

**ANALYSES OF SEX CHROMOSOME DOSAGE COMPENSATION IN
THE CODLING MOTH, *CYDIA POMONELLA*
(LEPIDOPTERA:TORTRICIDAE) YIELD INSIGHTS INTO SEX
CHROMOSOME EVOLUTION IN LEPIDOPTERA**

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Liuqi Gu

May 2015

© 2015 Liuqi Gu

ANALYSES OF SEX CHROMOSOME DOSAGE COMPENSATION IN THE CODLING MOTH,
CYDIA POMONELLA (LEPIDOPTERA:TORTRICIDAE) YIELD INSIGHTS INTO SEX
CHROMOSOME EVOLUTION IN LEPIDOPTERA

Liuqi Gu, Ph.D.

Cornell University 2015

Dosage problems of sex-linked genes accompany degenerating sex chromosomes, which animals deal with differently. Species with male heterogametic sex chromosomes (XX/XY) employ different chromosome-wide mechanisms to equalize X-linked gene expression between sexes (X~XX compensation: $X \approx XX$). Lepidopterans (moths and butterflies) have female heterogametic sex chromosomes (ZZ/ZW). Recognition of dosage compensation is controversial in this insect order due to differences in sequencing platforms, methods of analyses and tissues sampled. Among the other ZZ/ZW taxa examined so far, however, it is unanimously reported that overall Z-linked gene expression is reduced in females compared to males or the autosomes ($Z < ZZ \approx AA$), reflection of lack of global dosage compensation.

In this study, I assessed dosage compensation in the codling moth *Cydia pomonella* (Tortricidea), a basal moth species compared to the other lepidopterans that have been examined for dosage compensation. *C. pomonella* has a Z chromosomal segment (neo-Z) that arose from an autosome translocating to the ancestral Z chromosome (ancl-Z). I assembled the *C. pomonella* transcriptome *de novo* from RNAseq data and inferred the chromosome locations of identified transcripts by comparison to the *Bombyx mori*

reference genome. These data showed that in *C. pomonella* overall Z-linked gene expression was balanced between sexes but reduced relative to that of the autosomes ($Z \approx ZZ < AA$). The Z~ZZ compensation was complete on both the ancl-Z and neo-Z chromosomes in the head and midgut, suggesting the existence of a global compensation mechanism. The extent of Z~A compensation varied among tissues, being least in the ovary and most in male reproductive tissues. Comparison of the *C. pomonella* transcriptome with those of other Lepidopteran species with conserved karyotypes confirmed both complete Z~ZZ compensation and imperfect Z~A compensation in this species. Further, I found that the skewed gene content of the Z chromosome did not confound the analysis of dosage compensation. The pattern of dosage compensation in *C. pomonella* is consistent with that described for several other Lepidopteran species. However, a different pattern reported in *Plodia interpunctella* may reflect the use of whole body transcriptomes or problems with data processing in that study.

The dosage compensation pattern in Lepidoptera mirrors that in mammals and suggests the inactivation of one Z chromosome copy in male moths. This is the first report of dosage compensation analysis of a neo-sex chromosome in a female heterogametic (ZZ/ZW) species, and presents the Lepidoptera as a unique taxon to broaden our understanding of dosage compensation and sex chromosome evolution.

BIOGRAPHICAL SKETCH

Liuqi, also known as Aloy, was born as the only child to Zhongguo and Mifang, who works with a local rice wine brewery. He grew up without affluence but with a lot of love yet discipline. His versatile talents and autonomous obsession on art and science since grew with him since childhood. Liuqi's hometown, Zhoushan is well-known as the 'Thousand-Island City' in China. It is the largest Chinese archipelago comprised of nearly 2000 islands situated in the East China Sea. It is also the largest fishery in the country, fourth largest in the world as well as a renowned Buddhism attraction. The exquisite seafood and enjoyable island life did not withhold Liuqi's craving for the outside world. After graduation from high school, he travelled half the country to Beijing, the hub of rich cultures around the nation. He first studied entomology in China Agricultural University for bachelor's degree and then got master's degree from Chinese Academy of Sciences. In the summer of 2009, Liuqi travelled again, this time half the globe to United States of America, the melting pot of world cultures. At Cornell University, he enrolled in the Ph.D. program in the field of entomology. He is interested in pursuing a career in academia after graduation.

Liuqi's passion toward life and the world surrounding it brings along his broad spectrum of 'delights', from sports, sciences, humanities to philosophy. He is also open-minded and genuine. He always sees the good side in people and believes where there is sincerity, there is love. Because of that, he becomes a great asset of friendship to those who can bear his intense personality and high-maintenance.

This dissertation is fondly dedicated to my endearing parents, without whose unconditional love, understanding and support I would not be able to sail afar from an island amidst East China Sea to the Cayuga waters - in upstate New York.

仅以此篇献给我亲爱的爸妈。从祖国东海之滨的千岛之都到美国纽约指湖边的倚色佳小镇，唯有你们无私的奉献与默默的支持，儿得以重洋远越在世界名校摘得最高的学位桂冠。

ACKNOWLEDGMENTS

First of all, I cannot thank Dr. Douglas Knipple, Dr. David Soderlund and Dr. Ping Wang enough for serving on my committee. It is their meticulous mentoring that has brought my immense progress in the past half-decade. Doug has been a great supervisor and friend who supports me in so many ways. Under his encouragement I enjoy the freedom of exploring the unknown realm of science. Dave and Ping are among the most intelligent, scrupulous yet congenial scholars I have met. In particular, Dave dedicated much of his time during my finishing period to tutoring me in scientific writing, a key skill to becoming a qualified scientist. Ping is always eager to share his passion and curiosity in science and triggers interesting discussions. Often times I am enlightened by his insightful thoughts toward solving problems.

Next, I would like to thank Dr. James Walters at University of Kansas for his guidance in data analysis and encouragement in learning the R language. Our impassioned discussions on dosage compensation mutually advanced our understandings toward this cumbersome subject matter. I thank Dr. Qi Sun from Institute of Biotechnology at Cornell. His patient instruction was particularly helpful when I first shifted steers towards bioinformatics. I also thank Dr. Tom Burr, our most charismatic director of the Geneva Experiment Station. I appreciate his continuous support during and beyond my presidency of SAGES (Student Association of Geneva Experiment Station), which brightened my extracurricular student life in Geneva.

My cordial thanks also to my beloved friends at Cornell. Xiaozhao Song, Gaoyan Wang and Yang Bai, the all-girl gang of my fellow Chinese students in Geneva, during our Ph.D. years here, we shared our sweetness and bitterness together and helped each other in so many ways. Summer scholar Stephan Ireland from Michigan University and Chinese visiting student Jinda Wang from Nanjing Agricultural University both in our lab, they were my comrades with whom I shared benches as well as company when pursuing knowledge. My fellow students Dr. John Gottula from Plant Pathology and Dr. Kaixiong Ye from Nutritional Science, they helped me to jump start my 'wet' and 'dry' experiments, respectively. The Larsons (David, Beth, Hannah, Joe, Dan, Phil, Jesse) have become my 'adopted' American family since I first arrived at Cornell. They are the reason that my trips back to Ithaca always feel like going back home. Annetta Fotopoulos, a Ph.D candidate in Asian Studies, is the beloved sister that I never had. She is a complement to my unfulfilled desires in humanities and our many nightlong philosophical discussions are my sweet retreats from research. There are so many other wonderful people who helped me along the way and precious friends I made both on and off campus who makes my 'alien' life just as homey. Although I cannot enumerate every one of them here, I count myself as lucky and feel grateful with all my heart and soul.

Last but not the least, I dedicate my greatest appreciation and honor to my mom and dad. It is their relentless support and understanding that has made possible who I have become and what I have achieved today.

TABLE OF CONTENTS

	Biographical sketch	iii
	Acknowledgements	v
	List of figures	viii
	List of tables	x
	List of abbreviations	xi
Chapter I	Review of the literature and statement of the problem	1
Chapter II	<i>De novo</i> transcriptome assembly and gene expression profiling of <i>Cydia pomonella</i> adult tissues	14
Chapter III	Analysis of dosage compensation in <i>Cydia pomonella</i> based on <i>de novo</i> transcriptome data	29
Chapter IV	Analysis of dosage compensation in <i>Cydia pomonella</i> using a comparative transcriptome approach	46
Chapter V	Nonrandom genomic distribution of tissue-specific genes in <i>Cydia pomonella</i> : implications for dosage compensation	60
Chapter VI	Summary and conclusions	71
	References	78

LIST OF FIGURES

Fig. 1	Sex chromosome constitutions of two heterogametic systems.	2
Fig. 2	Models of dosage compensation in XX/XY systems.	4
Fig. 3	Summary of dosage compensation studies in Lepidoptera.	10
Fig. 4	'Ancl-Z' and 'neo-Z' chromosomes in <i>C. pomonella</i> .	12
Fig. 5	The male reproductive tissues of <i>C. pomonella</i> .	17
Fig. 6	Expression levels (\log_2 transformed FPKM) of all genes expressed (FPKM>0) in each tissue by chromosomal locations.	22
Fig. 7	Heat map of DE genes in <i>C. pomonella</i> tissues (FC > 1, FDR<0.05, <i>p</i> value for FDR<0.05).	24
Fig. 8	Z~ZZ compensation indicated by female over male expression (\log_2 transformed F:M FPKM).	33
Fig. 9	Z(Z):AA median expression ratios in <i>C. pomonella</i> tissues.	37
Fig. 10	Distribution of expression levels (\log_2 transformed FPKM) of expressed genes (FPKM>0) in seven tissues by chromosomal locations.	38
Fig. 11	Z~ZZ compensation of <i>C. pomonella</i> neo-Z genes using a comparative transcriptome analysis.	51
Fig. 12	Z~A compensation of <i>C. pomonella</i> neo-Z genes using a comparative transcriptome analysis.	53
Fig. 13	Median expression ratios relative to the autosomes of ancl-Z and neo-Z orthologous pairs identified in <i>C. pomonella</i> and <i>M. sexta</i> .	54

Fig. 14	Comparison of distributions of expression levels between neo-Z(Z) and neo-Z(Z).	55
Fig. 15	Genomic distributions of tissue-specific genes in <i>C. pomonella</i> .	63
Fig. 16	Expression levels (\log_2 transformed FPKM) of subset of genes in each tissue by chromosomal linkage.	65
Fig. 17	Z(Z):AA median expression ratios in <i>C. pomonella</i> tissues disregarding all tissue-specific genes.	67

LIST OF TABLES

Table 1	Summary of the total number of genes identified as expressed (FPKM > 0) in each tissue.	21
Table 2	Significance of differences in average expression between autosomes, ancl-Z and neo-Z across 7 tissues, assessed by Bonferroni corrected Mann-Whitney <i>U</i> test (two-sided).	35
Table 3	Significance of differences in distribution of expression levels between autosomes, ancl-Z and neo-Z across 7 tissues, assessed by Benjamini-Hochberg corrected Komolgorov-Smirnov test.	40

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
DE genes	differentially expressed genes (identified by EdgeR)
FDR	false discovery rate
FC	fold change
FPKM	fragment per kilobase per million mappable reads
F/M	female/male
MSCI	meiotic sex chromosome inactivation
MYA	million years ago
MYR	million years
mRNAseq	High throughput mRNA sequencing
PAR	pseudoautosomal region
RBH	reciprocal best hit
TMM	Trimmed Mean of <i>M</i> -values

Chapter I

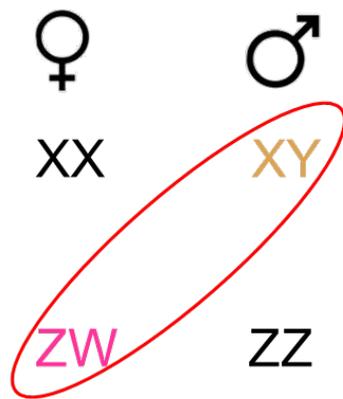
Review of the Literature and Statement of the Problem

Sex chromosome evolution and dosage compensation

In the animal kingdom, there are two predominant sex chromosome constitutions with opposing patterns. In XX/XY species, females are the homogametic sex (XX) and males are the heterogametic sex (XY) whereas in ZZ/ZW species, males are the homogametic sex (ZZ) and females are the heterogametic sex (ZW) (Fig. 1). Sex chromosomes evolved from autosomes (proto-sex chromosomes) independently many times among different lineages (Charlesworth, Charlesworth et al. 2005). Acquisition of the sex-determining function initially suppressed the recombination between proto-sex chromosomes in one sex, which resulted in progressive genetic degeneration of the chromosome copy bearing the sex-determining allele (Y or W), thus creating heteromorphic sex chromosome schemes in the heterogametic sex.

Most present-day Y and W chromosomes harbor very few, if any, functional genes (Berlin and Ellegren 2006, Berlin, Tomaras et al. 2007, Vitkova, Fukova et al. 2007). As a result, the sex-linked genes are effectively monoallelic in the heterogametic sex, which creates an inherent copy number (dosage) imbalance for genes residing on the sex chromosomes between sexes compared to the autosomal genes. Gene products (RNAs and proteins) are

A



B

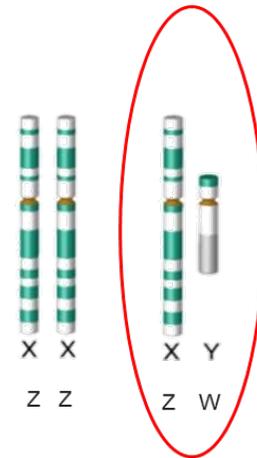


Fig. 1 Sex chromosome constitutions of two heterogametic systems. (A): In XX/XY species, such as flies, males are the heterogametic sex (XY); whereas in ZZ/ZW species, such as moths, females are the heterogametic sex (ZW) (both circled in red). (B) The Y or W chromosomes in the heterogametic sex are usually heavily heterochromatinized and bear few function genes. As a result, the sex chromosomes are effectively monoallelic in the heterogametic sex.

made in direct proportion to gene dosage. Although buffering mechanisms in the gene network may tolerate monoallelism for individual genes, the loss of hundreds of genes on a chromosome can substantially impair the health of individuals due to haplo-insufficiency.

In the course of sex chromosome evolution, dosage compensation evolved concomitantly to counteract the dosage problem (Charlesworth 1978). The two aspects of dosage compensation involve: 1) equalizing sex-linked gene expression between the male (XY or ZZ) and female (XX or ZW); and 2) balancing the gene expression from the sex chromosome (X(X) or Z(Z)) with that of autosomes (AA). For sake of simplicity, in this dissertation I refer to these two aspects of dosage compensation as X~XX | Z~ZZ compensation and X~A | Z~A compensation, respectively.

