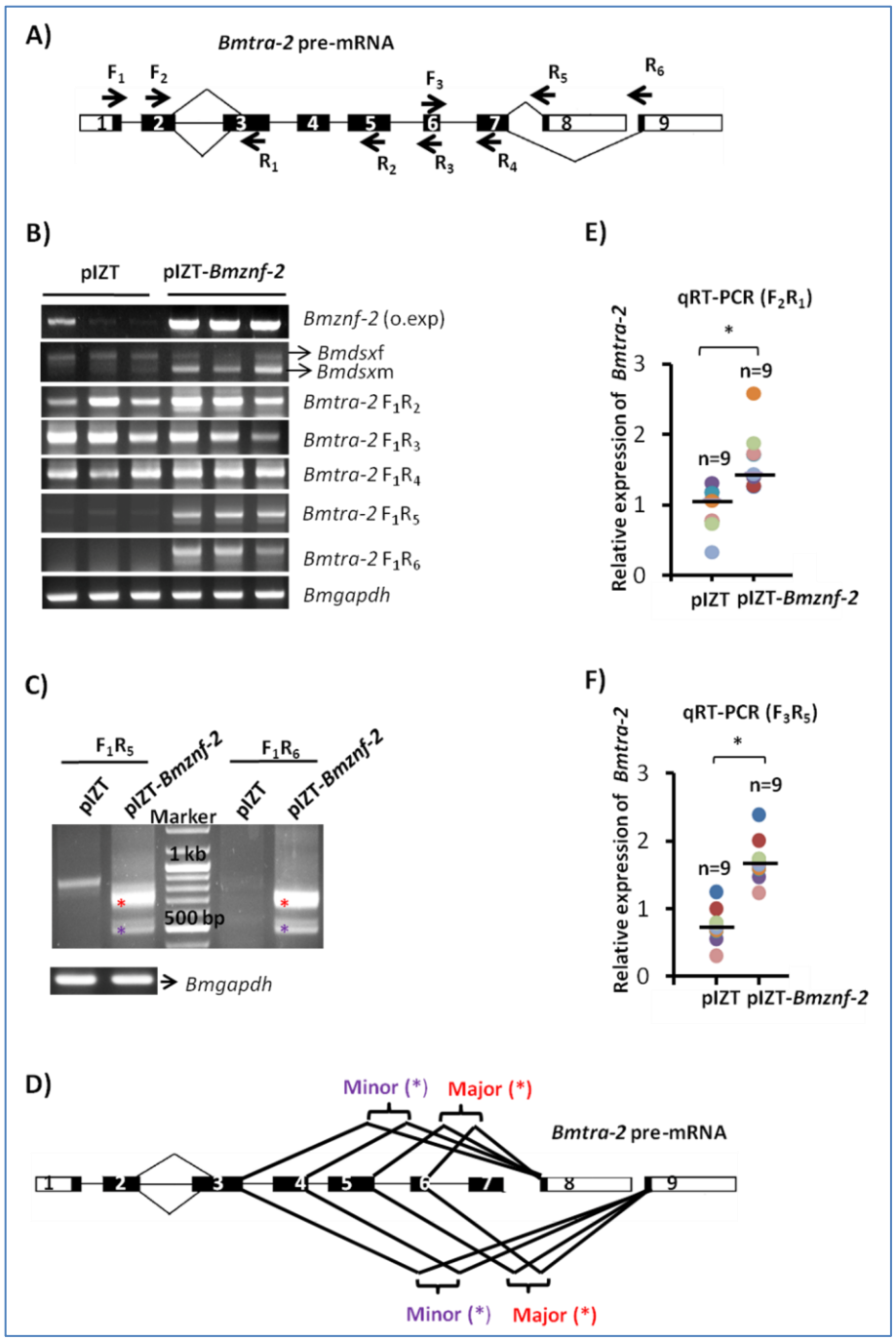


**Figure 12.** Altered splicing of *Bmtra-2* upon over expression of *Bmznf-2* in BmN cells. A) The gene structure of *Bmtra-2* showing 9 exons. B) Over expression of *Bmznf-2* enhances the *Bmtra-2* transcript levels. $F_1$ to $F_3$ and $R_1$ to $R_3$ represent the primers used for PCR amplification (both for semi-quantitative and qRT-PCR). C) *Bmtra-2* full length amplicons (with primers $F_1$-$R_2$ and $F_1$-$R_3$) obtained from control (pIZT vector treated) and *Bmznf-2* over-expressing (pIZT-Bmznf-2) BmN cells. The RT-PCR profiles indicate the differential splicing of *Bmtra-2* upon over expression of *Bmznf-2*. D) Schematic representation of the differentially spliced *Bmtra-2* isoforms upon overexpression of *Bmznf-2* gene. * indicates the major and minor splicing isoforms obtained from the sequencing of corresponding bands of C). E) and F) Real-time PCR to show the increased expression of *Bmtra-2* transcript level between control (pIZT, vector treated) and *Bmznf-2* over-expressing (pIZT-Bmznf-2) BmN cells.

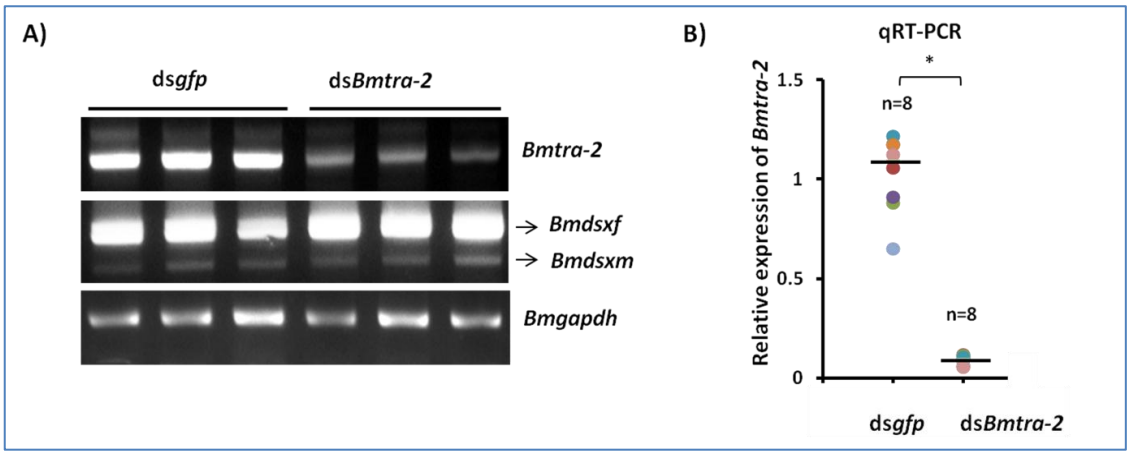


**Figure 13.** RNAi of *Bmtra-2* genes did not affect the differential splicing of *Bmdsx* in BmN cells. A) RNAi of *Bmtra-2* shown by RT-PCR and B) qRT-PCR to show the RNAi efficiency (75% to 90%) in BmN cells. The results suggest no role of this gene in female type of *Bmdsx* splicing (*Bmdsxf*).

Further, the essentiality of both CCCH motifs of *Bmznf-2* was shown in the unusual splicing events of *Bmtra-2* phenotype (Figure 14) by deploying the same site directed mutant clones used for evaluating the role of *Bmznf-2* in masculinization phenotype. Thus, the role of *Bmznf-2* CCCH motifs is apparent in both phenotypes. These observations reveal that *Bmznf-2*, by some unknown mechanisms, induces both masculinization and an enhanced expression coupled with an aberrant splicing of *Bmtra-2* pre-mRNA.

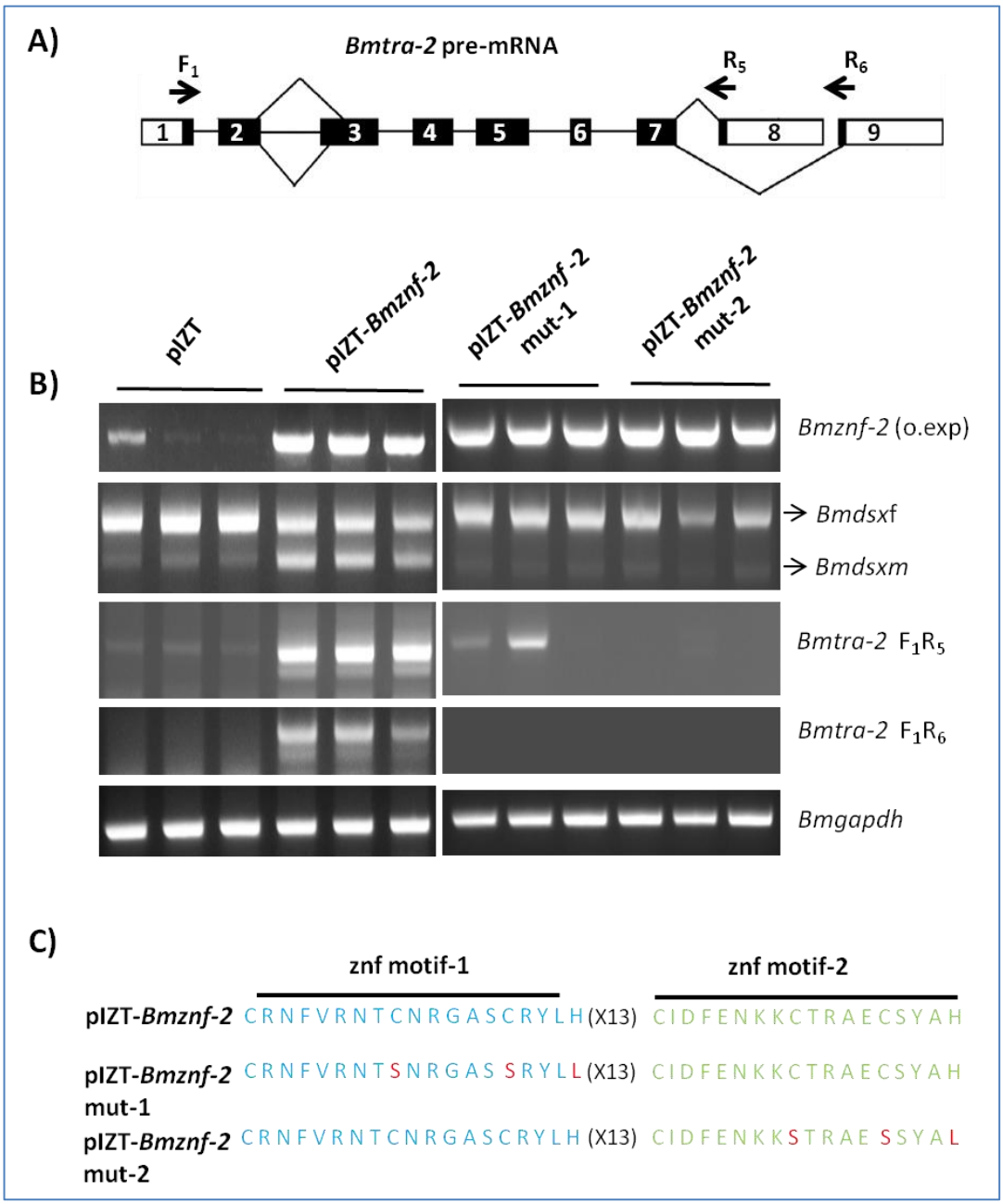


**Figure 14.** Site directed mutagenesis and *Bmtra-2* splicing. A) Schematic representation of the *Bmtra-2* splicing structure and the primers used. B) C) Point mutations (three cysteines to serines and one histidine to leucine) in atleast one CCCH motif out of two resulted in loss of masculinization and *Bmtra-2* aberrant splicing phenotypes. C) The sequences of the wild type and the mutated zinc finger motifs of the three clones used for transfection assays.

### 2.3.3 Existence of 5' UTR mediated negative regulation of *Bmznf-2* gene

The 5' UTR of *Bmznf-2* harbours 100% antisense match for two ovarian small RNAs 24319 and 37041. Therefore, it is expected that *Bmznf-2* might be a target for these two small RNAs. The endogenous expression of these small RNAs was detected by the stem loop PCR (Figure 15). To study the involvement of these ovarian small RNAs in *Bmznf-2*

regulation, dual-luciferase assays were performed with the 5' UTR sequence of *Bmznf-2* (Figure 16). The luciferase constructs containing entire 5' UTR region of the *Bmznf-2* (5'UTR) or the 5' UTR region with scrambled two ovarian small RNA target sites (5' UTR ov. target sites scrambled) were constructed using the cDNA prepared from RNA of the male larval fat body. The ratio of luciferase expression (firefly/renilla) for the constructs suggested a significant 5' UTR mediated translational repression (t-test, $p<0.05$) over *Bmznf-2* expression level in BmN cells (Figure 16). But an insignificant difference was observed between the constructs with and without ovarian small RNA target sites (t-test, $p>0.05$), thus ruling out the role of these two putative ovarian small RNAs 24319 and 37041 in the observed translational repression (Figure 16). From the above data, I infer a 5' UTR mediated translational repression for the gene *Bmznf-2* in BmN cells. But the role of these ovarian small RNAs is not clear as the scrambling of their target binding sites in 5'UTR region did not affect the negative regulation of *Bmznf-2* gene.

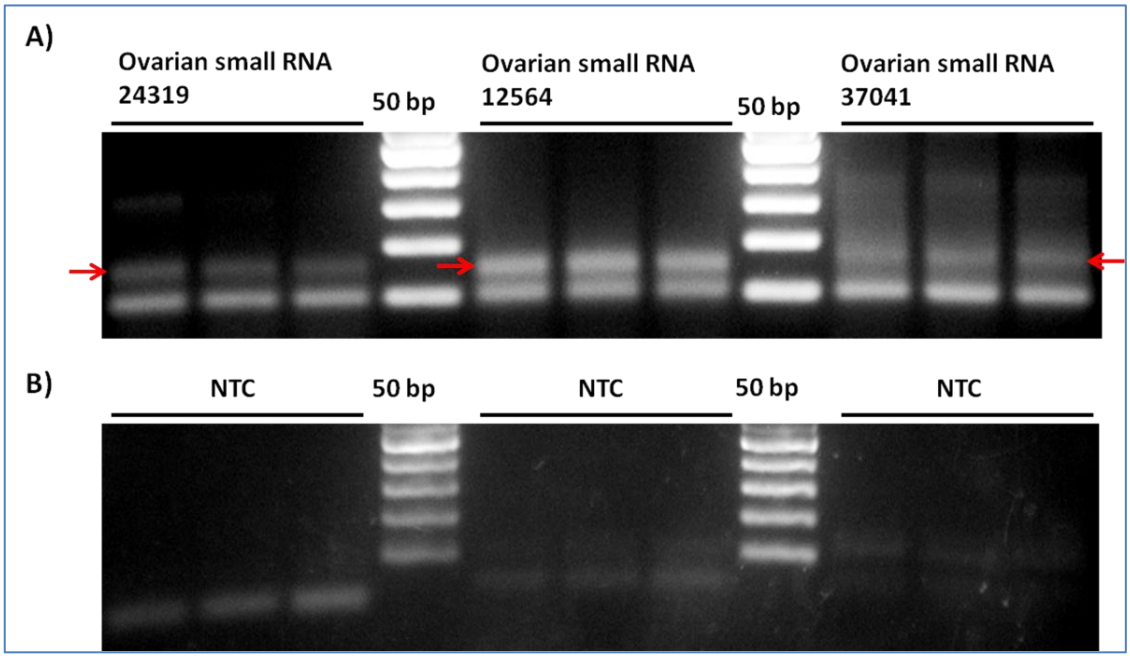


**Figure 15.** Endogenous expression of ovarian small RNAs in BmN cells. A) Stem loop PCR showing the expression of the ovarian small RNAs 24319, 12564, 3704 in BmN cells. Arrows indicate the corresponding small RNA amplicons. B) No template control (NTC) for the stem loop PCR.

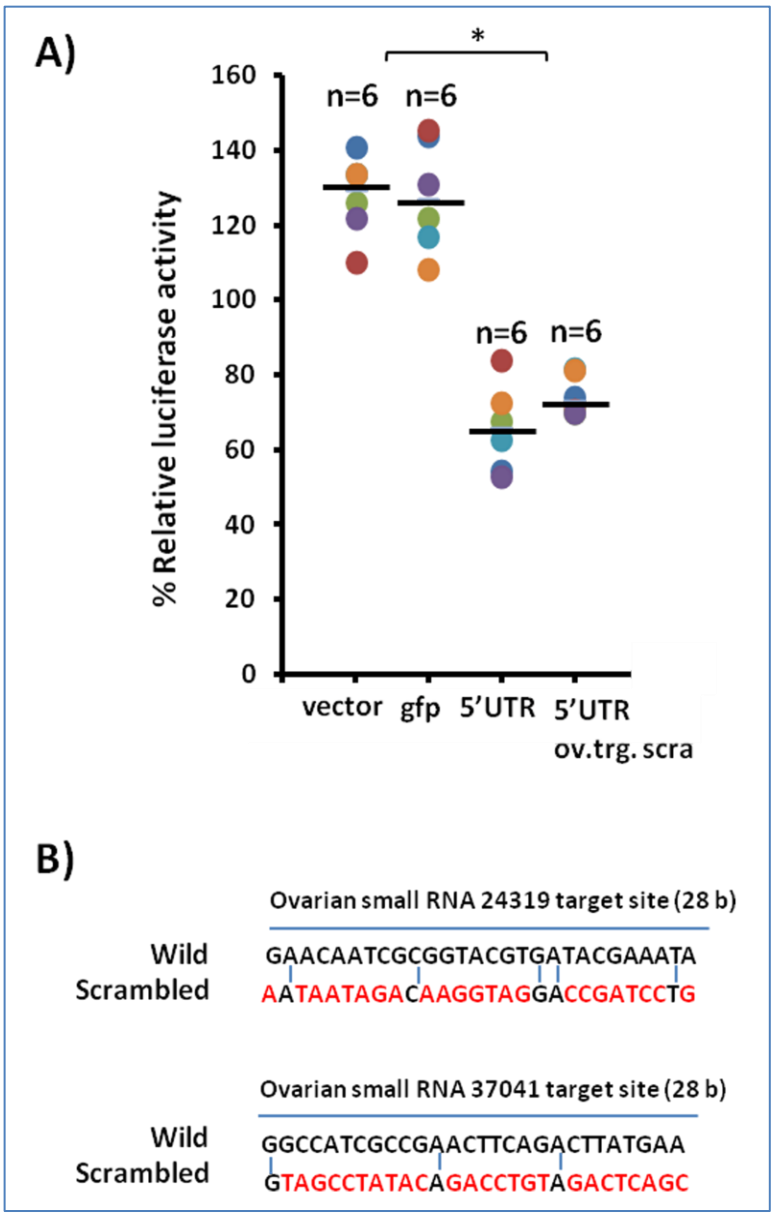


**Figure 16.** The existence of 5' UTR mediated negative regulation of *Bmznf-2* expression in BmN cells. A) The dark horizontal lines represent the median values of the data points. * indicates the statistically significant difference (*t*-test) observed between samples. B) The sequences of the wild and scrambled ovarian small RNA target sites for the ovarian small RNAs 24319 and 37041.

## 2.4 Discussion

In insects, *dsx* is considered to be the terminal gene in the sex determination pathway and is also treated as master regulator of sexual differentiation. Its pre-mRNA is sex specifically spliced, producing sex specific proteins, DSXM in males and DSXF in females. These two proteins share a common N-terminal zinc finger DNA recognition domain (DM) (ERDMAN and BURTIS 1993) and oligomerization domain 1 (OD1) but vary at their C-terminal oligomerization domain 2 (OD2), which leads to antagonistic properties in terms of gene expression. For instance, the genes transcriptionally activated by DSXM will be repressed by DSXF and vice versa, which results in insect sexual dimorphism (AN *et al.* 1996). The DSX orthologs are widespread across the animal kingdom. For example, in mouse, a DSX ortholog called DMRT1 is involved in testis differentiation (RAYMOND *et al.* 2000). The *Dmrt* genes are extensively studied as they are the transcription factors, which are mainly involved in sex determination and differentiation pathways (PICARD *et al.* 2015).

The fact that a large number of DSX target sites are present in *Drosophila* implies the significance of *dsx* in regulating somatic sexual differentiation. These target sites include transcription factors and signalling pathway components, which are directly or indirectly involved in sex biased expression. Further, these DSX target sequences of *Drosophila* overlap considerably with the DMRT1 target sites of Mouse (CLOUGH *et al.* 2014), denoting the functional conservation of DMRT across species. The upstream factors that regulate the differential splicing of *dsx* pre-mRNA are crucial in the mechanism of insect sexual dimorphism as *dsx* splicing is a very sensitive process that follows the "all or none principle." Hence, sex specific differential splicing of *dsx* pre-mRNA in insects is crucial and may be viewed as an essential regulatory mechanism, as this decides the sex of insects.

In *B. mori*, the *doublesex* gene (*Bmdsx*) is sex specifically spliced and its pre-mRNA splicing has been shown to be mainly regulated by two RNA binding proteins, BmPSI and BmIMP (SUZUKI *et al.* 2010; SUZUKI *et al.* 2008). These two RNA recognition motif (RRM) containing proteins have been shown to be physically interacting with the regulatory element CE1 sequence (20 nucleotides) on exon 4 of *Bmdsx* pre-mRNA and induce the skipping of exons 3 and 4. So far only these two proteins are reported to be the potent direct splicing regulators of the *Bmdsx* pre-mRNA splicing. Recent studies have

added the *masc* gene to be another potential factor, which may be indirectly promoting the *Bmdsxm* type of splicing by enhancing *Bmimp* transcript levels (KIUCHI *et al.* 2014; SAKAI *et al.* 2015a).

The present study clearly shows that BmZNF-2 is a nuclear localizing protein and is an important masculinization factor that upon its overexpression induces male specific splicing of *Bmdsx* pre-mRNA in BmN cells. Based on the following observations of BmZNF-2, 1) The similarity of CCCH zinc finger motifs to that of MBNL proteins, 2) Promoting the differential splicing of *Bmdsx* pre-mRNA and 3) Nuclear localization, I speculate *Bmznf*-2 to be a splicing regulatory protein that directs the alternative splicing of its target mRNAs in a similar fashion as that of MBNL proteins. But the interaction of *Bmznf*-2 with DNA should be tested by further experimentation as I cannot rule out its possibility as a transcription factor, activating many more genes that are crucial in sex determination. Altogether, from these observations, I group the gene *Bmznf-2* with splicing regulators like *Bmpsi*, *Bmimp* and *masc* genes, which determine the differential splicing of *Bmdsx* pre-mRNA, and in turn the *Bombyx* sex determination. But the interaction of *Bmznf*-2 with these proteins is an open question and needs to be investigated. Further, the growing list of *Bmdsx* sex specific splicing regulators (BmPSI, BmIMP, MASC and BmZNF-2), suggests a complex mechanism being operated for the regulation of *Bmdsx* differential splicing in *Bombyx*.

Studies in *Drosophila* suggest that the TRA-2 is a RNA binding protein involved in the alternative splicing of various genes. The *tra-2* orthologs are found across species from worms to mammals, and are essential in the alternative splicing of a variety of genes (BEST *et al.* 2014; SEONG *et al.* 2002; XIA *et al.* 2007), and its splicing enhancer activity is also linked to a few cancers (WATERMANN *et al.* 2006). In *Drosophila*, TRA-2 protein along with TRA forms a splicing enhancer complex, which plays a crucial role in the differential splicing of *dsx* (HOSHIJIMA *et al.* 1991), *fruitless* (RIDEOUT *et al.* 2007; SHIRANGI *et al.* 2006), *exuperantia* (*exu*) and *alternative-testis-transcript* (*att*) (HAZELRIGG and TU 1994; MADIGAN *et al.* 1996). Thus, *tra-2* plays an essential role in the *Drosophila* sex determination and development.

*Bmtra-2,* a *B. mori* ortholog of *tra-2,* is not sex specifically spliced and is expressed in both the sexes. The alternative splicing of *Bmtra-2* pre-mRNA results in 6 isoforms, namely *Bmtra-2* A-F (NIU *et al.* 2005), of which A, B, C share 8$^{th}$ exon and D, E, F share

9th exon for their stop codons. These isoforms are generated due to two 5' splicing acceptor sites on 3rd exon and by differential sharing of 8th or 9th exons. All the isoforms encode proteins that contain three domains namely N terminal arginine-serine rich (RS1), RRM and C terminal arginine-serine rich (RS2). Functionally, the RRM domain is involved in target binding and the RS domains are involved in protein-protein interactions (KOHTZ *et al.* 1994; WU and MANIATIS 1993). The RNAi studies for the functional analysis of *Bmtra-2* in *B. mori* embryos suggested the involvement of *Bmtra-2* in testis development but revealed no role in the sex specific splicing of *Bmdsx* (SUZUKI *et al.* 2012).

Here I present a complicated splicing structure of *Bmtra-2* pre-mRNA, resulting from differential skipping of exons. This is promoted by the over expression of *Bmznf-2* in BmN cells. The over expression of *Bmznf*-2 induces an unusual splicing pattern of *Bmtra-2* pre-mRNA, resulting in shorter isoforms, which get translated into shorter putative proteins that lack the RS2 domain. Insights from the literature suggest that the RRM domain (STEFL *et al.* 2005) of *Drosophila* TRA-2 is essential for the dsxRE (six exonic splicing enhancer sequences) binding, and RS1 and RS2 are involved in interacting with the auxiliary enhancer SR proteins that in turn promote the sex specific splicing of *dsx* in *Drosophila*. Deletion of RS2 domain in transgenic flies abolished the sex specific splicing of *dsx* pre-mRNA, whereas deletion of RS1 showed marginal effects, signifying the importance of RS2 domain (AMREIN *et al.* 1994). Based on these findings from TRA-2, I speculate that the RS2 domain of *Bmtra-2* may also have a crucial role in *B. mori* sex determination.

The present study indicated the role of *Bmznf-2* in regulating the alternative splicing of an essential splicing enhancer gene *Bmtra-2*, whose protein product may be involved in splicing of a variety of genes and expected to affect many pathways, especially in the germ cells. Further, I could not link the unusual splicing events of *Bmtra-2* with the masculinization phenotypes, but could explore the critical role of *Bmznf-2* in both.

Insights from functional analysis of long non-coding RNAs (lnRNAs) suggest the dynamic roles of these emerging regulators of gene expression in the field of molecular biology (MERCER *et al.* 2009). Reports indicate the role of lnRNAs in various post translational modifications. Their active participation in regulatory functions like disruption of translation, regulation of mRNA stability or transcriptional control, indicates the possible widespread functional significance of lnRNAs in gene regulation (BEST *et al.*

2014; FATICA and BOZZONI 2015; KUNG *et al.* 2013; SINGH and PRASANTH 2013). In the present study, I show the existence of *Bmznf-2* 5' UTR mediated negative regulation of this gene via a lnRNA(s)/unknown piRNA(s) or a regulatory protein that is involved in the probable suppression of the *Bmznf*-2 protein levels/activity. I speculate that this unknown factor may be encoded by W-chromosome or may be influenced by a putative W-factor operating via a gene regulatory mechanism (Figure 17). Further, I also assume that cell monitors the expression level/activity of the *Bmznf*-2 protein by the translational repression via 5'UTR based regulation, for avoiding masculinization.

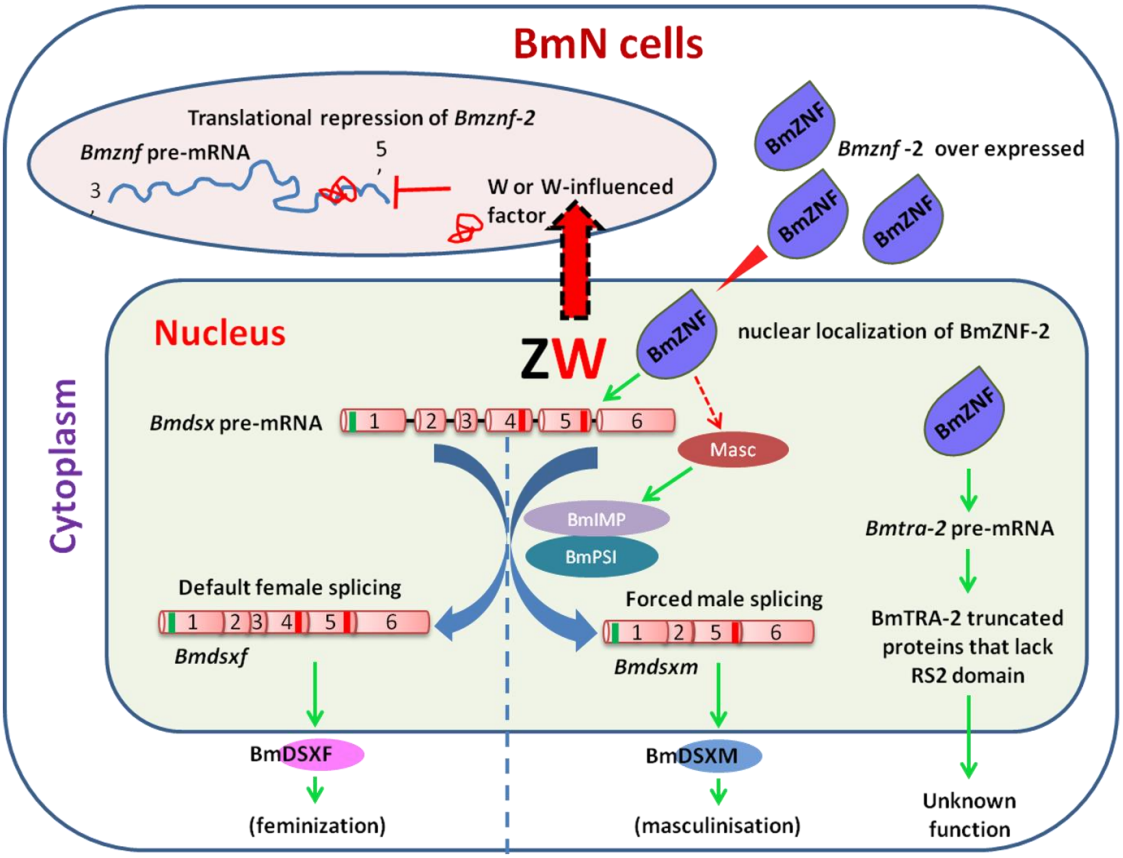


**Figure 17.** A model representing the influence of BmZNF-2 protein on splicing of *Bmdsx* and *Bmtra-2* pre-mRNAs. I propose that under normal condition, the *Bmznf-2* gene is translationally repressed by a W-factor or W influenced factor (probably a long non-coding RNA or protein). When this translational repression is overcome by overexpression of BmZNF-2 protein, it results in masculinization by inducing *Bmdsxm* type of splicing and also affects the *Bmtra-2* pre-mRNA by producing shorter splicing isoforms that code for putative BmTRA-2 proteins lacking the C-terminal RS2 domain. The function of these shorter proteins is not clear as their overexpression did not affect the *Bmdsx* differential splicing.

## 2.5 Summary

The results suggest that *Bmznf-2* might have evolved as an upstream sex determining factor by regulating sex specific splicing of *Bmdsx*. Our experiments in female cell lines infer the gain of function of *Bmznf-2* suggesting suppression of its function by a gene regulatory mechanism, most probably by a translational repression (as observed from the luciferase assays for the 5' UTR regions). In congruence with all the direct and indirect pieces of evidence obtained in this study, I propose *Bmznf-2* to be functionally active in males and not in females. Similar to the recently discovered *masc* gene, *Bmznf-2* also appears to be a redundant masculinization factor in the mechanism of *B. mori* sex determination. The presence of more than one upstream factor governing the sex specific splicing of *Bmdsx* pre-mRNA indicates the complexity behind the evolution of sexual differentiation in *B. mori*. The evolutionary reason for acquiring more than one upstream regulatory factor for *Bmdsx* differential splicing, and the exact mechanism by which the *Bmznf-2* and *masc* genes induce the male type of splicing and the mode of translational repression observed for *Bmznf-2* gene in female cells (BmN), are yet to be investigated. Thus, the mechanism of *B. mori* sex determination still has many open questions and awaits further investigation.