Dosage compensation models in XX/XY species

Species deploy different strategies to achieve the two aspects of dosage compensation. Among three extensively-studied XX/XY taxa (i.e. flies, mammals and nematodes), X~XX compensation involves orchestrated chromosome-wide regulation (Ferrari, Alekseyenko et al. 2014) (Fig. 2). In *Drosophila* the Dosage Compensation Complex mediates a two-fold hyper-transcription of the entire single X chromosome in the male, which equalizes its output relative both to that of the XX female and to the autosomes (AA) (Gelbart and Kuroda 2009, Lavery, Lucci et al. 2010, Conrad and Akhtar 2012). This process achieves complete X~XX compensation while obviating the need for- X~A compensation. By contrast, in mammals and *Caenorhabditis elegans* the homogametic sex (XX females) is

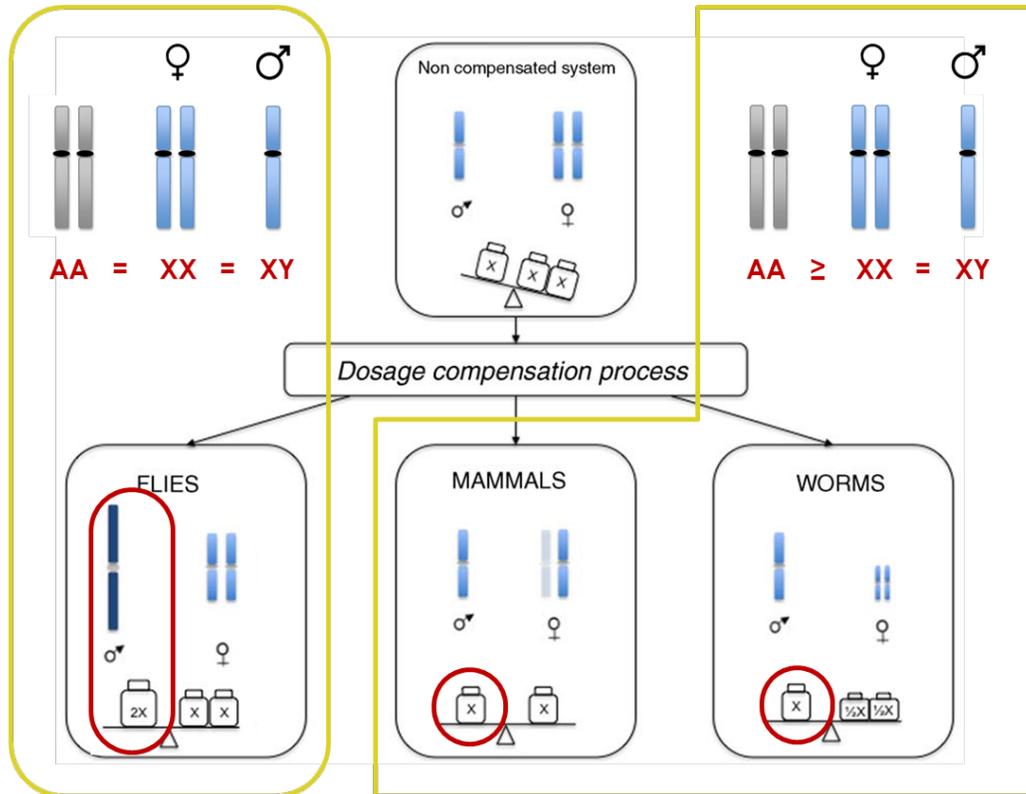


Fig. 2 Models of dosage compensation in XX/XY systems. In flies, the single X chromosome copy in male is hyper-transcribed by two-folds. In mammals, one of the two X chromosome copies in female is inactivated. In worms, both X chromosome copies in female are repressed by half. Adopted from (Lavery, Lucci et al. 2010) with modification.

targeted for X~XX compensation instead. In the female soma of placental mammals, one of the two X chromosome copies is mostly inactivated, a process known as X chromosome inactivation (Schulz and Heard 2013). In *C. elegans* hermaphrodites (XX), the overall gene expression from both X chromosome copies is reduced by approximately half (Meyer 2000, Meyer 2010). Both mechanisms result in effective monosomy of the X chromosome in both sexes, necessitating X~A compensation. Whereas full X~A compensation is evident in *C. elegans* (Deng, Hiatt et al. 2011), it is less so in the mammals. Ohno (1967) proposed that as the X-linked genes became monoallelic in the male as a result of the evolving sex chromosomes, up-regulation of X-linked genes (i.e. the X~A compensation) occurred in both sexes as the first step to match the transcriptional output from the diploid complement of autosomes, which subsequently drove the mammalian X chromosome inactivation as the solution to otherwise over-transcription of X-linked genes in the female. Ohno's theory was widely accepted for nearly half a century but could not be tested until the tools for genome-wide analysis became available recently.

The first wave of studies based on microarray data validated Ohno's hypothesis with regard to two-fold hyper-transcription of the single active X chromosome in mammals (Gupta, Parisi et al. 2006, Nguyen and Disteché 2006, Talebizadeh, Simon et al. 2006, Lin, Gupta et al. 2007, Sugimoto and Abe 2007). Several years later, the first study using RNAseq, which is considered superior to microarray analysis in estimating gene expression, reported a X(X):AA ratio ~0.5, calling Ohno's hypothesis into question (Xiong, Chen et al. 2010). Soon a spate of fierce debates followed, with some arguing in favor of compensatory up-regulation while others arguing against it (Deng, Hiatt et al. 2011, Kharchenko, Xi et al.

2011, Lin, Halsall et al. 2011, Yildirim, Sadreyev et al. 2012). On one hand, the root of these controversies lies in that assessing X~A compensation is more challenging than assessing X~XX compensation because it involves comparing the expression levels of two different sets of genes (autosomal versus X-linked), rather than the same set of X-linked genes that are both expressed in males and females. In fact, the present-day X chromosome has evolved new gene content that is quite different from the proto-X (Vicoso and Charlesworth 2006, Deng, Hiatt et al. 2011, Julien, Brawand et al. 2012). Thus X~A compensation should not be expected for the new genes that are gained after X-Y differentiation (Zhang, Vibranovski et al. 2010). On the other hand, conclusions drawn from different studies on mammalian X~A compensation using RNAseq are biased by the analytical approach employed (Jue, Murphy et al. 2013). In particular, many studies arbitrarily set different filtering thresholds to exclude data based on expression levels, which substantially biases the results. This is not only because functional genes expressed at any level may be subject to compensation, but also because expression estimates based on RNAseq data reflects relative rather than absolute gene expression levels. The measuring units vary among different transcriptome assemblies with nature of RNA source, sequencing platform, sequencing depth, and many other factors (Tarazona, García-Alcalde et al. 2011). Compilation of evidence from the most recent studies where both aspects mentioned above are properly considered and addressed (Julien, Brawand et al. 2012, Lin, Xing et al. 2012, Pessia, Makino et al. 2012) shows that only a minority of dosage-sensitive X-linked genes, mostly those encoding large protein complexes, are compensated for dosage while others are not. Although mechanisms underlying X~A compensation remain to be elucidated,

current evidence suggests that these mechanisms operate on a gene-by-gene basis, in contrast to the chromosome-wide X~XX compensation mechanisms at work.

Dosage compensation in ZZ/ZW systems

In contrast to the extent of dosage compensation research in XX/XY systems, investigations of ZZ/ZW systems are limited. Nevertheless, a different dosage compensation pattern has been reported in a broad spectrum of taxa where ZZ/ZW sex chromosomes have independently evolved. In birds (Ellegren, Hultin-Rosenberg et al. 2007, Itoh, Melamed et al. 2007, Itoh, Replogle et al. 2010, Wolf and Bryk 2011, Adolfsson and Ellegren 2013, Uebbing, Kunstner et al. 2013), *Schistosoma mansoni* (a trematode parasite) (Vicoso and Bachtrog 2011), snakes (Vicoso, Emerson et al. 2013) and *Cynoglossus semilaevis* (a flatfish) (Chen, Zhang et al. 2014)), overall gene expression from the Z chromosome is comparable to that from autosomes in males (ZZ) but much lower than in females (ZW), i.e., $Z < ZZ \approx AA$, reflecting the dosage imbalance between the sexes. In these species only a subset of female Z genes are locally regulated and a global dosage compensation mechanism across the entire chromosome is absent. This model of dosage compensation is considered to be 'incomplete', as opposed to the complete balance of expression at least between sexes in XX/XY animals.

Sex chromosome systems and dosage compensation in Lepidoptera

The Lepidoptera (moths and butterflies) includes many species of great economic value, such as the silkworm, as well as many important worldwide agricultural pests. It represents a very ancient female heterogametic taxon and is the only insect order to have a ZZ/ZW sex chromosomes system (Sahara, Yoshido et al. 2012). Despite over 150 million years of evolution, the Z chromosome is homologous among Lepidopteran lineages and its content is highly conserved (Sahara, Yoshido et al. 2012), much like that of the Z chromosome of birds or the X chromosome of mammals (Marshall Graves 2008, Graves 2014). In Lepidoptera the W chromosome occurs only in more derived lineages whereas basal species have a Z/ZZ constitution (Marec, Sahara et al. 2010). The W chromosome is replete with repetitive elements and has no genetic exchange with the Z chromosome. Thus even in those species with the ZW/ZZ constitution the sex chromosomes are virtually hemizygous in the female.

Information on dosage compensation in Lepidoptera is limited, and the few available studies have led to some controversy. Conclusions drawn from in early reports (Zha, Xia et al. 2009, Harrison, Mank et al. 2012) contributed to the popular speculation that incomplete dosage compensation might be a universal feature associated with female heterogamety (Graves and Disteché 2007, Mank 2009, Vicoso and Bachtrog 2009, Naurin, Hansson et al. 2010, Wolf and Bryk 2011, Mank 2013). The first study used microarray data to examine dosage compensation in the silk moth *Bombyx mori*. Results from that analysis showed the same 'incomplete' dosage compensation pattern ($Z < ZZ \approx AA$) reported in other ZZ/ZW systems (Zha, Xia et al. 2009). However this work was subsequently criticized

for lack of normalization; when properly normalized, the data actually support a $Z \approx ZZ < AA$ model (Walters and Hardcastle 2011). In a second moth species, the Indian meal moth *Plodia interpunctella*, the pattern of dosage compensation was also 'incomplete' in adult whole body (Harrison, Mank et al. 2012). These authors also pointed out that the previous microarray data from *B. mori* was itself too noisy to be compelling. Two more recent dosage compensation studies, involving *Manduca sexta* (Smith, Chen et al. 2014) and two *Heliconius* species (Walters et al, unpublished data), employed RNAseq data and provided new evidence supporting the $Z \approx ZZ < AA$ model as described in *B. mori* by Walters and Hardcastle (2011).

It is nevertheless worth noting that all these studies differ in the tissues sampled, analysis platforms (microarray vs RNAseq) or methods of normalization (Fig. 3). Thus a broader sampling of species and a more uniform platform of analysis are required to shed further light on dosage compensation in Lepidoptera and to resolve some of the controversy in this subject. In addition, since some studies have found the equal expression of sex-linked genes between males and females in Lepidoptera, which was previously described only in XX/XY systems, extended knowledge from this unique ZZ/ZW taxon should provide unique insights into understanding the evolutionary process of dosage compensation as well as sex chromosome evolution in general.

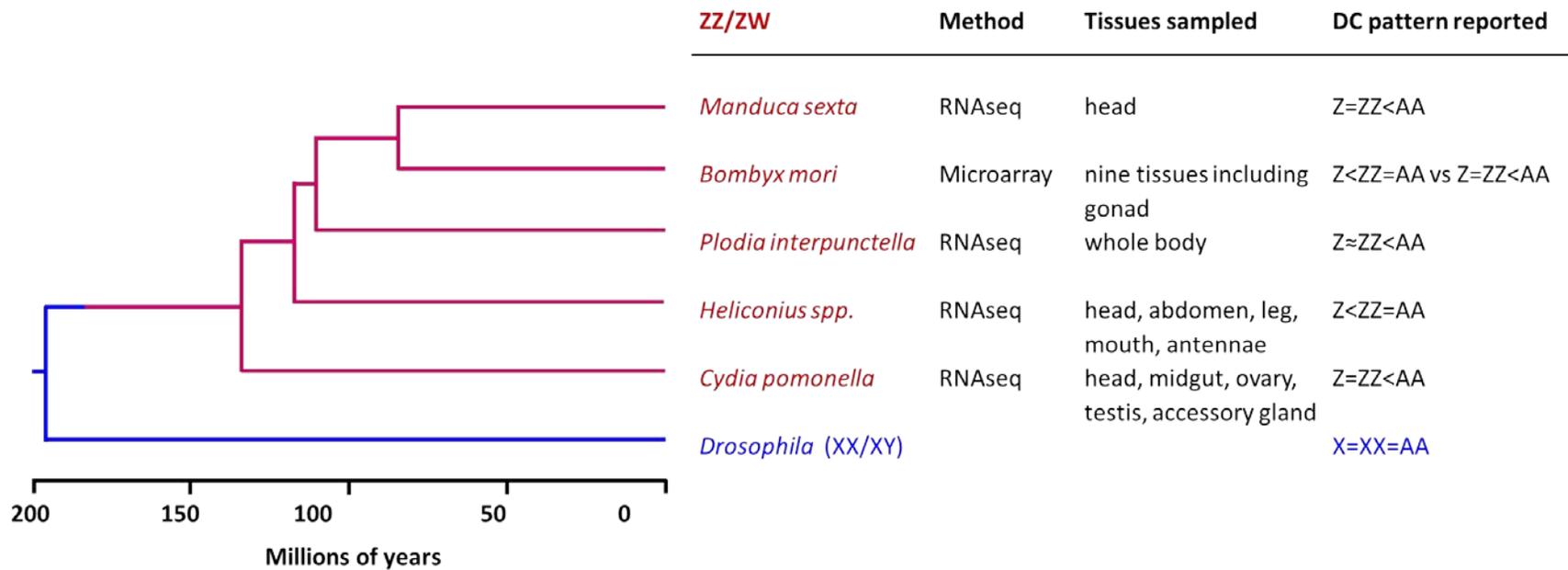


Fig. 3 Summary of dosage compensation studies in Lepidoptera (see text). Phylogenetic relationship among insect species mentioned in this study is based on Marec, Sahara et al. (2010) and Wahlberg, Wheat et al. (2013).

Neo-Z chromosome in Tortricids

Cydia pomonella is a worldwide pest infesting apples and other pome fruit. It is a good system for assessing lepidopteran dosage compensation because it represents a more basal phylogenetic group (Tortricidae) than others that have been examined for dosage compensation so far (Fig. 3). More importantly, *C. pomonella* has a secondary Z chromosome (neo-Z) that arose from an ancient translocation event between the ancestral Z chromosome (ancl-Z) and an autosome that is homologous to Chromosome 15 in *B. mori* (Nguyen, Sykorova et al. 2013) (Fig. 4). The Z-autosome fusion occurred in a common ancestor of the main tortricid subfamilies (Olethreutinae and Tortricinae) encompassing 97% of tortricids and about 700 pest species worldwide. Although the *C. pomonella* W chromosome is equally large in size as its Z partners, molecular and cytogenetic evidence show that it is extensively degenerate as in other Lepidoptera (Fukova, Traut et al. 2007, Nguyen, Sykorova et al. 2013, Sichova, Nguyen et al. 2013). Therefore, genes of autosomal origin on the neo-Z segment are brought under the same sex-linked inheritance, with the accompanying dosage problem, as the genes on the ancl-Z segment.

Examination of dosage compensation in *C. pomonella* presents the unique opportunity not only to address the question of dosage compensation in Lepidoptera from a more basal lineage, but also to analyze the evolution of gene expression on the neo-Z in response to sex-specific selective pressure.

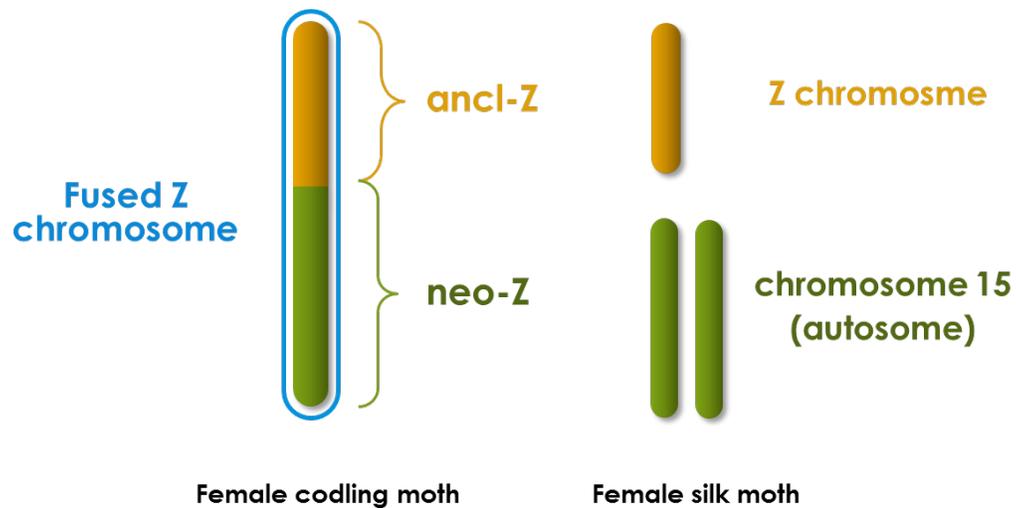


Fig. 4 'ancI-Z' and 'neo-Z' chromosomes in *C. pomonella* are homologous to the Z chromosome and Chromosome 15 in *B. mori*, respectively. Z chromosomes are hemizygotic in the females; W chromosomes contain no functional genes and are thus not drawn. Note that lepidopteran chromosomes are holocentric (Wolf 1996).

Statement of the problem and experimental approach

In this study I attempted to resolve the inconsistency in the status of dosage compensation in Lepidoptera and to gain further insights into dosage compensation and sex chromosome evolution. To that end, I used the codling moth *C. pomonella* as a model to study dosage compensation and evolution of sex-linked gene expression by employing a comprehensive protocol in combination with bioinformatic tools that allows comparisons with other relevant studies.

I first performed *de novo* transcriptome assembly and differential expression analysis among 7 *C. pomonella* adult tissue based on RNAseq data (Chapter II). I then assessed dosage compensation in *C. pomonella* using the transcriptome data and compared the patterns observed with those in other species. In particular, I examined and critiqued the inconsistent pattern of dosage compensation suggested in the *P. interpunctella* study (Harrison, Mank et al. 2012) (Chapter III). Next, I used comparative transcriptome analysis to confirm the dosage compensation pattern on the neo-Z, especially the ambiguity in Z~A compensation, in the head (Chapter IV). Furthermore, I analyzed the nonrandom genomic distribution of tissue-specific genes and addressed whether it confounded the tissue-specific Z~A compensation patterns identified earlier (Chapter V). Lastly, I summarize the conclusions drawn from the analyses and discuss my findings in the context of general questions regarding dosage compensation as well as sex chromosome evolution (Chapter VI).

Chapter II

***De novo* transcriptome assembly and gene expression profiling of *Cydia pomonella* adult tissues**

Introduction

Previous studies on dosage compensation in Lepidoptera differ in the tissue samples used, the experimental platforms employed to obtain transcriptome data (microarray versus RNAseq), and the methods used to normalize transcript levels (Fig. 1). In particular, the *P. interpunctella* study (Harrison, Mank et al. 2012) used the whole body of adult insects, which contains a substantial fraction of reproductive tissues. However, there has been no RNAseq-based expression analysis of the reproductive tissues in Lepidoptera, and dosage compensation studies in other taxa always exclude gonadal tissues. In order to shed further light on the status of dosage compensation in Lepidoptera, it is necessary to generate and prepare high quality data using comprehensive procedures that permit comparisons with parallel studies.

Dosage compensation is assessed by comparing the expression between sex-linked and autosomal genes in males and females. I sampled representative adult tissues from *C. pomonella* based on their levels of sexual divergence, in order to evaluate the possible association between dosage compensation and sexual dimorphism. Insect adult head is the olfactory center for courtship and mating and is thus presumed to be one of the most

sexually dimorphic tissues in the insect soma. In contrast, the digestive tissue of the adult midgut in the abdomen is predicted to have minimal sexual dimorphism because it plays a minor role during lepidopteran adulthood when feeding activity is limited. Reproductive tissues on the other hand, have highly differentiated sex-specific functions. Among the male reproductive tissues, testis is the gonadal counterpart of the female ovary, whereas the other reproductive components are of somatic origin and therefore represent sex-specific somatic tissues.

In this chapter, I describe the generation and preparation of the transcriptome data in *C. pomonella*. I will discuss the use of these data to assess dosage compensation in the following chapters. I performed high throughput RNA sequencing (RNAseq) to obtain transcriptome data from these tissue samples. RNAseq has preceded microarray as the preferred method to analyze gene expression. Compared to traditional microarray or EST (expressed sequence tags) analysis, it has much deeper genomic coverage, higher precision and repeatability, and the most straightforward protocol and normalization procedures. Deep RNAseq enables *de novo* transcriptome assembly, which is particularly useful when studying non-model species such as *C. pomonella*, for which a reference genome is not available.

I used *B. mori* reference genome to identify sex/autosomal linkage of *C. pomonella* genes based on the conserved synteny in Lepidoptera. Synteny (the conservation of gene content on chromosomes) in a taxon allows the inference of putative genomic locations of genes in non-model species by using linkage data from closely related model species, an approach that has been used in several of recent dosage compensation studies (Harrison, Mank et al.

2012, Uebbing, Kunstner et al. 2013, Smith, Chen et al. 2014, Walters et al., unpublished data).

Materials and Methods

Samples, RNASeq and de novo assembly

C. pomonella eggs were obtained commercially (Benzon Research Inc., Carlisle, PA). Insects were reared on artificial diet (Southland Products Inc., Lake Village, AR) in the growth chamber under constant temperature (25 ± 1 °C) and a 12:12 light:dark cycle.

The following seven tissue samples were dissected from young adult moths within one day after emergence: head (female), head (male), midgut (female), midgut (male), ovary (female), testis (male) and accessory gland (male). The testis was severed from the rest of the male reproductive tract, which comprises the accessory gland, seminal vesicles, ejaculatory duct and ejaculatory bulb (Fig. 5). In this study, the remaining male reproductive tract is regarded as a single tissue sample and referred to as 'accessory gland' for the purpose of simplicity. Dissection was performed on ice. Dissected tissues were rinsed twice with Ringer's solution and immediately stored in *RNAlater*TM RNA stabilizing solution (SIGMA, St. Louis, MO).

Tissues dissected from five individuals were pooled into one sequencing sample, and two sequencing samples (replicates) were used for each tissue. RNA was extracted separately from each sequencing sample using the QIAGEN RNeasy Kit (QIAGEN, Foster City, CA),

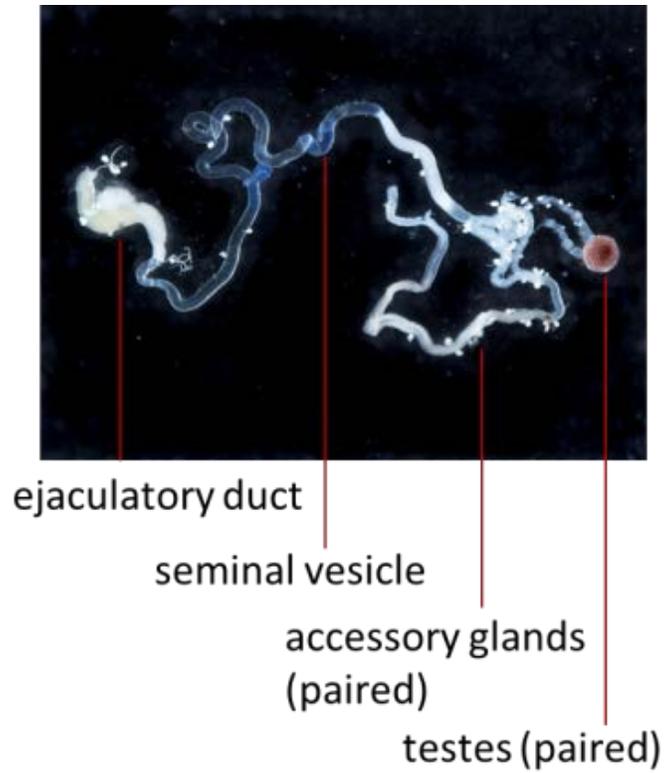


Fig. 5 The male reproductive tissues of *C. pomonella*. The testis and accessory gland, along with seminal vesicles, ejaculatory duct and ejaculatory bulb (not shown) comprise the adult male reproductive tract.

following the manufacturer's instructions and subsequently sent to Boyce Thompson Institute (Ithaca, NY) for the construction of barcoded libraries using standard Illumina library preparation protocols (Meyer and Kircher 2010). The resulting 14 libraries were then pooled into three lanes and sequenced on the Illumina HiSeq2000 platform at the Cornell University Biotechnology Resource Center (Ithaca, NY).

The resulting 150bp paired-end reads were first filtered and trimmed using Trimmomatic (V0.32) (Bolger, Lohse et al. 2014). The surviving reads - in all 14 sequencing samples were pooled for a single *de novo* assembly using the Trinity platform, which enables efficient and robust *de novo* transcriptome reconstruction from RNA-seq data (Grabherr, Haas et al. 2011, Haas, Papanicolaou et al. 2013). The finished assembly contained all 'Trinity components' (contigs) ≥ 200 bp in length, which represented the entire *C. pomonella* transcriptome from the 7 adult tissues.

Estimation of expressional intensity

Fragments per kilobase per million mappable reads (FPKM) were used to estimate gene expression intensity. Filtered read sets from each of the 14 sequencing samples were individually mapped back to the transcriptome assembly for FPKM calculation using RSEM (<http://deweylab.biostat.wisc.edu/rsem/rsem-calculate-expression.html>), which is included in the Trinity package. The resulting 14 FPKM sets represented gene expression in the 7 tissues, each with two replicates. In order to facilitate cross-tissue comparison, FPKM values were further normalized by the Trimmed Mean of *M*-values (TMM) method (Robinson and Oshlack 2010). Hereinafter, unless noted otherwise, all FPKM values mentioned are normalized by the TMM method.

Chromosome mapping

The 1:1 orthologs between *C. pomonella* and *B. mori* were identified using the reciprocal best-hit (RBH) approach with a Basic Local Alignment Search Tool (BLAST) algorithm e-value cut-off of 1E-5. *B. mori* gene sequences and a chromosome map with scaffold information were downloaded from SilkDB v2.0 (<http://silkdb.org/silkdb>). The contigs that mapped to *B. mori* chromosome 1 (Z) and chromosome 15 were identified as linked to the ancl-Z and neo-Z respectively, whereas those mapping to other chromosome locations were considered to be autosomal genes. Contigs that could not be mapped to a *B. mori* chromosome were excluded from subsequent analyses.

A subset of FPKM values corresponding to the contigs with identified chromosome locations was then extracted for all subsequent analysis. The reproducibility between the two replicates of each tissue sample was estimated with Pearson's ρ , an estimate of rank order correlation.

Differential expression analysis

Differentially expressed contigs (hereinafter 'DE genes') were identified using EdgeR (Bioconductor) imbedded in the Trinity package (Robinson, McCarthy et al. 2010). EdgeR uses false discovery rate (FDR) to determine differentially expressed contigs, which adjusts gene-specific p values for multiple tests (Storey and Tibshirani 2003). To reduce complexity, non sex-specific FPKM values were calculated for the head and the midgut by combining reads from respective female and male samples. A heat map was subsequently constructed using the TMM normalized FPKM of DE genes to plot changes across libraries.

Results

Illumina sequencing yielded 265 million pairs of total raw reads from 3 lanes. After filtering, 238 million pairs of reads were retained and subsequently assembled into 297,798 Trinity components, representing *C. pomonella* genes.

Among these genes, 10,616 were 1:1 orthologs to *B. mori*, with median length of 1,283bp. Chromosome mapping identified 447, 535 and 9,053 *C. pomonella* genes as ancl-Z, neo-Z and autosomal linked, respectively. Because the correlation of FPKM values between replicates was high for all the tissue samples (Pearson's ρ ranging 0.84 ~ 0.99, mean = 0.93), average FPKM values were used to measure levels of gene expression in all the following analyses. All genes with FPKM>0 values in a given tissue were considered to be actively expressed in that tissue. The total number of expressed genes by chromosome in each tissue is summarized in Table 1. The male reproductive tissues (especially accessory gland) had fewer total expressed genes compared to the other tissues. In the data set not normalized by TMM, the overall FPKM values were higher in the ovary but lower in the testis compared to the TMM-normalized data set. In contrast, FPKM values for somatic tissues did not differ before and after TMM normalization (Fig. 6).

The head and midgut contained only 20 and 8 DE genes, respectively, between the sex-specific samples. Therefore, non sex-specific FPKM values were re-calculated for these two tissues. Among the 10,035 *C. pomonella* genes with chromosome location information, 5,427 were DE genes ($p < 0.05$, FDR < 0.05). The heat map showed distinctive expression profiles in the testis and accessory gland compared to other tissues (Fig. 7). Most genes were down-regulated (green bars) in testis and accessory gland. In contrast, besides the

Table 1 Summary of the total number of genes identified as expressed (FPKM > 0) in each tissue. Testis and accessory gland have much smaller pools of expressed genes compared to the other tissues.

Chromosome location	ancl-Z	neo-Z	Neo Z (ancl-Z + neo-Z)	autosome	Total (neo Z + AS)
Total expressed genes (FPKM > 0 in at least one tissue)	447	535	982*	9,053	10,035*
Head (F)	366	463	829	7,744	8,573
Head (M)	327	442	769	7,363	8,132
Midgut (F)	318	419	737	7,002	7,739
Midgut (M)	335	447	782	7,255	8,037
Ovary	338	451	789	7,367	8,156
Testis	281	358	639	5,838	6,477
Accessory Gland	160	221	381	3,947	4,328

* All shaded values from this column are calculated by addition by the formula in the header.