# Chapter-III

## Transcriptome analysis of sexed embryonic stages and larval heads of *Bombyx mori*

## 3.1 Introduction

In *B. mori*, the W-chromosome confers femaleness. Translocation studies of W-fragments have resulted in identification of a portion of W-chromosome called "Fem" region. This region of W-chromosome is found to be linked to one of its 12 RAPD markers called Rikishi (ABE *et al.* 2008). In order to identify the factors originating from this region, researchers have compared the composition of small RNA between females of wild type strain with whole W-chromosome and the females of W-chromosome mutant strains (LY, DfZ-DfW, and MW), lacking various regions of W-chromosome. This analysis resulted in the profiling of a subset of piRNAs coming from "Fem" region (KAWAOKA *et al.* 2011). Further, transcriptome analysis of sexed embryos resulted in the identification of a piRNA called "*fem*" that targets a CCCH zinc finger encoding gene called *Masculinizer* (*masc*). This gene is involved in processes like sex determination by governing the sex specific differential splicing of *Bmdsx* and also in dosage compensation by reducing the expression of Z-linked genes in males (KIUCHI *et al.* 2014). But the growing list of the genes/factors being involved in the sex specific differential splicing of *Bmdsx* suggests that a few additional/novel factors may be involved in the sex-determination pathway of *B. mori*. Hence it will be interesting to explore for the additional/novel factors that are involved in the process of sex determination. Studies to sequence the whole W-chromosome are still at their preliminary stages due to the difficulty in assembling chromosome, as W-chromosome is full of repetitive elements and transposable elements. However, identification of additional factors involved in the *B. mori* sex determination may be achieved by generating RNA-seq data for sex specific tissues. In the current study, I have made such an attempt by deep sequencing of transcriptome of sexed early embryos at 78h, 96h and 120h stages of development along with larval heads at which a high level of somatic sexual differentiation is expected. The embryonic stages were selected based on the prior information of *Bmdsx* differential splicing by single embryo analysis, which suggests that the *Bmdsx* undergoes differential splicing in sexes at 96h stage, indicating that this stage is crucial in the sexual differentiation of the embryos. The sequence analysis resulted in the identification of a set of differentially expressed genes and thousands of unmapped transcripts, which could be precursors for the small RNAs. Unfortunately, I could not identify any protein coding gene from the pool of genome unmapped transcripts obtained from the embryonic and head samples, revealing the scanty nature of the W-chromosome for protein coding genes.

## 3.2 Materials and Methods

### 3.2.1 Sample collection, library preparation and RNA-seq

Two W-chromosome mutant strains of *B. mori* were used: 1) Japanese sex limited (JPSL) for the sexed embryo collection and 2) Sex limited strain (QGSLO) for the sexed larval heads. In JPSL, the translocation of chromosome 10 fragment harboring *kynurenine monooxygenase* gene on to the W-chromosome is believed to be responsible for the development of dark brownish serosal pigmentation, which acts as a visible marker to differentiate female embryos as early as 36 hours post oviposition (hpo). In the QGSLO strain, female larvae can be easily distinguished by distinct crescent shaped markings on dorsal side of larva from the $4^{th}$ instar stage. For the embryo collection, JPSL strain moths were set for 4 hrs mating at room temperature; transferred to 4˚C overnight, followed by depairing and were set in dark at room temperature for 2hrs for a uniform egg laying. Collected eggs were cold acid treated (one of the methods of breaking diapause) at an age of 20 hpo to break the diapause and were thoroughly washed under running tap water and incubated at 25˚C for the development to proceed. At 36 hpo, male and female eggs were segregated based on the serosal pigmentation. Sampling of 200 each of male and female embryos was done at 78 hpo, 96 hpo and 120 hpo. For the head tissue collection, larvae of (QGSLO) fifth instar $5^{th}$ day were numbed on ice for 30 min, decapitated and 10 male and 10 female larval heads were pooled separately and snap frozen in liquid nitrogen until use. For collecting BmN cells [regularly passaged in TC-100 (SIGMA) insect media with 10% FBS (Gibco)] log phase cells were selected, slogged, pelleted in PBS and stored in -80 ˚C until use (Please note that the BmN cell data was not used in this analysis, it is used for the analysis in Chapter IV of this thesis, it is included here for representing the read mapping statistics of all samples together) From the collected samples, total RNA was isolated and on-column DNAseI treated for the removal of genomic DNA contamination using Direct-Zol RNA isolation kit (Zymo Research). RNA libraries were prepared following the TruSeq RNA sample preparation kit v2 (Catalog No.: RS-122-2001) from Illumina using 1 µg of mRNA. Sequencing was performed on Illumina 1000 HiSeq platform (C-CAMP, Bangalore). In total, not less than ~60 million pair-end reads of 100 bp, for each of the sexed embryonic stages and head samples were generated. For simplicity, 78 hpo, 96 hpo and 120 hpo embryonic stages are referred to as 78h, 96h and 120h respectively throughout the thesis.

### 3.2.2 RNA-seq read quality filtration, mapping and data analyses

The RNA-seq read quality was assessed using the package FastQC (ANDREWS 2010). The adapter removal and quality trimming was performed using Trimmomatic version 0.35 (BOLGER *et al.* 2014). The leading and trailing low quality or N bases were removed below quality 3, reads were scanned with a 4 base sliding window, and clipped when the average quality per base dropped below 20, and read sequences below 30 bases in length were dropped. The filtered paired-end reads were mapped against *B. mori* genome sequence and its annotations [downloaded from Ensembl release 29 (2015) (GCA_000151625.1.29) (FLICEK *et al.* 2014)] using bowtie2 version 2.2.6 (LANGMEAD and SALZBERG 2012; LI and DEWEY 2011) with default parameters. The aligned reads were filtered to keep only uniquely mapped reads. SAM/BAM conversions, sorting, indexing and filtering were performed with SAMtools version 1.2. (LI *et al.* 2009). The alignment files (SAM format) so obtained were imported in to Seqmonk software (ANDREWS 2015). While importing the SAM files in to Seqmonk, the libraries were treated as pair-end and duplicate reads were eliminated. The log2 transformed FPKMs for genes were quantitated, by correcting for DNA contamination, transcript isoforms were merged, transcript length correction was made, besides excluding the genes with no or very low read counts (considered as noise in Seqmonk) to avoid bias in the data that might skew the analysis. The raw read counts table of Seqmonk analysis was exported to Excel (Microsoft) and further analyzed. Moreover, for *B. mori* as the GFF file from ensembl was not annotated with the chromosome number, which is required for current analysis. So to overcome this, the "description" code of each gene in GFF file was replaced with its corresponding chromosome number, based on the scaffold identity from kaikobase (SHIMOMURA *et al.* 2009) annotation data using custom shell scripts.

### 3.2.3 Differential gene expression (DGE) analysis and unmapped data analysis

The raw read counts data was subjected to Trimmed mean of M-values normalization method (TMM) and DGE analysis was done using edgeR package in R programme (ROBINSON *et al.* 2010). The genes with p-value and false discovery rate (FDR) <0.05 and logFC (female/male) values < -1 and > 1 (i.e., fold change >2) were considered as male biased and female biased genes. Further the embryonic samples were also analyzed using CLC Genomics Workbench 7.0 beta (CLC 2014) for the DGE analysis. In CLC analysis, the reads that did not map the genome were treated as unmapped data and these reads were de novo assembled for generating unmapped transcripts.

## 3.3 Results

### 3.3.1 Read mapping and sequencing output

RNA-Sequencing of three sexed embryonic stages (78h, 96h and 120h), 5$^{th}$ instar larval heads and BmN cells of *B. mori* resulted in 718M reads (359M paired) of 100 bp length from nine samples ranging from 66 to 126M paired end reads (Table 1). The use of the Trimmomatic software for the trimming of the adapter sequences resulted in 666M reads (333M paired). Technical replicates for all the samples (except for head samples) were included and the FPKM values for each gene was averaged across the technical replicates and used for analysis. Genome guided mapping using Bowtie 2.1.0 resulted in an average of 78.9% read alignment per sample (Table 1) of the quality filtered and adapter trimmed reads.

### 3.3.2 Differential gene expression in sexed embryos and larval heads

DGE analysis of the embryonic (78h, 96h and 120h) and head samples resulted in identification of an array of genes which showed significant sex biased expression (Figure 18A). At 78h stage, 520 genes were identified to be differentially expressed (FDR and p-value <0.05), out of which 350 genes (67%) were highly male biased (fold change $\geq$2) and only 15 genes (2.8%) were highly female biased (fold change $\geq$2). Genes that are known to be essential in development like *bicaudal C homolog 1-A* (logFC -5.537) and *nanos-M* (-2.768) were male biased (fold change $\geq$2). It is interesting to note the male biased expression (fold change $\geq$2) of silk proteins like Fibroin heavy chain (logFC -10.74) and Fibroin light chain (logFC -5.428) and Silk protein P25 (logFC -6.98) at 78h stage of development. Many zinc finger motif encoding genes showed a significant male biased expression at 78h, e.g., Zinc finger protein 608-like (logFC -3.61), Zinc finger protein OZF-like (logFC -2.768), Zinc finger protein 1 (logFC -2.014) and Zinc finger protein 729-like (logFC -19). Few genes that are highly female biased at 78h are 1) UDP-glucosyl transferase UGT48C1 (logFC 8.41), Superoxide dismutase [Cu-Zn]-like (logFC 1.96), scarlet (logFC, 1.38), Gastrula zinc finger protein (logFC 1.03), C108 diapause bioclock protein (ea4) mRNA (logFC 1.387) and DnaJ homolog subfamily C member 16 (logFC 1.28). Among the differentially expressed genes (FDR and p-value <0.05) at 78h, 86 genes encoded for uncharacterized proteins (annotation not found), 20 genes for kinases, 14 genes for zinc finger proteins, 14 genes for transcription factors or transcription related and 11 genes for receptor molecules.

**Table 1.** Number of paired reads before and after Trimmomatic

| Stage, Sample | Total paired end reads | Paired end reads after trimming |
|---|---|---|
| **Embryonic, 78h** | | |
| **Male, technical replicate 1** | 11050822 | 9914362 |
| **Male, technical replicate 2** | 31628333 | 30890006 |
| **Female, technical replicate 1** | 16233927 | 15815447 |
| **Female, technical replicate 2** | 2697763 | 2443184 |
| **Embryonic, 96h** | | |
| **Male, technical replicate 1** | 13884150 | 13260716 |
| **Male, technical replicate 2** | 26784132 | 26457089 |
| **Female, technical replicate 1** | 13342408 | 12649208 |
| **Female, technical replicate 2** | 16435822 | 16235564 |
| **Embryonic, 120h** | | |
| **Male, technical replicate 1** | 17455056 | 16955766 |
| **Male, technical replicate 2** | 17454043 | 16746448 |
| **Female, technical replicate 1** | 13499099 | 12938869 |
| **Female, technical replicate 2** | 20835824 | 20603602 |
| **Larval, Head** | | |
| **Male** | 49965704 | 43855957 |
| **Female** | 48243611 | 43641574 |
| **BmN cells** | | |
| **Biological replicate 1** | 28818725 | 24947548 |
| **Biological replicate 2** | 30766050 | 25902378 |
| **Total =** | **388651296** | **358250294** |

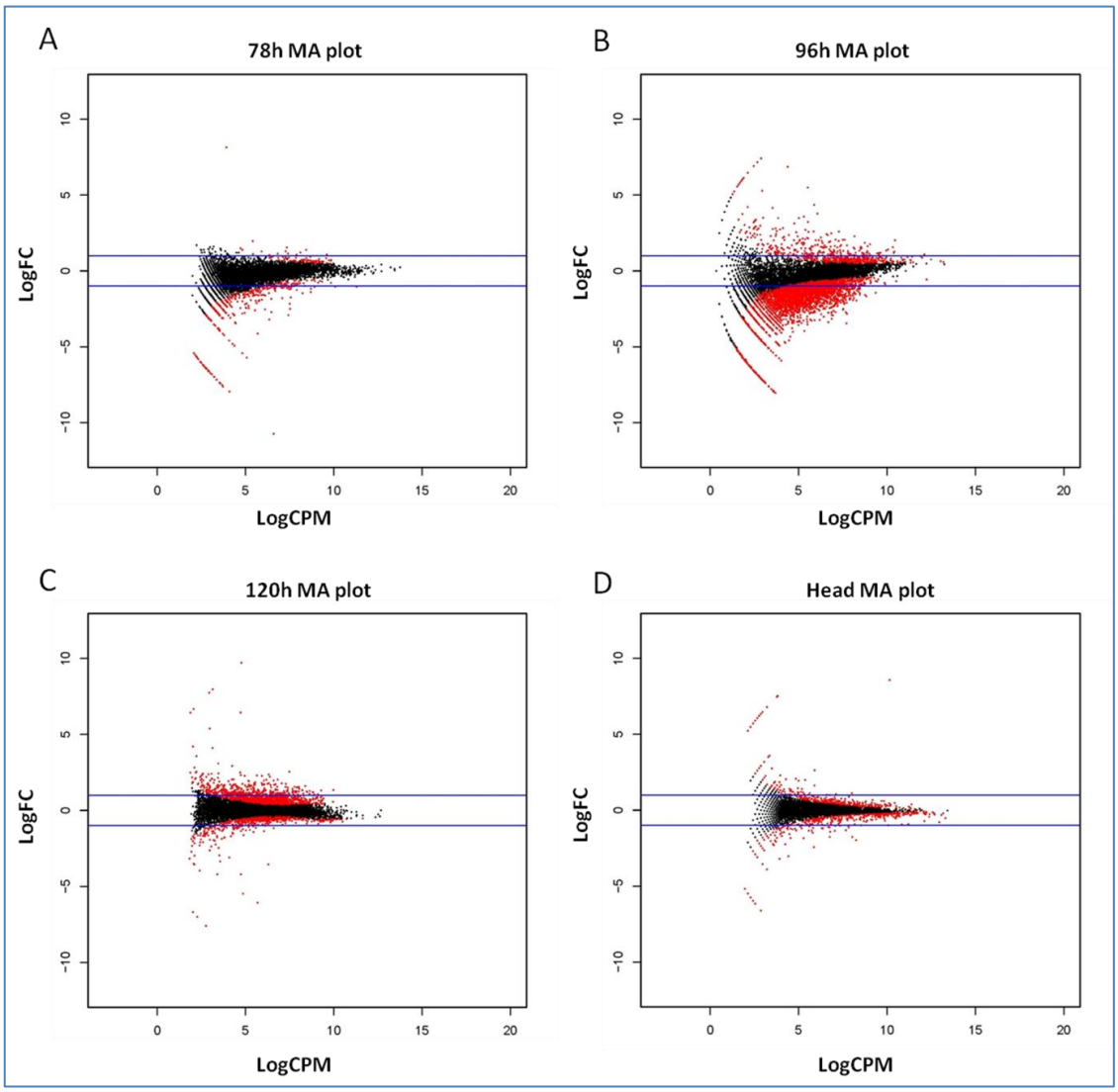


**Figure 18.** The MA plots [M(log ratios) and A (mean average)] of embryonic and head samples in male to female comparison. In these plots, logFC denotes the biased expression, female biased (+ve y-axis) and male biased (–ve y-axis) and logCPM represents the average expression strength of genes. A, B, C and D represents the MA plots for 78h, 96h, 120h and head samples respectively.

At 96h stage a very high number of genes 4,068 genes were differentially expressed (FDR and p-value <0.05) among them 2,641 genes (64.9%) were highly male biased (fold change ≥2) and 277 genes (6.8%) are highly female biased (fold change ≥2) (Figure 18B). Among the silk genes the silk protein P25 (logFC -5.65) showed a high male biased expression. The important developmental genes, *nanos-M* (logFC -5.4) along with *bicaudal D-related protein homolog* (logFC -4.44) exhibited a consistent male biased expression (*nanos-M* and a related gene, *bicaudal C homolog 1-A* were also male biased at 78h). The genes encoding zinc finger protein 1 (logFC -3.8) and Zinc finger protein 729-like (logFC -3.09) showed a male biased expression even at 96h (these genes are male biased even at 78h). The genes encoding UDP-glucosyl transferase UGT48C1 (logFC 8.41), Vestigial (logFC 7.421), Transcription factor BCFI (Gata-beta) (logFC 5.49), Vitellogenin (logFC 2.76), Transcription factor GATA-6 (2.06), Sprouty (3.775), Sid-1 related gene 3 (logFC 2.04), Krueppel-like factor 10 (logFC 1.65), Fork head domain transcription factor (logFC 1.51), DnaJ homology subfamily C member (logFC 1.49), development related genes- Homeobox protein Nkx-6.2 (logFC 5.05) and Homeobox protein vvab-15 like (logFC 2.69) a few important ones that were highly female biased. Among these genes UDP-glucosyl transferase UGT48C1 and DnaJ homology subfamily C member were consistently female biased from 78h to 96h stages. At 96h stage, 674 uncharacterized (annotation not found) genes, 172 genes encoding several kinases, 130 zinc finger motif encoding genes, 83 genes encoding receptor molecules, 77 transcription factors/ related genes, 63 ubiquitin associated genes, 45 polymerase genes, 33 histone encoding genes, 33 genes of helicases, 23 protease encoding genes, 23 genes involved in splicing mechanism, 19 RNA-binding proteins, 19 genes involved in signaling pathways, 17 genes of translation process and 11 genes encoding BTB/POZ domains were differentially expressed between sexes.

At 120h stage, 2,596 genes were differentially expressed (FDR and p-value <0.05), among them 119 genes (4.5%) were highly male biased (fold change ≥2) and 403 genes (15.5%) are highly female biased (fold change ≥2) (Figure 18C). This is in contrast to the findings from the above two stages of 78h and 96h, where the male biased genes were more in number compared to the female biased genes. Among the silk genes, only Fibroin heavy chain-like gene (-0.523) showed a male biased expression. The developmental gene *nanos-M* (logFC-3.508) was male biased at this stage (120h) similar to 78h and 96h stages. A few of the important protein coding genes that exhibited male biased expression at 120h

stage are Juvenile hormone epoxide hydrolase-like protein 1 (logFC -7.6), Sex specific storage protein (logFC -7.0), Cuticular protein gycing-rich (logFC -6.7), Microvitellogenin-like (logFC -3.566), Osris 9 (-1.78), Spatzle-like protein (logFC -1.203), Sprouty (-1.16), Transcription factor BCFI (gata-beta) (logFC -1.306), BTB/POZ domain-containing protein KCTD16 (logFC -1.016), LIM homeobox transcription factor 1-beta (logFC -1.27), WD repeat-containing protein 34-like (logFC -1.505) and Probable RNA-binding protein 46 (logFC -1.878). Interestingly there were no zinc finger motif encoding genes that showed a profound male biased expression at this stage, instead many zinc finger motif encoding genes, viz., Zinc finger protein 433-like (logFC 2.408), Zinc finger and BTB domain-containing protein (logFC 1.77), Zinc finger protein Xfin-like (logFC 1.732), Zinc finger protein 454 (logFC 1.37), Zinc finger protein 845-like (logFC 1.267) and Centrosome-associated zinc finger protein CP190 (logFC 1.22) exhibited female biased expression. A few important protein coding genes that showed female biased expression at this stage are UDG-glycosyl transferase UGT48C1 (logFC 9.7), Dual specificity protein kinase KNS1 (logFC 6.67), Vitellogenin (6.44), Oxysterol-binding protein-related protein-6 (logFC 6.43), Transcriptional adapter 2B (logFC 1.978), Forkhead box protein J1-A like (logFC 1.971), Argonaute 1 (logFC 1.71), Oxysterol-binding protein-related protein 8 (logFC 1.59), MSL1 (logFC 1.55) and Scarlet (logFC 1.506). At this stage (120h), 498 uncharacterized (not annotated) genes, 98 zinc finger motif encoding genes, 45 receptor molecule related genes, 43 transcription related, 42 ubiquitin associated, 24 polymerases, 22 helicase genes, 22 histone related, 17 protease encoding genes, 12 RNA-binding protein encoding genes, 11 genes involved in signaling pathways, 11 genes associated with translation process, 10 genes involved in splicing and 7 BTB domain containing genes were differentially expressed. The genes *nanos-M* and *UDP-glucosyl transferase UGT48C1* showed a consistent male and female biased expression respectively at 78h, 96h and 120h, denoting their distinctive sex biased expression in these developing embryonic stages.

In head samples, 1015 genes were differentially expressed (FDR and p-value <0.05) among them, 148 genes (14.6%) showed higher male biased expression (fold change $\geq$2) and 125 genes (12.3%) showed higher female biased expression (fold change $\geq$2) (Figure 18D). In head, the protein coding genes, GPI ethanolamine phosphate transferase3 (-6.612), DnaJ homolog subfamily C member 3 (logFC -5.178), and several transcription factors like Fork head domain transcription factor slp2-like (logFC -5.49),

General transcription factor IIF subunit 1 (logFC -5.49), Eukaryotic translation initiation factor 3 subunit M (logFC -3.206) and Fork head box protein C1 (logFC -2.91) showed male biased expression. The protein encoding genes, Sex-specific storage-protein (logFC 8.577), Basic juvenile hormone-suppressible protein 2 (logFC 7.525), Hermansky-pudlak syndrome 5 (logFC 6.475), Attacin (logFC 6.21), FAM117B-like (logFC 5.49), MSL2 (5.49), XK-related protein 6 (logFC 5.234) and Vitellogenin 2 like (logFC 1.849) showed a female biased expression. Among the Zinc finger protein encoding genes, Zinc finger protein 598 (logFC -5.49), Zinc finger protein 551-like (logFC -5.5), Zinc finger protein 117 like (-5.178) and Zinc finger protein 583-like (logFC -3.099) showed male biased expression. The Zinc finger protein encoding genes, Zinc finger protein 26-like (logFC 5.709), Zinc finger protein 1-like (logFC 5.709), Zinc finger protein 883-like (logFC 5.23), Zinc finger martin-type protein CG9776 (logFC 1.99) and Zinc finger Ran-binding domain-containing protein 2-like (logFC 1.655) showed female biased expression. In head, 155 uncharacterized (not annotated) genes, 29 kinase encoding genes, 18 receptor related genes, 14 zinc finger motif encoding genes, 11 protease genes, 9 translation related genes, 7 helicase encoding genes, 7 polymerase encoding genes, 6 transcription related and 6 ubiquitin associated genes are differentially expressed.

Altogether the profiling of sex biased gene expression in the early embryonic and head samples suggest that a male biased expression (significant) is observed at 78h (350 male biased, 15 female biased) and it increases significantly at 96h stage (2641 male biased, 277 female biased). But this significant male biased expression rapidly decreased at 120h stage (119 male biased, 403 female biased) and in head sample (148 male biased, 125 female biased) the sex biased genes are almost equally distributed among sexes (Table 2 and Figure 19). When the set of differentially expressed genes are compared among the three embryonic samples, 165 genes were found to be differentially expressed in all the three embryonic stages (Figure 20). This unique sub-set of genes among the differentially expressed genes at each stage is 138, 2327 and 1009 at 78h, 96h and 120h stages respectively (Figure 20). Among the differentially expressed genes, the sub-set of genes that are common between 78h and 96h stages are 178 and this number increases to 1384 between 96h and 120h stages, whereas this number is only 33 between 78h and 120h stages (Figure 20). This indicates the dynamic nature of the sex biased gene expression profile in the early developmental stages.

**Table 2.** Total number of genes that showed a significant (edgeR, FDR and p-value<0.05) and high (fold change >2) sex biased expression in each sample.

| Sample | Male biased genes | Female biased genes |
|---|---|---|
| **78h** | 350 | 15 |
| **96h** | 2641 | 277 |
| **120h** | 119 | 403 |
| **Head** | 148 | 125 |

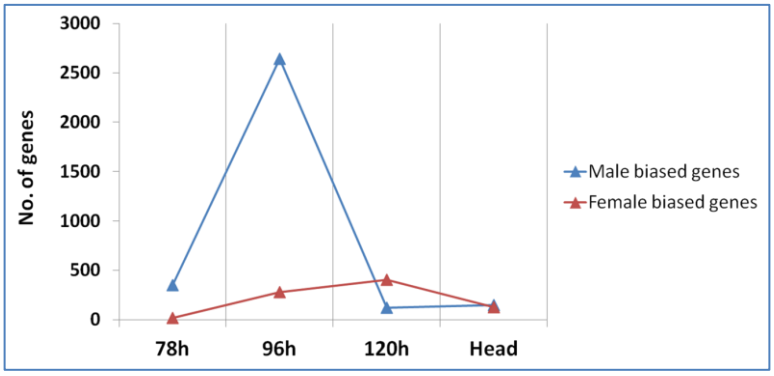


**Figure 19.** The total no. of genes that showed significant (edgeR, FDR and p-value <0.05) and highly (fold change>2) biased expression in sexes of the embryonic and head samples.

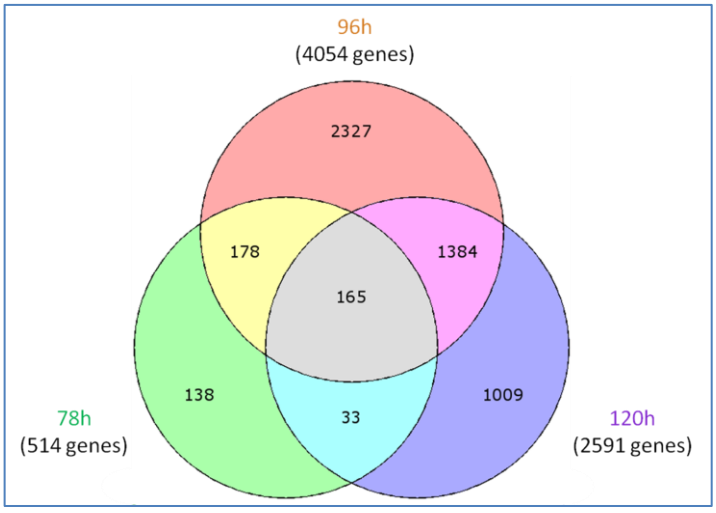


**Figure 20.** Venn diagram representing various sub-sets of genes among the differentially expressed gene sets (FDR and p-value <0.05) in the embryonic 78h, 96h and 120h stages.

### 3.3.3 *Bombyx* homologs of *Drosophila* genes involved in sexual development

Further, all the *Bombyx* homologs of various *Drosophila* sex determination, dosage compensation genes and sexual differentiation genes were identified by BLAST similarity search (Table 3). Among this sub-set the genes that showed significant biased expression (edgeR, p-value and FDR values of <0.05) were assessed for their most probable role in *B. mori* sex determination or differentiation. Six genes among this sub-set were Z-linked, namely *Masculinizer* (*Masc*), *Bmimp, Pale*, *Period, Unstream of N-ras* and *male specific lethal-1*. In 78h, 96h, 120h and head samples 3, 29, 6 and 5 genes respectively showed significant biased expression. The genes *daughterless* and *fem-1 homolog A-like* (*Apis* sex determination gene) genes showed a male biased expression at 78h and female biased at 96h (Table 4). A total of 23 genes of the sub-set (Table 4) showed a male biased expression at 96h compared to any other sample and 9 of these showed a female biased expression at 120h stage and no gene of this sub-set (except *Bmtra-2*) showed a male biased expression at this stage. In head samples, only 4 genes of the sub-set showed a female biased expression among them, *male specific lethal-2* showed a profound female biased expression. Only one gene, *Darkener of apricot* (*Doa*), involved in eye and embryonic development in *Drosophila* (YUN *et al.* 1994) exhibited a male biased expression (Table 4).

**Table 3.** The list of *Bombyx* homologs of *Drosophila* genes involved in sex determination, dosage compensation and sexual differentiation genes along with the *Bombyx* sex determination genes.

| ***Drosophila* gene** | ***Bombyx* homologs)** |
|---|---|
| **Establishing the X:A ratio** | |
| ***Daughterless*** | BGIBMGA003285 |
| | BGIBMGA003284 |
| ***Deadpan*** | BGIBMGA009932 |
| | BGIBMGA005390 |
| ***Degringolade*** | BGIBMGA006728 |
| ***Ovo*** | BGIBMGA000988 |
| ***Runt*** | BGIBMGA008906 |
| ***Scute*** | BGIBMGA001001 |
| ***Sex lethal*** | BGIBMGA013823 |
| | BGIBMGA012812 |
| | BGIBMGA005888 |
| ***Sisterless a*** | BGIBMGA009124 |
| **Somatic sex determination** | |
| ***Dissatisfaction*** | BGIBMGA010370 |
| ***Doublesex*** | BGIBMGA005108 |
| ***Fruitless*** | BGIBMGA006492 |
| ***Hermaphrodite*** | BGIBMGA011248 |
| ***Transformer*** | not found |
| ***Transformer 2*** | BGIBMGA009888 |
| ***Sans fille* (also known as *U1AsnRNP*)** | BGIBMGA012118 |
| ***Virilizer*** | BGIBMGA011914 |
| ***Wnt2*** | BGIBMGA013981 |

| | |
|---|---|
| **Germ-line sex determination and differentiation** | |
| ***Ornithine decarboxylase antizyme*** | BGIBMGA013339 |
| ***Ovarian tumor*** | BGIBMGA011335 |
| ***Ovo*** | BGIBMGA000988 |
| ***Sans fille* (also known as *U1AsnRNP*)** | BGIBMGA012118 |
| ***Sex lethal*** | BGIBMGA013823 |
| | BGIBMGA012812 |
| | BGIBMGA005888 |
| **Dosage compensation** | |
| ***Enhancer of bithorax/NURF301*** | BGIBMGA006818 |
| ***JIL-1*** | BGIBMGA011088 |
| ***Maleless*** | BGIBMGA010840 |
| ***Males absent on the first*** | BGIBMGA006995 |
| ***Male-specific lethal 1*** | BGIBMGA000539 |
| ***Male-specific lethal 2*** | BGIBMGA003686 |
| ***Male-specific lethal 3*** | BGIBMGA010146 |
| ***Painting of fourth*** | not found |
| ***RNA on the X-1*** | not found |
| ***RNA on the X-2*** | not found |
| ***Set2*** | not found |
| ***Suppressor of variegation 3-7*** | not found |
| ***Trithorax-like*** | BGIBMGA010221 |
| ***Upstream of N-ras*** | BGIBMGA000629 |
| ***Virilizer*** | BGIBMGA011914 |

| | |
|---|---|
| **Genes affecting courtship behavior** | |
| ***Btk family kinase at 29A/fickle*** | BGIBMGA012094 |
| ***Calcium/calmodulin-dependent protein Kinase II*** | BGIBMGA000408 |
| ***Dissatisfaction*** | BGIBMGA010370 |
| ***Doublesex*** | BGIBMGA005108 |
| ***Fruitless*** | BGIBMGA006492 |
| ***Cacophony*** | BGIBMGA007897 |
| ***Ovarian tumor*** | BGIBMGA011335 |
| ***Pale*** | BGIBMGA000563 |
| ***Period*** | BGIBMGA000486 |
| ***Protein C kinase 53E and Protein C kinase 98E*** | BGIBMGA014132<br>BGIBMGA000779 |
| ***Quick-to-court*** | BGIBMGA009004 |
| ***Sex lethal*** | BGIBMGA013823 |
| ***Slowpoke*** | BGIBMGA002400 |
| ***Spinster*** | BGIBMGA002400 |
| **Genes of *Bombyx* sex determination** | |
| ***BmImp*** | BGIBMGA000516 |
| ***Bmpsi*** | BGIBMGA004315 |
| ***Bmtra-2*** | BGIBMGA009888 |
| ***Bmznf-1/2*** | BGIBMGA004989 |
| ***Fruitless*** | BGIBMGA006492 |
| ***Sxl*** | BGIBMGA013823 |
| ***Ix*** | BGIBMGA010131 |

| | |
|---|---|
| ***Dsx*** | BGIBMGA005108 |
| ***Vitellogenin*** | BGIBMGA004585 |
| ***Masculinizer*** | BGIBMGA012300 |
| ***Fem-1 homolog A-like* (*Apis* sex determination gene)** | BGIBMGA010777 |
| | BGIBMGA010778 |
| ***Feminization 1 homolog b* (*Apis* sex determination gene)** | BGIBMGA001985 |
| ***Doa*** | BGIBMGA006461 |

### 3.3.4 Analysis of genome unmapped sequences

Additionally, all the three embryonic samples were also analyzed using CLC Genomics Workbench 7 beta version. The DGE results using CLC was comparable to that of the DGE results presented in section 3.3.1, above and hence the data is not shown here. Using CLC, even the genome unmapped reads were also analyzed for identifying the novel or unmapped transcripts. The de novo assembly of all the unmapped reads in male and female samples separately resulted in the assembly of 5,726 and 4,667 unmapped transcripts respectively. BLAST analysis of these transcripts against the small RNA database containing 38,493 entries (DDBJ/EMBL/GenBank accession numbers: AB386191-AB424683) resulted in hits for nearly 50% of these transcripts. Precisely, 2356 female unmapped transcripts could be precursors for 2292 small RNAs and 2591 male unmapped transcripts could be precursors for 2453 small RNAs. The percentage match identity is 100% between the small RNA sequences and the unmapped transcripts.

**Table 4**. List of *Bombyx* homologs that showed significant (edgeR, p-value and FDR<0.05) sex biased expression in embryonic and head samples.

| Sample | Sex biased | *Bombyx* homolog/ *Bombyx* gene involved in sex determination |
|---|---|---|
| **78h** | Male biased (logFC) | *Daughter less* (-1.96), *Fem 1 homolog A like* (-1.094), *Ca/calmodulin dependent PK II like* (-3.56) |
| | Female biased (logFC) | --------- |
| **96h** | Male biased (logFC) | *Deadpan* (-2.64), *Ca/calmodulin dependent PK II like* (-2.55), *Degringolade* (-2.2), *Ovo*, -2.28, *Sex-lethal* (-3.969), *Hermaphrodite* ( -1.02), *Masculinizer* (-5.92), *Bmznf-1/2* (-1.47), *Bmimp* (-1.105), *Spinster* (-1.22), *Slow poke* (-1.22), *Quick to court* (-1.897), *Period* (-5.09), *Btk-family kinase at 29A/fickle* (-2.017), *Upstream of N-ras* (-2.27), *Thorax-like* (-7.6), *Male specific lethal-3* (-0.8), *Male specific lethal-2* (-2.8), *Male specific lethal-1* (-2.33), *Males absent on the first* (-1.349), *JIL-1*(-5.26), *Enhancer of bithorax* (-2.13), O*varian tumor* (-2.88) |
| | Female biased (logFC) | *Daughter less* (2.57), *Fem 1 homolog A like* (1.629), *Pale* (6.13), *Bmpsi* (0.64), *Bmtra-2* (1.014), *Vitellogenin* (2.76) |
| **120h** | Male biased (logFC) | *Bmtra-2* (-0.38) |
| | Female biased (logFC) | *Ovo* (0.33), *Sex-lethal* (1.53), *Hermaphrodite* (0.56), *Thorax like* (2.54), *Upstream of N-ras* (0.359), *Ca/calmodulin dependent PK II like* (0.55), *Quick to court* (0.5), *Protein C kinase 53E and 98E* (1.14), *Period* (0.98) |
| **Head** | Male biased (logFC) | *Doa* (-0.55) |
| | Female biased (logFC) | *Bmznf-1/2* (1.937), *Ca/calmodulin dependent PK II like* (0.49), Upstream of N-ras (0.449), *Male specific lethal-2* (5.49) |

Further, in order to identify the W-transcriptome, from these unmapped transcripts, the commonly found transcripts were filtered based on BLAST between male and female unmapped transcripts. Thus only 862 transcripts were identified to be novel transcripts among them, 423 transcripts were from male embryos and 225 were from female embryos. Almost all of these transcripts ranged from 200 to 300 bases in size and are non-coding in nature. The BLAST analysis of the entire female unmapped transcripts against a W-BAC library suggested 62 of them to be presumably derived from W-chromosome. Among these transcripts, few are long ranging from 500 to 1500 bases. Four of these 62 transcripts gave BLAST hits (identity 98-100%) to *B. mori* ethanoaminephosphotransferase1 (GenBank accession: XM_012690691), *B. mori* DNA repair protein XRCC1 (GenBank accession: XM_012696723), *B. mori* mitogen activated protein kinase kinase kinase 10 like (GenBank accession: XM_012690189) and *B. mori* transcription factor CP2-like protein (GenBank accession: XM_004933314) in plus/minus strand orientation, suggesting that these transcripts could be long non-coding RNAs targeting the corresponding gene transcripts. Two among (female_contig_2036 and 3367) the 62 transcripts gave NCBI_BLAST hits (identity 98-100%) to *B. mori* zinc finger protein 260 (GenBank accession: XM_012690817) and *B. mori* peptidyl-prolyl cis-trans isomerase CWC27 homolog (GenBank accession: XM_004933211.2) respectively in plus/plus strand orientation, suggesting that these genes could have originated from W-chromosome. Interestingly, one of these transcripts, "female_contig_1149" showed 98% identity with the *Fem* (KIUCHI *et al.* 2014) non-coding RNA (GenBank accession: AB840787). This transcript also gave hits to five ovarian small RNAs with 100% identity (GenBank accessions: AB395669, AB412728, AB407068, AB420460, AB394669).

## 3.4 Discussion

The genes that show differential expression between females and males are considered as sex-biased genes. The evolution of sex-biased genes should be associated with the evolutionary forces natural selection, sexual selection and genetic drift (ELLEGREN and PARSCH 2007). Additional forces that shape and influence the sex-biased expression are sexual antagonism, gene duplication, dosage compensation and expression of sex-limited genes on Y or W chromosomes (ELLEGREN and PARSCH 2007). Further it is observed that the protein sequences by the sex-biased genes show a significant amino acid sequence divergence, especially the male biased genes between species compared to the expression of unbiased genes (ELLEGREN and PARSCH 2007). The differential gene expression between sexes is believed to be playing a key role in achieving various morphological, physiological, behavioral and somatic sexual differentiation of the organism. In the current context, the RNA-seq analysis of the embryonic stages and head samples has revealed a list of various differentially expressed genes. The number of sex-biased genes has increased greatly from 78h (520) to 96h (4,068) and this number decreased nearly to half at 120h (2,596) suggesting that 96h stage might represent a crucial developmental stage that exhibit a very high number of genes with sex biased expression. It is very interesting to note that the expression of the silk genes like *Fibroin heavy chain*, *Fibroin light chain*, *Silk protein-25* showed a higher male biased expression in the embryonic stages like 78h, 96h and 120h. The male biased expression of these silk genes in the early embryonic stages is dubious and awaits a logical explanation. This leads me to posit that the early male biased expression of silk genes could be physiologically linked to the property of higher silk yielding capacity of males, compared to that of females (the primary reason for this is that most of the protein is diverted to yolk synthesis in the females for egg production). It is surprising to note that a very important developmental gene, *nanos-M* (KOBAYASHI *et al.* 1996) is profoundly male biased in all the three embryonic stages that are analyzed (78h, 96h and 120h). In *Drosophila*, this gene is known to direct germline development and suppresses various somatic genes in germline progenitor cells (HAYASHI *et al.* 2004). The strong male biased expression of this developmental gene in the embryonic stages suggests that this protein may have a critical or very important role in the development of males or differentiation of males. Additionally bicaudal related genes also followed the same pattern of male biased expression in the early embryonic stages of 78h and 96h, similar to *nanos-M*. These bicaudal related genes are known to be directly involved in the localization of the

nanos proteins thus involved in the oocyte polarity and developmental fate of the embryos (WHARTON and STRUHL 1989).

Several important genes involved in various metabolisms exhibited a high male biased expression in the embryonic stages, especially many zinc finger motif encoding genes and transcription factors. It is very interesting to note that the zinc finger motif encoding genes that exhibited male biased expression at 78h and 96h stage are unique and none of them are male biased at 120h stage. At 120h stage, almost no zinc finger motif encoding gene exhibited a profound male biased expression, instead many zinc finger genes showed a female biased expression suggesting the dynamic expression profile of these zinc finger motif encoding genes which may be crucial in the development and sustainability of the embryos.

Among the genes that showed female biased expression, *UDP-glycosyl transferase UGT48C1* exhibited a profound female biased expression in all the three embryonic stages. By name, this gene is known to be involved in the process of glycosidation, which plays an important role in the process of inactivation and excretion of a variety of compounds. One class of UDP-glycosyl transferase (UGTs) is associated in this process. In insects, UGTs are involved in several processes like detoxication of plant allelochemicals, pigmentation, cuticle formation, and olfaction. In *B. mori*, 42 UGT genes were identified based on UGT signatures and homology similarity from other organisms. Their phylogenetic analysis revealed five major groups of UGTs (HUANG *et al.* 2008). Some UGTs are silkworm specific. The expression pattern analysis suggests that 36 UGT genes expressed in tissues with different patterns of expression profile, indicating that these genes might have different functions. In the current context, the *UDP-glycosyl transferase UGT48C1* showed a profound female biased expression in all the three embryonic stages and is not clustered with other silkworm UGT genes, indicating presumably a special function of this gene (HUANG *et al.* 2008). Further the gene *Vitellogenin* (*Vg*) is female biased even at 96h stage where a profound male biased expression of hundreds of genes is observed, this indicates the distinctive female biased expression of *Vg* gene. Additionally, the gene *Sex-specific storage protein 1* (*Sp1*) that accumulates in sex and stage specific manner (SAKURAI *et al.* 1988) is also female biased at 120h and in the head samples. From the hemolymph studies of larvae, it is suggested that the sex-dependent expression of *Vg* and *Sp1* is independent of sex related humoral factors and is genetically determined and developmentally regulated

(MINE *et al.* 1983), suggesting the female biased expression of *Vg* and *Sp1* in the embryos is obvious. In head sample, a relatively high number of kinases (29) showed differential expression, this indicates the high signaling activity in the brain, which is presumably involved in the physiological, morphological and behavioral differences between sexes. Additionally, the *Bombyx* homologs of various *Drosophila* sex determination, dosage compensation and sexual differentiation genes has yielded a mixed pattern and as expected (at 96h a very high number of genes are differentially expressed) there are relatively many genes of this sub-set that are male biased at 96h. Among the *Bombyx* homologs of sex determination, the genes *msl-2* and *Bmznf-1/2* showed a significant female biased expression in head. Interestingly two important genes of *Bombyx* sex determination, *Bmznf-1/2*, *Bmimp* and *Masc* are significantly male biased at 96h which might be essential in the development and sexual differentiation of the embryos.

In addition, the analysis of the unmapped transcripts against the available genome had resulted in the identification of thousands of small RNA precursors and presumably the W-chromosome derived transcripts mostly the long non-coding RNAs. The BLAST analysis of these unmapped transcripts resulted in the identification of hundreds of novel transcripts which are mostly non-coding in nature. Additionally, the BLAST analysis of these unmapped transcripts against a handful of W-BAC sequences revealed 62 transcripts, one of them include the W-derived "*fem*" non-coding RNA. This non-coding RNA is the precursor for the *fem* piRNA that targets the *masc* gene in *B. mori* sex determination. This lets me believe that these predicted W-derived transcripts could be a rich source of many such interesting piRNAs may have important roles in the process of sex determination and differentiation of *B. mori*. An analysis of the collected unmapped transcripts against the whole W-derived BAC library may result in the identification of the full range of the W-derived transcripts. This will add up to the W-transcriptome pool, which might unravel the completer structure and provide clear insights into its functional aspects of the W-chromosome in *B. mori* sex determination.

## 3.5 Summary

In the early stage of development, i.e., at 78h, hundreds of genes (520) showed a differential expression. This number surge to thousands at 96h (4068) and it decreases at 120h (2596). The DGE analysis suggested a very high male biased expression of many important genes of silk composition, developmental, transcription factors and many zinc finger genes which must have crucial roles in the process of development and sexual differentiation. In addition, the analysis of unmapped transcripts yielded thousands of precursors for the *B. mori* small RNAs and many non-coding transcripts that are presumably W-chromosome derived. Further analysis of these unmapped transcripts may help in uncovering the W-transcriptome and thus aid in a comprehensive understanding of the role of W-chromosome in *B. mori* sex determination.

# Chapter-IV

## Complete dosage compensation ensues later during the embryonic development in *B. mori*

## 4.1 Introduction

Sex chromosomes are believed to have evolved from autosomes through intermediate proto-sex chromosomes. Evolution of sex chromosome is thought to be initiated, once a chromosome from a pair acquires the sex determination gene (CARVALHO 2002). In the process of acquiring sex determination function (LAHN *et al.* 2001), the accumulation of sexually antagonistic mutations and repeat elements by Y (W) would have mostly contributed for the loss of homology with X/Z. Such loss of homology is believed to be the driving force for Y (W) recombination isolation (CHARLESWORTH 1978) and its degeneration via gene loss (CHARLESWORTH and CHARLESWORTH 2000) thus leaving exclusive sex determination function (SINCLAIR *et al.* 1990). Thus the evolution of Y for establishing two distinct sexes, males (XY) and females (XX) had resulted in "X chromosome aneuploidy" (BACHTROG 2013). This aneuploidy creates the dose difference in X linked genes, which could be deleterious if not compensated at the level of expression. Hence organisms have adopted a versatile gene regulatory mechanism called dosage compensation (DC), which is limited mostly to homogametic sex (X/Z) chromosomes in most species (MARIN *et al.* 2000).

Reports from flies to mammals suggest diverse epigenetic mechanisms for the phenomenon of DC in several male heterogametic species (XX/XY). In *Drosophila melanogaster*, the DC is achieved by direct upregulation of entire X-linked genes, selectively in males (XY) (LUCCHESI 1973; STRAUB *et al.* 2005). In *Caenorhabditis elegans* and Mammals, DC has been proposed to have evolved in a two step process (Ohno's hypothesis) (CHARLESWORTH 1996; MANK 2013; MANK *et al.* 2011; OHNO 1967; VICOSO and BACHTROG 2009). First there is a transcriptional upregulation of the X-linked genes in both sexes. Second, the down regulation of all the X linked genes in female sex (XX), to rescue from the detrimental effects of hyper transcription (CHARLESWORTH 1996; MANK 2013; MANK *et al.* 2011; OHNO 1967; VICOSO and BACHTROG 2009). In *Caenorhabditis elegans*, there is a transcriptional upregulation in males (XO) and hermaphrodites (XX) (DENG *et al.* 2011) where as in hermaphrodites an additional mechanism of transcriptional repression operates to rescue them from the deleterious effects of hyper expression of XX (ERCAN *et al.* 2007; MCDONEL *et al.* 2006; MENEELY and WOOD 1984). In case of mammals, there is a phenomenon of X inactivation in females and thus the single X in both sexes are upregulated to match the level of autosomes (DENG *et al.* 2011; KHARCHENKO *et al.* 2011; NGUYEN and DISTECHE 2006). In beetles, there is an

upregulation of X-linked genes in both males (XY) and females (XX) with no down regulation of XX in female sex (PRINCE *et al.* 2010). This leads to hyper expression of XX in females (PRINCE *et al.* 2010). But this hyper expression in females has been challenged recently (MAHAJAN and BACHTROG 2015), suggesting the existence of a proper DC in beetles. In *Anopheles stephensi*, DC was found to be complete (JIANG *et al.* 2015) with equally expressing X-linked loci between sexes but in *A. gambiae* it is incomplete (ROSE *et al.* 2016).

The patterns of DC in XX/XY species suggest that, the mechanism of DC operates in a chromosome wide manner and has evolved not only to 1) equalize the expression of sex chromosomes between sexes (first condition, $X_{male}=X_{female}$) but also 2) to abolish the expression disparity between autosomes and sex chromosomes in both sexes (second condition, X=A) (NGUYEN and DISTECHE 2006). As these two conditions were consistently observed in XX/XY species with DC, they have become the evolutionary precedent. Hence DC is viewed as complete and conventional, only when both these conditions are fulfilled in an organism. The first condition of DC is believed to surpass the deleterious effects of sex chromosome aneuploidy (CONRAD and AKHTAR 2012) and it appears to be fundamentally right as it implies the necessity of DC in a species. In contrast, the second condition of complete DC must have been set based on two things 1) insights from sex chromosome evolution (BULL 1985; CHARLESWORTH 1996) and 2) the general pattern of equally expressing autosomal and X-linked genes, in most of the male heterogametic species (XX/XY) studied for DC. Thus the second condition of complete DC appears to be more instinctively assigned. Thus following this assumption, the ancestral average expression of proto-X or proto-Z need not be necessarily equal to that of autosomes, as even the individual autosomes differ at their average expression levels. Indeed, the ancestral expression of proto-X or proto-Z will be determined by the gene constitution and their expression profile. Hence this ancestral expression of proto-X or proto-Z is a crucial factor in determining the path of DC evolution in a species.

Further, as the patterns of DC are highly variable, evolved independently and dynamic across sex determination systems and species (MANK 2013; MANK *et al.* 2011; MARIN *et al.* 2000), a variety of approaches with a high degree of flexibility are anticipated for attaining "DC" in a species. These approaches may not be following the evolutionary precedent of XX/XY system, for the approval of existence of DC in a species. Based on these arguments, a species can be considered to have DC and fulfill the primary objective

of DC if it satisfies the "first condition" ($X_{male}=X_{female}$). As the "second condition" (X=A) of complete DC is set by the evolutionary precedent, it may be limited to XX/XY systems and may not be universally applicable and readily extrapolated to female heterogametic systems like ZZ/ZW or ZZ/ZO.

Among female heterogametic species (ZZ/ZW), DC was first assessed in chicken (ELLEGREN *et al.* 2007; ITOH *et al.* 2007) and was found to be ineffective as the Z-linked genes showed ~1.4 fold higher expression in males. This was accompanied by similar results from *Bombyx mori*, a lepidopteran (ZHA *et al.* 2009) and a trematode parasite, *Schistosoma mansoi* (VICOSO and BACHTROG 2011), giving an impression that DC is generally absent or incomplete in female heterogametic (ZZ/ZW) system. But recent genome-wide studies based on RNA-sequencing in ZZ/ZW systems have yielded an unexpected result of complete DC in a lepidopteran, *Manduca sexta*, with equally expressing Z-linked genes between sexes and with an expression parity between autosomal and Z-linked genes (SMITH *et al.* 2014). This has renewed much interest in the phenomenon of DC in ZZ/ZW species. However, in another lepidopteran species *Plodia interpunctella*, (HARRISON *et al.* 2012) DC was demonstrated to be incomplete as it showed a just over half of average expression of female Z-linked genes to that of males. Recently, in Heliconius (a butterfly genera), dosage mechanism was reported to be imperfect or not complete, due to a consistent male biased expression of Z-linked genes in various tissues tested (WALTERS *et al.* 2015).Thus a mixed pattern of DC mechanisms can be observed in female heterogametic species.

DC in female heterogametic species is proposed to be specified to a subset of dose sensitive Z-linked genes (incomplete DC; (MANK 2013)) and in a few others it may be operating in a chromosome wide manner (SMITH *et al.* 2014). In addition, the ZZ/ZW systems would have adapted many auxiliary phenomena for assisting the primary objective of DC. For example the enrichment of the male biased genes on the Z-chromosome in a few female heterogametic species (ARUNKUMAR *et al.* 2009; KAISER and ELLEGREN 2006; ZHANG *et al.* 2010), which lead to sex chromosome biased expression (SCBE) of a set of Z-linked genes. The female heterogametic species may be taking an advantage of SCBE. One striking example would be the sex determination in ZZ/ZO system (in wild silkmoths, *Antheria assama* and *A. mylita*), where ZZ are males and ZO are females. Here the double dose of Z-chromosome stands as an essential criterion for male determination, presumably due to the male biased expression of at least a subset of Z-linked genes. Thus in this aspect, the sex biased expression observed in most of the ZW species may be a general norm and

not a disability, moreover their successful survival and proliferation supports this assumption. Thus, the homogametic sex chromosomes (X/Z) were considered to be highly influenced by two phenomena; 1) the SCBE and 2) the DC mechanism. It is not very clear whether SCBE and DC have co-evolved or DC is followed by SCBE or vice versa (VICOSO *et al.* 2013). The SCBE of X (Z) chromosome, achieved by an "in and out" gene trafficking (ELLEGREN 2011; WANG *et al.* 2012b) may strengthen the sexually dimorphic traits; besides this can be perceived as a counter attacking force of DC.

The initial reports for DC in *Bombyx mori* (lepidopteran), based on expression profile of a set of a few Z-linked genes has suggested an incomplete DC, where males showed higher expression of Z-linked genes, (SUZUKI *et al.* 1998; SUZUKI *et al.* 1999). This was further confirmed by global, microarray analysis (ZHA *et al.* 2009). However, another study using reanalysis of the same microarray data draws a clue for the possibility of a globally operating DC mechanism in *B. mori* (WALTERS and HARDCASTLE 2011). Recently, RNA-sequencing was proven to be an efficient tool in addressing DC on a genome-wide scale (HARRISON *et al.* 2012; JIANG *et al.* 2015; ROSE *et al.* 2016; SMITH *et al.* 2014; WALTERS *et al.* 2015). Two recent reports (based on RNA-Sequencing), one suggesting the involvement of a Z-linked gene called *masc* in *B. mori* DC (KIUCHI *et al.* 2014) and another for the existence of an incomplete dosage in very early stage and most of the Z-linked genes are dosage compensated by 78h of development (KAWAMOTO *et al.* 2015). Further the process of sex determination is governed by the differential splicing of Bombyx *doublesex* gene, *Bmdsx.* It has two splicing isoforms, *Bmdsxf* in females and *Bmdsxm* in males. These isoforms produce differential proteins, having antagonistic functions in sexes thus inducing sexual differentiation. Based on the *Bmdsx* splicing in eggs at various developmental stages, it is found that *Bmdsxf* splicing isoform is predominant at 12h of development. This could be due to the maternal deposition of *Bmdsxf* pre-mRNA (SAKAI *et al.* 2014). In the same study, at 24h of development, there is a shift of splicing from female to male form, indicating the endogenous expression of *Bmdsx* mRNA(SAKAI *et al.* 2014). However in this study few eggs showed an equal expression of *Bmdsxf* and *Bmdsxm* isoforms, probably these are female eggs. As the development progresses, there is a shift in splicing from the predominant or equally expressing *Bmdsxm* isoform to *Bmdsxf* isoform significantly in female eggs between 29h to 32h (SAKAI *et al.* 2014). All these stages were studied in non-diapause strains (SAKAI *et al.* 2014).

However in the current study, we did RNA-sequencing for three different embryonic stages of bivoltine strain (78h, 96h, 120h), 5$^{th}$ instar larval heads and BmN cells and analyzed the relative expression of Z-linked genes to that of autosomal (Z:A) and the Z-linked genes between sexes (M:F). In bivoltine strain, the eggs undergo diapause and are not hatched in 10 days hence the eggs have to be acid treated to break the diapause. It is also known that the development of these eggs is comparatively slower than the non-diapause eggs; hence there is a possibility of delayed development in the eggs. In diapause eggs we have found a similar pattern of *Bmdsx* splicing shift in male and female embryos, however at later stages of development, i.e., at 78h, 96h and 120h. We have used 5$^{th}$ instar larval heads as a reference sample (DC is expected to be established) for the egg samples. Based on the differential splicing of *Bmdsx*, we consider the 78h and 96h stages to be before sex determination stages and 120h to be after sex determining stage. Hence the sex is determined in between 96h and 120h stage of development in diapause eggs of *B. mori* (Figure 21A). However in both the studies (diapause and non-diapause strains) have found that it is the differential splicing of *Bmdsx* that occurs first followed by the advent of dosage compensation in *B. mori*. From these observations, we infer that though the time points for *Bmdsx* differential splicing and DC vary between non-diapause and diapause strains, the patterns of the occurrence of these processes is sequential and are comparably similar.

## 4.2 Materials and Methods

### 4.2.1 Sample collection, library preparation and RNA-seq

Two W-chromosome mutant strains of *B. mori* were used: 1) Japanese sex limited (JPSL) for the sexed embryo collection and 2) Sex limited strain (QGSLO) for the sexed larval heads. In JPSL, the translocation of chromosome 10 fragment harboring *kynurenine monooxygenase* gene on to the W-chromosome is believed to be responsible for the development of dark brownish serosal pigmentation, which acts as a visible marker to differentiate female embryos as early as 36 hours post oviposition (hpo). In the QGSLO strain, female larvae can be easily distinguished by distinct crescent shaped markings on dorsal side of larva from the $4^{th}$ instar stage. For the embryo collection, JPSL strain moths were set for 4 hrs mating at room temperature; transferred to 4˚C overnight, followed by depairing and were set in dark at room temperature for 2hrs for a uniform egg laying. Collected eggs were cold acid treated at an age of 20 hpo to break the diapause and were thoroughly washed under running tap water and incubated at 25˚C for the development to proceed. At 36 hpo, male and female eggs were segregated based on the serosal pigmentation. Sampling of 200 each of male and female embryos was done at 78 hpo, 96 hpo and 120 hpo. For the head tissue collection, larvae of (QGSLO) fifth instar $5^{th}$ day were numbed on ice for 30 min, decapitated, 10 male and 10 female larval heads were pooled and snap frozen in liquid nitrogen until use.

For collecting BmN cells, regularly passaged in TC-100 (SIGMA) insect media with 10% FBS (Gibco) log phase cells were selected, slogged, pelleted in PBS and stored in -80 ˚C until use. From the collected samples, total RNA was isolated and on-column DNAseI treated for the removal of genomic DNA contamination using Direct-Zol RNA isolation kit (Zymo Research). RNA libraries were prepared following the TruSeq RNA sample preparation kit v2 (Catalog No.: RS-122-2001) from Illumina using 1 µg of mRNA. Sequencing was performed on Illumina 1000 HiSeq platform (C-CAMP, Bangalore). In total, not less than ~60 million pair-end reads of 100 bp, for each of the sexed embryonic stages and head samples were generated. For simplicity, 78 hpo, 96 hpo and 120 hpo embryonic stages are referred to as 78h, 96h and 120h respectively throughout the thesis.

### 4.2.2 RNA-seq read quality filtration, mapping and data analyses

The RNA-seq read quality was assessed using the package FastQC (ANDREWS 2010). The adapter removal and quality trimming was performed using Trimmomatic version 0.35 (BOLGER *et al.* 2014). The leading and trailing low quality or N bases were removed below quality 3, reads were scanned with a 4 base sliding window, and clipped when the average quality per base dropped below 20, and read sequences below 30 bases in length were dropped. The filtered paired-end reads were mapped against *B. mori* genome sequence and its annotations [downloaded from Ensembl release 29 (2015) (GCA_000151625.1.29) (FLICEK *et al.* 2014)] using bowtie2 version 2.2.6 (LANGMEAD and SALZBERG 2012; LI and DEWEY 2011) with default parameters. The aligned reads were filtered to keep only uniquely mapped reads. SAM/BAM conversions, sorting, indexing and filtering were performed with SAMtools version 1.2. (LI *et al.* 2009). The alignment files (SAM format) so obtained were imported in to Seqmonk software (ANDREWS 2015). While importing the SAM files in to Seqmonk, the libraries were treated as pair-end and duplicate reads were eliminated. The log2 transformed FPKMs for genes were quantitated, by correcting for DNA contamination, transcript isoforms were merged, transcript length correction was made, besides excluding the genes with no or very low read counts (considered as noise in Seqmonk) to avoid bias in the data that might skew the analysis. The raw read counts table and quantitated genes report (FPKM values) of Seqmonk analysis was exported to Excel (Microsoft) and further analyzed. Moreover, for *B. mori* as the GFF file from ensembl was not annotated with the chromosome number, which is required for current analysis. So to overcome this, the "description" code of each gene in GFF file was replaced with its corresponding chromosome number, based on the scaffold identity from kaikobase (SHIMOMURA *et al.* 2009) annotation data using custom shell scripts.

### 4.2.3 Statistics for Z dosage and data representation

Based on the scaffold mapping, the genes were grouped into autosomal (A) and Z-linked (Z) genes. The unmapped genes were excluded from further analysis. Similar to a conventional dosage analysis, the dosage in *B. mori* was tested by two estimates. They are 1) Z:A ratios (autosomal relative expression of Z-linked genes) and 2) M:F ratios (sex biased expression of A and Z-linked gene expressions). To assess the Z dosage effects in the samples, the ratio of autosomal and Z-linked gene expression (Z:A) was calculated within each sample (male, female) and this ratio was compared between the sexes. The true expression (in terms of FPKM) of all the genes were estimated by choosing the option,

"Don't quantitate probes with no counts" in Seqmonk while probe quantitation. The Z:A ratios were estimated using this "true expression dataset", resulted in mapping of ~10,000 genes for embryo samples and ~7,000 genes for head samples. The mean and median Z:A ratios were estimated which represents the relative expression level of Z-linked genes compared to that of autosomal. Bootstrapping (10K) of log transformed FPKM data was performed to find the 95% confidence interval for the Z:A point estimate of the median, using the online web tool STATKEY, (ROBIN *et al.* 2015). We compared the median level expression differences between autosomal and Z-linked genes within the sex for each sample individually by nonparametric Mann Whitney U test (Wilcox rank sum test), conducted in R package (TEAM 2010). Further the Male: Female (M:F) ratio distributions were calculated from FPKM values and raw counts data of genes which showed expression in both sexes (true expressed dataset). The raw reads were further subjected to TMM normalization for the estimation of M:F ratio distributions. The log2 (M:F) density distributions were generated using Wessa.net histogram, online web tool (WESSA 2015) to reveal the overall picture of the sex biased distribution of the autosomal and Z-linked genes in the samples. The median level differences between the M:F distributions for autosomal and Z-linked genes were also tested by MWU test . In order to compare and explore any discrepancies found between male and female median Z-linked genes at various levels of gene expressions, an unpaired comparison for the Z-linked genes expression data between sexes was conducted. For this the Z-linked log2 FPMK distribution data for both the sexes were sorted independently in descending order and were divided in to quartiles (Q1-Q4), each representing different magnitudes of gene expression (High-Q4, medium-Q3, low-Q2 and very low-Q1 expressing genes). The median expression difference within quartiles between sexes was tested by MWU test. Additionally to view the profile of Z-linked genes expression, all the Z-linked genes were filtered and saved as annotation track from which a heat map was generated based on clustering. For generating heat map in Seqmonk, even the loci with no true expressions were also quantitated.

### 4.2.4 Quantitative RT-PCR

The relative expression of selected genes was validated through reverse transcription quantitative PCR (ABI 7500). The reaction was set using SYBR Premix Ex Taq (Tli RNaseH Plus) from Takara Bio Inc. The reaction mixture included cDNA sample of 3 µl (diluted to 10 ng/µl), 0.2 µM primers in a final volume of 20 µl of master mix. Reaction conditions were: 95 for 30 sec, 95 for 5 sec and 60 for 34 sec. The standard curve analysis

was done using ABI SDS software version 1.2.3. The reactions were carried out in triplicates and the relative expression was determined using ΔCt analysis. Fold change (FC) values for male samples relative to female samples (calibrator) were obtained by normalizing the gene expression values to the *rp49* as endogenous/internal reference control separately (TENG *et al.* 2011). For primer sequences, see Annexure V.

## 4.3 Results

### 4.3.1 The Z:A ratios; relative expression of Z-chromosome

The Z:A ratio, is informative on how DC is achieved in organisms. It provides an autosomal relative expression of Z-linked genes (MANK 2013). The Z:A ratio values of 0.5, 1 and 2 correspond to half, equal and double expression of Z-linked genes respectively to that of autosomes. The mean (using FPKM data) and median based (using log2 transformed FPKM) Z:A ratios were estimated from the "true expression dataset" (Figure 21B and Table 5). Thus estimated mean and median Z:A ratios ranged from ~0.4 to ~0.6 in both sexes suggesting a significantly low expressing Z-linked genes compared to that of autosomes, in all the analyzed samples (embryonic stages of 78h, 96h, 120h; larval head and BmN cells). The nonparametric Mann Whitney U tests (A≠Z) statistically supports the significant difference observed between A and Z expression levels (Figure 22 and Table 6).

At 78h, the median Z:A ratios for male (0.39) and female (0.42) were substantially lower than 0.5, suggesting an initially lesser than half expression of Z-linked genes to that of autosomal (Figure 22 and Table 6). At 96h stage, this ratio was higher in male (0.51) compared to female (0.37) embryos, depicting a relatively (autosomal) increased expression of Z-linked genes in males. This scenario was reversed at 120h stage where the male (0.47) and female (0.54) ratios indicate a relatively (compared to that of autosomes), increased expression of Z-linked genes in females. These median Z:A ratio profiles of the embryonic samples (78h to 120h stages) present a dynamic picture of relative (compared to that of autosomes) expression of Z-linked genes in the process of acquisition of DC (Figure 22 and Table 6). Head (male=0.57, female=0.61) and BmN (0.65) cells showed a median Z:A ratio of ~0.6 suggesting a relative just over half expression of Z-linked genes to that of autosomes (Figure 22 and Table 6).

Quartile based analysis was also done using the "true expression dataset" and the results were consistent at various magnitudes of expression (Figure 23). The point estimate of Z:A ratios and the non-parametric MWU tests together suggests a substantially reduced expression of Z-linked genes to that of autosomal in both sexes.