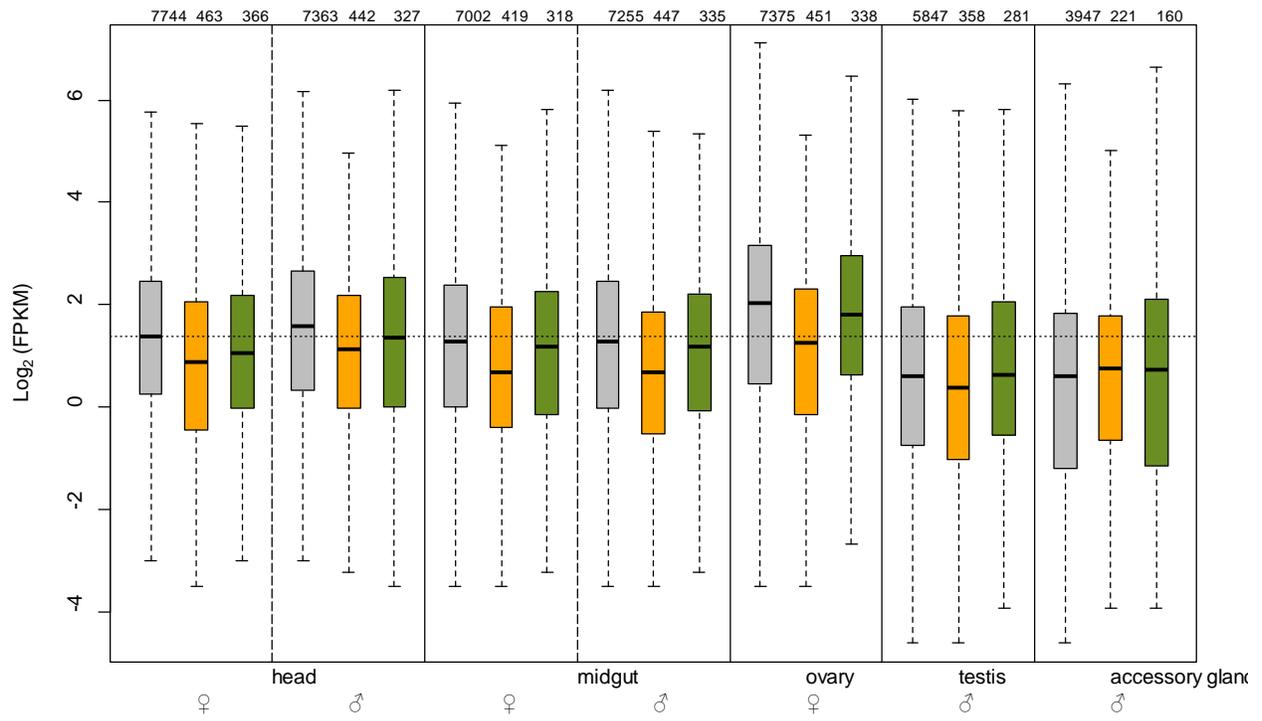
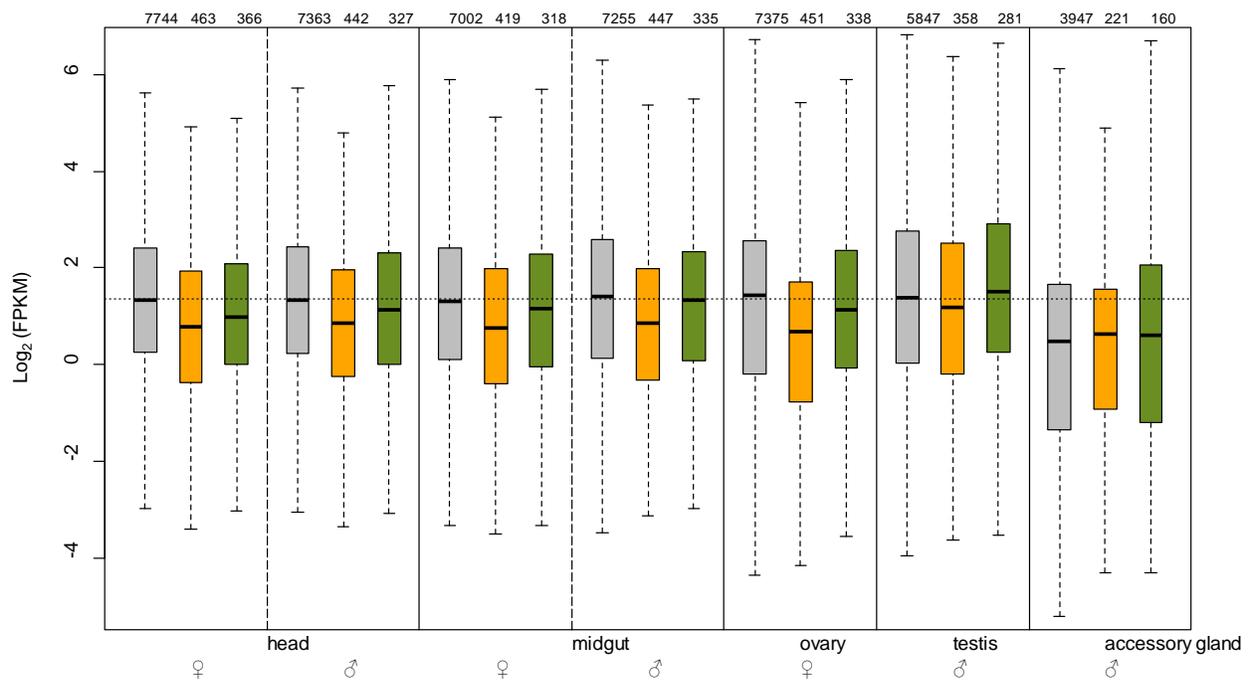
A**B**

Fig. 6 Expression levels (\log_2 transformed FPKM) of all genes expressed (FPKM>0) in each tissue by chromosomal locations. (A): FPKM without normalization; (B): FPKM normalized with TMM. Grey: autosomal genes; orange: ancl-Z genes; green: neo-Z genes. The number of total observations is indicated above each boxplot. Plot boxes represent the median and interquartile range of expression levels. The whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range. The dotted horizontal line across the whole plotting area denotes the average autosomal expression in the head and migut. The overall FPKM values changed significantly before and after TMM normalization in the ovary and testis but not in the other tissues.

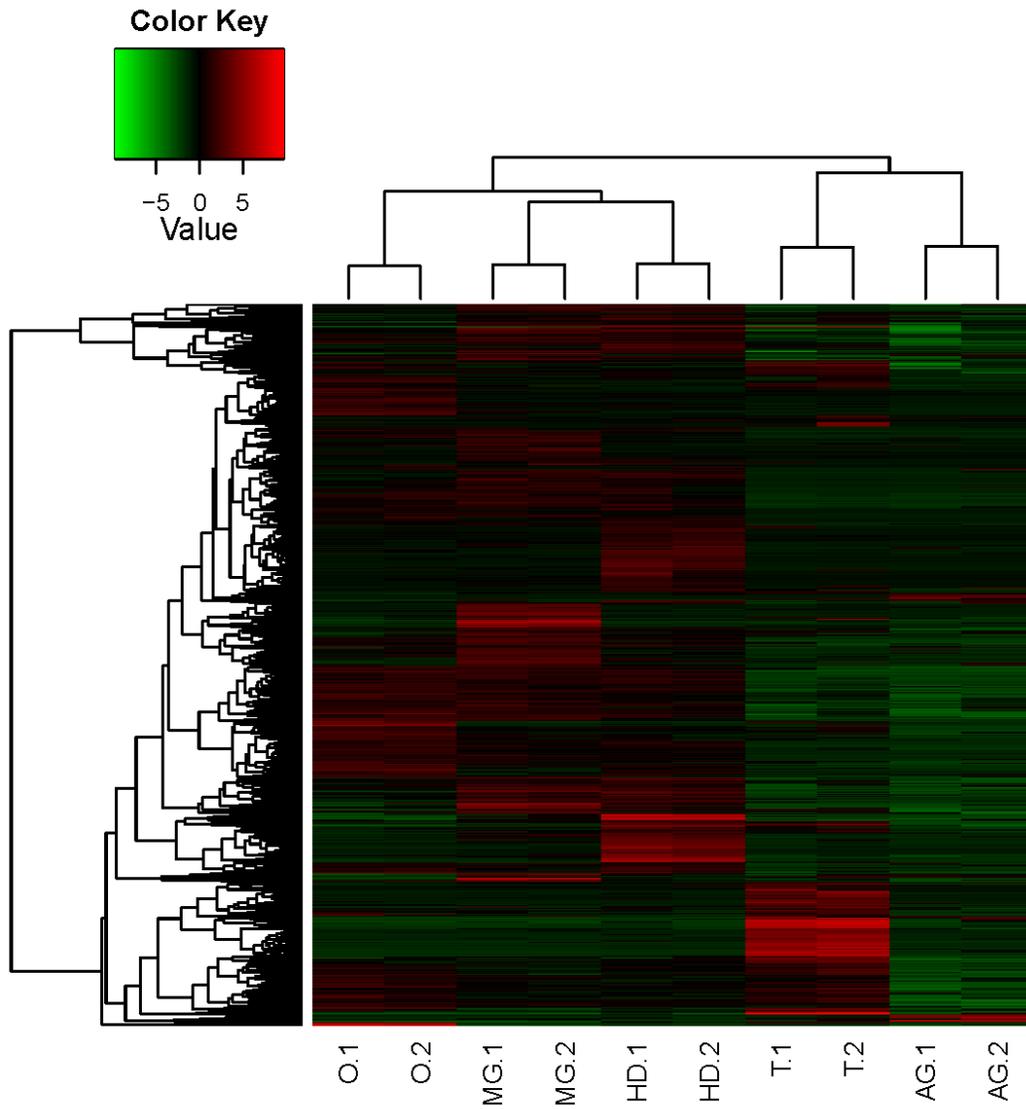


Fig. 7 Heat map of the relative expression levels of transcripts (rows) in *C. pomonella* tissues (column) ($FC > 1$, $FDR < 0.05$, p value for $FDR < 0.05$). Each tissue has two replicates: O, ovary; HD, head; MG, midgut; T, testis; AG, accessory gland. Color bar legend indicates the degree of \log_2 FPKM change among tissues. On the left is a dendrogram of genes hierarchically clustered according related expression patterns.

accessory gland, up-regulated genes (red bars) were most abundant in the testis but most scarce in the ovary.

Discussion

Conserved synteny in Lepidoptera

The conservation of synteny in Lepidoptera supports the validity of using the *B. mori* reference genome to assign chromosomal locations to *C. pomonella* genes. Linkage mapping studies first revealed that chromosomes in the Lepidoptera exhibit high levels of synteny with frequent local rearrangement of gene orders (Yasukochi, Ashakumary et al. 2006, Pringle, Baxter et al. 2007, Beldade, Saenko et al. 2009, Yasukochi, Tanaka-Okuyama et al. 2009, d'Alencon, Sezutsu et al. 2010, Baxter, Davey et al. 2011, Van't Hof, Nguyen et al. 2013). The characterization of increasing genomes of moth and butterfly species confirms the high level of synteny within this order in a more comprehensive scope (Xia, Zhou et al. 2004, International Silkworm Genome 2008, Zhan, Merlin et al. 2011, Heliconius Genome 2012, You, Yue et al. 2013, Ahola, Lehtonen et al. 2014). Furthermore, the degree of synteny among representatives of the Lepidoptera that diverged ~160 MYA is comparable to that found within the single, 65-MYR-old genus *Drosophila* (Labandeira, Dilcher et al. 1994, Singh, Larracuenta et al. 2009, Van't Hof, Nguyen et al. 2013).

Distinctive gene expression profiles in C. pomonella adult reproductive tissues

Normalization is a critical procedure to ensure the accurate estimation of gene expression as well as the detection of differential expression by reducing the systematic technical effects on the data to the minimal extent. FPKM is the method most commonly used to normalize RNA-seq data for gene expression analyses (Mortazavi, Williams et al. 2008). In addition to standardizing gene length, FPKM also standardizes data between samples by scaling the number of reads in a given lane or library to a common value across all sequenced libraries in the experiment. However, the number of reads expected to map to a gene is dependent not only on the expression level and length of the gene, but also on the composition of the RNA population that is being sampled. Thus, if a large number of genes are unique to, or highly expressed in, one experimental condition or specific tissue, the sequencing capacity available for the remaining genes in that sample is decreased. Therefore, TMM normalization is important when comparing data from different tissue samples that have different pools of expressed genes (Robinson and Oshlack 2010). Across *C. pomonella* tissues, FPKM values not normalized by TMM were higher in the ovary but lower in the testis, whereas the values for somatic tissues did not differ before and after TMM normalization. By comparison, TMM normalization does not affect the results of dosage compensation analysis in *M. sexta* head (Smith, Chen et al. 2014). These observations suggest that the testis transcriptome contains a much higher portion of highly expressed genes that are unique compared to the other tissues. Conversely, because of the high level of gene expression associated with oogenesis, the ovary transcriptome would be expected to contain the smallest number of unique transcripts. The data set for the

accessory gland stood out with overall low FPKM values that did not change with normalization but this tissue also has the smallest number of total expressed genes. Thus it is inferred that genes expressed in the accessory gland comprise a unique and small repertoire which finds little overlap with the entire transcriptome. The distinct DE profiles in the reproductive tissues further corroborated these implications from TMM normalization.

The primary contents of insect male reproductive tract (accessory gland) are seminal fluid proteins, which are transferred to the female during mating and play an important role in a wide range of post-mating responses (Avila, Sirot et al. 2011). Further, in *Drosophila* the male reproductive genes have undergone rapid sequence evolution (Haerty, Jagadeeshan et al. 2007) and the proteins expressed in the testis and accessory gland are much more divergent than other somatic proteins (Thomas and Singh 1992, Civetta and Singh 1995).

In light of the unique gene expression patterns in *C. pomonella* reproductive tissues, caution must be taken when including these tissues in the analysis of dosage compensation. Germline-specific regulation on the sex chromosomes is quite common. Examples include the restoration of the silenced X chromosome in female mammals (Sugimoto and Abe 2007), the absence of dosage compensation in the *Drosophila* male germline (Meiklejohn, Landeen et al. 2011), the silencing of the X chromosomes in the germline of both *C. elegans* sexes (Kelly, Schaner et al. 2002), and meiotic sex chromosome inactivation in *mammals*, *Drosophila* and *C. elegans* (Monesi 1965, Bean, Schaner et al. 2004, Turner 2007, Schoenmakers, Wassenaar et al. 2009, Guioli, Lovell-Badge et al. 2012, Vibranovski 2014). Any of these mechanisms would potentially complicate the assessment of dosage

compensation and could bias the results. Therefore, dosage compensation studies in other taxa typically avoid the inclusion of gonadal tissue.

Chapter III

Analysis of dosage compensation in *Cydia pomonella* based on *de novo* transcriptome data

Introduction

Information on dosage compensation in Lepidoptera is scarce. Conclusions drawn from two early reports on two lepidopteran species (Zha, Xia et al. 2009, Harrison, Mank et al. 2012) contributed to the popular speculation that incomplete dosage compensation might be a universal feature associated with female heterogamety (Graves and Disteche 2007, Mank 2009, Vicoso and Bachtrog 2009, Naurin, Hansson et al. 2010, Wolf and Bryk 2011, Harrison, Mank et al. 2012, Mank 2013). The first study used microarray data to examine dosage compensation in the silk moth *Bombyx mori* (Zha, Xia et al. 2009). The authors showed that the expression levels of Z-linked genes in all five tissues examined are significantly higher in the male than the female, and concluded that this pattern is consistent with lack of dosage compensation found in other ZZ/ZW systems. However, a subsequent study concluded that the initial analysis reflects artifacts of microarray normalization without examining the global F:M expression ratio for the autosomal loci as control (Walters and Hardcastle 2011). Revisiting the same data set using different normalization methods revealed no F:M parity and that global Z chromosome expression

was significantly reduced relative to that of autosomes, consistent with a $Z \approx ZZ < AA$ model. Authors of a third study suggested that the previous microarray data from *B. mori* was itself too noisy to be compelling, and used RNAseq data from whole insects to demonstrate incomplete Z chromosome dosage compensation in the Indian meal moth *Plodia interpunctella* (Harrison, Mank et al. 2012). In this species female Z-linked genes exhibited slightly over half the expression level of male and autosomal genes ($Z \ll ZZ \approx AA$). Based on those results the authors suggested that the Lepidoptera and possibly all female heterogametic taxa lack global dosage compensation. Nevertheless, two recent analyses, also based on RNAseq data, from *Manduca sexta* and two *Heliconius* species provided new evidence for the $Z \approx ZZ < AA$ model (Smith, Chen et al. 2014, Walters et al., unpublished data). However, the *M. sexta* study only sampled the adult head whereas the *Heliconius* study separately analyzed the adult head and rest of the body.