**Table 5.** Summary of average expression of Z-linked and autosomal loci across embryonic, head and BmN samples

| Statistic | 78h | | 96h | | 120h | | Head | | BmN |
|---|---|---|---|---|---|---|---|---|---|
| | **Male** | **Female** | **Male** | **Female** | **Male** | **Female** | **Male** | **Female** | **Ovary** |
| **Mean autosomal FPKM** | 26.34339 | 32.68268 | 19.31366 | 22.9903 | 18.73881 | 17.45362 | 30.64596 | 30.85957 | 20.1034 |
| **Mean Z-linked FPKM** | 13.55054 | 16.68408 | 12.42164 | 12.30837 | 11.8242 | 11.28153 | 19.1806 | 19.69835 | 11.65642 |
| **Mean Z:A ratio** | **0.514381** | **0.510487** | **0.643153** | **0.535372** | **0.631001** | **0.646372** | **0.625877** | **0.638322** | **0.579823** |
| **Median autosomal FPKM** | 3.99718 | 3.490582 | 7.354102 | 5.405826 | 8.613708 | 9.700895 | 3.579668 | 3.996381 | 3.569715 |
| **Median Z-linked FPKM** | 1.567086 | 1.480323 | 3.662872 | 1.958461 | 4.0324 | 5.25235 | 2.053293 | 2.389728 | 2.320983 |
| **Median Z:A ratio** | **0.392048** | **0.424091** | **0.498072** | **0.362287** | **0.468138** | **0.541429** | **0.573599** | **0.597973** | **0.650187** |
| **MWU p-value: A ≠Z-linked** | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 |
| **No. Z-linked loci** | 399 | 321 | 435 | 379 | 460 | 467 | 382 | 361 | 398 |
| **No. A linked loci** | 9511 | 8235 | 9925 | 9224 | 10125 | 9990 | 7825 | 7841 | 9248 |

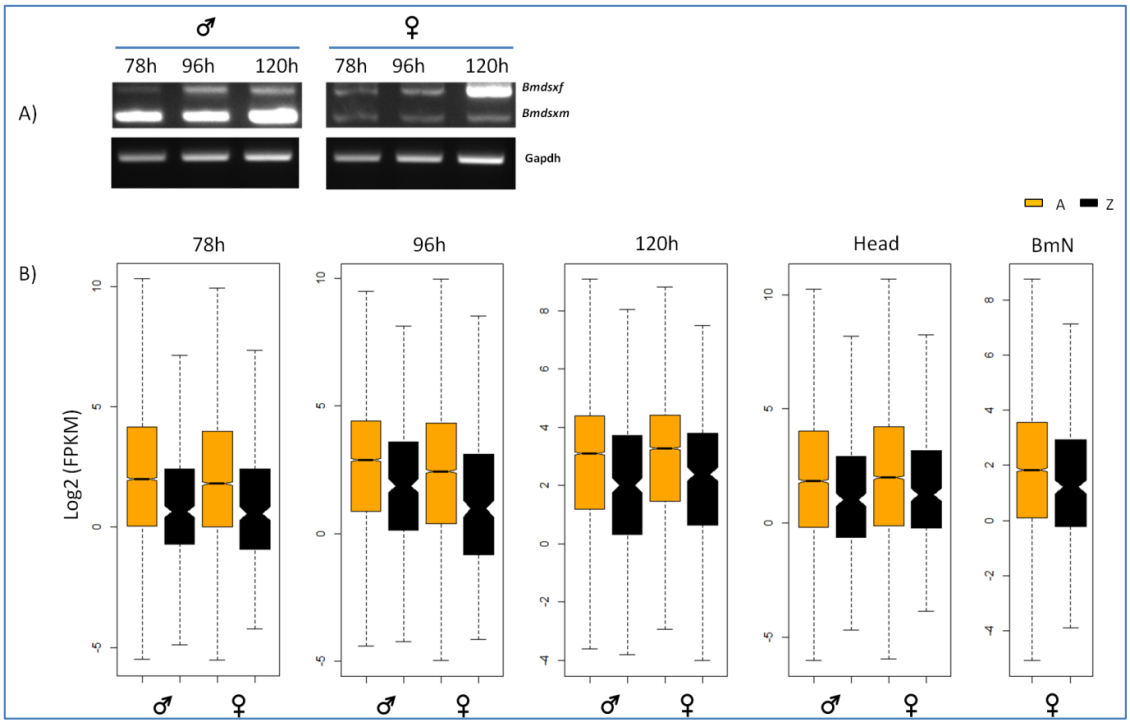


**Figure 21.** A) The sex specific differential splicing pattern of *Bmdsx* in the embryonic stages of 78h, 96h and 120h. B) Box plot of the $\log_2$ transformed FPKM distribution of autosomal (A; yellow) and Z-linked (Z; black) truly expressed genes of male and female samples. The boxes represent the inter quartile range (IQR), the notch of box plots represents median (horizontal line) expression with a 95% confidence interval and the outliers were not plotted. The data indicates a significantly reduced expression of Z-linked genes when compared to that of the autosomes by nonparametric tests and mean, median point estimates.

**Table 6.** The bootstrapped Z:A ratios (median) and their confidence intervals (lower-CI and upper-CI)

| Sample | MEDIAN | LOWER-CI | UPPER-CI |
|---|---|---|---|
| **m 78h** | **0.389312** | 0.331252 | 0.430176 |
| **f 78h** | **0.418994** | 0.366021 | 0.480964 |
| **m 96h** | **0.50698** | 0.40669 | 0.607097 |
| **f 96h** | **0.370874** | 0.307573 | 0.472701 |
| **m 120h** | **0.474671** | 0.417544 | 0.559419 |
| **f 120h** | **0.542239** | 0.485654 | 0.618566 |
| **m head** | **0.570777** | 0.463294 | 0.688725 |
| **f head** | **0.609628** | 0.53812 | 0.718968 |
| **BmN** | **0.646624** | 0.587231 | 0.719966 |

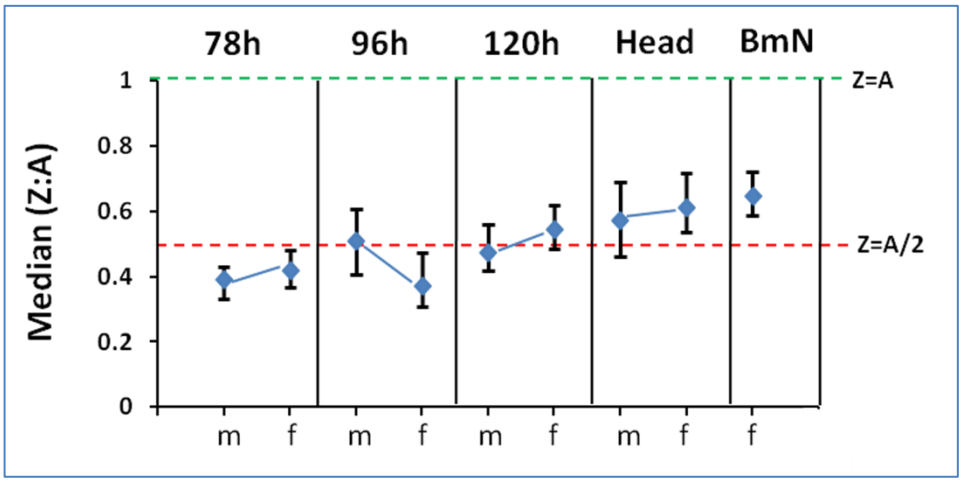


**Figure 22.** The bootstrapped median of autosomes to Z-chromosome (Z:A) gene expression ratios for the samples. The dashed line corresponds to the relative expression of Z to that of autosomes. Z:A of 1 and 0.5 indicate an equal (Z=A) and half expression (Z=A/2) of Z-chromosome to that of autosomes respectively. m=male and f=female. Error bars represent the 95% confidence intervals for the median as estimated by 10,000 bootstrap replicates.

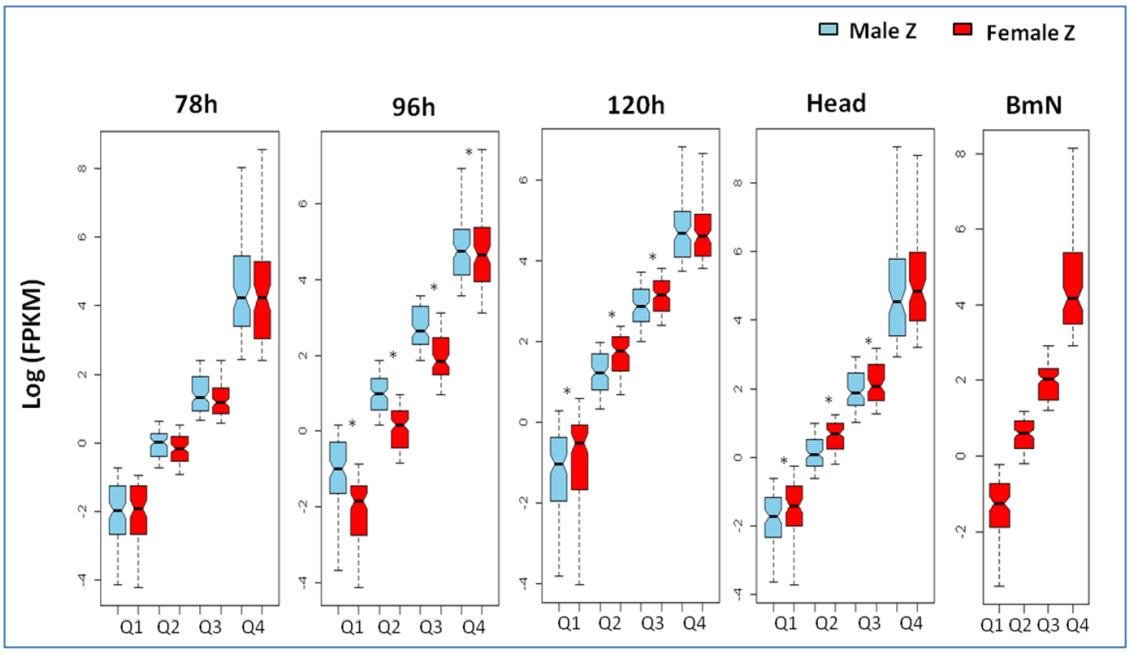


**Figure 23.** Comparison of quartile expressions for Z-linked genes dataset used for Z:A analysis. The $\log_2$ FPKM expression data for the Z-linked genes were segregated as quartiles based on independent binning. The boxes represent the inter quartile range (IQR), the notch of box plots represents median (horizontal line) expression with a 95% confidence interval and the outliers were not plotted. An asterisk (*) indicates a significant difference between quartile expressions (MWU, $p < 0.05$).

### 4.3.2 Male: Female (M:F) ratio distributions - sex biased expression of Z-chromosome

The M:F ratio distributions indicate the sex biased expression of autosomal and Z-linked genes. These distributions are the direct comparison for testing dosage in a species. Three things were assessed from M:F ratio distributions. They are i) Sex biased expression of Z-linked genes, ii) Sex biased expression of autosomal genes and iii) The difference in the M:F distributions of A and Z-linked genes. The Z-linked genes showed a significant male biased expression at 78h (MWU, p=0.00056) and 96h (MWU, p=2.127E-09) embryonic stages, whereas such difference was not observed at 120h (MWU, p=0.08008) embryonic stage and also in head samples (MWU, p=0.5517) (Figure 24). Based on these findings it is assumed that DC (sex chromosome dosage compensation) would have initiated after 96h and became established at 120h stage of embryonic development. Interestingly, the male biased expression of Z-linked genes at 78h (MWU, p=0.00056) and 96h (MWU, p=2.127E-09) stages is also coupled with a significant male biased expression of autosomal genes at 78h (MWU, p=<2.2E-16) and 96h (MWU, p=<2.2E-16) samples. This could be due to the dosage uncompensated effects of Z-linked genes (male biased) at 78h and 96h stages which can significantly influence the expression of autosomal genes (Figure 25 and Table 7). But a significant difference was also observed for the autosomal expressions between sexes at 120h (MWU, p=0.01172) and head (MWU, p=0.006069) samples where the effect of DC are visible (Figure 25 and Table 7), this could be due to the local effects being established by the process of sexual differentiation.

**Table 7.** Summary of average M:F gene expression ratios of Z-linked and autosomal loci in embryonic and head samples

| Sample | Z-linked loci | Autosomal loci | Median Z-Linked log2 (M:F) | Median Autosomal log2(M:F) | Z:A ratio of Medians (Not log2) | MWU p-values; Male Z ≠ Female Z | MWU p-values; Male A ≠ Female A | MWU p-values; Autosomal M:F ≠ Z-linked M:F |
|---|---|---|---|---|---|---|---|---|
| **78h** | 310 | 8102 | 0.617812 | 0.483265 | 1.097748 | 0.0005608 | <2.2E-16 | 0.005881 |
| **96h** | 365 | 8983 | 1.02793 | 0.44333 | 1.499623 | 2.127E-09 | <2.2E-16 | <2.2E-16 |
| **120h** | 442 | 9632 | -0.17748 | -0.00368 | 0.886504 | 0.08008 | 0.01172 | 1.588E-09 |
| **Head** | 339 | 7210 | -0.09608 | -0.12558 | 1.020661 | 0.5517 | 0.006069 | 0.4517 |

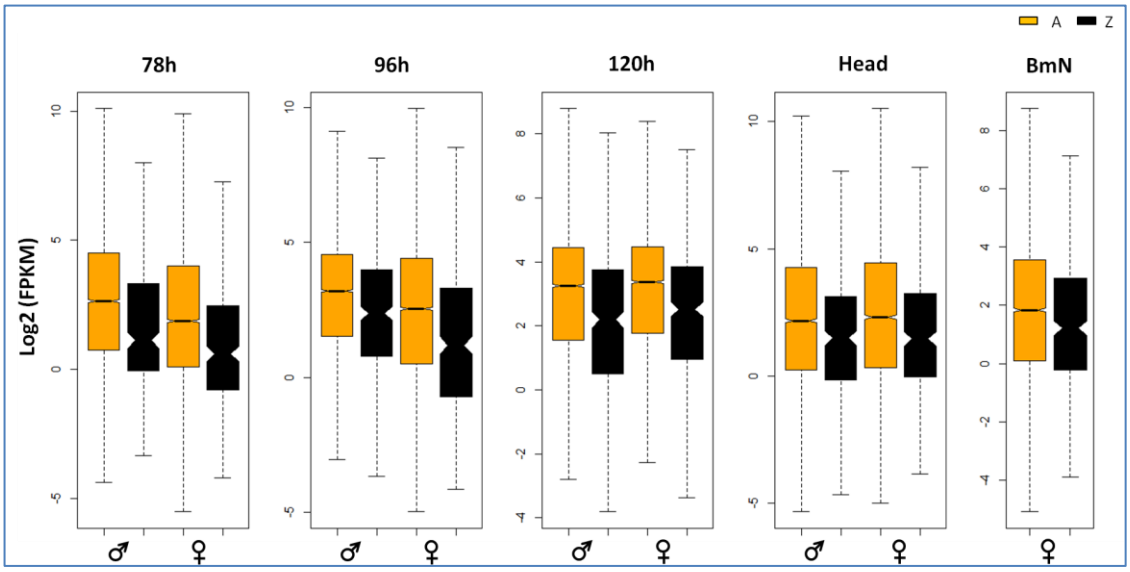


**Figure 24.** Box plot of the $\log_2$ transformed FPKM distribution of autosomal (A; yellow) and Z-linked (Z; black) genes that showed true expression in both male and female samples. The boxes represent the inter quartile range (IQR), the notch of box plots represents median (horizontal line) expression with a 95% confidence interval and outliers were not plotted. The data indicates a significant male biased expression of autosomal and Z-linked genes at the early embryonic stages of 78h and 96h stages.

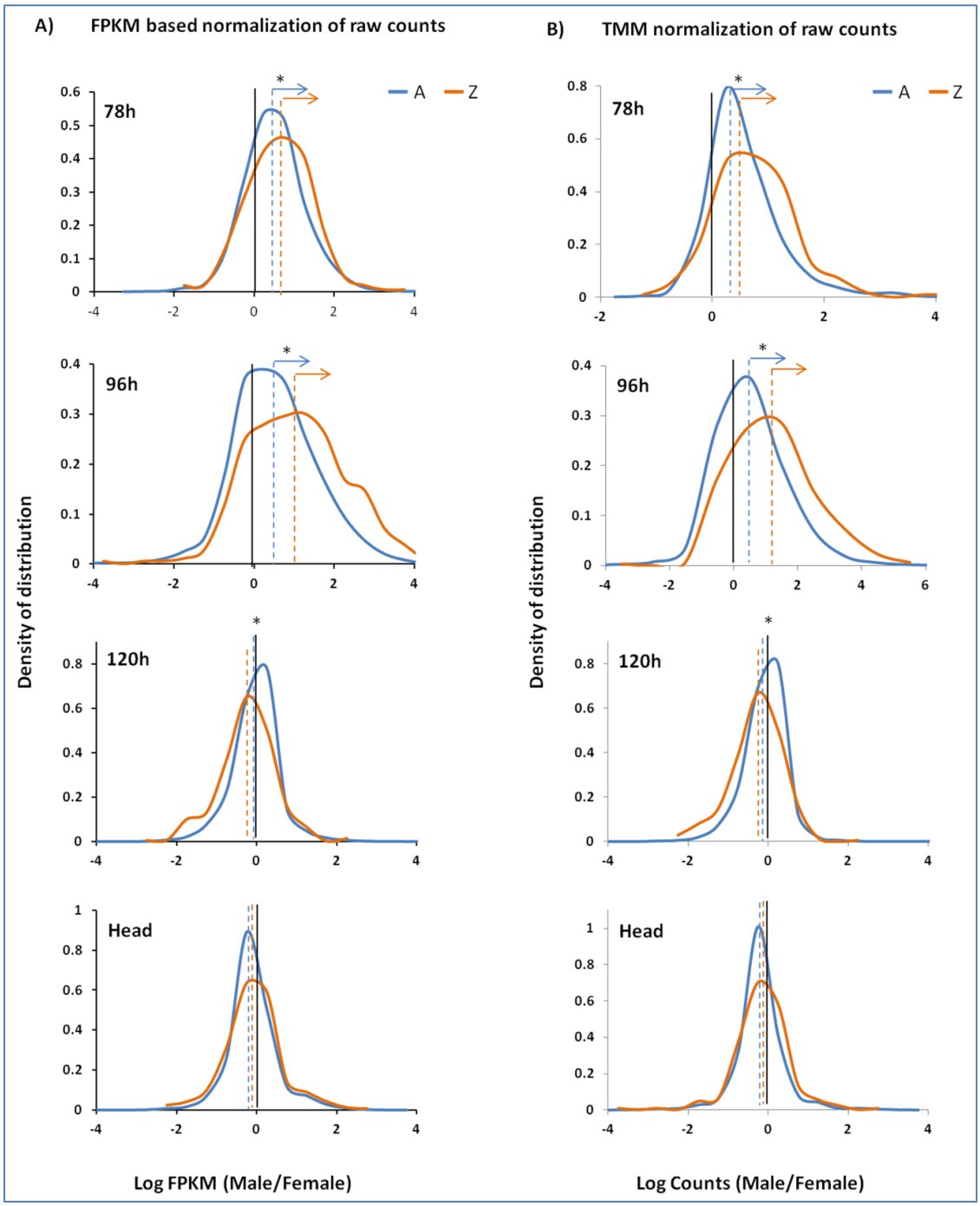


**Figure 25.** The male to female autosomal and Z-linked gene expression ratio (M:F) distributions for the samples. A) The M:F distributions from FPKM based normalization of raw counts B) The M:F distributions from TMM normalization of raw counts. The dashed vertical lines represent the median of the frequency distributions (shown in dashed lines). An asterisk (*) represents the significant difference (MWU, $p<0.01$) observed between autosomal and Z-linked M:F distributions. At 78h and 96h stages, the arrows indicate the profound male biased expression of autosomal (MWU, $p<2.2E-16$) and Z-linked (MWU, 78h- $p=0.0005$, and 96h- $p=2.127E-9$) genes.

Further the difference between the M:F distributions for A and Z-linked genes were estimated. A significant difference was observed between these distributions at 78h (MWU, p=0.005881), 96h (MWU, p=<2.2E-16) and 120h (MWU, p=1.588E-09) but not for the head samples (MWU, p=0.4517), suggesting that DC is apparent in head sample. The differences in the M:F distribution plots were also reflected in the Log2 (M:F) median Z:A ratios (Table 7), this ratio at 78h (1.1) represents a slight male biased expression of Z-linked genes, and a profoundly male biased expression at 96h (1.5); female biased expression at 120h (0.89) and an almost unbiased expression in head (1.02) samples (Figure 25 and Table 7). Additionally a quartile-based analysis for the Z-linked genes was used for M:F distribution analysis (Figure 26) and showed a consistency in the observed biased expressions of Z-linked genes at various magnitudes of gene expression (Figure 26).

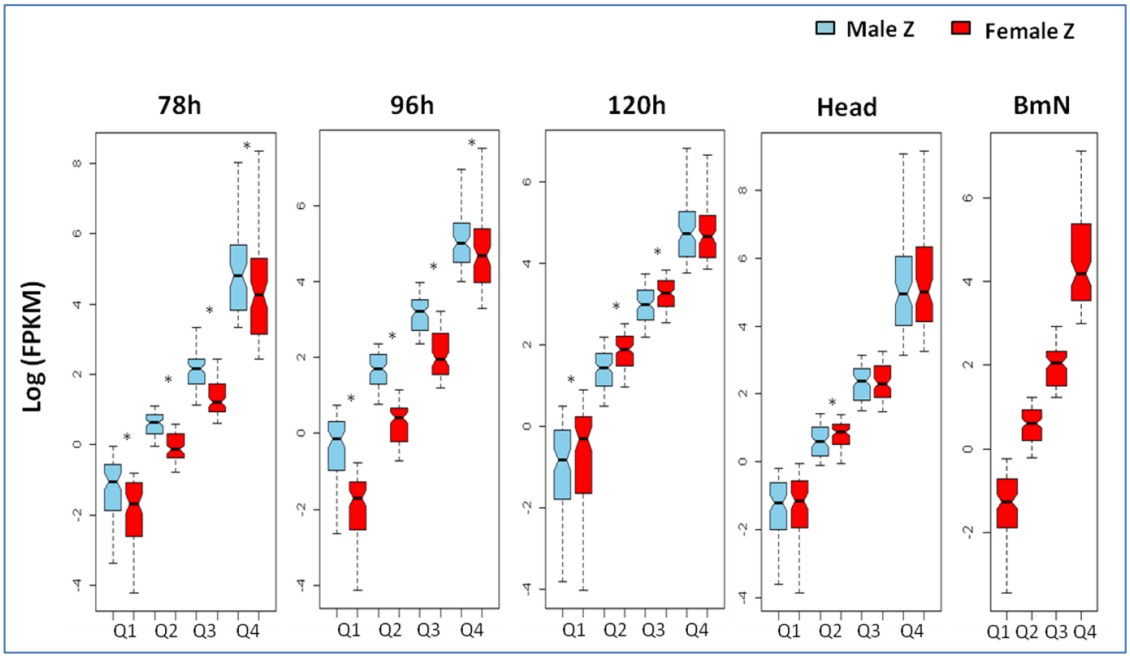


**Figure 26.** Comparison of quartile expressions for Z-linked genes dataset used for M:F analysis. The $\log_2$ FPKM expression data for the Z-linked genes were segregated as quartiles based on independent binning. The boxes represent the inter quartile range (IQR), the notch of box plots represents median (horizontal line) expression with a 95% confidence interval and the outliers were not plotted. An asterisk (*) indicates a significant difference between quartile expressions (MWU, p <0.05).

The expression of Z-linked genes (log2 FPKM) was compared between sexes by using a heat map (Figure 27). The genes with a relatively higher expression level showed a male biased expression at early developmental stages, 78h and 96h (cluster-9, 212 genes). The lowly expressing genes showed a female biased expression (cluster-2, 222 genes),

exclusively at 78h and 96h stages, and also showed an overall male biased to unbiased expression at 120h stage. The cluster of genes having a very high level of expression (cluster-4, 61 genes) showed an almost unbiased expression in all the samples. Almost all the clusters appeared to be compensated by having similar level of expression between sexes at 120h stage, except for the mosaic cluster (cluster-7, 125 genes). In the head samples, almost all of the Z-linked genes were found to be dosage compensated except for a very few local effects.

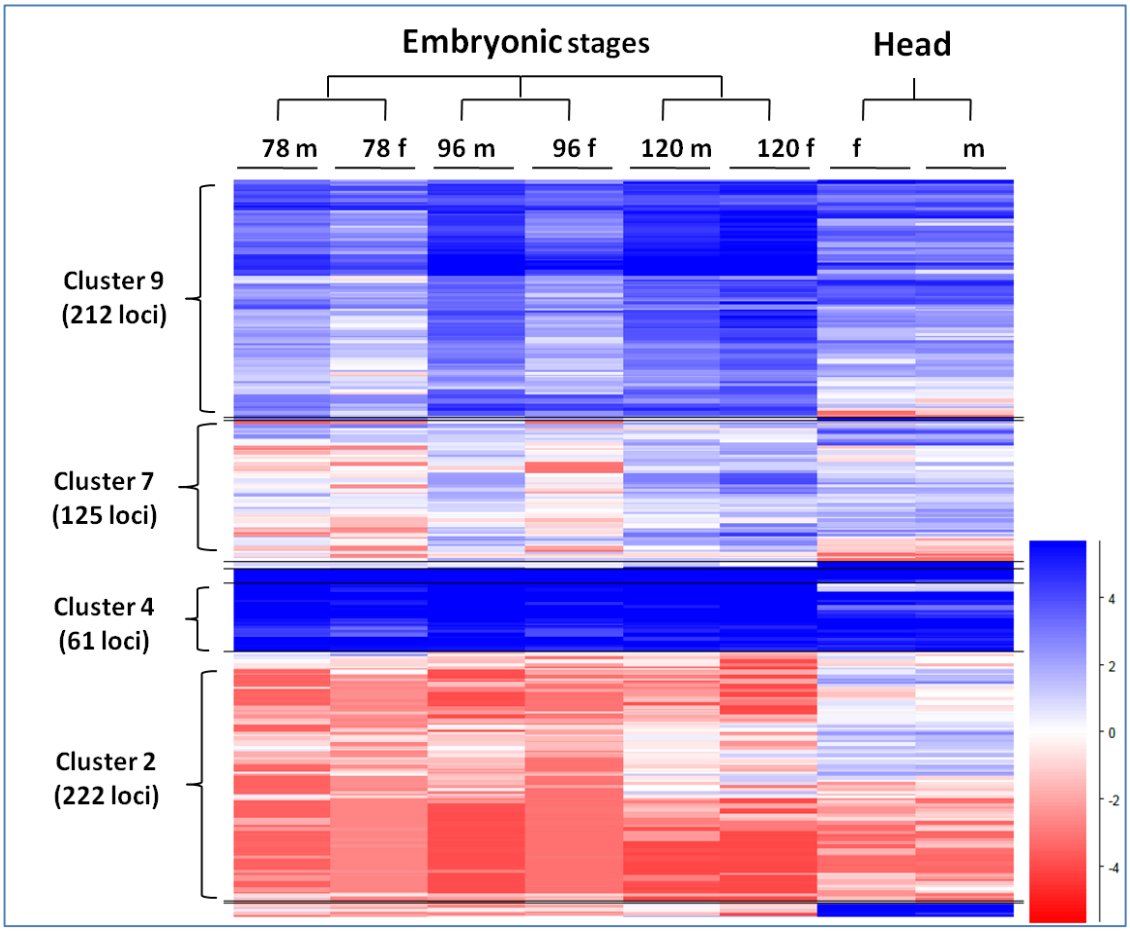


**Figure 27.** Schematic representation of Z-linked gene expression. The genes mapped to Z-linked scaffolds were segregated and their $\log_2$ FPKM values were plotted as a heat map. The heat map depicts a male biased cluster of 212 genes with high expression, a female biased cluster of 222 genes with low expression and an unbiased cluster of 61 genes.

Finally, to evaluate the profound male biased expression at 96h stage from the embryonic RNA-seq data, five selected Z-linked and autosomal genes expressions from different chromosomal locations were analyzed through quantitative PCR (qRT-PCR)

using *rp49* as an endogenous control as this is the best endogenous reference gene, with a least stability index of 0.083 in *B. mori* when compared to other endogenous reference genes like *E2F, actinA1, actinA3, G3PDH* and *GAPDH* (TENG *et al.* 2011). Autosomal genes of embryonic stages showed an overall equal expression in all three stages (Figure 28A). All the five Z-linked genes showed a relatively higher fold expression at least in 96h stage (Figure 28B). For the head samples, the selected 4 autosomal (Figure 28C) and 4 Z-linked genes (Figure 28D) (based on FPKM values) were shown to be almost equally expressing in both sexes in qRT-PCR. The qRT-PCR data of selected genes supports the FPKM based relative expression from RNA-seq data analysis.

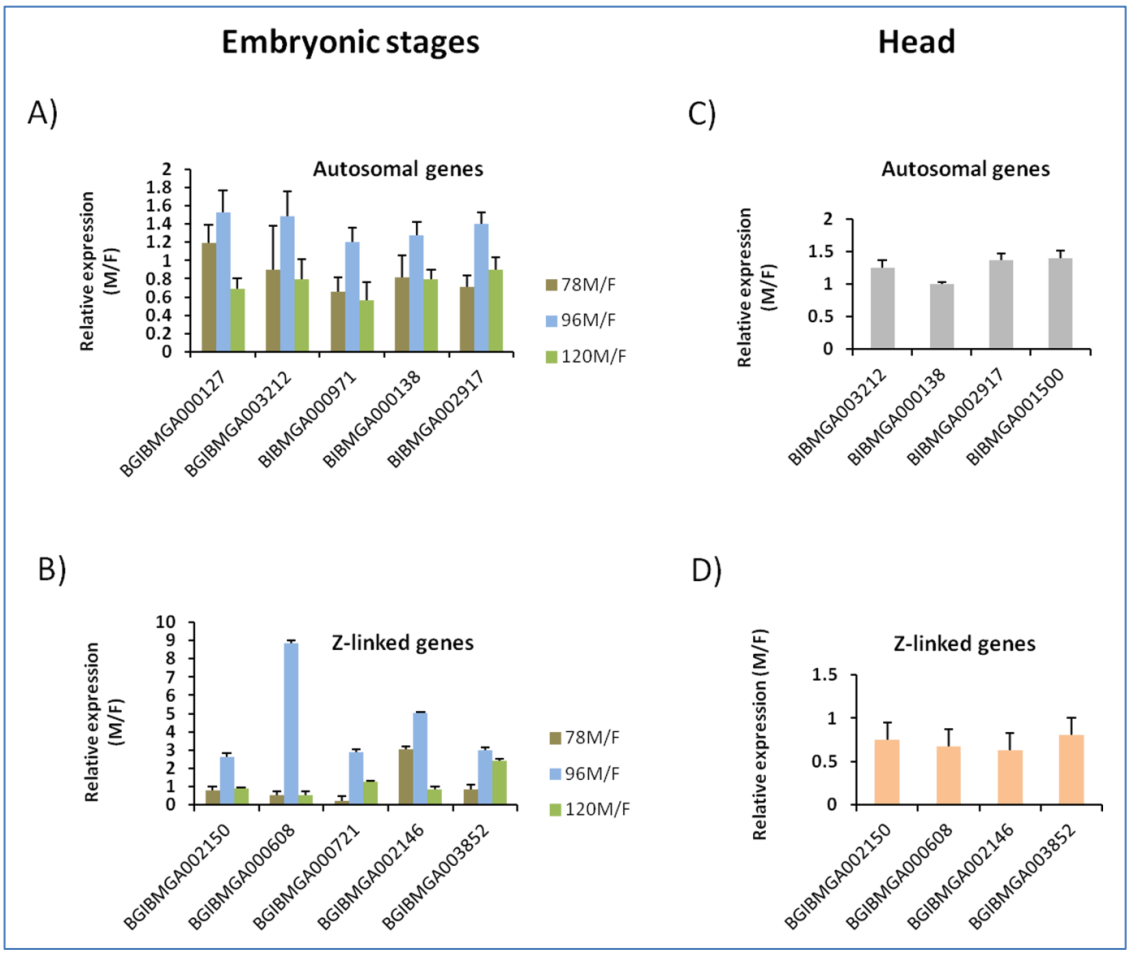


**Figure 28.** Validation of RNA-seq results of embryonic stages (78h, 96h and 120h) and head samples by qRT-PCR. *rp49* was used as an endogenous reference expression. Female samples were used as calibrators; hence the relative expression (M/F) of females can be taken as 1 for all the genes. The relative expression of selected five autosomal (a) and five Z-linked genes (b) for the three embryonic stages and four equally expressing (M/F=~1) autosomal (c) and four Z-linked genes (d) were shown in the head samples.

## 4.4 Discussion

Our study showed that in *B. mori*, Z:A ratio is lesser than 1 in both the sexes, representing reduced expression of Z-linked genes to that of autosomes. A plausible explanation for this unusually reduced Z:A ratio is the low expression of Z-linked genes, compared to a major proportion of autosomes (12, which is 48.06% of the genome) (Figure 29 and Table 8). The case of *B. mori* is similar to *Heliconius* species, where the expression of Z-linked genes is substantially reduced to that of autosomes (WALTERS *et al.* 2015). But from an evolutionary perspective, there is no currently available hypothesis to explain this reduced Z:A ratio in these species. The mechanisms of DC evolution can provide insights to answer this. It has been proposed that DC has evolved in a two step process (Ohno's hypothesis) in *Caenorhabditis* and Mammals (CHARLESWORTH 1996; MANK 2013; MANK *et al.* 2011; OHNO 1967; VICOSO and BACHTROG 2009), whereas in *Drosophila*, it's a direct upregulation of X-linked genes by dosage compensation complex (DCC) in males (XY) (LUCCHESI 1973; STRAUB *et al.* 2005) (see introduction). I propose that the basic element of driving force for the evolution of such diverse mechanisms with a single objective of achieving DC should be the "ancestral expression of proto-sex chromosomes". Here, the ancestral expression refers to the fixed point of average gene expression for proto-sex chromosomes in the course of sex chromosome evolution. For instance, in *Drosophila*, *Caenorhabditis* and Mammals the value of ancestral expression of proto-X might be fixed as 1 (CHARLESWORTH 1996; MANK 2013; MANK *et al.* 2011; OHNO 1967; VICOSO and BACHTROG 2009). So in males (XY) the expression level of X is 1 and in females (XX) it is 2. Hence in these species, the DC force would have evolved by choosing one of the possible/suitable paths like "DCC mediated X upregulation" or "two step process" (Ohno's hypothesis) to attain the destined expression level of 2 for X chromosome(s) in males XY=2 and females XX=2 which matches to the value of autosomes (AA=2). If the ancestral expression of proto-X/Z gets fixed as ~0.6 or so in a species, then the DC evolutionary pressures may choose a different path like, "repression of ZZ expression in homogametic sex" for achieving DC. For example, in case of *B. mori* the Z:A ratios were ~0.6 in both sexes (all samples) (Table 6) but the M:F ratios for both A and Z-linked genes are ~1 (in head), representing the equally expressing A and Z-linked genes in both sexes (Table 9). Based on these observations, we speculate that the DC evolutionary pressures in *B. mori* would have selected for the repression of Z in homogametic sex, in order to achieve the DC faster. Thus in short, the "ancestral expression of proto-sex chromosomes" to be one of the crucial determinants of the X:A or Z:A ratio, which in turn drives the DC

evolutionary pressures to choose the path of DC mechanism (ancestral expression of proto-sex chromosomes > X:A or Z:A ratio > DC path).