The available studies of dosage compensation in Lepidoptera differ in sampling methods and types of transcriptome data (e.g., microarray versus RNAseq). In particular, there has been no RNAseq-based analysis specifically of gonadal tissue. Therefore, a broader sampling of species with a comparable protocol of data analysis is needed to resolve the controversies and shed further light on dosage compensation in Lepidoptera.

To address these issues, I performed a dosage compensation analysis in *C. pomonella* using the *de novo* transcriptome data generated from Chapter II. In particular, I examined the somatic tissues and gonads individually. I then compared my results with other lepidopteran dosage compensation studies. I concluded from this comparison and that the

conclusion drawn from the *P. interpunctella* study is called into question by the inclusion of gonads and problematic data treatment.

Materials and Methods

The transcriptome data from *C. pomonella* tissues which were generated and prepared in Chapter II were used for dosage compensation analysis. Both aspects of dosage compensation were assessed: whether the expression of Z linked genes is equal between sexes (i.e., 'Z~ZZ compensation') and whether Z expression is balanced with autosomal expression (i.e., 'Z~A compensation').

To assess Z~ZZ compensation, F:M (Z:ZZ) expression parity was evaluated for all genes expressed in both sexes in the somatic (head and midgut) and gonadal (ovary-testis) tissues. In the absence of Z~ZZ compensation, the Z:ZZ ratio is expected to be 0.5; under complete Z~ZZ compensation, the Z:ZZ ratio is expected to be 1, the same as the autosomal baseline (AA:AA).

To assess 'Z~A compensation', all genes expressed (FPKM > 0) in each of the 7 tissues were compared between the ancl-Z, neo-Z and the autosomes in average expression, median expression and distribution of expression levels. The Z to autosome median expression ratios in female (Z:AA) and male (ZZ:AA) are hereinafter referred to as 'Z(Z):AA ratio'. Under full Z~A compensation, the Z(Z):AA ratio is expected to be close to 1 (considering the natural variation among chromosomes), whereas a Z(Z):AA ratio of 0.5 would indicate monoallelic Z expression without Z~A compensation.

Results

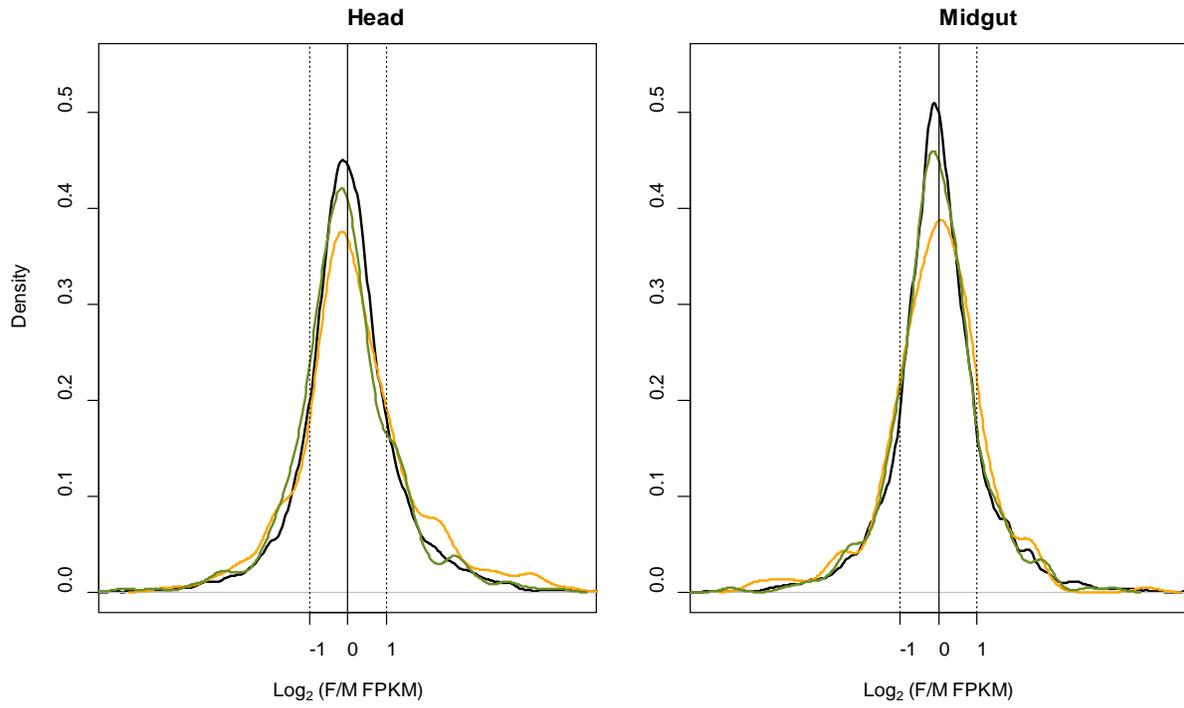
Z~ZZ compensation: comparison of Z-linked gene expression between male and female

In both head and midgut tissues from *C. pomonella*, there were no significant differences in either autosomal or Z-linked gene expression between the males and females (FDR corrected $p > 0.05$, Mann–Whitney U test), indicating complete Z~ZZ compensation (Fig. 8A). In gonadal tissues, however, autosomal gene expression *per se* did not exhibit the expected sex parity (F:M), being overall higher in the ovary than in the testis (Fig. 8B). Even compared to the autosomal F:M baseline, neither ancl-Z nor neo-Z gene expression showed the same sex parity. In addition, the correlation of gene expression between the ovary and the testis was extremely poor (Pearson's $\rho = 0.12$), compared to the head (Pearson's $\rho = 0.92$) or the midgut (Pearson's $\rho = 0.73$).

Z~A compensation: comparison of gene expression between the Z chromosome and autosomes

Evaluations based on average expression, median expression, and the distribution of expression levels yielded consistent results, indicating a variable extent of Z~A compensation both between the ancl-Z and neo-Z segments and among different tissues. In general, the overall expression of neo-Z genes was higher than that of ancl-Z genes in all tissues except for the accessory gland (Fig. 6B). Table 2 summarizes the p values for significant differences in the average expression levels between the Z chromosome and the autosomes based on Bonferroni-corrected Mann–Whitney U tests (two-sided). For ancl-Z genes, average expression levels were significantly lower than those of the autosomes in

A



B

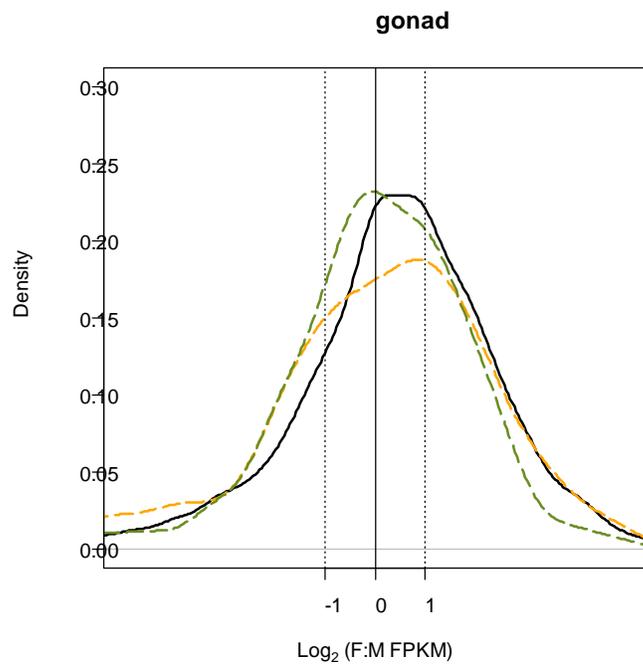


Fig. 8

Z~ZZ compensation indicated by female over male expression (\log_2 transformed F:M FPKM) on autosomes (black), ancl-Z (orange) and neo-Z (green) in (A): head and midgut and (B): gonad (ovary~testis), respectively. Key reference \log_2 ratios are marked: -1 (two-fold male over female expression); 0 (equal expression between sexes) and 1 (two fold female over male expression).

Table 2

Significance of differences in average expression between autosome, ancl-Z and neo-Z across 7 tissues, assessed by Bonferroni corrected Mann–Whitney *U* test (two-sided).

$P > 0.01$, $P > 0.1$ values (no significant difference) are marked as * and **, respectively.

	AA~ancl-Z(Z)	AA~neo-Z(Z)	ancl-Z(Z)~neo-Z(Z)
Head (F)	7.36E-08	1.05E-03	1.50E-01**
Head (M)	1.49E-05	2.00E-02*	2.81E-01**
Midgut (F)	3.21E-06	5.88E-01**	1.04E-02*
Midgut (M)	4.09E-06	3.96E-01**	1.45E-02*
Ovary	1.78E-09	6.69E-01**	6.40E-05
Testis	4.54E-02*	2.26E-01**	6.04E-03
Accessory gland	9.95E-01**	9.53E-01**	1.00E+00**

every tissue except for testis and accessory gland. For neo-Z genes, average expression levels were significantly lower than those of the autosomes only in the head. Both the ancl-Z(Z):AA and neo-Z(Z):AA ratios were highest in the accessory gland (1.15 and 1.13, respectively) and the testis (0.82 and 1.14, respectively), with the confidence intervals all encompassing 1 (Fig. 9). In the other tissues, ancl-Z(Z):AA ratios ranged from 0.48-0.61 and the confidence intervals all encompassed 0.5, whereas neo-Z(Z):AA ratios ranged from 0.70-0.93 and the confidence intervals encompassed 1 or fell in the 0.5-1 range.

Testis and accessory gland were the only tissues where the distributions of gene expression levels on both ancl-Z and neo-Z did not differ significantly from each other or from the autosomes (Fig. 10, Table 3). In all the other tissues, the distributions of ancl-Z gene expression differed significantly from the autosomes (AA) but not from the computationally halved autosomal expression (A). For neo-Z gene expression, the distributions differed significantly from the autosomes (AA) in most but not all tissues, but did not differ from the halved autosomal expression (A) in any tissue.

Discussion

Data from *C. pomonella* and *B. mori* show disparate patterns of dosage compensation between the soma and the gonad.

In *C. pomonella*, Z~ZZ compensation was complete on both the ancl-Z and neo-Z segments in the head and midgut, whereas the extent of Z~A compensation varied among tissues and generally appears to be more complete on the neo-Z than on the ancl-Z segments. In tissues

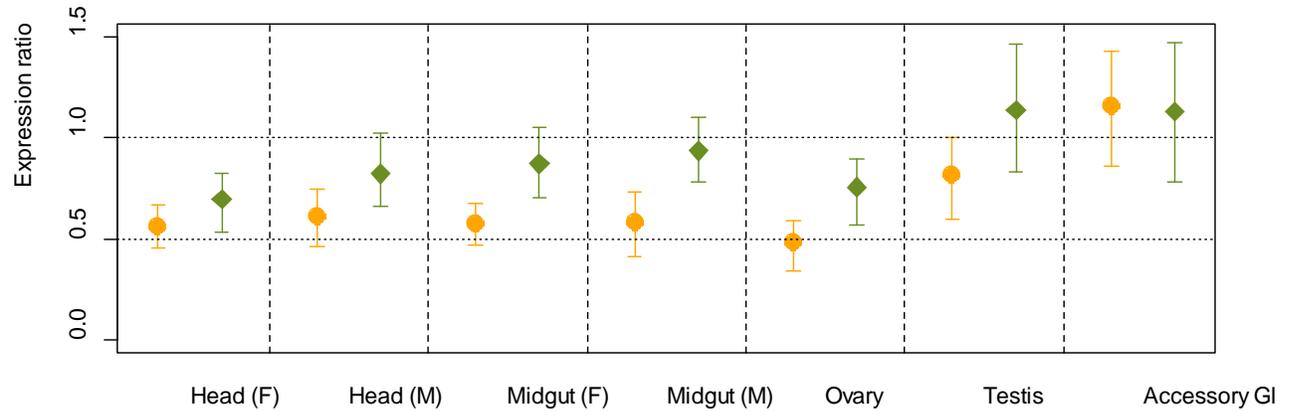


Fig. 9

Z(Z):AA median expression ratios in *C. pomonella* tissues. Orange circles: ancl-Z; green diamond: neo-Z. Error bars show 95% confidence intervals estimated by 1000 bootstrap replicates. The horizontal dotted lines denote the reference ratios: 0.5 suggests equal Z(Z) expression to monoallelic autosomal expression (A); 1 suggests equal Z(Z) expression to biallelic autosomal expression (AA) .

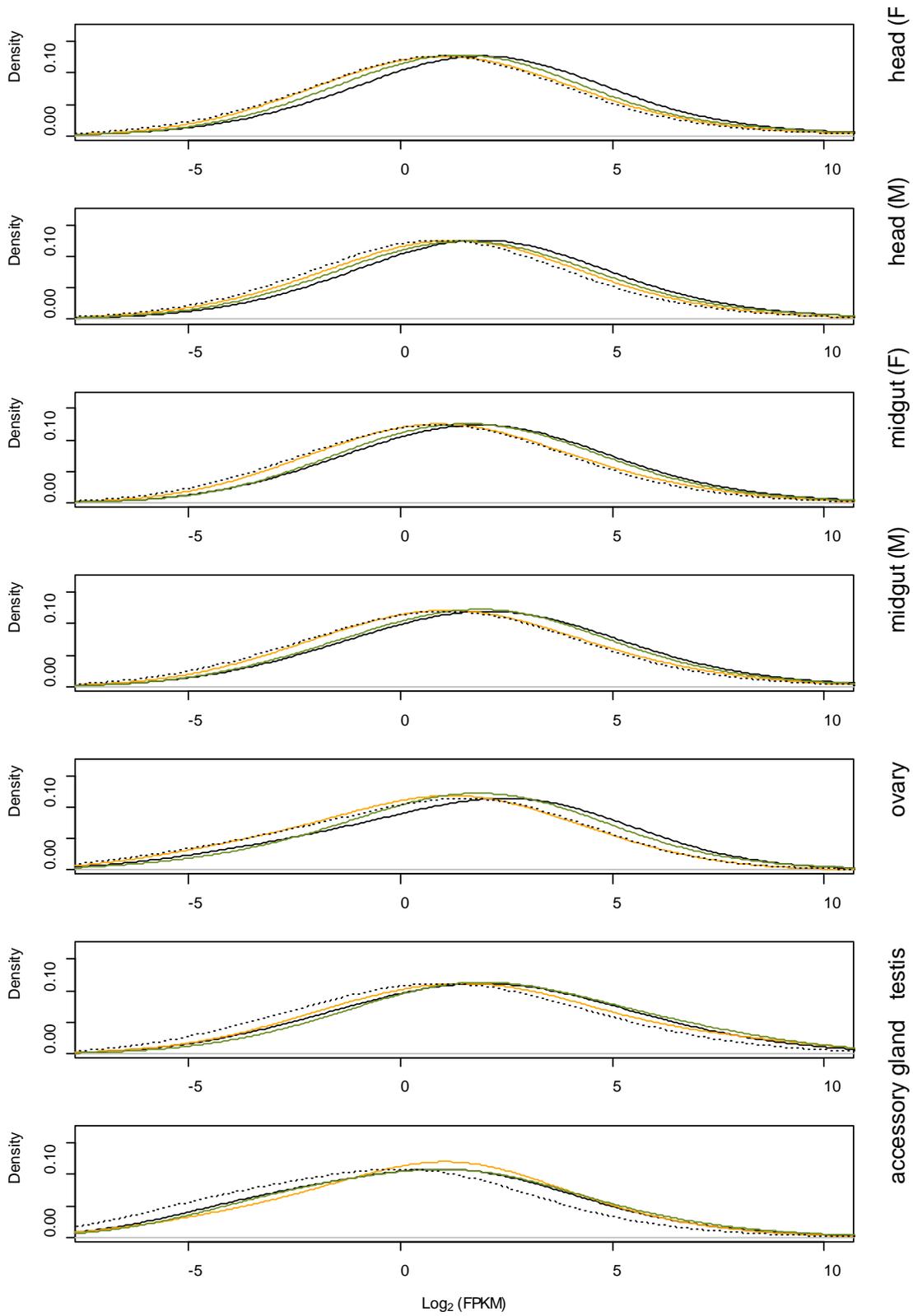


Fig. 10

Distribution of expression levels (\log_2 transformed FPKM) of expressed genes (FPKM>0) in seven tissues by chromosomal locations: autosomes (black), ancl-Z (orange) and neo-Z (green). Note the dotted black line denotes the computationally halved FPKM values from autosomal gene, conferring the expression from a single allele.