**Table 8.** Comparison of Z expression level to that of autosomes, from Figure 29

| | **Chromosomes with higher average expression than Z (12)** | **Chromosomes with lesser average expression than Z (7)** | **Chromosomes with equal average expression as Z (8)** |
|---|---|---|---|
| **Chromosomes** | 3,4,5,8,9,10,11,13,15,17,18 & 22 | 2,7,14,20,26,27 & 28 | 6,12,16,19,21,23, 24 & 25 |
| **Size in Mb** | **201.83** | 75.43 | 122.37 |
| **% of genome size (419.98 Mb)** | 48.06 | 17.96 | 29.14 |

The autosomes were grouped in to three, based on their relative Z expression level (Figure 29) and their percentages of genomic sizes were calculated to show that a major proportion of genome show on an average higher expression than that of Z-chromosome.

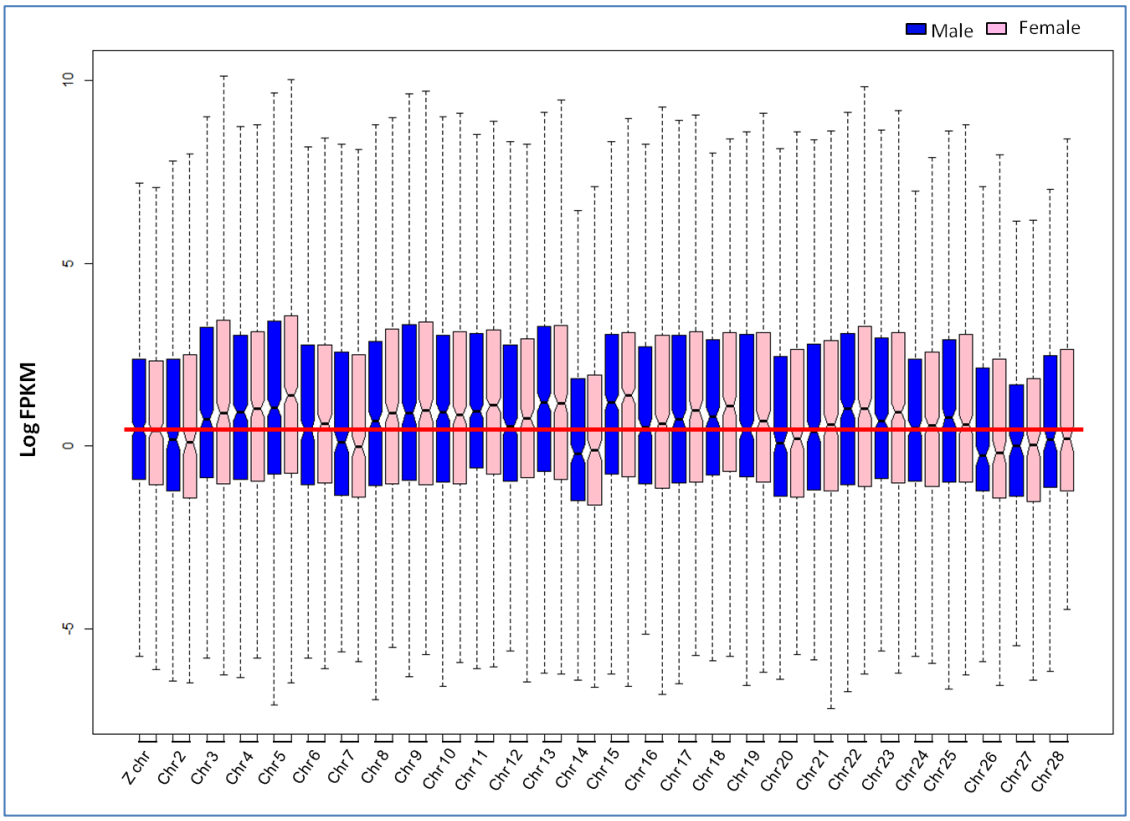


**Figure 29.** The Z-chromosome expression in dosage compensated head samples compared to autosomes. The solid blue line represents the Z-chromosome expression level in comparison with autosomes. Z-chromosome expression is roughly lesser than 12, higher than 7 and almost equal to 9 autosomes. The outliers were not plotted and the notch of the box plots represents the median expression with a 95% confidence interval.

**Table 9.** The male to female Autosomal and Z expression ratios (M:F) for the head samples.

| | **Autosomal (7210 genes)** | **Z-linked (339 genes)** |
|---|---|---|
| **Head (median, M:F ratio)** | 0.904* | 1.019* |
| **Median FPKM (Male/Female)** | 4.4969/4.973 | 2.8172/2.763 |

*These ratios of ~1 represent an equal expression of the autosomal and the Z-linked genes between sexes.

The M:F ratio distributions imply the sex biased expression of genes for the samples. In our study, these ratios suggest a clear male biased expression of Z-linked genes at 96h and a female biased expression at 120h (Table 7). A stark transition of Z from a male biased expression at 96h to a female biased expression at 120h probably indicates the onset of dosage compensation effect probably initiated at 96h and executed at the later stage, 120h of development. A significant male biased expression of A and Z-linked genes being observed at the early stages of 78h and 96h gets normalized at a later stage of 120h, suggesting the advent of DC mechanism. This scenario clearly indicates the primary objective of DC mechanism to equalize the expression differences of A and Z-linked genes between sexes (NGUYEN and DISTECHE 2006). In head samples, Z-linked genes showed an almost unbiased expression suggesting its compensation. In contrast, a slightly male biased expression (statistically significant, Table 7) of Z at 78h stage may be initial and I presume this stage might not be representing a relative full-fledged Z-linked gene expression, based on comparatively lesser Z:A ratios at 78h stage (Figure 22 and Table 6 ). However, from the heat map it is evident that a large number of genes with higher expressions showed a male biased expression at 78h and 96h stages. These genes shift from a male biased to an unbiased expression (based on similar color schema observed between sexes) at the later stage of 120h. This indicates the dosage compensatory effect over these genes at this stage and also suggests that the onset of DC can be considered to initiate post 96h. In the head samples, except for a very few sub-set of genes, almost all the genes showed a comparable level of expression, indicating the established DC.

From this study, an embryonic stage (120h) is identified, at which the effect of DC comes in to action (Figure 22). Gene-wise comparisons at 96h showed a profound male biased Z-linked gene expression, which gets counter attacked by DC at 120h, exhibiting a

slight female biased Z-linked gene expression at this stage. The male biased expression of Z-linked genes in the early stages of development (78h and 96h) could be due to the functionally inactive, putative DCC; whose initiation (post 96h) and advent from the later stage (120h) would have established the DC in *B. mori* tissues. The results suggest that 96h as a crucial developmental stage both for DC, based on relatively higher expression of *masc,* a dosage regulator gene in males (Table 10) and also in sex determination, due to the initiation of *dsx* sex specific differential splicing; Ajimura et al., unpublished data (AJIMURA *et al.* 2006). From our results, the complete DC is apparent in *B. mori,* depicting its essentiality for sexual fitness in this species. The complete DC is represented by an overall Z-linked gene expression parity between the sexes with a relatively reduced expression compared to autosomes, a unique trend generally not seen in the dosage compensated taxa (VICOSO and BACHTROG 2009). This type of reduced expression of ZZ in *B. mori* males is analogous to the down regulation of both XX transcription in hermaphrodite *C. elegans,* probably by increased chromosome condensation (MELLER 2000). The speculated DC mechanism of *C. elegans* promoted by *XO lethal-1* (*xol-1*) is similar to that of the *masc* (KIUCHI *et al.* 2014)in the *B. mori*; both the genes regulate the sex determination and the DC. Both the dosage mechanisms result in the hypo-expression of the XX/ZZ chromosomes.

**Table 10.** Expression of *masc* gene in the embryonic and head samples.

| Sample | *Masc* FPKM in males | *Masc* FPKM in females | *Masc* gene M/F Fold Change |
|---|---|---|---|
| **78h** | 0.255 | 0 | -- |
| **96h** | 0.92 | 0.14 | 6.45* |
| **120h** | 2.01 | 2.11 | 0.95 |
| **Head** | 0.67 | 0 | -- |

**masc* gene showed male biased expression at 96h, speculated to be a crucial stage for sex determination and DC. M=male and F=female.

The emerging evidences suggest that the patterns of DC are highly variable across sex-determination system and species (MANK *et al.* 2011). The DC observed in *B. mori* stand as a unique mechanism, which is achieved mostly through the hypo expression of the ZZ chromosomes in males. In a few ZW species studied, male ZZ:AA expression ratios in

general were reported to be ~1, eg, in aves it is 1.01 (ELLEGREN *et al.* 2007; ITOH *et al.* 2007), in nematode *S. mansoni* it is 1.06 (VICOSO and BACHTROG 2011), and in lepidopteran *P. interpunctella* it is 0.95 (HARRISON *et al.* 2012). Whereas in females ZW/AA ratios have been reported to be ~0.5, suggesting IDC in these species. In *Heliconius* species, the dosage compensation is reported to be imperfect or incomplete based on a consistent male biased expression of Z-linked genes in various tissues tested for dosage (WALTERS *et al.* 2015). In contrast, these ratios were almost equal between the sexes and have a value of <1 in the lepidopterans *M. sexta* (ZZ:AA = 0.83 and ZW/AA = 0.81) (SMITH *et al.* 2014) and *B. mori* [ZZ:AA = ~0.6 (mean) and ZW/AA = ~0.6 (mean)] suggesting the CDC.

The median Z:A ratios (Figure 22) suggest that in male embryos, the Z-linked gene expression reaches the relative expression point of ~0.5 first compared to females (observed first at 96h stage). But at a later stage of development (120h), the Z-linked gene expression was found to be slightly higher in females compared to males (Figure 22), contrasting the double dose (copy number) of Z-chromosomes in males. This can be explained by these possibilities 1) the hyper expression of single Z-chromosome in females or 2) by the hypo expression of ZZ in males or 3) by a combinatorial effect of both hyper and hypo expressions of Z-linked genes.

A recent report suggests the involvement of a Z-linked CCCH zinc finger gene, *masc* in the DC, as the RNAi of *masc* resulted in up regulation of Z-linked genes (KIUCHI *et al.* 2014). In this study, a *masc* gene expression level-driven suppression of Z-linked genes is observed in males (ZZ). The overall Z-linked gene expression difference between sexes at 96h was found to be significantly different (Figure 22). The *masc* gene at 96h stage was found to express 6.45 fold higher in males (Table 10). Being hypothesized to be involved in the mechanism of DC, the increase in the fold change of *masc* at 96h suggests its crucial involvement in the hypo-expression of ZZ in males to that of the Z in females (KIUCHI *et al.* 2014). The male biased expression of *masc* especially at the dosage uncompensated stage of 96h denotes its probable involvement in the mechanism of dosage compensation and this male biased *masc* expression would have led to the hypo expression of ZZ at the immediate next embryonic stage of 120h where the Z dosage was compensated (Figure 27). As the *masc* gene was also shown to be involved in the mechanism of DC by suppression of ZZ in males, its male biased expression at 96h stage of development (Table 10) probably suggests its crucial involvement in DC (KIUCHI *et al.* 2014).

The overall Z-linked gene expression being slightly female biased at 120h stage may be due to "dosage over compensation effect", an exact opposing phenomenon to "inverse dosage effect", observed in *D. melanogaster* (PHILIP and STENBERG 2013; SUN *et al.* 2012). I presume this effect as temporary, being mediated by the gain of ZZ repression in males due to a tight transcriptional downregulation of the freely expressing ZZ chromosomes in males (KIUCHI *et al.* 2014). Besides, this effect may also be treated as over representation of single female Z-chromosome expression in the background of ZZ repression (males). Because of this effect, presumably the freely expressing female Z-chromosome at 120h stage seems to be apparently higher than its male counterpart. The M:F ratio distributions (Figure 25) of A and Z clearly indicate the initial absence (96h) and probable establishment of DC in the later stage (120h) of embryo development. Based on this finding, I speculate that the reason for lower Z:A ratio observed in 120h male could be due to the suppressed expression of Z-chromosome at this stage; high level of *masc* expression at 96h may be suppressing ZZ expression in males, thereby bringing down the Z expression at 120h stage. All these findings suggest and support the hypothesis of probable suppression of Z-linked genes in males (KIUCHI *et al.* 2014).

Altogether the growing data and analyses, especially using RNA-seq, in various species present a dynamic picture of patterns of DC suggesting the initial selection of a highly diverse mechanisms being adapted by the evolutionary forces with a focused objective of achieving the DC. A re-evaluation of DC in mammals and *C. elegans* using RNA-seq data contradicts the Ohno's hypothesis, questioning our current knowledge of the sex chromosome evolution and DC mechanisms (XIONG *et al.* 2010). Hence, by taking the advanced and accurate RNA-seq data into consideration, there is a necessity of critical revision of current evolutionary theories on DC. A new theory should emerge in order to explain the reasons for lower expression of X or Z to that of autosomes to gain a comprehensive understanding of the sex chromosome evolution and DC mechanisms.

## 4.5 Summary

In this study, I have compared dosage of Z-linked genes in different embryonic stages between male and female sexes and showed that in early embryonic stages, Z-linked gene dosage was not compensated. As the embryo ages and after the up-regulation of *masc* gene, Z-linked genes in males show a lower expression, compensating with the dosage of females. I speculate that DC emerges after 96h in male silkworm. In the embryo samples, 96h is considered as a crucial developmental stage, at which the sex determination and differentiation are most widely tuned in the embryos. To my knowledge, this is the first report showing the initiation of DC in embryonic stages using RNA-seq data. I also show complete DC in the larval stage of *B. mori* by comparing male and female transcriptomes of the head tissue. In *B. mori*, the type of complete DC observed is ZZ:AA=Z:AA<1, which is very distinct to that of XX/AA=X/AA=1. Further studies have to be conducted to confirm whether this compensation is through silencing of one of two Z-chromosomes in males or through the reduced expression of genes in both the copies of Z-chromosome.

# Chapter-V

## Summary and future prospects

## 5.1 Summary

The study was an extension of the ongoing quest in the laboratory to understand the mechanism of sex determination in the domestic silkworm, *Bombyx mori.* In *B. mori,* females are ZW and males are ZZ i.e. female heterogamy is observed. A previous study in our lab had identified a novel group of zinc finger genes (CCCH type and C2H2 type) on the W-chromosome. My study began with the objective of functional characterization of CCCH type zinc finger genes, *B. mori zinc finger-1* (*Bmznf-1*) present on W-chromosome and its homologous copy of *B. mori zinc finger-2* (*Bmznf-2*) on the 25$^{th}$ chromosome (SATISH *et al.* 2006). Initially these genes were thought to be involved in determining the female sex as the *Bmznf-1* gene is linked to the *feminizer* region (the region responsible for femaleness) on the W-chromosome (ABE *et al.* 2008). Experiments were carried out to identify the effect of this gene on the differential splicing of the *B. mori doublesex* (*Bmdsx*) gene. This gene undergoes sex specific differential splicing and produces male and female specific isoforms (*Bmdsxm* in males and *Bmdsxf* in females) which produces differential proteins (BmDSXM in males and BmDSXF in females) that have antagonistic roles in respective sexes (CHRISTIANSEN *et al.* 2002; XU *et al.* 2017b). Two genes, *B. mori P-element somatic inhibitor* (*Bmpsi*) and *B. mori homolog of IGF-II mRNA binding protein* (*Bmimp*) were shown to promote the differential splicing of *Bmdsx* gene. In males the proteins, BmPSI and BmIMP were shown to interact with each other and form a strong splicing inhibitory complex that bind the CE1 element of 4$^{th}$ exon and promotes the skipping of 3$^{rd}$ and 4$^{th}$ exons resulting in the male isoform, *Bmdsxm* (SUZUKI *et al.* 2010; SUZUKI *et al.* 2008). Whereas in females this type of splicing is absent and *Bmdsxf* isoforms were produced. The differential splicing of *Bmdsx* is the key regulatory process as it governs the sexual development of organism.

In my study, the overexpression of *Bmznf-1* or *Bmznf-2* in the female ovary derived BmN cells resulted in the abnormal splicing shift of female specific *Bmdsxf* isoform to *Bmdsxm* isoform, called masculinisation (GOPINATH *et al.* 2016). This affect of *Bmznf-1* or *Bmznf-2* on *Bmdsx* differential splicing is similar to that of the splicing inhibitory complex of BmPSI and BmIMP. This suggests that the genes *Bmznf-1* or *Bmznf-2* are promoting the male development or differentiation. Further it is also identified that the overexpression of *Bmznf-1* or *Bmznf-2* also resulted in the aberrant splicing of an interesting splicing enhancer gene called *Bombyx mori transformer-2* (*Bmtra-2*) (GOPINATH *et al.* 2016). The aberrant splicing of *Bmtra-2* includes the short isoforms that resulted in the skipping of a few exons that ultimately produces short BmTRA-2 putative proteins whose function is obscure. However previous studies have shown that the gene *Bmtra-2* is not involved in the somatic sex

determination pathway and RNAi studies have shown the role of *Bmtra-2* in the testis development (SUZUKI *et al.* 2012). However in the current situation the research conducted by the Japanese group had resulted in the identification of two more key players called *Masculinizer* (*Masc*), a CCCH type zinc finger protein and a W-chromosome encoded *feminizer* (*fem*) piRNA in the pathway (KIUCHI *et al.* 2014). The gene *Masc* is shown to induce the male specific *Bmdsxm* isoforms and is silent in the females as it is negatively regulated by the fem piRNA. Interestingly the CCCH type zinc finger motifs of *Masc* are not involved in the masculinising activity instead Cys-301 and Cys-304 of the C-terminal 288 residues were important (KATSUMA *et al.* 2015). The LNA-DNA gapmers mediated silencing of *fem* piRNA had resulted in producing *Bmdsxm* forms i.e. masculinization and the silencing of *Masc* had resulted in *Bmdsxf* forms and a dramatic reduction of a male specific *Bmimp*(M) forms in males (SAKAI *et al.* 2015b). The *Masc* gene orthologue in a lepidopteran *Trilocha varians* is also shown to be having similar function (LEE *et al.* 2015). Further the transgenic analysis of *fem* piRNA resistant *Masc* gene was performed to reveal the sex reversal potential of this gene in the female genetic background (SAKAI *et al.* 2016). Results showed that the female lethality and only partial female to male sex reversal. Assumption is that as the gene *Masc* is involved in dosage compensation along with sex determination, this gene may not be suitable for the sex reversal (SAKAI *et al.* 2016). However a very recent study conducted by Chinese group using CRISPR-Cas9 mediated gene disruption had resulted in a very interesting finding that in the pathway, it is not the *Masc* on the top of the pathway but it is the *Bmpsi* the key auxiliary factor in male sex determination (XU *et al.* 2017a). In addition, my research on *Bmznf-1* and *Bmznf-2* had also revealed the induction of masculinization whose molecular mechanism is unknown and hence in the current scenario the exact pathway of *B. mori* sex determination is unclear and many new players are still emerging (GOPINATH *et al.* 2016).

As the second objective of my work, I wanted to explore the W-derived genes or regulatory elements like piRNA precursors by analyzing the early sex biased transcriptome in embryos and the somatic sexual differences in heads of larvae and also the patterns of dosage compensation in all these samples. For this, the deep sequencing was performed for three sexed early embryonic stages of 78h, 96h, 120h, larval heads and ovary derived BmN cells. The embryonic stages were selected based on the observation that *dsx* gene exhibits sex specific differential splicing at 96h (based on a previous observation in our lab). Hence, a stage before (78h) and a stage after (120h) 96h were selected for analysis. At 78h (in the early stage of development) 520 genes were differentially expressed between sexes. This number

had increased to 4,068 at 96h and decreases to 2,596 at 120h. The differential gene expression (DGE) analysis suggested a very high male biased expression of many important genes of silk composition, developmental, transcription factors and many zinc finger genes which must be having crucial roles in the process of development and sexual differentiation. Further the BLAST analysis of the unmapped transcriptome to the genome yielded thousands of precursors for the *B. mori* small RNAs and many long non-coding RNAs which presumably include W-derived transcripts. However no W-chromosome derived protein coding genes were identified in the study. Further analysis of the unmapped transcripts may help in uncovering the W-transcriptome and thus aid in a comprehensive understanding of the role of W-chromosome in *B. mori* sex determination.

Further in the study, I have subjected the RNA-seq data for dosage compensation analysis. Dosage compensation (DC) refers to the average equal expression of sex linked genes (**X** or **Z** chromosomal expression) to that of average expression of autosomal genes (autosomal expression) i.e. **XX**=**AA**=**X**Y (or) **ZZ**=**AA**=**Z**W devoid of the unequal distribution of the sex chromosomes in sexes (**XX** in females and **X**Y in males or **ZZ** in males and **Z**W in females). There are various ways in which DC can be achieved in XX/XY species (please refer to section 4.1 in this thesis for complete details) and is generally observed (**XX**=**AA** in females and **X**Y**=AA** in males) in almost all the XX/XY species studied for DC. However it is found to be incomplete or inefficient (**Z**W<**AA** in females and **ZZ**=**AA** in males) in the studied ZZ/ZW species for the existence of DC (MANK 2013). It also gave an impression that the phenomenon of Dosage compensation is more common in XX/XY species than in ZZ/ZW species. In contrast to this assumption, the DC was observed to be complete in *Manduca sexta,* a lepidopteran, in which the average expression of Z-linked genes are lesser than the autosomal in both sexes, without a significant difference between the average Z-linked gene expression between sexes (SMITH *et al.* 2014). In addition to this, the reanalysis of previously analysed microarray data showed the probable existence of DC in *B. mori* (WALTERS and HARDCASTLE 2011). In my current study, the dosage analyses of the RNA-seq data have shown the existence of DC in *B. mori.* This study provides a valuable insight that a myriad ways of achieving the complete dosage compensation in ZZ/ZW species. This study clear outs the previously raised confusions that the dosage in *B. mori* could incomplete. It suggests that despite the dose of Z-chromosomes i.e., their number (ZZ in males and Z in females) the average expression of Z-linked genes is half to that of the average expression of autosomal genes. This pattern of complete dosage compensation observed in *B. mori* represents an unconventional and a unique type of DC mechanism.

## 5.2 Future prospectus

In the multimillion dollar silk industry, sex is money. It is know that the male silkworms exhibit higher resistance to disease, lower food consumption and weave cocoons with a relatively higher amount and superior quality of silk compared to that of females (TRAUT *et al.* 2007b). The molecular reason for this factor is that the protein resource in females is directed towards the yolk formation in the eggs thus the protein share towards the silk production gets obviously affected. Hence it is long been sought to breed only males. In *B. mori*, using classical genetics, several strains were developed for facilitating the selective breeding of only males but these approaches are laborious and time consuming (STRUNNIKOV 1975). Using modern female specific conditional lethal transgenic approach, strains are developed by employing the tetracycline-repressible transactivator (tTAV) system to achieve it (TAN *et al.* 2013). However such approach is limiting as it is expensive and lack practicality due to the usage of tetracycline. My proposal is aimed at 1) functional characterization of *Bmznf-2 in vivo*, which is shown to be capable of inducing male type of splicing of the key gene, *Bmdsx* in sexual development and 2) producing all male progeny by taking the advantage of miss expression of the potential gene, *Bmznf-1* or *Bmznf-2* that govern the differential splicing of *Bmdsx* pre-mRNA for the male only production. Using this approach, all male progeny may be achieved in a simple manner.

In the study deep sequencing of embryonic samples resulted in the identification of thousands of long non-coding RNAs (lnRNAs) in which 862 lnRNAs are novel. It is evident that the fem piRNA that blocks the expression of *Masc* gene is derived from a precursor Fem lnRNA (KIUCHI *et al.* 2014). Thus in *B. mori* it is possible that many vital genes might be regulated by these lnRNAs. Nearly 50% of the identified lnRNAs were also shown to be potential precursors for many reported ovarian small RNAs based on 100% sequence alignment in BLAST search. However in my analysis only few transcripts 62 were identified as W-chromosome derived and most of them are not encoding any reported small RNA or protein coding genes. Screening of these transcripts by taking the support of in silico predictions may provide further clues in the mechanism of sex determination. A significant research has been carried out and is advancing on Bombyx lnRNAs. A comparative analysis to identify in the silk glands of domestic and wild silkworm and have resulted in the identification of 3 lnRNAs to be differentially expressing and 2 of them are linked to post translational regulation of silk protein (ZHOU *et al.* 2017). Their study revealed the role of lnRNAs in silk production and also signifies commercial applications. Systemic identification and characterization of Bombyx lnRNAs presents the first comprehensive genome wide

analysis of silkworm lnRNAs and provides an invaluable resource for genetic, evolutionary, and genomic studies of *B. mori* (WU *et al.* 2016). A comprehensive study on these lnRNAs using bioinformatics analysis may also make a very interesting piece of work in exploring the functional aspects of these lnRNAs in *B. mori*. A comprehensive database of Bombyx lnRNAs and miRNAs was created by collecting the publicly available Bombyx RNA-seq data and public unigenes (ZHOU *et al.* 2016). All this information provides an excellent opportunity to pursue even more exciting research in the area of lnRNAs and their regulation in Bombyx.

From the dosage analysis in my thesis, it is clear that complete dosage compensation occurs in Bombyx. This study also supports the already proposed theory of reduced expression of ZZ in males induced by the *Masc* gene (KIUCHI *et al.* 2014). However the molecular mechanism of how this type of dosage compensation is achieved is not studied. Hence future studies conducted in this direction may yield very interesting results in the subject of dosage compensation. For example isolating the interacting partners of MASC protein by pull down assays may unravel even exciting players in both the sex determination and dosage compensation of *B. mori* and can extend our current knowledge in these pathways.

Overall, the study provides several new leads which can be pursued to further understand the exact in-depth mechanism of Bombyx sex determination and dosage compensation pathways. Such studies would obviously lead to identification of various key players in the pathways that can be utilized to design further strategies in pest management by transgenic approaches.

# Annexures

**Annexure I.** The ClustalO alignment of the *Bmznf-1* and *Bmznf-2* transcripts along with the CCCH zinc finger clones (F2, F3, F4, F6, F7, F8, F9 and F10) obtained from female fat body samples.

```
Bmznf-2      ------------------------------------------------------ATGAAAA  7
F7           -----------GGGAAGGAATTTGTTCGAATTAAGCTTGGTACCGAGCTCGGGATGAAAA  49
F3           ----------TGAGAGCCATCTGTTTGAATTTAAGCTTGGTACCGAGCTCGGGATGAAAA  50
F6           ------------CGAAAGCATGGGCTTCGATTTAGCTTGGTACCGAGCTCGGGATGAAAA  48
F4           -----------AGGAATATCTTGTTCGAATTAAAGCTTGGTACCGAGCTCGGGATGAAAA  49
Bmznf-1      ------------------------------------------------------ATGAAAA  7
F2           GAAGTGGATTTTGGAGGGGGGGACTGTCGATTGAAGCTTGGTCCGAGCTCGGGATGAAAA  60
F8           ----------AGGGAAGAACATGGGTTCGATTTAGGCTTGGTCCGAGCTCGGGATGAAAA  50
F9           -----------CGCGGAGTCTCGGTTCGATTTAAGCTTGGTACCGAGCTCGGGATGAAAA  49
F10          -------------TAACAATCGGGTTGAATTTAAGCTTGGTACCGAGCTCGGGATGAAAA  47
                                                                 *******

Bmznf-2      AATACTTAAAATACAAGATCGGGGAGGAGAGTCACTCGCCCCAATATCATGGAGTAGCAG  67
F7           AATACTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAG  109
F3           AATACTTAAAATACAAGTTCGGGGAGGAG-AGGACACGCCCCAATATCATGGAGTAGCAG  109
F6           AATACTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAG  108
F4           AATACTTAAAATACAAGTTCGGGGAGGAG-AGCACTCGCCCCAATATCATGGAGTAGCAG  108
Bmznf-1      AATACTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAG  67
F2           AATACTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAG  120
F8           AGTACTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAG  110
F9           AATACTTAAAATACAAGTTCGGGGAGGAGA-GCACTCGCCCCAATATCATGGAGTAGCAG  108
F10          AATACTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAG  107
             * *************** ***********    ** ************************

Bmznf-2      AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  127
F7           AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  169
F3           AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  169
F6           AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  168
F4           AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  168
Bmznf-1      AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  127
F2           AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  180
F8           AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  170
F9           AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  168
F10          AAATCCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATC  167
             ***** ************************************************************

Bmznf-2      CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAAAA  187
F7           CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  229
F3           CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  229
F6           CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  228
F4           CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  228
Bmznf-1      CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  187
F2           CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  240
F8           CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  230
F9           CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  228
F10          CACCATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATA  227
             ************************************************************ *

Bmznf-2      AGGAAATCAACGACGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  247
F7           AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  289
F3           AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  289
F6           AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  288
F4           AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  288
Bmznf-1      AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  247
F2           AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  300
F8           AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  290
F9           AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  288
F10          AGGAAATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAA  287
             ********** ** ******************** ************************

Bmznf-2      CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  307
F7           CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  349
```

```
F3          CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  349
F6          CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  348
F4          CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  348
Bmznf-1     CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  307
F2          CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  360
F8          CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  350
F9          CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  348
F10         CTGAGAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTA  347
            ************************************************************

Bmznf-2     ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  367
F7          ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  409
F3          ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  409
F6          ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  408
F4          ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  408
Bmznf-1     ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  367
F2          ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  420
F8          ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  410
F9          ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  408
F10         ACACTTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTA  407
            ************************************************************

Bmznf-2     AAGGAGTTTACAAATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  427
F7          AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  469
F3          AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  469
F6          AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  468
F4          AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  468
Bmznf-1     AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  427
F2          AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  480
F8          AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  470
F9          AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  468
F10         AAGGAGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTT  467
            ************ ********************************************

Bmznf-2     CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAC  487
F7          CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  529
F3          CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  529
F6          CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  528
F4          CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  528
Bmznf-1     CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  487
F2          CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  540
F8          CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  530
F9          CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  528
F10         CATATGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCAT  527
            ************************************************************

Bmznf-2     CAAATACATTGAGCCATATTAAGAAAAAAACAGTCCT-ACCTCCAGCTAAAACCAAAGAA  546
F7          CACATACATTGAGCCATATTAAGAAAAAAACAGTCCA-ACCTCCAGCTAAAACCAAAGAA  588
F3          CACATACATTGAGCCATATTAAGAAAAAAACAGTCCA-ACCTCCAGCTAAAACCAAAGAA  588
F6          CACATACATTGAGCCATATTAAGAAAAAAACAGTCCAAACCTCCAGCTAAAACCAAAGAA  588
F4          CACATACATTGAGCCATATTAAGAAAAAAACAGTCCA-ACCTCCAGCTAAAACCAAAGAA  587
Bmznf-1     CACATACATTGAGCCATATTAAGAAAAAAACAGTCCA-ACCTCCAGCTAAAACCAAAGAA  546
F2          CACATACATTGAGCCATATTAAGAAAAAAACAGTCCA-ACCTCCAGCTAAAACCAAAGAA  599
F8          CACATACATTGAGCCATATTAAGAAAAAAACAGTCCA-ACCTCCAGCTAAAACCAAAGAA  589
F9          CACATACATTGAGCCATATTAAGAAAAAAACAGTCCA-ACCTCCAGCTAAAACCAAAGAA  587
F10         CACATACATTGAGCCATATTAAGAAAAAAACAGTCCA-ACCTCCAGCTAAAACCAAAGAA  586
            ** ********************************  **********************

Bmznf-2     ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  606
F7          ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  648
F3          ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  648
F6          ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  648
F4          ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  647
Bmznf-1     ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  606
F2          ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  659
F8          ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  649
F9          ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  647
F10         ACATCATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCT  646
```

```
 ************************************************************

Bmznf-2      GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCTTACGAGAAGCAA  666
F7           GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAA  708
F3           GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAA  708
F6           GTTCCGATGTACACAAATTCTTCGAATATAAGCATGATGCAAAATCCGTACTAGAAGCAA  708
F4           GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAA  707
Bmznf-1      GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAA  666
F2           GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAA  719
F8           GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAA  709
F9           GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAA  707
F10          GTTCCGATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAA  706
             ****************** **************************** *** ********

Bmznf-2      TGGCCGGAAATGGAAGACAAAACGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  726
F7           TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  768
F3           TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  768
F6           TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  768
F4           TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  767
Bmznf-1      TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  726
F2           TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  779
F8           TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  769
F9           TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  767
F10          TGGCCGGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCT  766
             ********************* ************************************

Bmznf-2      TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  786
F7           TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  828
F3           TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  828
F6           TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  828
F4           TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  827
Bmznf-1      TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  786
F2           TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  839
F8           TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  829
F9           TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  827
F10          TATCCTCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAA  826
             ************************************************************

Bmznf-2      AATCCATACGACACTGTAATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCC  846
F7           AATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCC  888
F3           AATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCC  888
F6           AATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCC  888
F4           AATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCC  887
Bmznf-1      AATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCC  846
F2           AATCCATACGCCACTGTGATGTCTCCTACTAAGAGGCAGTGGACAGAAATGGAAGAACCC  899
F8           AATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAAGAAATGGGAGACCC  889
F9           AATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGACA-GAAATGGAAGAACCC  886
F10          AATCCATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCC  886
             ********** ****** *********** ***********      **  *    ****

Bmznf-2      GTGAGTCAACCCACGCACGAGTACGCAGAGTACGCTGAAACGGCTGGATCCAAGAAATGC  906
F7           GTGAGTCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGC  948
F3           GTGAGTCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGC  948
F6           GTGAGTCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGC  948
F4           GTGAGTCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGC  947
Bmznf-1      GTGAGTCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGC  906
F2           GTGAGTCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGC  959
F8           GGTGGAGTCACCCACGCCACGAGTACGGGATACGCTGAACGGCTGGATCCAAGAAATG-C  948
F9           GTGAGTCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCAAGAAATG-C  945
F10          GTGAGTCAACCCACGCACGAGTACGCGGA-TACGCTGAAACGGCTGGATCCAGAAATG-C  944
             *   *     **     *  *    * *  *********  *   *   * *  **   *

Bmznf-2      AGAAACTGTGATGTCAACGAGTTCAGATTCCAGCATAACAAAAACAAAATCATGAAAATG  966
F7           AGAAACTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATG 1008
F3           AGAAACTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATG 1008
F6           AGAAACTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATG 1008
F4           AGAAACTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATG 1007
Bmznf-1      AGAAACTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATG  966
```

```
F2          AGAAACTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATG  1019
F8          AGAACTG-TGATGTCACGGAGTTCAGATTCAGCCCAACAAAACAAATCTG----GAAATG  1003
F9          AGAACCT-GTGATGTCACGAGTTCAGATTCCAGCCAC--AAAACAAATCT----GAAATG  998
F10         AGAA-CT-GTGATGTCACGAGTTCAGATTCAGCCCCACAAAAACAAATCT----GAAATG  998
            ****             ************   *     ***   *         *****

Bmznf-2     ATAAAAGACACAGACGACTTGAATCATCGCGTG-GGGCAAATAACTAAAAAGAACACTAA  1025
F7          ATAAAAGACACAGACGACTTGAATTATGGGGAGGGGACAAATAACTAAAAAGAACACTAA  1068
F3          ATAAAAGACACAGACGACTTGAATTATCGCGTG-GGACAAATAACTAAAAAGAACACTAA  1067
F6          ATAAAAGACACAGACGACTTGAATTATCGCGTG-GGACAAATAACTAAAAAGAACACTAA  1067
F4          ATAAAAGACACAGACGACTTGAATTATCGCGTG-GGACAAATAACTAAAAAGAACACTAA  1066
Bmznf-1     ATAAAAGACACAGACGACTTGAATTATCGCGTG-GGACAAATAACTAAAAAGAACACTAA  1025
F2          ATAAAAGACACAGACGACTTGAATTATCGCGTG-GGACAAATAACTAAAAAGAACACTAA  1078
F8          ATAAAGA--CCGGACGACTGATTT---ATCGCGGTGAACAATACTAAA----AGAACCTA  1054
F9          ATAAAGA--CCGACGACTGATTAT---CCGCGGGACAAA--TACTTAA----AGACCTAC  1047
F10         ATAAAGA--ACCGACGACTGATAT---CCGCGGGACAAA--TAACTAA----AGAACCTT  1047
            *****     *            *        *        **   **    *   *

Bmznf-2     ATTGAATGAAGTACTTGTGCTGCTCGTCGAAGTTGTGATAAGACGCACTTTTCGCTAG--  1083
F7          ATTGAATGAAATACTTGTGTTGCTCGTCGAAGTTTTGATACTCTTTTTTTATATGGCTAG  1128
F3          ATTGAATGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACG--TATCGCTAG  1125
F6          ATTGAATGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACG--TATCGCTAG  1125
F4          ATTGAATGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACG--TATCGCTAG  1124
Bmznf-1     ATTGAATGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACG--TATCGCTAG  1083
F2          ATTGAATGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACG--TATCGCTAG  1136
F8          AGTGAATGAAATACT--TGGGTGGTGGCTCC----GTCCGAA------GT----------  1092
F9          TGATGAAATACTTGTGTTGCTCGTCGAGTTT----GTTTAGAC----------------  1086
F10         ACTGATTGAATACTGGGTGCTCGTCGAGTTT----GTATAAGACGCCCGTATCGCTTA--  1101
                     *       **      *

Bmznf-2     ------------------------------------------------------------  1083
F7          GCGGCCGCTCGAGTCTAGAGGGCCCGCTGCCCTTTAA-AACCTATCCCTAACCCTCTCCT  1187
F3          GCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGTTATAAGCCTATCCCTAACCCTCTCCT  1185
F6          GCGGCCGCTCGAGTCTAGAGGGCCCCCGGTTCGTTGTAACCTAT-CCCTAACCCTCTCCT  1184
F4          GCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGTTGTAAGCCTATCCCTAACCCTCTCCT  1184
Bmznf-1     ------------------------------------------------------------  1083
F2          GCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGTAGTAAGCCTATCCCTAACCCTCTCCT  1196
F8          ------------------------------------------------------------  1092
F9          ------------------------------------------------------------  1086
F10         ------------------------------------------------------------  1101

Bmznf-2     ------------------------------------------------------------  1083
F7          CGGGGGGGGATTTACGGTACCGGCCACAACACCTTCGCCATTAGTTGACCCCAGACCCC-  1246
F3          CGGTCTCGATTCTACGCGTACCGGTCATCATCACCTCACCAATAAAGGGA----------  1235
F6          CGGTCTCGATTCTACGCGACCGGTCATCTCACCTCACCATAGGGGCGTTCCCCTGAGCC-  1243
F4          CGTCTCGATTCTACGCGTACCGGTCATCATCCCTCCATAACGCCTAACTAAACTGCCACT  1244
Bmznf-1     ------------------------------------------------------------  1083
F2          CGGGCTCGATTCTACGCGTACCGTCATCATCACCTCACCATAAGAACCGTTAAACAAAAA  1256
F8          ------------------------------------------------------------  1092
F9          ------------------------------------------------------------  1086
F10         ------------------------------------------------------------  1101

Bmznf-2     --------------------------------     1083
F7          -------CTACCCA-----------------     1253
F3          --------------------------------     1235
F6          --------------------------------     1243
F4          AGATGGTATAGACTATTATGCGTCTGCGTGG     1275
Bmznf-1     --------------------------------     1083
F2          ACAACA--------------------------     1262
F8          --------------------------------     1092
F9          --------------------------------     1086
F10         --------------------------------     1101
```

**Annexure II.** The CLUSTAL alignment of the *Bmznf-1* and *Bmznf-2* transcripts along with the CCCH zinc finger clones (Clone A, B, C, D, E and F) obtained from male fat body samples. The nucleotides highlighted in red show variations and in green indicate similarities.

```
Bmznf-2      ----------------------------------------------ATGAAAAAATACTTAA  16
CloneA       ----------------------------------------------ATGAAAAAATACTTAA  16
CloneC       ----------------------------------------------ATGAAAAAATACTTAA  16
CloneB       ----------------------------------------------ATGAAAAAATACTTAA  16
CloneE       -AACCCATATTGGTTTGGAATTTAAGCTTGGTACCGAGCTCGGGATGAAAAAATACTTAA  59
CloneF       --CGCAATCTGGGTCCGGATTTAAAGCTTGGTACCGAGCTCGGGATGAAAAAATACTTAA  58
Bmznf-1      ----------------------------------------------ATGAAAAAATACTTAA  16
CloneD       AAGGACACCTGGTTTGGGATTTAAAGCTTGGTACCGAGCTCGGGATGAAAAAATACTTAA  60
                                                           ****************

Bmznf-2      AATACAAGATCGGGGAGGAGAGTCACTCGCCCCAATATCATGGAGTAGCAGAAATCTCAG  76
CloneA       AATACAAGATCGGGGAGGAGAGTCACTCGCCCCAATATCATGGAGTAGCAGAAATCTCAG  76
CloneC       AATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATAACATGGAGTAGCAGAAATCCCAG  76
CloneB       AATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAATCCCAG  76
CloneE       AATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAATCCCAG  119
CloneF       AATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAATCCCAG  118
Bmznf-1      AATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAATCCCAG  76
CloneD       AATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAATCCCAG  120
             ******** ************* ************** ****************** ***

Bmznf-2      AAACAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACCATTAC  136
CloneA       AAACAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACCATTAC  136
CloneC       AAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACCATTAC  136
CloneB       AAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACCATTAC  136
CloneE       AAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACCATTAC  179
CloneF       AAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACCATTAC  178
Bmznf-1      AAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACCATTAC  136
CloneD       AAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACCATTAC  180
             ** ************************************************************

Bmznf-2      CTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAAAAAGGAAATCA  196
CloneA       CTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAAAAAGGAAATCA  196
CloneC       CTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGAAATCA  196
CloneB       CTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGAAATCA  196
CloneE       CTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGAAATCA  239
CloneF       CTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGAAATCA  238
Bmznf-1      CTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGAAATCA  196
CloneD       CTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGAAATCA  240
             ************************************************** **********

Bmznf-2      ATGATGACGAGGAGACACCTTCAATAACCATAAATATCAAGAAAGATCCAACTGAGAATT  256
CloneA       ATGATGACGAGGAGACACCTTCAATAACCATAAATATCAAGAAAGATCCAACTGAGAATT  256
CloneC       ACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGAGAATT  256
CloneB       ACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGAGAATT  256
CloneE       ACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGAGAATT  299
CloneF       ACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGAGAATT  298
Bmznf-1      ACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGAGAATT  256
CloneD       ACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGAGAATT  300
             * ** ******************** ************************************

Bmznf-2      ATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACACTTGCA  316
CloneA       ATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACACTTGCA  316
CloneC       ATGAAAAGCCTAATGACGTCGAAGAGGCCATGTGTCGTAATTTTGTGCGTAACACTTGCA  316
CloneB       ATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACACTTGCA  316
CloneE       ATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACACTTGCA  359
CloneF       ATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACACTTGCA  358
Bmznf-1      ATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACACTTGCA  316
CloneD       ATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACACTTGCA  360
             ************************** ************************************

Bmznf-2      ATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGGAGTTT  376
```

```
CloneA      ATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGGAGTTT  376
CloneC      ATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGGAGTTT  376
CloneB      ATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGGAGTTT  376
CloneE      ATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGGAGTTT  419
CloneF      ATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGGAGTTT  418
Bmznf-1     ATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGGAGTTT  376
CloneD      ATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGGAGTTT  420
            ************************************************************

Bmznf-2     ACAAATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTC  436
CloneA      ACAAATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTC  436
CloneC      ACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTC  436
CloneB      ACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTC  436
CloneE      ACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTC  479
CloneF      ACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTC  478
Bmznf-1     ACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTC  436
CloneD      ACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTC  480
            *** ********************************************************

Bmznf-2     ACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCACCAAATACAT  496
CloneA      ACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCACCAAATACAT  496
CloneC      ACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACATACAT  496
CloneB      ACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACATACAT  496
CloneE      ACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAGCGGGCTATTTACCATCACATACAT  539
CloneF      ACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACATACAT  538
Bmznf-1     ACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACATACAT  496
CloneD      ACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACATACAT  540
            ********************************** ********** ** ******

Bmznf-2     TGAGCCATATTAAGAAAAAAACAGTCCTACCTCCAGCTAAAACCAAAGAAACATCATCCG  556
CloneA      TGAGCCATATTAAGAAAAAAACAGTCCTACCTCCAGCTAAAACCAAAGAAACATCATCCG  556
CloneC      TGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATCATCCG  556
CloneB      TGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATCATCCG  556
CloneE      TGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATCATCCG  599
CloneF      TGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATCATCCG  598
Bmznf-1     TGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATCATCCG  556
CloneD      TGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATCATCCG  600
            ************************** *********************************

Bmznf-2     AGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCCGATGT  616
CloneA      AGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCCGATGT  616
CloneC      AGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCCGATGT  616
CloneB      AGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCCGATGT  616
CloneE      AGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCCGATGT  659
CloneF      AGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCCGATGT  658
Bmznf-1     AGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCCGATGT  616
CloneD      AGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCCGATGT  660
            ************************************************************

Bmznf-2     ACACAAATGCTTCGAATATAAGCATGATGCAAAATCCATACGAGAAGCAATGGCCGGAAA  676
CloneA      ACACAAATGCTTCGAATATAAGCATGATGCAAAATCCTTACGAGAAGCAATGGCCGGAAA  676
CloneC      ACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCCGGAAA  676
CloneB      ACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCCGGAAA  676
CloneE      ACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCCGGAAA  719
CloneF      ACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCCGGAAA  718
Bmznf-1     ACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCCGGAAA  676
CloneD      ACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCCGGAAA  720
            ************************************ ***********************

Bmznf-2     TGGAAGACAAAAACGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCCTCACA  736
CloneA      TGGAAGACAAAAACGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCCTCACA  736
CloneC      TGGAAGACAAAACGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCCTCACA  736
CloneB      TGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCCTCACA  736
CloneE      TGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCCTCACA  779
CloneF      TGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCCTCACA  778
Bmznf-1     TGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCCTCACA  736
CloneD      TGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCCTCACA  780
            ********** ***************************************************
```

```
Bmznf-2      GTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCCATACG  796
CloneA       GTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCCATACG  796
CloneC       GTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCCATACG  796
CloneB       GTACGAGCACAGGTTAACAAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCCATACG  796
CloneE       GTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCCATACG  839
CloneF       GTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCCATACG  838
Bmznf-1      GTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCCATACG  796
CloneD       GTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCCATACG  840
             ********************* * ****************************************

Bmznf-2      TCACTGTAAATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAGTCAAC  856
CloneA       ACACTGTAAATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAGTCAAC  856
CloneC       CCACTGTAAATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAGTCAAC  856
CloneB       CCACTGTGATGTCACCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAGTCAAC  856
CloneE       CCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAGTCAAC  899
CloneF       CCACTGTGATGTCTCCTACGAAGAGGCAGTGGACGGAAATGGAAGAACCCGTGAGTCAAC  898
Bmznf-1      CCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAGTCAAC  856
CloneD       CCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAGTCAAC  900
              ****** ***** ******************** *************************

Bmznf-2      CCACGCACGAGTACGCAGAGTACGCTGAAACGGCTGGATCCAAGAAATGCAGAAACTGTG  916
CloneA       CCACGCACGAGTACGCTGAGTACGCTGAAACGGCTGGATCCAAGAAATGCAGAAACTGTG  916
CloneC       CCACGCACGAGTACGCGGAGTACGCTGAAACGGCTGGATCCAAGAAATGCAGAAACTGTG  916
CloneB       CCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAACTGTG  916
CloneE       CCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAACTGTG  959
CloneF       CCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAACTGTG  958
Bmznf-1      CCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAACTGTG  916
CloneD       CCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAACTGTG  960
             ***************** ** ****************************************

Bmznf-2      AAGTCAACGAGTTCAGATTCCAGCATAACAAAAACAAAATCATGAAAATGATAAAAGACA  976
CloneA       AAGTCAACGAGTTCAGATTCCAGCATAACAAAAACAAAATCATGAAAATGATAAAAGACA  976
CloneC       ACGTCAACGAGTTCAGATTCCAGCATAACAAAAACAAAATCATGAAAATGATAAAAGACA  976
CloneB       ATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAAAGACA  976
CloneE       ATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAAAGACA 1019
CloneF       ATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAAAGACA 1018
Bmznf-1      ATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAAAGACA  976
CloneD       ATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAAAGACA 1020
             * ************************ ************************************

Bmznf-2      CAGACGACTTGAATCATCGCGTGGGGCAAATAACTAAAAAGAACACTAAATTGAATGAAG 1036
CloneA       CAGACGACTTGAATCATCGCGTGGGGCAAATAACTAAAAAGAACACTAAATTGAATGAAG 1036
CloneC       CAGACGACTTGAATCATCGCGTGGGGCAAATAACTAAAAAGAACACTAAATTGAATGAAG 1036
CloneB       CAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAATGAAA 1036
CloneE       CAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAATGAAA 1079
CloneF       CAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAATGAAA 1078
Bmznf-1      CAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAATGAAA 1036
CloneD       CAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAATGAAA 1080
             ************** ********** ********************************

Bmznf-2      TACTTGTGCTGCTCGTCGAAGTTGTGATAAGACGCACTTTTCGCTAG------------- 1083
CloneA       TACTTGTGCTGCTCGTCGAAGTTGTGATAAGACGCCCGTATCGCGGCGGCCGCTCGAGTC 1096
CloneC       TACTTGTGCTGCTCGTCGAAGTTGTGATAAGACGCCCGTATCGCGGCGGCCGCTCGAGTC 1096
CloneB       TACTTGTGTTGCTCGTCGAAGTTATGATAAGACGCCCT-CTCGCGGCGGCCGCTCGAGTC 1095
CloneE       CACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCGGCGGCCGCTCGAGTC 1139
CloneF       TACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCGGCGGCCGCTCGAGTC 1138
Bmznf-1      TACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCTAG------------- 1083
CloneD       TACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCGGCGGCCGCTCGAGTC 1140
              ******* *************** *********** *   ****

Bmznf-2      ------------------------------------------------------------ 1083
CloneA       TAGAGGGCCCGCGGCC-----GTCTACCTATCCCTAACCCTCTCCTCGGTCTCGATTCTA 1151
CloneC       TAGAGGGCCCGCGGCC-----GTCTACCTATCCCTAACCCTCTCCTCGGTCTCGATTCTA 1151
CloneB       TAGAGGGCCCGCGGG----GGACATTCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTA 1151
CloneE       TAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTA 1199
CloneF       TAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTA 1198
Bmznf-1      ------------------------------------------------------------ 1083
```

```
CloneD      TAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTA  1200


Bmznf-2     ------------------------------------------------------------  1083
CloneA      CGCGACCGTCTCTCCCTCCTCTGTACCTAACCCTAAAAAAAAAAAAAAAATAAATAGGGTG  1211
CloneC      CGCGACCGTCTCTCCCTCCTCTGTACCTAACCCTAAAAAAAAAAAAAAAATAAATAGGGTG  1211
CloneB      CGCGTACCGGTCATCATCACCTCACCATGAGAAAAAACTAAAAAA-ACCGACCC------  1204
CloneE      CGCGACCGGTCTCTCCCTCCCCTGTCATCTGAAAAAAAAAAAAAA-AAAAAAATAAACAG  1258
CloneF      CGCGTACCGGTCATCATCACACAATCACCCCATTAGTTT----------------------  1237
Bmznf-1     ------------------------------------------------------------  1083
CloneD      CGCGTACCGGTCATCATCACCATCAACCCATTAAAAAGC----------------------  1239


Bmznf-2     --------------------------------     1083
CloneA      TAGTGTACGCGCGCGATATTATTAGTAAG--      1240
CloneC      TAGTGTACGCGCGCGATATTATTAGTAAG--      1240
CloneB      --------------------------------     1204
CloneE      TGGAGTACACACAGTGTCGTATAACCTAGAA      1289
CloneF      --------------------------------     1237
Bmznf-1     --------------------------------     1083
CloneD      --------------------------------     1239
```

**Annexure III.** The CLUSTAL alignment of the *Bmznf-1* and *Bmznf-2* transcripts with a new set of the CCCH zinc finger clones (M5, M6, M9, M12 and M14) obtained from male fat body samples. All the clones showed similarity to *Bmznf-1* transcripts at all the 25 positions (highlighted in red) that vary between *Bmznf-1* and *Bmznf-2* CDS regions. The start and stop codons are highlighted in green and yellow respectively.

```
Bmznf-2      ------------------------------------------------------------  0
M12          ----------------------------------------------------GCCTTTT  7
M6           ---------------------------------------------------GCGAAATGA  9
M19          ----------------------------------------------------GGAAGTC  7
M14          ---------------------------------------------------GGCGGGGA  8
M9           TGGATTGATAGAACACGTTTCTAGTATGTTTGAGTGGGAAAAAACCCTCTCATCCATCTT  60
Bmznf-1      ------------------------------------------------------------  0
M5           ----------------------------------------------GGGATCCAGCAA  12

Bmznf-2      ------------------------------------------------ATGAAAAAATA  11
M12          TTCGCAAAAAAAAGACGGGACTGGTCGATTTAGCTTGGTCCGAGCTCGGGAGAAAAAATA  67
M6           CTAAAATCCCGTCCGCAACTGTCGATTTAAGCTTGGTACCGAGCTCGGGATGAAAAAATA  69
M19          AGTGAAGAGGAGGTCAAGAACGTTGAATTAGCTTGGTACCGAGCTCGGGATGAAAAAATA  67
M14          TCGATAGTTTTATAACCGAACGTCGAATTAAGCTTGGTCCGAGCTCGGGATGAAAAAATA  68
M9           TTCTTTCGAGGAAAATAGAAGAAATGCTGATTAGCTGGTACGACTCGGGATGAAAAAATA  120
Bmznf-1      ------------------------------------------------ATGAAAAAATA  11
M5           -----CTAATCACCGCGGGCCGGCTGATTTAGCTTGGTCCGAGCTCGGGATGAAAAAATA  67
                                                              *********

Bmznf-2      CTTAAAATACAAGATCGGGGAGGAGAGTCACTCGCCCCAATATCATGGAGTAGCAGAAAT  71
M12          CTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAAT  127
M6           CTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAAT  129
M19          CTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAAT  127
M14          CTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAAT  128
M9           CTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAAT  180
Bmznf-1      CTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAAT  71
M5           CTTAAAATACAAGTTCGGGGAGGAGAGCCACTCGCCCCAATATCATGGAGTAGCAGAAAT  127
             ************* ************* ********************************

Bmznf-2      CTCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACC  131
M12          CCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACC  187
M6           CCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACC  189
M19          CCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACC  187
M14          CCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACC  188
M9           CCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACC  240
Bmznf-1      CCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACC  131
M5           CCCAGAAGCAGAAGCGATGAAAAGAAAGAAACATGAGCAGACAGAGAAGGTTCATCCACC  187
             * ************************************************************

Bmznf-2      ATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAAAAAGGA  191
M12          ATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGA  247
M6           ATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGA  249
M19          ATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGA  247
M14          ATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGA  248
M9           ATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGA  300
Bmznf-1      ATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGA  191
M5           ATTACCTAACCGCCCACCGCTGCCGCCGCCGGAACCACTCCTTGCAAATTTCAATAAGGA  247
             ************************************************** *****

Bmznf-2      AATCAATGATGACGAGGAGACACCTTCAATAACCATAAATATCAAGAAAGATCCAACTGA  251
M12          AATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGA  307
M6           AATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGA  309
M19          AATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGA  307
M14          AATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGA  308
M9           AATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGA  360
Bmznf-1      AATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGA  251
M5           AATCAACGAAGACGAGGAGACACCTTCAATCACCATAAATATCAAGAAAGATCCAACTGA  307
             ****** ** ******************** ****************************

Bmznf-2      GAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACAC  311
```

```
M12          GAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACAC  367
M6           GAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACAC  369
M19          GAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACAC  367
M14          GAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACAC  368
M9           GAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACAC  420
Bmznf-1      GAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACAC  311
M5           GAATTATGAAAAGCCTAATGACGTCGAAGAGACCATGTGTCGTAATTTTGTGCGTAACAC  367
             ************************************************************

Bmznf-2      TTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGG  371
M12          TTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGG  427
M6           TTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGG  429
M19          TTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGG  427
M14          TTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGG  428
M9           TTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGG  480
Bmznf-1      TTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGG  371
M5           TTGCAATCGTGGTGCGTCATGTAGATATCTCCACAAAATAATACATTCCCAACTTAAAGG  427
             ************************************************************

Bmznf-2      AGTTTACAAATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATA  431
M12          AGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATA  487
M6           AGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATA  489
M19          AGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATA  487
M14          AGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATA  488
M9           AGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATA  540
Bmznf-1      AGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATA  431
M5           AGTTTACAGATTCTGTATTGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATA  487
             ******** ***************************************************

Bmznf-2      TGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCACCAAA  491
M12          TGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACA  547
M6           TGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACA  549
M19          TGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACA  547
M14          TGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACA  548
M9           TGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACA  600
Bmznf-1      TGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACA  491
M5           TGCTCACGCCACAGTGCACGAGAAAGAACATTTCTTCAGAACGGGCTATTTACCATCACA  547
             ************************************************************ ** *

Bmznf-2      TACATTGAGCCATATTAAGAAAAAAACAGTCCTACCTCCAGCTAAAACCAAAGAAACATC  551
M12          TACATTGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATC  607
M6           TACATTGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATC  609
M19          TACATTGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATC  607
M14          TACATTGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATC  608
M9           TACATTGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATC  660
Bmznf-1      TACATTGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATC  551
M5           TACATTGAGCCATATTAAGAAAAAAACAGTCCAACCTCCAGCTAAAACCAAAGAAACATC  607
             ******************************** ***************************

Bmznf-2      ATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCC  611
M12          ATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCC  667
M6           ATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCC  669
M19          ATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCC  667
M14          ATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCC  668
M9           ATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCC  720
Bmznf-1      ATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCC  611
M5           ATCCGAGCTGTCGGCGAATTATACCGGCACTACATACCCTCATAGCACCACTTCTGTTCC  667
             ************************************************************

Bmznf-2      GATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCATACGAGAAGCAATGGCC  671
M12          GATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCC  727
M6           GATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCC  729
M19          GATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCC  727
M14          GATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCC  728
M9           GATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCC  780
Bmznf-1      GATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCC  671
M5           GATGTACACAAATGCTTCGAATATAAGCATGATGCAAAATCCGTACGAGAAGCAATGGCC  727
             ****************************************** ******************
```

```
Bmznf-2     GGAAATGGAAGACAAAACGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCC  731
M12         GGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCC  787
M6          GGGAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCC  789
M19         GGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCC  787
M14         GGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCC  788
M9          GGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCC  840
Bmznf-1     GGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCC  731
M5          GGAAATGGAAGACAAATCGTCGCCCGAGATGGCATATTATAGTGGGCATCCATCTTATCC  787
            ** ************** ****************************************

Bmznf-2     TCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCC  791
M12         TCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCC  847
M6          TCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCC  849
M19         TCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCC  847
M14         TCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCC  848
M9          TCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCC  900
Bmznf-1     TCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCC  791
M5          TCACAGTACGAGCACAGGTTACCCAAACGCTCCGCCCAATACGCCTATAGTGCAAAATCC  847
            ************************************************************

Bmznf-2     ATACGTCACTGTAATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAG  851
M12         ATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAG  907
M6          ATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAG  909
M19         ATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAG  907
M14         ATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAG  908
M9          ATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAG  960
Bmznf-1     ATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAG  851
M5          ATACGCCACTGTGATGTCTCCTACGAAGAGGCAGTGGACAGAAATGGAAGAACCCGTGAG  907
            ***** ****** ************************************************

Bmznf-2     TCAACCCACGCACGAGTACGCAGAGTACGCTGAAACGGCTGGATCCAAGAAATGCAGAAA  911
M12         TCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAA  967
M6          TCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAA  969
M19         TCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAA  967
M14         TCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAA  968
M9          TCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAA  1020
Bmznf-1     TCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAA  911
M5          TCAACCCACGCACGAGTACGCGGAATACGCTGAAACGGCTGGATCCAAGAAATGCAGAAA  967
            ********************** ** ************************************

Bmznf-2     CTGTGATGTCAACGAGTTCAGATTCCAGCATAACAAAAACAAAATCATGAAAATGATAAA  971
M12         CTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAA  1027
M6          CTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAA  1029
M19         CTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAA  1027
M14         CTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAA  1028
M9          CTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAA  1080
Bmznf-1     CTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAA  971
M5          CTGTGATGTCAACGAGTTCAGATTCCAGCACAACAAAAACAAAATCATGAAAATGATAAA  1027
            ****************************** ******************************

Bmznf-2     AGACACAGACGACTTGAATTATCGCGTGGGTCAAATAACTAAAAAGAACACTAAATTGAA  1031
M12         AGACACAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAA  1087
M6          AGACACATACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAA  1089
M19         AGACACAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAA  1087
M14         AGACACAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAA  1088
M9          AGACACAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAA  1140
Bmznf-1     AGACACAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAA  1031
M5          AGACACAGACGACTTGAATTATCGCGTGGGACAAATAACTAAAAAGAACACTAAATTGAA  1087
            ******* *********** ********** ******************************

Bmznf-2     TGAAGTACTTGTGGTGCTCGTCGAAGTTTGTGATAAGACGCACTTTTCGCTAG--------  1083
M12         TGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCTAGGCGGCCGC  1147
M6          TGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCTAGGCGGCCGC  1149
M19         TGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCTAGGCGGCCGC  1147
M14         TGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCTAGGCGGCCGC  1148
M9          TGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCTAGGCGGCCGC  1200
Bmznf-1     TGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCTAG--------  1083
```

```
M5          TGAAATACTTGTGTTGCTCGTCGAAGTTTTGATAAGACGCACGTATCGCTAGGCGGCCGC  1147
            **** ******** ************** ************* * *******

Bmznf-2     ------------------------------------------------------------  1083
M12         TCGAGTCTAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGTCTCG  1207
M6          TCGAGTCTAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTC  1209
M19         TCGAGTCTAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTC  1207
M14         TCGAGTCTAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTC  1208
M9          TCGAGTCTAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTC  1260
Bmznf-1     ------------------------------------------------------------  1083
M5          TCGAGTCTAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTC  1207

Bmznf-2     ------------------------------------------------------------  1083
M12         TTCACGCTCCGTCTCTCCCCCTCCTATTTCTGAAAAACTAAAAAAAATAAAACGAGATGA  1267
M6          GATTCTACGCGTACCGGTCTCTCACCTCACCATATTAGTCGCTTGAAATAAACCCCG---  1266
M19         GATTCTACGCGTACCGGTCTCTCACCTCACCATAGAGCACCTCAAAAAAAAAGCCCG---  1264
M14         GATTCTACGCGTACCGGTCCTCTCACCTCGCCATAACCCTATCAATAAAAACCGTA----  1264
M9          GATTCTACGCGACCGTCTCTCACCTCACCA----TATCATAATTCAATAGGCCC------  1310
Bmznf-1     ------------------------------------------------------------  1083
M5          GATTCTACGCGTACCGGTCATCATCACCTCACCATAACCCGCTAAAAAAACGGCCGCC--  1265

Bmznf-2     --------------------------- 1083
M12         CACACTCTCAACACCCACAAAACGCCA 1294
M6          --------------------------- 1266
M19         --------------------------- 1264
M14         --------------------------- 1264
M9          --------------------------- 1310
Bmznf-1     --------------------------- 1083
M5          --------------------------- 1265
```