Table 3

Significance of differences in distribution of expression levels between ancl-Z/neo-Z and autosomes (AA and A) across 7 tissues assessed by Benjamini-Hochberg corrected Komolgorov-Smirnov test.

0.05 < P < 0.2 and P > 0.2 values which denote no significant difference are marked as * and **, respectively.

	AA~ancl-Z(Z)	A~ancl-Z(Z)	AA~neo-Z(Z)	A~neo-Z(Z)
Head (F)	3.49E-07	3.17E-01**	4.02E-03	2.27E-04
Head (M)	1.35E-04	2.14E-01**	3.42E-02	8.14E-06
Midgut (F)	4.39E-07	3.61E-01**	1.66E-01*	3.38E-07
Midgut (M)	2.92E-05	3.72E-01**	8.32E-02*	2.84E-08
Ovary	3.04E-12	1.97E-01*	1.29E-02	1.46E-05
Testis	3.43E-01**	1.11E-03	3.43E-01***	2.44E-10
Accessory gland	5.46E-01**	1.35E-06	5.46E-01**	1.35E-06

other than the male reproductive tissues (head, midgut and ovary), the ancl-Z gene expression was similar to the monoallelic autosomal gene expression, implying that Z~A compensation is absent on the ancl-Z segment. By contrast neo-Z gene expression fell between the monoallelic and biallelic autosomal gene expression, indicating that Z~A compensation is incomplete. In the male reproductive tissues (both testis and accessory gland), both ancl-Z and neo-Z gene expression (ZZ) did not differ significantly from autosomal gene expression (AA), a finding consistent with constitutive expression in these tissues (ZZ=AA vs Z=ZZ<AA in the head and midgut). This result is in good agreement with those from *B. mori*, where the ZZ:AA ratio in testis is close to 1 and also the highest among all the tissues examined, although the rest of the male reproductive tract not surveyed (Walters and Hardcastle 2011). Thus the evidence from both *B. mori* and *C. pomonella* suggests that the male reproductive systems are exempt from dosage compensation in lepidopterans. By comparison, in female mammals, one of the two X chromosome copies that is inactivated early in embryonic development is later restored in the germline (Sugimoto and Abe 2007). Dosage compensation, i.e. the doubling of gene expression on the single X chromosome, is also absent in the *Drosophila* male germline (Meiklejohn, Landeen et al. 2011).

The lepidopteran Z chromosome has a larger fraction of weakly-expressed genes than the autosomes.

In *M. sexta* head, the Z(Z):AA ratio is ~0.8 and average Z expression is significantly lower than that of the autosomes (Smith, Chen et al. 2014). However, the authors also showed that the Z(Z)~AA disparity observed in the unfiltered data (FPKM>0) diminishes with

increasing FPKM filtering threshold. Nevertheless, it is worth noting that in that study, the minimal threshold at which the difference starts to appear significant is 4, which is not trivial compared to the median values (~ 10 and ~ 12 for Z and autosomes respectively). In fact, filtering removes nearly twice as many Z genes as autosomal genes which indicates that a substantially larger proportion of Z-linked genes are expressed at low levels compared to the autosomal genes, since systematic or technical errors should not yield such bias

It is noteworthy that in general for two data groups with different distributions, removing lower values in the pair would compress the medians of both groups and eventually make them statistically indistinguishable (Walters and Hardcastle 2011). This effect is greater when the group with the lower mean also has a larger proportion of low values, as found to be the case between Z and autosomal FPKM values. Therefore, filtering would mask the difference between Z and autosomal expression due to the disproportionate reduction of low FPKM values.

By comparison, the *C. pomonella* data followed the same trend, and the cutoff values under which Z(Z) \sim AA disparity becomes insignificant vary both among tissues and between the ancl-Z and neo-Z segments, as would be expected. It is also important to assess the distributions of gene expression, because weakly-expressed genes may bias comparisons based on median values (Deng, Hiatt et al. 2011, Julien, Brawand et al. 2012). Indeed, in the *C. pomonella* data the Z(Z):AA ratios were corroborated by the distributions of gene expression. For example, in tissues with Z(Z):AA ≈ 0.5 , the Z expression resembled the proxy of autosomal expression from one allele rather than from both.

Together, these observations suggest that in lepidopterans, a larger proportion of genes are expressed at low levels in the Z chromosome compared to the autosomes in tissues other than the male reproductive tissues.

Inconsistencies between *P. interpunctella* and other lepidopteran species

The pattern of $Z \ll ZZ \approx AA$ reported in *P. interpunctella* differs from all the other lepidopteran species ($Z \approx ZZ < AA$) examined for dosage compensation so far, and instead appears to be in good agreement with the other ZZ/ZW taxa (Harrison, Mank et al. 2012). However, I suggest that this interpretation is open to criticism for several reasons.

First, the analysis of *P. interpunctella* was based solely on gene expression from adult whole bodies, which include a substantial fraction of reproductive tissues. As shown in both *B. mori* (Walters and Hardcastle 2011) and *C. pomonella* (this study), gonadal tissues exhibit a different pattern from soma regarding both Z~ZZ compensation and Z~A compensation. Further, in *C. pomonella* gene expression was poorly correlated between male and female gonadal tissues. The higher level of autosomal expression in the ovary compared to the testis may reflect the more active transcriptional activity in the female germline. These observations are consistent with disparate expression profile for the reproductive tissues described in Chapter II. Therefore conclusions about dosage compensation based on data sets that contain gonadal tissues must be made with caution.

Second, the FPKM values for *P. interpunctella* were not normalized using the TMM method. As shown in Chapter II, the FPKM values for reproductive tissues from *C. pomonella* differed significantly from those for the soma when the dataset was not normalized by

TMM, with ovary having overall higher values and testis having overall lower values. In *P. interpunctella*, median autosomal expression appears to be higher in the female whole body than in the male whole body, a result that is consistent with the differential impact of expression in ovary and testis on the whole body data.

Third, the *P. interpunctella* data were filtered using a procedure that removes contigs with FPKM values less than 4 in at least two out of four samples (two replicates each for male and female). Because the genes expressed in the ovary have overall higher non-normalized FPKM values than those expressed in the male reproductive tissues, this filtering method would 'selectively' remove more contigs with lower values in the two male replicates but not in the two female replicates. On the other hand, filtering weakly-expressed loci *per se* diminishes the observed Z:A disparity, as discussed earlier. As a result, at a certain filtering threshold the Z:A disparity would appear insignificant in the male but still significant in the female. Curiously, the arbitrary filtering threshold was set at FPKM = 4 in the *P.*

interpunctella study (Harrison, Mank et al. 2012); this is the same filtering threshold at which the Z:A disparity can be eliminated in the *M. sexta* study (Smith, Chen et al. 2014).

Finally, the *P. interpunctella* study employed mean FPKM values rather than median values to assess dosage compensation. This may further mask the difference between Z and autosomal gene expression because of the different extent of data variance in the two groups. For example, in the *C. pomonella* data sets estimates based on mean FPKM values reflect an even greater Z:AA expression disparity in the ovary (ancl-Z:AA mean expression ratio = 0.32, compared to the median ratio = 0.48).

To summarize, the conclusion that *P. interpunctella* exhibits ‘incomplete’ dosage compensation is weakened both in experiment design and statistical bias. Unfortunately the raw RNAseq data for *P. interpunctella* is not publicly available and therefore not open to comprehensive re-evaluation. Moreover, the Z chromosome is highly conserved among Lepidopteran lineages (Sahara, Yoshido et al. 2012) and dosage compensation patterns reported are typically conserved within individual taxa. Therefore the inconsistent pattern reported for *P. interpunctella* seems particularly surprising because both a more basal species (*C. pomonella*) and several more-derived species (*B. mori*, *M. sexta* and *H. spp.*) share a common pattern that differs from that found in *P. interpunctella*.

Chapter IV

Analysis of dosage compensation in *Cydia pomonella* using a comparative transcriptome approach

Introduction

The essence of X|Z~A compensation is to restore sex-linked gene expression to the ancestral output of the proto sex chromosome before its differentiation, which cannot be directly measured (Ohno, Kaplan et al. 1959). Therefore, X|Z~A compensation is typically assessed indirectly by comparing the gene expression of the present-day sex chromosome to that of present-day autosomes, under the implicit assumption that ancestral global expression levels were approximately equal across the chromosomes. However, expression levels are not necessarily uniform among present-day individual autosomes *per se* (Walters and Hardcastle 2011, Julien, Brawand et al. 2012). Thus the average expression of present-day autosomes may not accurately reflect proto-sex chromosome expression.

Present-day sex chromosomes have already evolved with different functionality and gene content compared to autosomes over the course of differentiation (Vicoso and Charlesworth 2006). For example, sex-biased or germline-specific genes are found to be disproportionally distributed on the sex chromosomes in both XX/XY and ZZ/ZW species (Saifi and Chandra 1999, Lercher, Urrutia et al. 2003, Parisi, Nuttall et al. 2003, Khil, Smirnova et al. 2004, Reinke, Gil et al. 2004, Arunkumar, Mita et al. 2009). In mammals,

these genes have overall low expression levels in the soma and lead to disproportionate reductions of estimates on the somatic expression level for the X chromosome (Deng, Hiatt et al. 2011, Julien, Brawand et al. 2012). In addition, a subset of autosomal genes interacting with sex-linked genes may also be subject to down-regulation as a result of the differentiating sex chromosome, adding another layer of complexity, as has been reported in placental mammals (Julien, Brawand et al. 2012) and *C. elegans* (Meyer 2010).

These considerations confound the analysis of X|Z~A compensation if average autosomal expression is used as the reference and may lead to erroneous conclusions. Therefore, it is important to detect differences between expression levels of the same set of genes on both the present-day sex chromosome and the proto-sex chromosome because these differences should provide the greatest insight into Z~A compensation. Although the expression of genes on the proto-sex chromosome cannot be measured directly, it can be inferred by the expression of their orthologs which remain autosomal-linked in other species (He, Chen et al. 2011, Julien, Brawand et al. 2012, Lin, Xing et al. 2012). For example, protracted debates regarding mammalian X(X):AA compensation were finally resolved by two studies comparing the expression of X-linked genes in mammals with their autosomal orthologs in birds despite ~300 MYR of divergence (Julien, Brawand et al. 2012, Lin, Xing et al. 2012). Similarly, a dosage compensation analysis of a neo-X chromosome in *Drosophila pseudoobscura* quantified the varying levels of X~X compensation among different tissues and developmental stages (Nozawa, Fukuda et al. 2014). Most recently, another study comparing the expression of differentially located (X versus A) orthologs among five

nematode species revealed the dichotomy of X~A compensation between strongly and weakly expressed X-linked genes (Albritton, Kranz et al. 2014).

In a common ancestor leading to the two subfamilies of Tortricidae (Tortricinae and Olethreutinae), a chromosome translocation event caused the fusion of an ancestral autosome to the ancestral Z chromosome (Sichova, Nguyen et al. 2013). Thus orthologs of *C. pomonella* neo-Z genes are autosomal-linked in closely related out-group Lepidopterans with the conserved ancestral karyotype. In Chapter III I demonstrated that the expression of neo-Z genes was reduced to a lesser extent than ancl-Z genes using autosomal gene expression as a reference. In this chapter, using RNAseq data available from two parallel studies (*M. sexta* (Smith, Chen et al. 2014) and *H. Melpomene* (Walters et al, unpublished data), I performed an interspecific transcriptome analysis to quantify the extent to which the neo-Z gene expression is reduced.

Materials and Methods

RNAseq data (*de novo* transcriptome assembly and raw reads) from *M. sexta* head samples (4 replicates for each sex) was accessed from published data (Smith, Chen et al. 2014). FPKM values were calculated in this study as described in Chapter II. For unpublished *H. melpomene* data, FPKM values were provided by James Walters (University of Kansas, Lawrence, KS, unpublished data). Both FPKM data sets were normalized using the TMM method within the species and averaged among the replicated data sets as described in Chapter III.

The orthologous pairs between *C. pomonella* and *M. sexta* and their chromosomal locations were identified by mapping to the *B. mori* reference genome. The *H. melpomene* reference genome (<http://www.butterflygenome.org/>) was used directly to identify orthologs between *C. pomonella* and *H. melpomene* and to infer their chromosome locations using the protocol described in Chapter III. *H. melpomene* chromosome 11 corresponds to *B. mori* chromosome 15 and the two chromosomes share a high level of synteny (Heliconius Genome 2012).

FPKM values cannot be directly compared between different species due to length differences between orthologous gene pairs, different transcriptome compositions, and technical variations between different sequencing platforms. Therefore, I employed the scaling procedure described by Lin, Xing et al. (2012). Specifically, all FPKM values from one species were linearly adjusted by a common factor in each sex so that the median FPKM values for autosomal genes were the same between the two species to be compared. To avoid infinite values when calculating ratios without changing the medians, all FPKM = 0 values were converted to FPKM = 0.01.

Z~ZZ compensation was assessed using the equation described by Nozawa, Fukuda et al. (2014):

$$R_{dc} = \frac{F_{Cp}}{M_{Cp}} / \frac{F_{ref}}{M_{ref}}$$

Here, the F:M ratio of the reference species is introduced to account for any possible changes in expression between sexes during evolution in the absence of dosage compensation, thus allowing a more accurate and sensitive detection of dosage

compensation. The rate of dosage compensation (R_{dc}) is expected to be 1 under perfect dosage compensation and 0.5 in absence of dosage compensation. Values for R_{dc} were also calculated for autosomes and the ancl-Z segment as controls.

Z~A compensation was assessed by comparing the expression of neo-Z genes with that of their autosomal orthologs in the reference species, which are denoted as

neo-ZZ. The genes in the reference species that are orthologous to *C. pomonella* ancl-Z and autosomal genes are used as controls and denoted as ancl-Z(Z) and AA, respectively.

Results

Expression levels between interspecific orthologous pairs were strongly correlated (Pearson's $\rho = 0.55$, p -value $< 2.2e-16$ for the *C. pomonella*~*M. sexta* pair and $\rho = 0.4$, p -value $< 2.2e-16$ for the *C. pomonella*~*H. melpomene* pair). The results from the *C. pomonella*~*H. melpomene* pair were comparable to those from the *C. pomonella*~*M. sexta* pair (Fig. 11). In light of closer phylogenetic relationship and stronger correlation in expression between *C. pomonella* and *M. sexta*, this data set was employed for all further analyses.

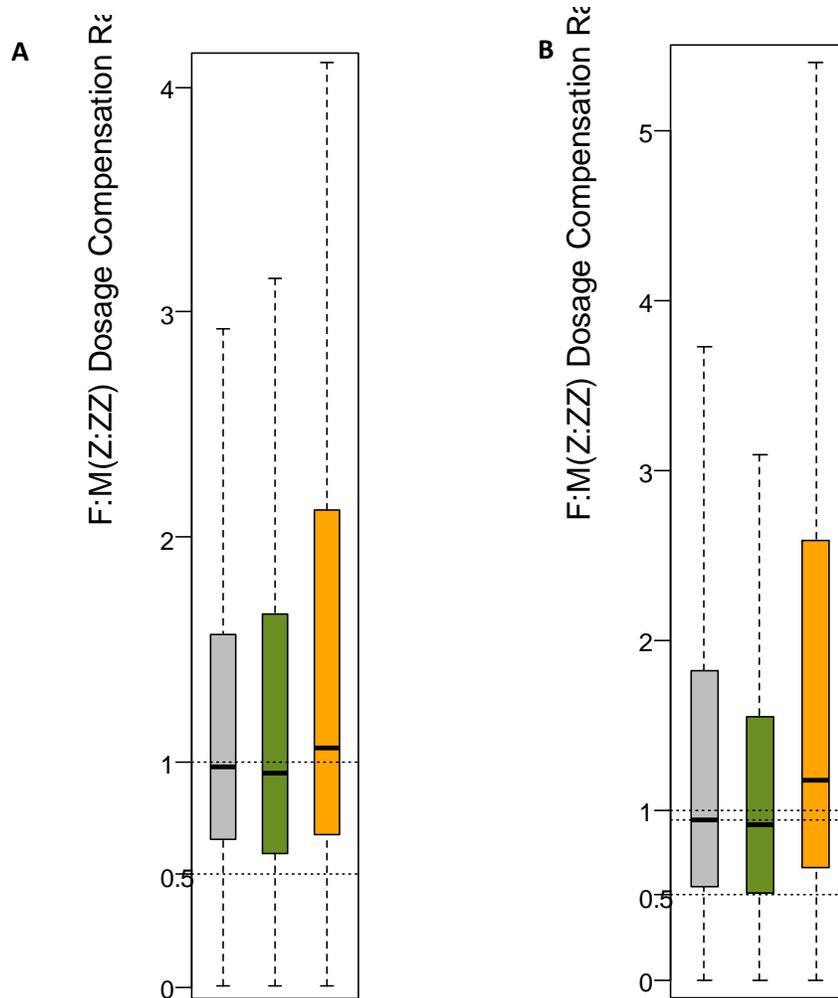


Fig. 11

$Z\sim ZZ$ compensation of *C. pomonella* neo-Z genes (green), indicated in R_{dc} using *M. sexta* (A) and *H. melpomene* (B) as references. Autosomal (grey) and ancl-Z (orange) genes are both plotted as controls. Plot boxes represent the median and interquartile range of R_{dc} . The whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range. Key reference values 1 (complete $Z\sim ZZ$ compensation) and 0.5 (absence of $Z\sim ZZ$ compensation) are denoted by dotted horizontal lines.

The median R_{dc} value provides an index of Z~ZZ compensation for neo-Z genes. This value (~0.95) was not significantly different from the autosomal control ($p = 0.44$, FDR corrected Mann–Whitney U test) (Fig. 11A). The median R_{dc} value for the ancl-Z control (~1.06) was not significantly different from either the autosomes or the neo-Z segment ($p = 0.16|0.16$, FDR corrected Mann–Whitney U test).

Neo-Z(Z) : $\boxed{\text{neo-ZZ}}$ ratios denote the extent of expression reduction of neo-Z genes in *C. pomonella* (Fig. 12). The neo-Z(Z) : $\boxed{\text{neo-ZZ}}$ median ratios were ~0.7 in both sexes, significantly lower than those of the autosomes (AA : $\boxed{\text{AA}}$) ($p < 5e-05$, Bonferroni-corrected Mann–Whitney U test). Further, the $\boxed{\text{neo-ZZ}}$: $\boxed{\text{AA}}$ median ratios were approximately 1.1 and significantly higher than those of neo-Z(Z):AA (bootstrap tests) (Fig. 13). The distribution of $\boxed{\text{neo-ZZ}}$ expression levels did not differ significantly from that of $\boxed{\text{AA}}$ (Fig. 14). By contrast, the distributions of neo-Z(Z) expression levels shifted significantly towards lower values compared to $\boxed{\text{neo-ZZ}}$, whereas these differences diminished for genes with $\log_2(\text{FPKM})$ values great than ~ 5, which comprised ~21-25 % of all neo-Z(Z)~ $\boxed{\text{neo-ZZ}}$ pairs. However, comparison between neo-Z(Z) and AA did not show the same trend.

By comparison, the ancl-Z(Z) : $\boxed{\text{ancl-Z(Z)}}$ median ratios were also found to be ~0.7 (Fig. 12), and $\boxed{\text{ancl-Z(Z)}}$: $\boxed{\text{AA}}$ median ratios were significantly higher than those of ancl-Z(Z):AA (bootstrap tests). Thus, the expression of ancl-Z genes in *C. pomonella* is reduced compared to the Z-linked genes in the reference species (Fig. 13).

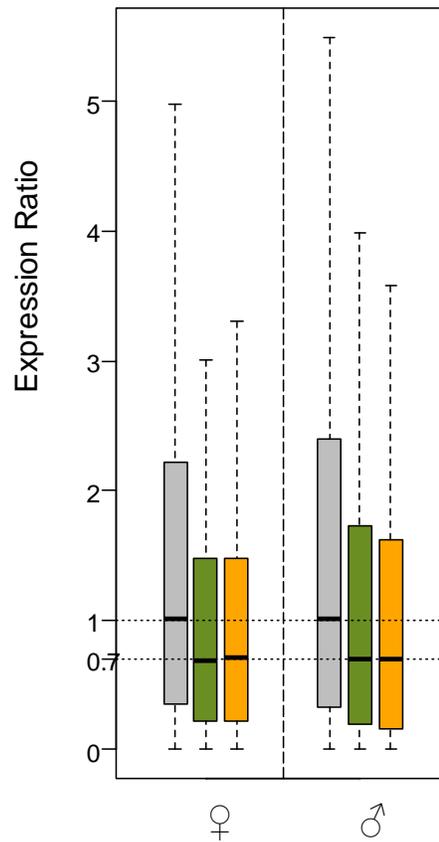


Fig. 12

Z~A compensation of *C. pomonella* neo-Z (green), as indicated by neo-Z(Z) : $\boxed{\text{neo-ZZ}}$ ratios (*C. pomonella*~*M. sexta*). Autosomal (grey) and ancl-Z (orange) genes are both plotted as controls. Plot boxes represent the median and interquartile range of ratios. The whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range. Both neo-Z(Z) and ancl-Z(Z) show 30% reduction in expression compared to $\boxed{\text{neo-ZZ}}$ and ancl-Z(Z), respectively.

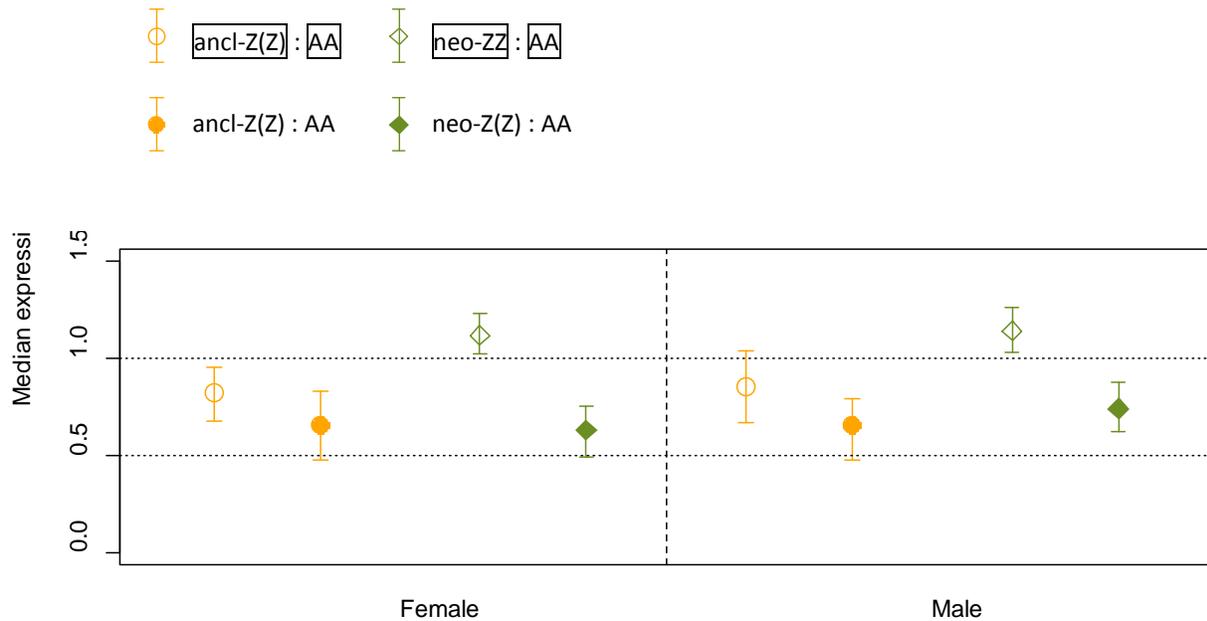


Fig. 13

Median expression ratios relative to the autosomes of ancl-Z (orange) and neo-Z (green) orthologous pairs identified in *C. pomonella* and *M. sexta*. Open circles and diamonds denote the expression ratios in *M. sexta*, while filled ones denote the expression ratios in *C. pomonella*. Error bars show 95% confidence intervals estimated by 1,000 bootstrap replicates. In both sexes, neo-Z(Z):AA ratios are significantly lower than $\boxed{\text{neo-ZZ}}:\boxed{\text{AA}}$.

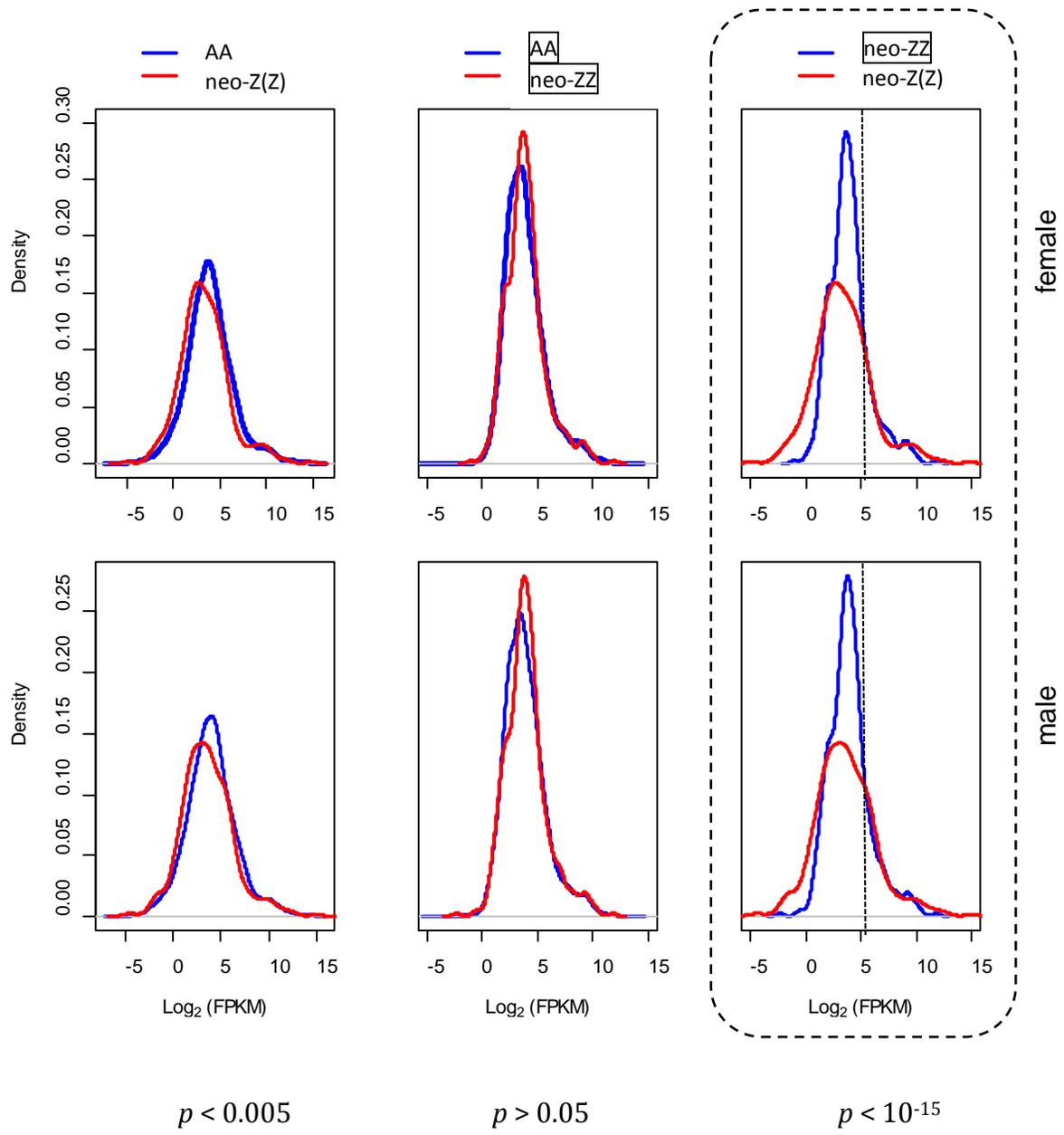


Fig. 14

Comparison of distributions of expression levels between neo-Z(Z) and neo-Z(Z). Bonferroni corrected p values from Komolgorov-Smirnov tests are shown below each category. There is no significant difference in distributions of expression levels between the neo-ZZ and AA (center column). However in both sexes, the distributions are shifted significantly toward neo-Z(Z) genes with \log_2 (FPKM) values <5 compared to the neo-ZZ distributions (right column).

Discussion

Comparative transcriptome analysis provides further confirmation of the dosage compensation pattern determined in *C. pomonella* head ($Z \approx ZZ < AA$), as described in Chapter III. In particular, a median expression ratio of 0.7 between neo-Z genes and their autosomal orthologs revealed approximate 30% reduction in overall expressional output of neo-Z compared to the ancestral level of its autosomal progenitor, consistent with imperfect Z~A compensation. Comparisons of distributions of expression levels suggested that the ~21~25% of genes with the highest expression levels receive compensation while other do not.

Although proto-Z gene expression for the conserved lepidopteran Z chromosome cannot be inferred as for the *C. pomonella* neo-Z segment, comparison of transcriptional output between the *C. pomonella* ancl-Z and the *M. sexta* Z chromosomes also showed ~30% reduction in *C. pomonella*, the more basal species. Curiously, comparison of results from separate studies showed that ancl-Z(Z):AA ratios from the same tissue (head) are also the lowest in *C. pomonella* (0.56|0.62), the most basal species examined, whereas the Z(Z):AA ratios are 0.761|0.766 in *B. mori* (Walters and Hardcastle 2011), 0.8|0.83 in *M. sexta* (Smith, Chen et al. 2014) and 0.63,0.69 in *H. melpomene* (Walters et al., unpublished data). Furthermore, in *C. pomonella* ancl-Z(Z):AA ratios were also generally lower than neo-Z(Z):AA ratios. These two lines of evidence indicate that in Lepidoptera, gene expression is reduced to a lesser extent for younger Z chromosomes compared to older ones. Thus I suggest that the Z~A compensation might be an ongoing evolutionary process in this order.