**Annexure IV.** Primers used for cloning and qPCR in Chapter II

| Sc. no | Primer name | Sequence/Used for |
|---|---|---|
| | | Endogenous control, for *Bmgapdh* PCR. |
| 1 | G3PDHF | GCCTTCTAATGTTGTTGCTGTTCC |
| 2 | G3PDHR | ACCACCTTCGGCTATATCAAATGC |
| | | Over lapping primers used for generating 5' UTR with two scrambled ovarian small RNA target sites and cloning it in pmirGLO vector |
| 3 | 5'UTR scramF1 | TATAGAGCTCAATAATAGACAAGGTAGGACCGATCCTG |
| 4 | 5'UTR scramF2 | GGTAGGACCGATCCTGAATTTGACGTTTTGCGATCG |
| 5 | 5'UTR scramR1 | TATATCTAGAGCTGAGTCTACAGGTCTGTATAGGCTAC |
| 6 | 5'UTR scramR2 | GGTCTGTATAGGCTACGAAAATTGACAGAATCTTC |
| | | Cloning the entire 5' UTR sequence of *Bmznf* gene |
| 7 | 5'UTR_LUCF | TATAGAGCTCGAACAATCGCGGTACGTGATAC |
| 8 | 5'UTR_LUCR | TATATCTAGATTCATAAGTCTGAAGTTCGGCG |
| | | Specific amplification of *Bmdsxm* or *Bmdsxf* splicing isoforms. And Also used for real-time PCR for *Bmdsxm* |
| 9 | dsx F forward | ATCTACGAAGGGAAGATGATCG |
| 10 | dsx M forward | ATCTACGAAGGTTATTGGATG |
| 11 | dsx Reverse | CTGTATCGGCGCGCAGTGTCGTCGC |
| | | Amplifying the m-cherry sequence and cloning it in to pIZT-Bmznf, for expression of fusion protein in BmN cells. |
| 12 | pIZTz1mchF | AAAATGCGGCCGCATGGTGAGCAAGGGCGAG |
| 13 | pIZTz1mchR | TATAATCTAGATTACTTGTACAGCTCGTCCATG |

| | | |
|---|---|---|
| | | Cloning and amplification of *Bmtra-2* splicing isoforms. |
| 14 | tra2_cloF _EcorI | TATAGAATTCATGTCTGATCGAGAGAGAAGTC |
| 15 | Tra2-realex8R _XbaI | TATATCTAGATCAGCGTCGCGAACTTACTAATAAC |
| 16 | Tra2-ex9 | GCCCTTTCGAAACGGCGGTCGAAAACCCGG |
| | | Site directed mutagenesis of zinc finger motifs or *Bmznf* using PCR. |
| 17 | Zmut-pair1F | AACACTTCCAATCGTGGTGCGTCATCTAGATATCTCCTCAAAATA |
| 18 | Zmut-pair2R | TGTGGCGAGAGCATATGAAGATTCTGCACGTGTGGACTTTTT |
| | | Producing short dsRNA for the RNAi of *Bmtra-2* gene. |
| 19 | Tra2KD_A | GAGGACATGGAAGATGCTAAGATTGCAAAGAATGAATTTTCCTGTCTC |
| 20 | Tra2KD_B | TTCATTCTTTGCAATCTTAGCATCTTCCATGTCCTCTTTTCCTGTCTC |
| | | Amplifying both *Bmdsxm* and *Bmdsxf* splicing isoforms. |
| 21 | DSX_2F | CCGTCCCCTCGGAGACGCTTGTG |
| 22 | DSX_5R | CTGTATCGGCGCGCAGTGTCGTCGC |
| | | Cloning the Bmznf gene in pIZT vector. |
| 23 | Zn2-F-SacI | GCGCGGGAGCTCGGGATGAAAAAATACTTAAAATAC |
| 24 | Zn2-R-NotI | TATTAGCGGCCGCCGCGATACGTGCGTCTTATC |
| | | Sequencing of pIZT clones. |
| 25 | OpIE2 F | CGCAACGATCTGGTAAACAC |
| 26 | OpIE2 R | GACAATACAAACTAAGATTTAGTCAG |
| | | Producing short dsRNA for the RNAi of Bmznf gene. |
| 27 | TKDZ123F | TGATTTTGAAAATAAAAAGTGCACACGTGCAGAATGTTCATATGCTCACGCCACAGTGCACGAGAAAGAATTCCTGTCTC |
| 28 | TKDZ123R | TTCTTTCTCGTGCACTGTGGCGTGAGCATATGAACATTCTGCACGTGTGCACTTTTTATTTTCAAAATCATTCCTGTCTC |

| | | |
|---|---|---|
| | | RT primer for stem loop PCR |
| 29 | Stmlp_RT12564 | GTCGTATCCAGTG CAGGGTCCGAGGTATTCG CACTGGATACGACATTACGAC |
| | | Forward primer |
| 30 | Stmlp_F12564 | GCGGCGGATGACGTCGAAGAGACCATGT |
| | | RT primer for stem loop PCR |
| 31 | Stmlp_RT24319 | GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGAACAATC |
| | | Forward primer |
| 32 | Stmlp_F24319 | GCGGCGGTATTTCGTATCACGTACCGC |
| | | RT primer for stem loop PCR |
| 33 | Stmlp_RT37041 | GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAAGGCCAT |
| | | Forward primer |
| 34 | Stmlp_F37041 | GCGGCGGTTCATAAGTCTGAAGTTCGGCG |
| | | Reverse primer for all Stem loop PCRs |
| 35 | Stmlp_R | GTGCAGGGTCCGAGGT |
| | | Primers used for Bmtra-2 splicing and qRT-PCR analysis ($F_1$ to $R_6$) |
| 36 | $F_1$ | TATAGAATTCCGATTTGCGTCTTCAACGTGC |
| 37 | $F_2$ | CTCGCAACGGTTCTCGCGAGCCGG |
| 38 | $F_3$ | GGCGATAATGGATACGACAGGCGCCGC |
| 39 | $R_1$ | GGCGCGAGCGGTAGCCGCGACG |
| 40 | $R_2$ | GTAGGTTTTCCCATGTAGATGCCCGG |
| 41 | $R_3$ | GCGGCGCCTGTCGTATCCATTATCGCC |
| 42 | $R_4$ | GGGGCGAGTAGGAGCGCTCCCGCTCG |
| 43 | $R_5$ | TATATCTAGATCAGCGTCGCGAACTTACTAATAAC |
| 44 | $R_6$ | CCGGGTTTTCGACCGCCGTTTCGAAAGGGC |

**Annexure V.** Primers used for the qRT-PCR in Chapter IV

| Sc.No | Primer_Name | Primer |
|---|---|---|
| 1 | BGIBMGA000127_F | GCATCTAGCACAGACAGTGA |
| 2 | BGIBMGA000127_R | ATCTGACATCGTTAATACCA |
| 3 | BGIBMGA003212_F | TGAAACACCCGTTTGGCAAG |
| 4 | BGIBMGA003212_R | TAGCTTGTTCCTTGAACGCC |
| 5 | BGIBMGA000971_F | ACAAGACCTAAACTAGGTTT |
| 6 | BGIBMGA000971_R | CAAGTGATCTCTCTATGACA |
| 7 | BGIBMGA000138_F | GGGCTTGGAGGAACAACATA |
| 8 | BGIBMGA000138_R | ACCTGCATCACTGCCGTCAA |
| 9 | BGIBMGA002917_F | AAGTCCCTGCAACGAAATTA |
| 10 | BGIBMGA002917_R | CTCCGCCCGGTATTATCCTA |
| 11 | BGIBMGA002150_F | AAAGCAGAGCTTGCTGTAAA |
| 12 | BGIBMGA002150_R | CACTTTCCTCTTCTATATCG |
| 13 | BGIBMGA000608_F | CTAAATATGTGGAGGACCTTAT |
| 14 | BGIBMGA000608_R | CATCATATATAAAGTAGCGAAG |
| 15 | BGIBMGA000721_F | ATACGTTGTGTACGCGCCGCTG |
| 16 | BGIBMGA000721_R | CGGCTTGGCTGTGGTACTACAC |
| 17 | BGIBMGA002146_F | TGAACTACCAGTAAATTAC |
| 18 | BGIBMGA002146_R | AAGACAGCTAGTTAAATAG |
| 19 | BGIBMGA003852_F | GCCACCGAGAACATGAAAGCT |
| 20 | BGIBMGA003852_R | TCTTTCTTCCGTGACGACCAG |
| 21 | BGIBMGA001500_F | ATTGTCAACCAGGAGGTGAG |
| 22 | BGIBMGA001500_R | TCCAGTTGGTCAACGCTCAA |

# References

# References


ABE, H., T. FUJII, N. TANAKA, T. YOKOYAMA, H. KAKEHASHI *et al.*, 2008 Identification of the female-determining region of the W chromosome in Bombyx mori. Genetica **133:** 269-282.

ABE, H., M. SEKI, F. OHBAYASHI, N. TANAKA, J. YAMASHITA *et al.*, 2005 Partial deletions of the W chromosome due to reciprocal translocation in the silkworm Bombyx mori. Insect Mol Biol **14:** 339-352.

ADDEPALLI, B., and A. G. HUNT, 2007 A novel endonuclease activity associated with the Arabidopsis ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor. Nucleic Acids Res **35:** 4453-4463.

AJIMURA, M., K. SAHARA, H. ABE, T. TAMURA, T. SHIMADA *et al.*, 2006 Are the zinc-finger motif genes, zl and z20, located in the W chromosome involved in the sex-determination of the domesticated silkworm, Bombyx mori? abstracts from Seventh International Workshop on the Molecular Biology and Genetics of the Lepidoptera. J. Insect Sci **7:** 4-5.

AMACK, J. D., and M. S. MAHADEVAN, 2004 Myogenic defects in myotonic dystrophy. Dev Biol **265:** 294-301.

AMREIN, H., M. L. HEDLEY and T. MANIATIS, 1994 The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. Cell **76:** 735-746.

AN, W., S. CHO, H. ISHII and P. C. WENSINK, 1996 Sex-specific and non-sex-specific oligomerization domains in both of the doublesex transcription factors from Drosophila melanogaster. Mol Cell Biol **16:** 3106-3111.

ANDREWS, S., 2010 FastQC: a quality control tool for high throughput sequence, pp. http://www.citeulike.org/user/nailest/article/11583827.

ANDREWS, S., 2015 A tool to visualise and analyse high throughput mapped sequence data, pp.

ARAVIN, A. A., R. SACHIDANANDAM, A. GIRARD, K. FEJES-TOTH and G. J. HANNON, 2007 Developmentally regulated piRNA clusters implicate MILI in transposon control. Science **316:** 744-747.

ARNOLD, A. P., 2012 The end of gonad-centric sex determination in mammals. Trends in genetics : TIG **28:** 55-61.

ARNOLD, A. P., X. CHEN, J. C. LINK, Y. ITOH and K. REUE, 2013 Cell-autonomous sex determination outside of the gonad. Developmental dynamics : an official publication of the American Association of Anatomists **242:** 371-379.

ARUNKUMAR, K. P., K. MITA and J. NAGARAJU, 2009 The silkworm Z chromosome is enriched in testis-specific genes. Genetics **182:** 493-501.

BACHTROG, D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet **14:** 113-124.

BAKER, R. H., and R. K. SAKAI, 1976 Male determining factor on chromosome 3 in the mosquito, Culex tritaeniorhynchus. J Hered **67:** 289-294.

BALL, C. B., K. F. RODRIGUEZ, D. J. STUMPO, F. RIBEIRO-NETO, K. S. KORACH *et al.*, 2014 The RNA-binding protein, ZFP36L2, influences ovulation and oocyte maturation. PLoS One **9:** e97324.

BARBOSA, C., I. PEIXEIRO and L. ROMAO, 2013 Gene expression regulation by upstream open reading frames and human disease. PLoS genetics **9:** e1003529.

BEDO, D. G., and G. G. FOSTER, 1985 Cytogenetic mapping of the male-determining region of Lucilia cuprina (Diptera: Calliphoridae). Chromosoma **92:** 344–350.

BERTA, P., J. R. HAWKINS, A. H. SINCLAIR, A. TAYLOR, B. L. GRIFFITHS *et al.*, 1990 Genetic evidence equating SRY and the testis-determining factor. Nature **348:** 448-450.

BEST, A., K. JAMES, C. DALGLIESH, E. HONG, M. KHEIROLAHI-KOUHESTANI *et al.*, 2014 Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons. Nat Commun **5:** 4760.

BEUKEBOOM, L. W., 1995 Sex determination in Hymenoptera: a need for genetic and molecular studies. BioEssays : news and reviews in molecular, cellular and developmental biology **17:** 813-817.

BEYE, M., C. SEELMANN, T. GEMPE, M. HASSELMANN, X. VEKEMANS *et al.*, 2013 Gradual molecular evolution of a sex determination switch through incomplete penetrance of femaleness. Current biology : CB **23:** 2559-2564.

BLACKSHEAR, P. J., 2002 Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. Biochem Soc Trans **30:** 945-952.

BOLGER, A. M., M. LOHSE and B. USADEL, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:** 2114-2120.

BUCHNER, D. A., A. CHARRIER, E. SRINIVASAN, L. WANG, M. T. PAULSEN *et al.*, 2015 Zinc finger protein 407 (ZFP407) regulates insulin-stimulated glucose uptake and glucose transporter 4 (Glut4) mRNA. J Biol Chem **290:** 6376-6386.

BULL, J. J., 1985 Sex determining mechanisms: an evolutionary perspective. Experientia **41:** 1285-1296.

CAI, Y., Q. ZHOU, C. YU, X. WANG, S. HU *et al.*, 2012 Transposable-element associated small RNAs in Bombyx mori genome. PloS one **7:** e36599.

CALVO, S. E., D. J. PAGLIARINI and V. K. MOOTHA, 2009 Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proceedings of the National Academy of Sciences of the United States of America **106:** 7507-7512.

CARVALHO, A. B., 2002 Origin and evolution of the Drosophila Y chromosome. Curr Opin Genet Dev **12:** 664-668.

CHARLESWORTH, B., 1978 Model for evolution of Y chromosomes and dosage compensation. Proc Natl Acad Sci U S A **75:** 5618-5622.

CHARLESWORTH, B., 1996 The evolution of chromosomal sex determination and dosage compensation. Curr Biol **6:** 149-162.

CHARLESWORTH, B., and D. CHARLESWORTH, 2000 The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci **355:** 1563-1572.

CHENG, A. W., J. SHI, P. WONG, K. L. LUO, P. TREPMAN *et al.*, 2014 Muscleblind-like 1 (Mbnl1) regulates pre-mRNA alternative splicing during terminal erythropoiesis. Blood **124:** 598-610.

CHOI, Y. J., W. S. LAI, R. FEDIC, D. J. STUMPO, W. HUANG *et al.*, 2014 The Drosophila Tis11 protein and its effects on mRNA expression in flies. J Biol Chem **289:** 35042-35060.

CHRISTIANSEN, A. E., E. L. KEISMAN, S. M. AHMAD and B. S. BAKER, 2002 Sex comes in from the cold: the integration of sex and pattern. Trends Genet **18:** 510-516.

CLC, 2014 https://www.qiagenbioinformatics.com pp.

CLOUGH, E., E. JIMENEZ, Y. A. KIM, C. WHITWORTH, M. C. NEVILLE *et al.*, 2014 Sex- and tissue-specific functions of Drosophila doublesex transcription factor target genes. Dev Cell **31:** 761-773.

CONCHA, C., and M. J. SCOTT, 2009 Sexual development in Lucilia cuprina (Diptera, Calliphoridae) is controlled by the transformer gene. Genetics **182:** 785-798.

CONRAD, T., and A. AKHTAR, 2012 Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat Rev Genet **13:** 123-134.

CONSORTIUM, I. S. G., 2008 The genome of a lepidopteran model insect, the silkworm Bombyx mori. Insect biochemistry and molecular biology **38:** 1036-1045.

CUTTING, A., J. CHUE and C. A. SMITH, 2013 Just how conserved is vertebrate sex determination? Developmental dynamics : an official publication of the American Association of Anatomists **242:** 380-387.

DE, J., W. S. LAI, J. M. THORN, S. M. GOLDSWORTHY, X. LIU *et al.*, 1999 Identification of four CCCH zinc finger proteins in Xenopus, including a novel vertebrate protein with four zinc fingers and severely restricted expression. Gene **228:** 133-145.

DENG, X., J. B. HIATT, D. K. NGUYEN, S. ERCAN, D. STURGILL *et al.*, 2011 Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. Nat Genet **43:** 1179-1185.

DRAPER, B. W., C. C. MELLO, B. BOWERMAN, J. HARDIN and J. R. PRIESS, 1996 MEX-3 is a KH domain protein that regulates blastomere identity in early C. elegans embryos. Cell **87:** 205-216.

EFIMENKO, E., M. B. PADUA, N. L. MANUYLOV, S. C. FOX, D. A. MORSE *et al.*, 2013 The transcription factor GATA4 is required for follicular development and normal ovarian function. Dev Biol **381:** 144-158.

ELLEGREN, H., 2011 Emergence of male-biased genes on the chicken Z-chromosome: sex-chromosome contrasts between male and female heterogametic systems. Genome Res **21:** 2082-2086.

ELLEGREN, H., L. HULTIN-ROSENBERG, B. BRUNSTROM, L. DENCKER, K. KULTIMA *et al.*, 2007 Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. BMC Biol **5:** 40.

ELLEGREN, H., and J. PARSCH, 2007 The evolution of sex-biased genes and sex-biased gene expression. Nature reviews. Genetics **8:** 689-698.

ERCAN, S., P. G. GIRESI, C. M. WHITTLE, X. ZHANG, R. D. GREEN *et al.*, 2007 X chromosome repression by localization of the C. elegans dosage compensation machinery to sites of transcription initiation. Nat Genet **39:** 403-408.

ERDMAN, S. E., and K. C. BURTIS, 1993 The Drosophila doublesex proteins share a novel zinc finger related DNA binding domain. EMBO J **12:** 527-535.

ERICKSON, J. W., and J. J. QUINTERO, 2007 Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila. PLoS Biol **5:** e332.

FATICA, A., and I. BOZZONI, 2015 Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet **15:** 7-21.

FLICEK, P., M. R. AMODE, D. BARRELL, K. BEAL, K. BILLIS *et al.*, 2014 Ensembl 2014. Nucleic Acids Res **42:** D749-755.

GAMBLE, T., and D. ZARKOWER, 2012 Sex determination. Current biology : CB **22:** R257-262.

GOPINATH, G., K. P. ARUNKUMAR, K. MITA and J. NAGARAJU, 2016 Role of Bmznf-2, a Bombyx mori CCCH zinc finger gene, in masculinisation and differential splicing of Bmtra-2. Insect biochemistry and molecular biology **75:** 32-44.

GRONDIN, B., M. BAZINET and M. AUBRY, 1996 The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix. J Biol Chem **271:** 15458-15467.

GUEDES, S., and J. R. PRIESS, 1997 The C. elegans MEX-1 protein is present in germline blastomeres and is a P granule component. Development **124:** 731-739.

GUO, J., T. WU, J. ANDERSON, B. F. KANE, D. G. JOHNSON *et al.*, 2000 Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. J Virol **74:** 8980-8988.

HAGELE, K., 1985 Identification of a polytene chromosome band containing a male sex determiner of Chironomus thummi thummi. Chromosoma **91:** 167-171.

HAGEN, G., S. MULLER, M. BEATO and G. SUSKE, 1994 Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J **13:** 3843-3851.

HALL, A. B., S. BASU, X. JIANG, Y. QI, V. A. TIMOSHEVSKIY *et al.*, 2015 SEX DETERMINATION. A male-determining factor in the mosquito Aedes aegypti. Science **348:** 1268-1270.

HAN, G., M. WANG, F. YUAN, N. SUI, J. SONG *et al.*, 2014 The CCCH zinc finger protein gene AtZFP1 improves salt resistance in Arabidopsis thaliana. Plant Mol Biol **86:** 237-253.

HARRISON, P. W., J. E. MANK and N. WEDELL, 2012 Incomplete sex chromosome dosage compensation in the Indian meal moth, Plodia interpunctella, based on de novo transcriptome assembly. Genome Biol Evol **4:** 1118-1126.

HASHIMOTO, H., 1933 The role of the W chromosome for sexdetermination in the silkworm, Bombyx mori. Jap J. Genet **8:** 245-258.

HAYASHI, Y., M. HAYASHI and S. KOBAYASHI, 2004 Nanos suppresses somatic cell fate in Drosophila germ line. Proceedings of the National Academy of Sciences of the United States of America **101:** 10338-10342.

HAZELRIGG, T., and C. TU, 1994 Sex-specific processing of the Drosophila exuperantia transcript is regulated in male germ cells by the tra-2 gene. Proc Natl Acad Sci U S A **91:** 10752-10756.

HEDIGER, M., C. HENGGELER, N. MEIER, R. PEREZ, G. SACCONE *et al.*, 2010 Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. Genetics **184:** 155-170.

HOSHIJIMA, K., K. INOUE, I. HIGUCHI, H. SAKAMOTO and Y. SHIMURA, 1991 Control of doublesex alternative splicing by transformer and transformer-2 in Drosophila. Science **252:** 833-836.

HTTP://WWW.SIRNAWIZARD.COM/DESIGN.PHP, 2016 Invivogen, innovation with research, pp.

HUANG, F. F., C. L. CHAI, Z. ZHANG, Z. H. LIU, F. Y. DAI *et al.*, 2008 The UDP-glucosyltransferase multigene family in Bombyx mori. BMC Genomics **9:** 563.

HUANG, N. N., and C. P. HUNTER, 2015 The RNA binding protein MEX-3 retains asymmetric activity in the early Caenorhabditis elegans embryo in the absence of asymmetric protein localization. Gene **554:** 160-173.

HUANG, N. N., D. E. MOOTZ, A. J. WALHOUT, M. VIDAL and C. P. HUNTER, 2002 MEX-3 interacting proteins link cell polarity to asymmetric gene expression in Caenorhabditis elegans. Development **129:** 747-759.

HUDSON, B. P., M. A. MARTINEZ-YAMOUT, H. J. DYSON and P. E. WRIGHT, 2004 Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. Nat Struct Mol Biol **11:** 257-264.

IACONO, M., F. MIGNONE and G. PESOLE, 2005 uAUG and uORFs in human and rodent 5'untranslated mRNAs. Gene **349:** 97-105.

ICHIDA, Y., Y. UTSUNOMIYA, T. YASUDA, K. NAKABAYASHI, T. SATO *et al.*, 2015 Functional Domains of ZFP809 Essential for Nuclear Localization and Gene Silencing. PLoS One **10:** e0139274.

ITOH, Y., E. MELAMED, X. YANG, K. KAMPF, S. WANG *et al.*, 2007 Dosage compensation is less effective in birds than in mammals. J Biol **6:** 2.

JIANG, X., J. K. BIEDLER, Y. QI, A. B. HALL and Z. TU, 2015 Complete Dosage Compensation in Anopheles stephensi and the Evolution of Sex-Biased Genes in Mosquitoes. Genome Biol Evol **7:** 1914-1924.

KAISER, V. B., and H. ELLEGREN, 2006 Nonrandom distribution of genes with sex-biased expression in the chicken genome. Evolution **60:** 1945-1951.

KATSUMA, S., M. KAWAMOTO and T. KIUCHI, 2014 Guardian small RNAs and sex determination. RNA biology **11:** 1238-1242.

KATSUMA, S., Y. SUGANO, T. KIUCHI and T. SHIMADA, 2015 Two Conserved Cysteine Residues Are Required for the Masculinizing Activity of the Silkworm Masc Protein. The Journal of biological chemistry **290:** 26114-26124.

KAWAMATA, M., H. INOUE and K. NISHIMORI, 2007 Male-specific function of Dmrt7 by sexually dimorphic translation in mouse testis. Sex Dev **1:** 297-304.

KAWAMATA, M., and K. NISHIMORI, 2006 Mice deficient in Dmrt7 show infertility with spermatogenic arrest at pachytene stage. FEBS Lett **580:** 6442-6446.

KAWAMOTO, M., H. KOGA, T. KIUCHI, K. SHOJI, S. SUGANO *et al.*, 2015 Sexually biased transcripts at early embryonic stages of the silkworm depend on the sex chromosome constitution. Gene **560:** 50-56.

KAWAOKA, S., N. HAYASHI, S. KATSUMA, H. KISHINO, Y. KOHARA *et al.*, 2008 Bombyx small RNAs: genomic defense system against transposons in the silkworm, Bombyx mori. Insect biochemistry and molecular biology **38:** 1058-1065.

KAWAOKA, S., N. HAYASHI, Y. SUZUKI, H. ABE, S. SUGANO *et al.*, 2009 The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. RNA **15:** 1258-1264.

KAWAOKA, S., K. KADOTA, Y. ARAI, Y. SUZUKI, T. FUJII *et al.*, 2011 The silkworm W chromosome is a source of female-enriched piRNAs. RNA **17:** 2144-2151.

KAY, B. K., M. P. WILLIAMSON and M. SUDOL, 2000 The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J **14:** 231-241.

KAYMAK, E., and S. P. RYDER, 2013 RNA recognition by the Caenorhabditis elegans oocyte maturation determinant OMA-1. J Biol Chem **288:** 30463-30472.

KHARCHENKO, P. V., R. XI and P. J. PARK, 2011 Evidence for dosage compensation between the X chromosome and autosomes in mammals. Nat Genet **43:** 1167-1169; author reply 1171-1162.

KIUCHI, T., H. KOGA, M. KAWAMOTO, K. SHOJI, H. SAKAI *et al.*, 2014 A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature **509:** 633-636.

KLUVER, N., M. KONDO, A. HERPIN, H. MITANI and M. SCHARTL, 2005 Divergent expression patterns of Sox9 duplicates in teleosts indicate a lineage specific subfunctionalization. Dev Genes Evol **215:** 297-305.

KOBAYASHI, S., M. YAMADA, M. ASAOKA and T. KITAMURA, 1996 Essential role of the posterior morphogen nanos for germline development in Drosophila. Nature **380:** 708-711.

KOHTZ, J. D., S. F. JAMISON, C. L. WILL, P. ZUO, R. LUHRMANN *et al.*, 1994 Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature **368:** 119-124.

KONIECZNY, P., E. STEPNIAK-KONIECZNA and K. SOBCZAK, 2014 MBNL proteins and their target RNAs, interaction and splicing regulation. Nucleic Acids Res **42:** 10873-10887.

KOZAK, M., 1987 An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic acids research **15:** 8125-8148.

KOZAK, M., 1991 An analysis of vertebrate mRNA sequences: intimations of translational control. The Journal of cell biology **115:** 887-903.

KREBS, C. J., and D. M. ROBINS, 2010 A pair of mouse KRAB zinc finger proteins modulates multiple indicators of female reproduction. Biol Reprod **82:** 662-668.

KULATHINAL, R. J., L. SKWAREK, R. A. MORTON and R. S. SINGH, 2003 Rapid evolution of the sex-determining gene, transformer: structural diversity and rate heterogeneity among sibling species of Drosophila. Mol Biol Evol **20:** 441-452.

KUNG, J. T., D. COLOGNORI and J. T. LEE, 2013 Long noncoding RNAs: past, present, and future. Genetics **193:** 651-669.

LAHN, B. T., N. M. PEARSON and K. JEGALIAN, 2001 The human Y chromosome, in the light of evolution. Nat Rev Genet **2:** 207-216.

LAI, W. S., and P. J. BLACKSHEAR, 2001 Interactions of CCCH zinc finger proteins with mRNA: tristetraprolin-mediated AU-rich element-dependent mRNA degradation can occur in the absence of a poly(A) tail. J Biol Chem **276:** 23144-23154.

LANGMEAD, B., and S. L. SALZBERG, 2012 Fast gapped-read alignment with Bowtie 2. Nat Methods **9:** 357-359.

LEE, J., T. KIUCHI, M. KAWAMOTO, T. SHIMADA and S. KATSUMA, 2015 Identification and functional analysis of a Masculinizer orthologue in Trilocha varians (Lepidoptera: Bombycidae). Insect molecular biology **24:** 561-569.

LI, B., and C. N. DEWEY, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12:** 323.

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079.

LI, Z., and T. L. THOMAS, 1998 PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis. Plant Cell **10:** 383-398.

LUCCHESI, J. C., 1973 Dosage compensation in Drosophila. Annu Rev Genet **7:** 225-237.

LUTFALLA, G., H. ROEST CROLLIUS, F. G. BRUNET, V. LAUDET and M. ROBINSON-RECHAVI, 2003 Inventing a sex-specific gene: a conserved role of DMRT1 in teleost fishes plus a recent duplication in the medaka Oryzias latipes resulted in DMY. J Mol Evol **57 Suppl 1:** S148-153.

MADIGAN, S. J., P. EDEEN, J. ESNAYRA and M. MCKEOWN, 1996 att, a target for regulation by tra2 in the testes of Drosophila melanogaster, encodes alternative RNAs and alternative proteins. Mol Cell Biol **16:** 4222-4230.

MADL, J. E., and R. K. HERMAN, 1979 Polyploids and sex determination in Caenorhabditis elegans. Genetics **93:** 393-402.

MAHAJAN, S., and D. BACHTROG, 2015 Partial dosage compensation in Strepsiptera, a sister group of beetles. Genome Biol Evol **7:** 591-600.

MANK, J. E., 2013 Sex chromosome dosage compensation: definitely not for everyone. Trends Genet **29:** 677-683.

MANK, J. E., D. J. HOSKEN and N. WEDELL, 2011 Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict. Evolution **65:** 2133-2144.

MARIN, I., M. L. SIEGAL and B. S. BAKER, 2000 The evolution of dosage-compensation mechanisms. Bioessays **22:** 1106-1114.

MASUDA, A., H. S. ANDERSEN, T. K. DOKTOR, T. OKAMOTO, M. ITO *et al.*, 2012 CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay. Sci Rep **2:** 209.

MATSUDA, M., T. SATO, Y. TOYAZAKI, Y. NAGAHAMA, S. HAMAGUCHI *et al.*, 2003 Oryzias curvinotus has DMY, a gene that is required for male development in the medaka, O. latipes. Zoolog Sci **20:** 159-161.

MATTHEWS, J. M., K. KOWALSKI, C. K. LIEW, B. K. SHARPE, A. H. FOX *et al.*, 2000 A class of zinc fingers involved in protein-protein interactions biophysical characterization of CCHC fingers from fog and U-shaped. Eur J Biochem **267:** 1030-1038.

MCDONEL, P., J. JANS, B. K. PETERSON and B. J. MEYER, 2006 Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. Nature **444:** 614-618.

MELLER, V. H., 2000 Dosage compensation: making 1X equal 2X. Trends Cell Biol **10:** 54-59.

MENEELY, P. M., and W. B. WOOD, 1984 An autosomal gene that affects X chromosome expression and sex determination in Caenorhabditis elegans. Genetics **106:** 29-44.

MERCER, T. R., M. E. DINGER and J. S. MATTICK, 2009 Long non-coding RNAs: insights into functions. Nat Rev Genet **10:** 155-159.

MINE, E., S. IZUMI, M. KATSUKI and S. TOMINO, 1983 Developmental and sex-dependent regulation of storage protein synthesis in the silkworm, Bombyx mori. Developmental biology **97:** 329-337.

MOLNAR, G., A. CROZAT and A. B. PARDEE, 1994 The immediate-early gene Egr-1 regulates the activity of the thymidine kinase promoter at the G0-to-G1 transition of the cell cycle. Mol Cell Biol **14:** 5242-5248.

MON, H., Z. LI, I. KOBAYASHI, S. TOMITA, J. LEE *et al.*, 2012 Soaking RNAi in Bombyx mori BmN4-SID1 cells arrests cell cycle progression. J Insect Sci **13:** 155.

MORRIS, D. R., 1995 Growth control of translation in mammalian cells. Progress in nucleic acid research and molecular biology **51:** 339-363.

MORRIS, D. R., and A. P. GEBALLE, 2000 Upstream open reading frames as regulators of mRNA translation. Molecular and cellular biology **20:** 8635-8642.

NGUYEN, D. K., and C. M. DISTECHE, 2006 Dosage compensation of the active X chromosome in mammals. Nat Genet **38:** 47-53.

NIIMI, T., K. SAHARA, H. OSHIMA, Y. YASUKOCHI, K. IKEO *et al.*, 2006 Molecular cloning and chromosomal localization of the Bombyx Sex-lethal gene. Genome **49:** 263-268.

NIU, B. L., Z. Q. MENG, Y. Z. TAO, S. L. LU, H. B. WENG *et al.*, 2005 Cloning and alternative splicing analysis of Bombyx mori transformer-2 gene using silkworm EST database. Acta Biochim Biophys Sin (Shanghai) **37:** 728-736.

O'NEIL, M. T., and J. M. BELOTE, 1992 Interspecific comparison of the transformer gene of Drosophila reveals an unusually high degree of evolutionary divergence. Genetics **131:** 113-128.

OGO, O. A., J. TYSON, S. J. COCKELL, A. HOWARD, R. A. VALENTINE *et al.*, 2015 The zinc finger protein ZNF658 regulates the transcription of genes involved in zinc homeostasis and affects ribosome biogenesis through the zinc transcriptional regulatory element. Mol Cell Biol **35:** 977-987.

OHNO, S. (Editor), 1967 *Sex Chromosomes and Sex Linked Genes*, Springer Verlag, Berlin, Germany.

OKANO, H., T. IMAI and M. OKABE, 2002 Musashi: a translational regulator of cell fate. Journal of cell science **115:** 1355-1359.

OSANAI-FUTAHASHI, M., Y. SUETSUGU, K. MITA and H. FUJIWARA, 2008 Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, Bombyx mori. Insect biochemistry and molecular biology **38:** 1046-1057.

OSBORNE, R. J., X. LIN, S. WELLE, K. SOBCZAK, J. R. O'ROURKE *et al.*, 2009 Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. Hum Mol Genet **18:** 1471-1481.

PACHECO, T. R., M. B. COELHO, J. M. DESTERRO, I. MOLLET and M. CARMO-FONSECA, 2006 In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. Mol Cell Biol **26:** 8183-8190.

PHILIP, P., and P. STENBERG, 2013 Male X-linked genes in Drosophila melanogaster are compensated independently of the Male-Specific Lethal complex. Epigenetics Chromatin **6:** 35.

PICARD, M. A., C. COSSEAU, G. MOUAHID, D. DUVAL, C. GRUNAU *et al.*, 2015 The roles of Dmrt (Double sex/Male-abnormal-3 Related Transcription factor) genes in sex determination and differentiation mechanisms: Ubiquity and diversity across the animal kingdom. C R Biol **338:** 451-462.

PIELER, T., and E. BELLEFROID, 1994 Perspectives on zinc finger protein function and evolution--an update. Mol Biol Rep **20:** 1-8.

PIERACCIOLI, M., S. NICOLAI, A. ANTONOV, J. SOMERS, M. MALEWICZ *et al.*, 2015 ZNF281 contributes to the DNA damage response by controlling the expression of XRCC2 and XRCC4. Oncogene.

POULOS, M. G., R. BATRA, M. LI, Y. YUAN, C. ZHANG *et al.*, 2013 Progressive impairment of muscle regeneration in muscleblind-like 3 isoform knockout mice. Hum Mol Genet **22:** 3547-3558.

PRINCE, E. G., D. KIRKLAND and J. P. DEMUTH, 2010 Hyperexpression of the X chromosome in both sexes results in extensive female bias of X-linked genes in the flour beetle. Genome Biol Evol **2:** 336-346.

RAMOS, S. B., D. J. STUMPO, E. A. KENNINGTON, R. S. PHILLIPS, C. B. BOCK *et al.*, 2004 The CCCH tandem zinc-finger protein Zfp36l2 is crucial for female fertility and early embryonic development. Development **131:** 4883-4893.

RANUM, L. P., and T. A. COOPER, 2006 RNA-mediated neuromuscular disorders. Annu Rev Neurosci **29:** 259-277.

RAYMOND, C. S., M. W. MURPHY, M. G. O'SULLIVAN, V. J. BARDWELL and D. ZARKOWER, 2000 Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes Dev **14:** 2587-2595.

REESE, K. J., M. A. DUNN, J. A. WADDLE and G. SEYDOUX, 2000 Asymmetric segregation of PIE-1 in C. elegans is mediated by two complementary mechanisms that act through separate PIE-1 protein domains. Mol Cell **6:** 445-455.

RIDEOUT, E. J., J. C. BILLETER and S. F. GOODWIN, 2007 The sex-determination genes fruitless and doublesex specify a neural substrate required for courtship song. Curr Biol **17:** 1473-1478.

ROBIN, H. L., F. L. PATTI, L. M. KARI, F. L. ERIC and F. L. DENNIS, 2015 Statistics: Unlocking the Power of Data by Lock, Lock, Lock, Lock, and Lock, pp.

ROBINSON, M. D., D. J. MCCARTHY and G. K. SMYTH, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26:** 139-140.

ROSE, G., E. KRZYWINSKA, J. KIM, L. REVUELTA, L. FERRETTI *et al.*, 2016 Dosage Compensation in the African Malaria Mosquito Anopheles gambiae. Genome Biol Evol **8:** 411-425.

RUIZ, M. F., A. MILANO, M. SALVEMINI, J. M. EIRIN-LOPEZ, A. L. PERONDINI *et al.*, 2007 The gene transformer of anastrepha fruit flies (Diptera, tephritidae) and its evolution in insects. PLoS One **2:** e1239.

SAKAI, H., F. AOKI and M. G. SUZUKI, 2014 Identification of the key stages for sex determination in the silkworm, Bombyx mori. Development genes and evolution **224:** 119-123.

SAKAI, H., H. SAKAGUCHI, F. AOKI and M. G. SUZUKI, 2015a Functional analysis of sex-determination genes by gene silencing with LNA-DNA gapmers in the silkworm, Bombyx mori. Mech Dev **137:** 45-52.

SAKAI, H., H. SAKAGUCHI, F. AOKI and M. G. SUZUKI, 2015b Functional analysis of sex-determination genes by gene silencing with LNA-DNA gapmers in the silkworm, Bombyx mori. Mechanisms of development **137:** 45-52.

SAKAI, H., M. SUMITANI, Y. CHIKAMI, K. YAHATA, K. UCHINO *et al.*, 2016 Transgenic Expression of the piRNA-Resistant Masculinizer Gene Induces Female-Specific Lethality and Partial Female-to-Male Sex Reversal in the Silkworm, Bombyx mori. PLoS genetics **12:** e1006203.

SAKURAI, H., T. FUJII, S. IZUMI and S. TOMINO, 1988 Structure and expression of gene coding for sex-specific storage protein of Bombyx mori. The Journal of biological chemistry **263:** 7876-7880.

SATISH, V., J. N. SHUKLA and J. NAGARAJU, 2006 CCCH-type zinc finger genes: candidate regulators of sex-determination pathway in the silkworm, Bombyx mori. Abstracts from Seventh International Workshop on the Molecular Biology and Genetics of the Lepidoptera. J. Insect Sci **7:** 40.

SCHUBERT, C. M., R. LIN, C. J. DE VRIES, R. H. PLASTERK and J. R. PRIESS, 2000 MEX-5 and MEX-6 function to establish soma/germline asymmetry in early C. elegans embryos. Mol Cell **5:** 671-682.

SEONG, J. Y., J. HAN, S. PARK, W. WUTTKE, H. JARRY *et al.*, 2002 Exonic splicing enhancer-dependent splicing of the gonadotropin-releasing hormone premessenger ribonucleic acid is mediated by tra2alpha, a 40-kilodalton serine/arginine-rich protein. Mol Endocrinol **16:** 2426-2438.

SHEARMAN, D. C., and M. FROMMER, 1998 The Bactrocera tryoni homologue of the Drosophila melanogaster sex-determination gene doublesex. Insect Mol Biol **7:** 355-366.

SHIMADA, M., H. KAWAHARA and H. DOI, 2002 Novel family of CCCH-type zinc-finger proteins, MOE-1, -2 and -3, participates in C. elegans oocyte maturation. Genes Cells **7:** 933-947.

SHIMOMURA, M., H. MINAMI, Y. SUETSUGU, H. OHYANAGI, C. SATOH *et al.*, 2009 KAIKObase: an integrated silkworm genome database and data mining tool. BMC Genomics **10:** 486.

SHIRANGI, T. R., B. J. TAYLOR and M. MCKEOWN, 2006 A double-switch system regulates male courtship behavior in male and female Drosophila melanogaster. Nat Genet **38:** 1435-1439.

SINCLAIR, A. H., P. BERTA, M. S. PALMER, J. R. HAWKINS, B. L. GRIFFITHS *et al.*, 1990 A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature **346:** 240-244.

SINGH, D. K., and K. V. PRASANTH, 2013 Functional insights into the role of nuclear-retained long noncoding RNAs in gene expression control in mammalian cells. Chromosome Res **21:** 695-711.

SINGH, J., C. P. SINGH, A. BHAVANI and J. NAGARAJU, 2010 Discovering microRNAs from Bombyx mori nucleopolyhedrosis virus. Virology **407:** 120-128.

SMITH, G., Y. R. CHEN, G. W. BLISSARD and A. D. BRISCOE, 2014 Complete dosage compensation and sex-biased gene expression in the moth Manduca sexta. Genome Biol Evol **6:** 526-537.

SOARES, L. M., K. ZANIER, C. MACKERETH, M. SATTLER and J. VALCARCEL, 2006 Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. Science **312:** 1961-1965.

SPRIGGS, K. A., M. BUSHELL and A. E. WILLIS, 2010 Translational regulation of gene expression during conditions of cell stress. Molecular cell **40:** 228-237.

STEFL, R., L. SKRISOVSKA and F. H. ALLAIN, 2005 RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. EMBO Rep **6:** 33-38.

STRAUB, T., G. D. GILFILLAN, V. K. MAIER and P. B. BECKER, 2005 The Drosophila MSL complex activates the transcription of target genes. Genes Dev **19:** 2284-2288.

STRUNNIKOV, V., 1975 Sex control in silkworms. Nature **255:** 111-113.

SUN, L., H. R. FERNANDEZ, R. C. DONOHUE, J. LI, J. CHENG *et al.*, 2012 Male-specific lethal complex in Drosophila counteracts histone acetylation and does not mediate dosage compensation. Proc Natl Acad Sci U S A **110:** E808-817.

SUZUKI, M. G., S. IMANISHI, N. DOHMAE, M. ASANUMA and S. MATSUMOTO, 2010 Identification of a male-specific RNA binding protein that regulates sex-specific splicing of Bmdsx by increasing RNA binding activity of BmPSI. Mol Cell Biol **30:** 5776-5786.

SUZUKI, M. G., S. IMANISHI, N. DOHMAE, T. NISHIMURA, T. SHIMADA *et al.*, 2008 Establishment of a novel in vivo sex-specific splicing assay system to identify a trans-acting factor that negatively regulates splicing of Bombyx mori dsx female exons. Mol Cell Biol **28:** 333-343.

SUZUKI, M. G., F. OHBAYASHI, K. MITA and T. SHIMADA, 2001 The mechanism of sex-specific splicing at the doublesex gene is different between Drosophila melanogaster and Bombyx mori. Insect Biochem Mol Biol **31:** 1201-1211.

SUZUKI, M. G., T. SHIMADA and M. KOBAYASHI, 1998 Absence of dosage compensation at the transcription level of a sex-linked gene in a female heterogametic insect, Bombyx mori. Heredity (Edinb) **81 ( Pt 3):** 275-283.

SUZUKI, M. G., T. SHIMADA and M. KOBAYASHI, 1999 Bm kettin, homologue of the Drosophila kettin gene, is located on the Z chromosome in Bombyx mori and is not dosage compensated. Heredity (Edinb) **82 ( Pt 2):** 170-179.

SUZUKI, M. G., K. SUZUKI, F. AOKI and M. AJIMURA, 2012 Effect of RNAi-mediated knockdown of the Bombyx mori transformer-2 gene on the sex-specific splicing of Bmdsx pre-mRNA. Int J Dev Biol **56:** 693-699.

SUZUKI, Y., D. ISHIHARA, M. SASAKI, H. NAKAGAWA, H. HATA *et al.*, 2000 Statistical analysis of the 5' untranslated region of human mRNA using "Oligo-Capped" cDNA libraries. Genomics **64:** 286-297.

TAN, A., G. FU, L. JIN, Q. GUO, Z. LI *et al.*, 2013 Transgene-based, female-specific lethality system for genetic sexing of the silkworm, Bombyx mori. Proceedings of the National Academy of Sciences of the United States of America **110:** 6766-6770.

TAZIMA, Y., 1954 Mechanisms of the sex-determination in the silkworm, Bombyx mori. In Proceedings of the 9th International Congress of Genetics. Caryologia 6 (suppl.)**:** 958–960.

TE KRONNIE, G., H. STROBAND, H. SCHIPPER and J. SAMALLO, 1999 Zebrafish CTH1, a C3H zinc finger protein, is expressed in ovarian oocytes and embryos. Dev Genes Evol **209:** 443-446.

TEAM, R. D. C., 2010 R: A language and environment for the statistical analysis and computing, pp. Austria: R foundation for statistical computing. Retrieved from http://www.R-project.org., Vienna.

TENENHAUS, C., K. SUBRAMANIAM, M. A. DUNN and G. SEYDOUX, 2001 PIE-1 is a bifunctional protein that regulates maternal and zygotic gene expression in the embryonic germ line of Caenorhabditis elegans. Genes Dev **15:** 1031-1040.

TENG, X., Z. ZHANG, G. HE, L. YANG and F. LI, 2011 Validation of reference genes for quantitative expression analysis by real-time rt-PCR in four lepidopteran insects. J Insect Sci **12:** 60.

TRAUT, W., 1994 Sex determination in the fly Megaselia scalaris, a model system for primary steps of sex chromosome evolution. Genetics **136:** 1097-1104.

TRAUT, W., and F. MAREC, 1996 Sex chromatin in lepidoptera. Q Rev Biol **71:** 239-256.

TRAUT, W., T. NIIMI, K. IKEO and K. SAHARA, 2006 Phylogeny of the sex-determining gene Sex-lethal in insects. Genome **49:** 254-262.

TRAUT, W., K. SAHARA and F. MAREC, 2007a Sex chromosomes and sex determination in Lepidoptera. Sex Dev **1:** 332-346.

TRAUT, W., K. SAHARA and F. MAREC, 2007b Sex chromosomes and sex determination in Lepidoptera. Sexual development : genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation **1:** 332-346.

VAGIN, V. V., A. SIGOVA, C. LI, H. SEITZ, V. GVOZDEV *et al.*, 2006 A distinct small RNA pathway silences selfish genetic elements in the germline. Science **313:** 320-324.

VAN DE ZANDE, L., and E. C. VERHULST, 2014 Genomic imprinting and maternal effect genes in haplodiploid sex determination. Sex Dev **8:** 74-82.

VERHULST, E. C., L. W. BEUKEBOOM and L. VAN DE ZANDE, 2010 Maternal control of haplodiploid sex determination in the wasp Nasonia. Science **328:** 620-623.

VICOSO, B., and D. BACHTROG, 2009 Progress and prospects toward our understanding of the evolution of dosage compensation. Chromosome Res **17:** 585-602.

VICOSO, B., and D. BACHTROG, 2011 Lack of global dosage compensation in Schistosoma mansoni, a female-heterogametic parasite. Genome Biol Evol **3:** 230-235.

VICOSO, B., V. B. KAISER and D. BACHTROG, 2013 Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc Natl Acad Sci U S A **110:** 6453-6458.

WALTERS, J. R., and T. J. HARDCASTLE, 2011 Getting a full dose? Reconsidering sex chromosome dosage compensation in the silkworm, Bombyx mori. Genome Biol Evol **3:** 491-504.

WALTERS, J. R., T. J. HARDCASTLE and C. D. JIGGINS, 2015 Sex Chromosome Dosage Compensation in Heliconius Butterflies: Global yet Still Incomplete? Genome Biol Evol **7:** 2545-2559.

WANG, D., Y. GUO, C. WU, G. YANG, Y. LI *et al.*, 2008 Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice. BMC Genomics **9:** 44.

WANG, E. T., N. A. CODY, S. JOG, M. BIANCOLELLA, T. T. WANG *et al.*, 2012a Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell **150:** 710-724.

WANG, J., M. LONG and M. D. VIBRANOVSKI, 2012b Retrogenes moved out of the z chromosome in the silkworm. J Mol Evol **74:** 113-126.

WANG, W., B. LIU, M. XU, M. JAMIL and G. WANG, 2015 ABA-induced CCCH tandem zinc finger protein OsC3H47 decreases ABA sensitivity and promotes drought tolerance in Oryza sativa. Biochem Biophys Res Commun **464:** 33-37.

WATERMANN, D. O., Y. TANG, A. ZUR HAUSEN, M. JAGER, S. STAMM *et al.*, 2006 Splicing factor Tra2-beta1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene. Cancer Res **66:** 4774-4780.

WELLS, M. L., S. N. HICKS, L. PERERA and P. J. BLACKSHEAR, 2015 Functional Equivalence of an Evolutionarily Conserved RNA Binding Module. J Biol Chem **290:** 24413-24423.

WESSA, P., 2015 Histogram (v1.0.15) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education.

WHARTON, R. P., and G. STRUHL, 1989 Structure of the Drosophila BicaudalD protein and its role in localizing the the posterior determinant nanos. Cell **59:** 881-892.

WILKINS, A. S., 1995 Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. BioEssays : news and reviews in molecular, cellular and developmental biology **17:** 71-77.

WILLHOEFT, U., and G. FRANZ, 1996 Identification of the sex-determining region of the Ceratitis capitata Y chromosome by deletion mapping. Genetics **144:** 737-745.

WILLIAMSON, M. P., 1994 The structure and function of proline-rich regions in proteins. Biochem J **297 ( Pt 2):** 249-260.

WU, H. H., and B. SU, 2008 Adaptive evolution of SCML1 in primates, a gene involved in male reproduction. BMC Evol Biol **8:** 192.

WU, J. Y., and T. MANIATIS, 1993 Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell **75:** 1061-1070.

WU, Y., T. CHENG, C. LIU, D. LIU, Q. ZHANG *et al.*, 2016 Systematic Identification and Characterization of Long Non-Coding RNAs in the Silkworm, Bombyx mori. PloS one **11:** e0147147.

XIA, Q., D. CHENG, J. DUAN, G. WANG, T. CHENG *et al.*, 2007 Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, Bombyx mori. Genome Biol **8:** R162.

XIONG, Y., X. CHEN, Z. CHEN, X. WANG, S. SHI *et al.*, 2010 RNA sequencing shows no dosage compensation of the active X-chromosome. Nat Genet **42:** 1043-1047.

XU, J., S. CHEN, B. ZENG, A. A. JAMES, A. TAN *et al.*, 2017a Bombyx mori P-element Somatic Inhibitor (BmPSI) Is a Key Auxiliary Factor for Silkworm Male Sex Determination. PLoS genetics **13:** e1006576.

XU, J., S. ZHAN, S. CHEN, B. ZENG, Z. LI *et al.*, 2017b Sexually dimorphic traits in the silkworm, Bombyx mori, are regulated by doublesex. Insect biochemistry and molecular biology **80:** 42-51.

YANG, C., S. HUANG, X. WANG and Y. GU, 2015 Emerging Roles of CCCH-Type Zinc Finger Proteins in Destabilizing mRNA Encoding Inflammatory Factors and Regulating Immune Responses. Crit Rev Eukaryot Gene Expr **25:** 77-89.

YOSHIDA, H., S. Y. PARK, T. ODA, T. AKIYOSHI, M. SATO *et al.*, 2015 A novel 3' splice site recognition by the two zinc fingers in the U2AF small subunit. Genes Dev **29:** 1649-1660.

YOSHIMOTO, S., E. OKADA, H. UMEMOTO, K. TAMURA, Y. UNO *et al.*, 2008 A W-linked DM-domain gene, DM-W, participates in primary ovary development in Xenopus laevis. Proc Natl Acad Sci U S A **105:** 2469-2474.

YU, H., J. K. CHEN, S. FENG, D. C. DALGARNO, A. W. BRAUER *et al.*, 1994 Structural basis for the binding of proline-rich peptides to SH3 domains. Cell **76:** 933-945.

YUN, B., R. FARKAS, K. LEE and L. RABINOW, 1994 The Doa locus encodes a member of a new protein kinase family and is essential for eye and embryonic development in Drosophila melanogaster. Genes & development **8:** 1160-1173.

ZHA, X., Q. XIA, J. DUAN, C. WANG, N. HE *et al.*, 2009 Dosage analysis of Z chromosome genes using microarray in silkworm, Bombyx mori. Insect Biochem Mol Biol **39:** 315-321.

ZHANG, S. O., S. MATHUR, G. HATTEM, O. TASSY and O. POURQUIE, 2010 Sex-dimorphic gene expression and ineffective dosage compensation of Z-linked genes in gastrulating chicken embryos. BMC Genomics **11:** 13.

ZHOU, Q. Z., S. M. FANG, Q. ZHANG, Q. Y. YU and Z. ZHANG, 2017 Identification and comparison of long non-coding RNAs in the silk gland between the domestic and wild silkworms. Insect science.

ZHOU, Q. Z., B. ZHANG, Q. Y. YU and Z. ZHANG, 2016 BmncRNAdb: a comprehensive database of non-coding RNAs in the silkworm, Bombyx mori. BMC bioinformatics **17:** 370.

# Publications

# Publications

1. **Gopinath G***, Arunkumar KP*, Mita K and Nagaraju J (2016) Role of *Bmznf-2*, a *Bombyx mori* CCCH zinc finger gene, in masculinisation and differential splicing of *Bmtra-2*. ***Insect Biochemistry and Molecular Biology*** 75: 32-44. (Corresponding authors)

2. Sawanth SK*, **Gopinath G***, Sambrani N* and Arunkumar KP* (2016) The autoregulatory loop: A common mechanism of regulation of key sex determining genes in insects. ***Journal of Biosciences*** 41: 283-294. (*equal contribution)

3. Nagaraju J, **Gopinath G**, Sharma V and Shukla JN (2014) Lepidopteran sex determination: A cascade of surprises. ***Sexual Development*** 8: 104-112

4. **Gopinath G**, Kuchi S, Tomar A, Sekhar SM and Arunkumar KP (2016) RNA sequencing reveals a complete but unconventional dosage compensation in *Bombyx mori* (Under revision, ***Royal Society Open Science***).

5. Ajimura M, Venkatesan S, Shukla JN, **Gopinath G**, Sahara K, Abe H, Tamura T, Shimada T, Xia Q, Arunkumar KP, Goldsmith MR, Mita K, Nagaraju J (2016) W-chromosome linked CCCH and C2H2 zinc-finger motif gene pair required for female development in the silkworm, *Bombyx mori*. (Manuscript under preparation).

Insect Biochemistry and Molecular Biology 75 (2016) 32–44

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

# Role of *Bmznf-2*, a *Bombyx mori* CCCH zinc finger gene, in masculinisation and differential splicing of *Bmtra-2*

Gajula Gopinath [a, **], Kallare P. Arunkumar [a, *], Kazuei Mita [b], Javaregowda Nagaraju [a, †]

[a] Centre of Excellence for Genetics and Genomics of Silkmoths, Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad 500001, India
[b] State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400716, China

ARTICLE INFO

*Article history:*
Received 29 October 2015
Received in revised form
26 May 2016
Accepted 31 May 2016
Available online 1 June 2016

*Keywords:*
*Bombyx mori*
Zinc finger gene
Femaleness
Splicing
Masculinisation
Doublesex
Transformer

ABSTRACT

Deciphering the regulatory factors involved in *Bombyx mori* sex determination has been a puzzle, challenging researchers for nearly a century now. The pre-mRNA of *B. mori doublesex* (*Bmdsx*), a master regulator gene of sexual differentiation, is differentially spliced, producing *Bmdsxm* and *Bmdsxf* transcripts in males and females respectively. The putative proteins encoded by these differential transcripts orchestrate antagonistic functions, which lead to sexual differentiation. A recent study in *B. mori* illustrated the role of a W-derived *fem* piRNA in conferring femaleness. In females, the *fem* piRNA was shown to suppress the activity of a Z-linked CCCH type zinc finger (*znf*) gene, *Masculiniser* (*masc*), which indirectly promotes the *Bmdsxm* type of splicing. In this study, we report a novel autosomal (Chr 25) CCCH type znf motif encoding gene *Bmznf-2* as one of the potential factors in the *Bmdsx* sex specific differential splicing, and we also provide insights into its role in the alternative splicing of *Bmtra2* by using ovary derived BmN cells. Over-expression of *Bmznf-2* induced *Bmdsxm* type of splicing (masculinisation) with a correspondingly reduced expression of *Bmdsxf* type isoform in BmN cells. Further, the site-directed mutational studies targeting the tandem CCCH znf motifs revealed their indispensability in the observed phenotype of masculinisation. Additionally, the dual luciferase assays in BmN cells using 5′ UTR region of the *Bmznf-2* strongly implied the existence of a translational repression over this gene. From these findings, we propose *Bmznf-2* to be one of the potential factors of masculinisation similar to *Masc*. From the growing number of *Bmdsx* splicing regulators, we assume that the sex determination cascade of *B. mori* is quite intricate in nature; hence, it has to be further investigated for its comprehensive understanding.



## 1. Introduction

Sex determination is a fundamental biological process that determines two distinct sexes. A variety of sex determination mechanisms is observed in animal species, most of which follow the chromosomal/genetic sex determination, except in a few cases where the sex is determined by environmental factors like temperature (e.g. crocodiles, alligators and few lizards). Among insects, the mechanism of sex determination is well understood in *Drosophila* and serves as a reference for all insects. In *Drosophila*, (XX is female and XY is male) the sex is determined by the dose of X-linked signalling elements (XSE) (XSE are four transcription factors *Scute*, *SisA*, *Runt* and *Unpaired*) (Erickson and Quintero, 2007), which in turn is determined by the number of X chromosomes. XSE, whose expression threshold can be reached only in female embryos, confines the production of the Sex-lethal (SXL) protein to females. SXL produced in this way directs the female specific splicing of pre-mRNA of *transformer* (*tra*) gene resulting in functional TRA protein. The TRA interacts with non sex specific transformer2 (TRA2) protein and this complex binds to the *doublesex* repeat element (*dsx*RE) in the middle of fourth exon and forces the female specific splicing of *doublesex* (*dsx*) mRNA, producing the female DSX protein. These two proteins have been shown to exhibit antagonistic functions in the process of sexual differentiation (Christiansen et al., 2002). In a few insect species like *Megaselia scalaris* (Traut, 1994), *Ceratitis capitata* (Willhoeft and

* Corresponding author.
** Corresponding author.
*E-mail addresses:* gnath@cdfd.org.in (G. Gopinath), arun@cdfd.org.in (K.P. Arunkumar).
† Deceased.

http://dx.doi.org/10.1016/j.ibmb.2016.05.008
0965-1748/

*Review*

# The autoregulatory loop: A common mechanism of regulation of key sex determining genes in insects

SURESH KUMAR SAWANTH†, GAJULA GOPINATH†, NAGRAJ SAMBRANI† and KALLARE P ARUNKUMAR*

*Centre of Excellence for Genetics and Genomics of Silkmoths, Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad 500 001, India*

**Corresponding author (Email, arun@cdfd.org.in)*

*†These authors contributed equally to the work.*

Sex determination in most insects is structured as a gene cascade, wherein a primary signal is passed through a series of sex-determining genes, culminating in a downstream double-switch known as *doublesex* that decides the sexual fate of the embryo. From the literature available on sex determination cascades, it becomes apparent that sex determination mechanisms have evolved rapidly. The primary signal that provides the cue to determine the sex of the embryo varies remarkably, not only among taxa, but also within taxa. Furthermore, the upstream key gene in the cascade also varies between species and even among closely related species. The order Insecta alone provides examples of astoundingly complex diversity of upstream key genes in sex determination mechanisms. Besides, unlike key upstream genes, the downstream double-switch gene is alternatively spliced to form functional sex-specific isoforms. This sex-specific splicing is conserved across insect taxa. The genes involved in the sex determination cascade such as *Sex-lethal* (*Sxl*) in *Drosophila melanogaster*, *transformer* (*tra*) in many other dipterans, coleopterans and hymenopterans, *Feminizer* (*fem*) in *Apis mellifera*, and *IGF-II mRNA-binding protein* (*Bmimp*) in *Bombyx mori* are reported to be regulated by an autoregulatory positive feedback loop. In this review, by taking examples from various insects, we propose the hypothesis that autoregulatory loop mechanisms of sex determination might be a general strategy. We also discuss the possible reasons for the evolution of autoregulatory loops in sex determination cascades and their impact on binary developmental choices.

[Sawanth SK, Gopinath G, Sambrani N and Arunkumar KP 2016 The autoregulatory loop: A common mechanism of regulation of key sex determining genes in insects. *J. Biosci.*] **DOI** 10.1007/s12038-016-9609-x

## 1. Introduction

Sex determination is a complex developmental program that involves the fine-tuned action of numerous genes required to direct the developing embryo to either a male or female pathway. It is an essential and universal phenomenon among a majority of metazoans. Organisms show an astounding diversity in the mechanisms regulating sex determination. The primary signal that provides the cue for determining the sexual fate of an organism varies remarkably not only among taxa but also within taxa, suggesting rapid evolution of sex-determining mechanisms. Insects have evolved a variety of primary signals in sex determination pathways. A cascade of genes act upon one another to carry the information from primary signal to terminal differentiation. The three important components of the sex determination cascade are primary signal, key gene and terminal double-switch gene. The primary signal and the key gene vary across different insects but the terminal gene is highly conserved (Graham *et al.* 2003). In the presence or absence of a primary signal, a few splicing regulators interact and lead to regulated splicing in one sex and default splicing in the other. The terminal double-switch gene transcripts so formed are sex-specific and responsible for secondary sexual characters.

**Keywords.** Autoregulation; convergent evolution; doublesex; sex determination; splicing; transformer

Sexual Development

Sex Dev
DOI: 10.1159/000357483

Published online: January 21, 2014

# Lepidopteran Sex Determination: A Cascade of Surprises

J. Nagaraju[a] G. Gopinath[a] V. Sharma[a] J.N. Shukla[b]

[a]Centre of Excellence for Genetics and Genomics of Silkmoths, Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; [b]Department of Entomology, College of Agriculture, University of Kentucky, Lexington, Ky., USA

**Key Words**
Alternative splicing · *Bombyx mori* · Doublesex · Sex determination · Transformer

**Abstract**
Sex determination is a developmental pathway that fixes the sexual fate (male or female) of an individual at early stages of embryonic development. This pathway is ideally suited for evolutionary studies given the astoundingly diverse mechanisms found in the animal kingdom. In particular, insects use multiple different cues to specify the sexual fate of an individual. In this review, we focus on genes and genetic interactions involved in the sex determination of insect species belonging to the order Lepidoptera. Unique features of the lepidopteran sex determination system are discussed.



Global crop losses by insects are estimated to be 13% per annum despite the usage of multiple pesticides. The insect order Lepidoptera (Lepido = scales, ptera = wing) includes a major fraction of crop pests and is divided into 4 suborders: Zeugloptera, Aglossata, Heterobathmiina, and Glossata, containing 15,578 genera and 157,424 species [van Nieukerken et al., 2011]. Lepidopterans undergo complete metamorphosis; life cycles contain 3 distinct stages, i.e. caterpillar/larva, chrysalis/pupa, and moth/adult stages. Generally, larval stages are crop destroyers that include defoliators, shoot/root borers, and seed predators causing a significant agricultural loss. Use of pesticides is not eco-friendly, and risk of developing resistance is an alarming threat from insects of this order. Strategies utilizing the release of sterile males, e.g. sterile insect technique, is a promising approach for pest control. Sterile insect technique achieved success in controlling the New World screwworm fly, Mediterranean fruit fly, and melon fly [Lindquist et al., 1992; Hendrichs et al., 1995; Koyama et al., 2004]. The productive female sex of insects could be targeted using modern techniques of genetic engineering, e.g. release of insects carrying a dominant lethal gene or genetic system [Thomas et al., 2000; Alphey and Andreasen, 2002; Jin et al., 2013]. A major improvement in sterile insect technique and its application can be envisaged by an understanding of molecular mechanisms

This article is dedicated to Dr. Jawaregowda Nagaraju, who unexpectedly passed away during the preparation of this paper. G.G. and V.S. contributed equally to this paper.

KARGER

E-Mail karger@karger.com
www.karger.com/sxd


1661–5425/14/0000–0000$39.50/0

Jayendra Nath Shukla
Department of Entomology, College of Agriculture, University of Kentucky
S-225 Agriculture Science Center North
Lexington, KY 40546 (USA)
E-Mail jayendrashukla@gmail.com

# Report of Plagiarism check

# CHARACTERIZATION OF SEX DETERMINATION GENES IN LEPIDOPTERANS

*by* Gajula Gopinath

| | | | |
|---|---|---|---|
| FILE | 10LBPH02.PDF (4.3M) | | |
| TIME SUBMITTED | 29-JUN-2016 12:52PM | WORD COUNT | 24627 |
| SUBMISSION ID | 687016140 | CHARACTER COUNT | 134867 |

# CHARACTERIZATION OF SEX DETERMINATION GENES IN LEPIDOPTERANS

ORIGINALITY REPORT

| 10% | 4% | 8% | 1% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | www.sdbonline.org<br>Internet Source | 1% |
| 2 | spandidos-publications.com<br>Internet Source | <1% |
| 3 | Masataka T. Suzuki. "Sex determination: insights from the silkworm", Journal of Genetics, 09/2010<br>Publication | <1% |
| 4 | Jiang, X., J. K. Biedler, Y. Qi, A. B. Hall, and Z. J. Tu. "Complete dosage compensation in Anopheles stephensi and the evolution of sex-biased genes in mosquitoes", Genome Biology and Evolution, 2015.<br>Publication | <1% |
| 5 | Parsch, John, and Hans Ellegren. "The evolutionary causes and consequences of sex-biased gene expression", Nature Reviews Genetics, 2013.<br>Publication | <1% |
| 6 | Walters, James R., Thomas J. Hardcastle, and Chris D. Jiggins. "Sex chromosome | <1% |

dosage compensation in Heliconius butterflies: global yet still incomplete?", Genome Biology and Evolution, 2015.
Publication

| | | |
|---|---|---|
| 7 | Jayendra Nath Shukla. "Novel female-specific splice form of dsx in the silkworm, Bombyx mori", Genetica, 08/18/2010<br>Publication | <1% |
| 8 | Warke, Vishal G., Sandeep Krishnan, Madhusoodana P. Nambiar, Donna L. Farber, George C. Tsokos, and Henry K. Wong. "IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES IN HUMAN MEMORY (CD45RO+) CD4+ T LYMPHOCYTES", Immunological Investigations, 2001.<br>Publication | <1% |
| 9 | Cheng, Tingcai, Bohua Fu, Yuqian Wu, Renwen Long, Chun Liu, and Qingyou Xia. "Transcriptome Sequencing and Positive Selected Genes Analysis of Bombyx mandarina", PLoS ONE, 2015.<br>Publication | <1% |
| 10 | M. G. Suzuki. "Identification of a Male-Specific RNA Binding Protein That Regulates Sex-Specific Splicing of Bmdsx by Increasing RNA Binding Activity of BmPSI", Molecular and Cellular Biology, 12/15/2010<br>Publication | <1% |
| 11 | Uebbing, S., A. Kunstner, H. Makinen, and H. | |

| | | |
|---|---|---|
| | Ellegren. "Transcriptome Sequencing Reveals the Character of Incomplete Dosage Compensation Across Multiple Tissues in Flycatchers", Genome Biology and Evolution, 2013.<br>Publication | <1% |
| 12 | Barbosa, Cristina, Isabel Peixeiro, and Luísa Romão. "Gene Expression Regulation by Upstream Open Reading Frames and Human Disease", PLoS Genetics, 2013.<br>Publication | <1% |
| 13 | Sawanth, Suresh Kumar, Gajula Gopinath, Nagraj Sambrani, and Kallare P Arunkumar. "The autoregulatory loop: A common mechanism of regulation of key sex determining genes in insects", Journal of Biosciences, 2016.<br>Publication | <1% |
| 14 | mro.massey.ac.nz<br>Internet Source | <1% |
| 15 | g3journal.org<br>Internet Source | <1% |
| 16 | ecommons.usask.ca<br>Internet Source | <1% |
| 17 | Submitted to University of Seoul<br>Student Paper | <1% |
| 18 | Advances in Experimental Medicine and Biology, 2013. | <1% |