In *B.mori*, the ZZ:AA ratio in the testis is approximately 1, the highest among all the tissues surveyed (Walters and Hardcastle 2011). This value is consistent with the expected proto-ZZ:AA ratio of 1, suggesting the constitutive Z expression in the testis. Similarly, in *C. pomonella* the testis and accessory gland also had the highest ancl-Z(Z):AA and ancl-Z(Z):AA ratios compared to the other tissues. On one hand, the ancl-ZZ:AA ratio (0.82) in the testis was lower than that in *B. mori*, consistent with the afore-suggested evolutionary progression. On the other hand, neo-ZZ:AA ratios in the testis and accessory gland (1.13 and 1.14, respectively) were in in good agreement with the expected proto-ZZ:AA inferred by $\frac{\text{neo-ZZ}}{\text{AA}}$ ratios (1.12|1.14 for female|male) in the head. If the relative expression of proto-Z and autosomal genes is uniform across tissues then in *C. pomonella* both ancl-Z and neo-Z genes are expressed at constitutive levels in the male reproductive tissues.

Furthermore, it is worth noting that the $\frac{\text{neo-ZZ}}{\text{AA}}$ median ratios (~ 1.1) were significantly higher (bootstrap tests) than the expected proto-ZZ:AA value (1). This provides further evidence for the substantial variation in median expression among autosomes (Walters and Hardcastle 2011, Julien, Brawand et al. 2012) and strengthens the conclusion that it is more precise and reliable to compare Z expression with inferred proto-Z expression than with present-day autosomal expression.

The interspecific transcriptome approach allows the direct comparison of gene expression between neo-Z and its autosomal progenitor. It is thus not confounded by the disproportionately large number of weakly-expressed genes on the Z chromosome compared to the autosomes, which causes variation in Z(Z):AA ratios depending on data filtering stringency (Chapter III). This approach further enables the quantitative evaluation

of Z~A compensation. This is especially helpful in lepidopteran species where the Z~A compensation tends to fall between the complete (1) and absent (0.5) states. In particular, this method distinguished the complete compensation of highly expressed genes in *C. pomonella*, which was not detected through the comparison with present-day autosomes (Chapter III).

Variable Z(Z):AA ratios in different tissues suggest that Z~A compensation acts in a tissue-dependent manner. Unfortunately, so far RNAseq data is available only from the head for use in comparative assessments of transcriptome-based dosage compensation. The generation of additional datasets from other tissues in other lepidopterans will provide a more complete picture of Z~A compensation in tissues in which a small neo-Z(Z):AA disparity is predicted (e.g., the midgut).

Chapter V

Nonrandom genomic distribution of tissue-specific genes in *Cydia pomonella*: implications for dosage compensation

Introduction

In Chapter IV I proposed that in *C. pomonella* testis and accessory gland both ancl-Z and neo-Z genes were expressed at constitutive levels compared to their respective proto-Z chromosomes. Compared to the autosomes, however, sex chromosomes have different complements of genes that are expressed predominantly in one sex (sex-biased genes), especially in the germline (Vicoso and Charlesworth 2006). In mammals, these genes typically have low expression levels and no obvious function in the soma. Thus sex-biased genes contribute to the disproportionate reduction of Z chromosome expression in the soma (Deng, Hiatt et al. 2011, Julien, Brawand et al. 2012). In *B. mori*, the Z chromosome is enriched in testis-specific genes (Arunkumar, Mita et al. 2009, Walters and Hardcastle 2011), which is hypothesized by Walters and Hardcastle (2011) to account for the high ZZ:AA ratio in testis compared to the other tissues examined. Nevertheless, the authors did not test this.

The skewed gene content of the Z chromosome, if it also occurs in *C. pomonella* and other lepidopteran species, could further complicate the analysis of dosage compensation. On one hand, the high Z(Z):AA ratios in the reproductive tissues may reflect enrichment on the Z

chromosome of genes expressed predominantly in these tissues rather than dosage compensation, as suggested in *B. mori*. On the other hand, these genes may also result in the disproportionate reductions of the ancl-Z(Z):AA/neo-Z(Z):AA ratios observed in the head, midgut and ovary (Chapter III).

Therefore, it is important to explore the relationship between dosage compensation and the distributions of tissue-specific genes in the *C. pomonella* genome, especially testis- and accessory gland-specific genes. To address this question, I used DE analysis to identify tissue-specific genes in *C. pomonella* and compared the distribution of these tissue-specific genes among the ancl-Z and neo-Z segments and the autosomes. Furthermore, I tested the possible influence of tissue-specific gene expression on the dosage compensation patterns which were described in the previous chapters.

Materials and Methods

Results of differential expression via pairwise comparisons between five *C. pomonella* adult tissues (head, midgut, ovary, testis and accessory gland) from Chapter II were used. Tissue-specific genes are identified as having consensus up-regulation in one tissue relative to the other four tissues at a given FC cutoff (FDR < 0.05, *p* value for FDR < 0.05). The higher the FC cutoff, the higher bias in expression and therefore the more specialized function. The numbers of identified tissue-specific genes were then normalized to the total numbers of expressed genes (Table 1, Chapter II) on the ancl-Z and neo-Z segments and autosomes in

each tissue. Fisher's exact test was used to assess the significance of differences between proportions of tissue-specific genes on ancl-Z, neo-Z and the autosomes.

Results

*Nonrandom distribution of tissue-specific genes in the *C. pomonella* genome*

In general, both the Z chromosome and the autosomes harbored more testis-specific genes than other tissue-specific genes (Fig. 15). However, the ancl-Z segment had a larger proportion of testis-specific genes than the autosomes at any FC cutoff value ($p < 0.05$, Fisher's exact test). This effect was greater at higher cutoff values. At FC =128, in particular, the proportion of testis-specific genes was twice as high on the ancl-Z segment as on the autosomes. In contrast, head- and midgut-specific genes were significantly underrepresented on the ancl-Z segment ($p < 0.05$, Fisher's exact test). Ovary-specific genes appeared to be neither over- nor under-represented on both the ancl-Z and the neo-Z segments.

Accessory gland-specific genes appeared to be enriched on both the ancl-Z and the neo-Z segments relative to the autosomes, but the differences were not statistically significant. Similarly, for testis-, head- and midgut-specific genes the neo-Z segment appeared to follow the same pattern as the ancl-Z segment but to a lesser and statistically insignificant extent.

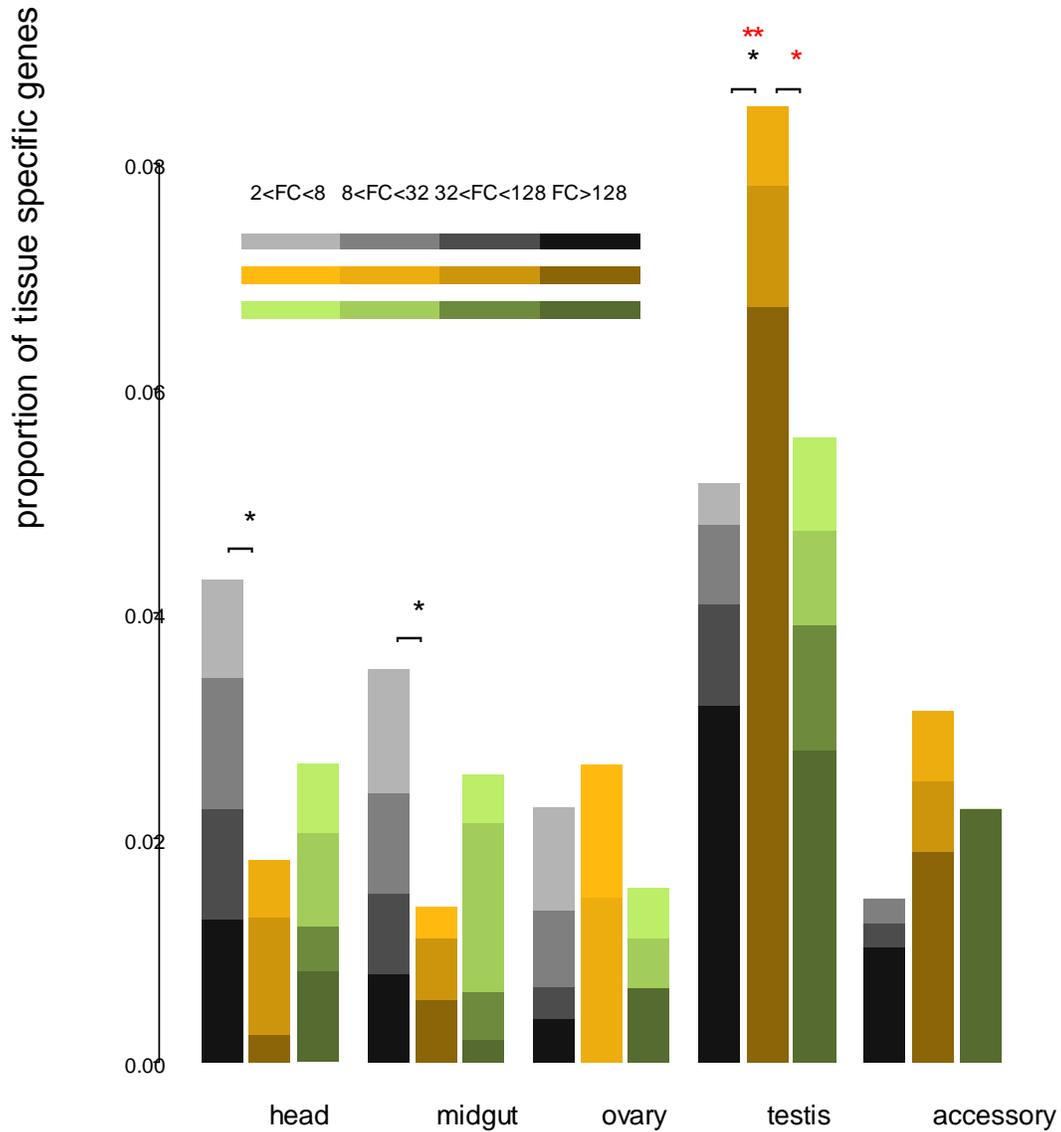


Fig. 15

Proportions of tissue-specific genes on *C. pomonella* autosomes (grey), the ancl-Z (orange) and the neo-Z (green) identified at different FC cutoffs from 2 to 128. Significance of differences assessed by two tailed Fisher's exact test are indicated: * $p < 0.05$; ** $p < 0.01$ | * FC cutoff = 2; **, ** FC cutoff = 128.

Assessment of dosage compensation is not confounded by the unusual gene content on the Z chromosome

Highly biased expression usually also coincides with high expression levels. Indeed, testis-specific genes were among the most highly-expressed genes in the testis (Fig. 16). However, despite the disparate distribution patterns of tissue-specific genes across the genome, overall expression of these genes was comparable between the Z chromosome and the autosomes. Therefore, computational removal of all the identified 1,247 tissue-specific genes at FC=2 (Fig. 17) did not appreciably change the ZZ:AA ratios as described previously in Chapter II (see Fig.9).

Discussion

The distribution of tissue-specific genes in the *C. pomonella* genome was non-random, so that different patterns were found on the ancl-Z and neo-Z chromosomal segments and the autosomes. In addition, the large proportion of testis-specific genes with high expression and small proportion of ovary-specific genes generally in the genome were in good agreement with the conclusions drawn from the expression profiles in Chapter II.

Nevertheless, the pattern of Z(Z):AA ratios across *C. pomonella* tissues was unchanged when tissue-specific genes were excluded from the analysis. This ensures that the pattern of dosage compensation described in previous chapters was not confounded by the unique gene content of the Z chromosome. In particular, Z expression levels were confirmed to be constitutive in the reproductive tissues, indicating the exempt from dosage compensation.

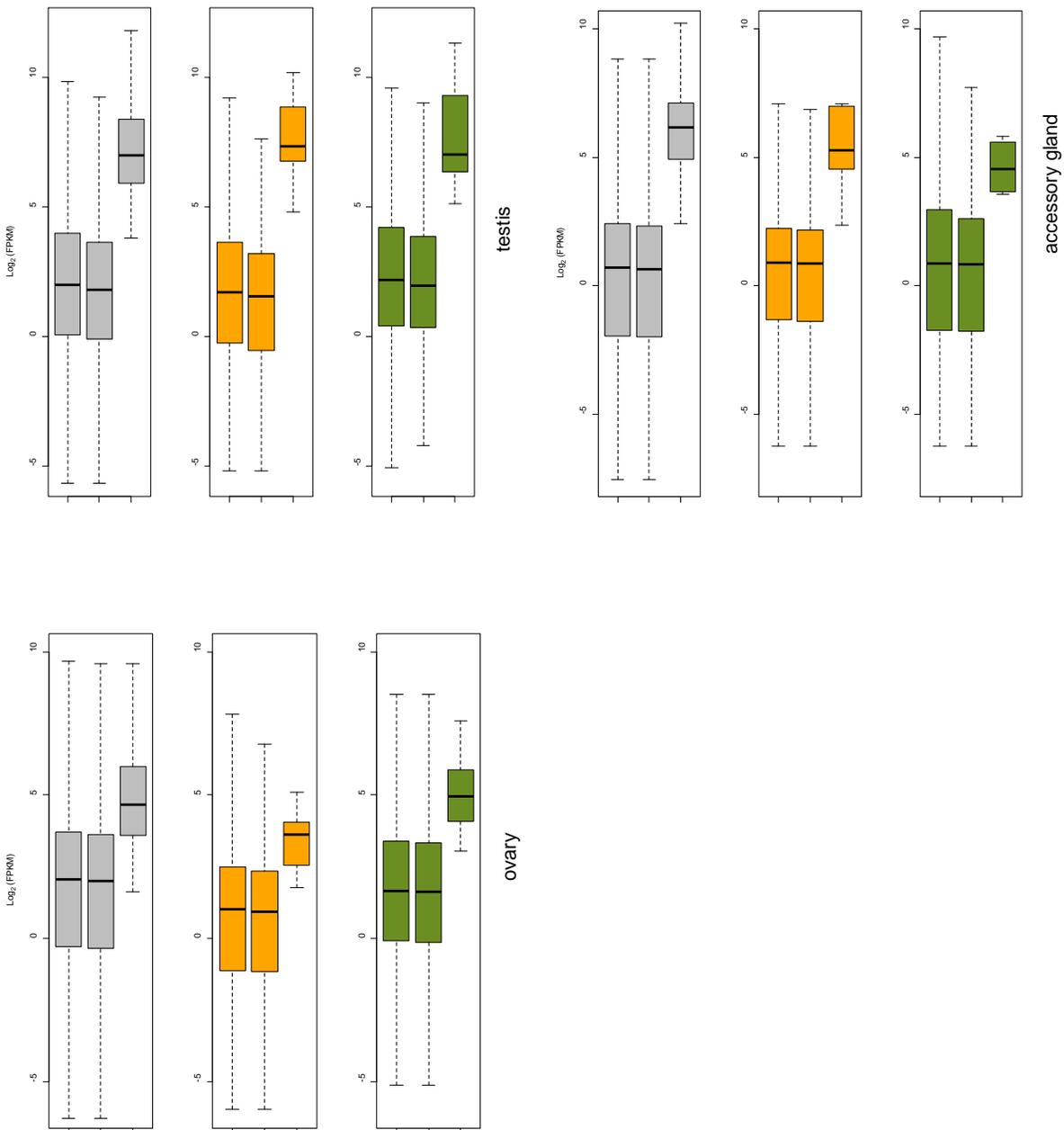
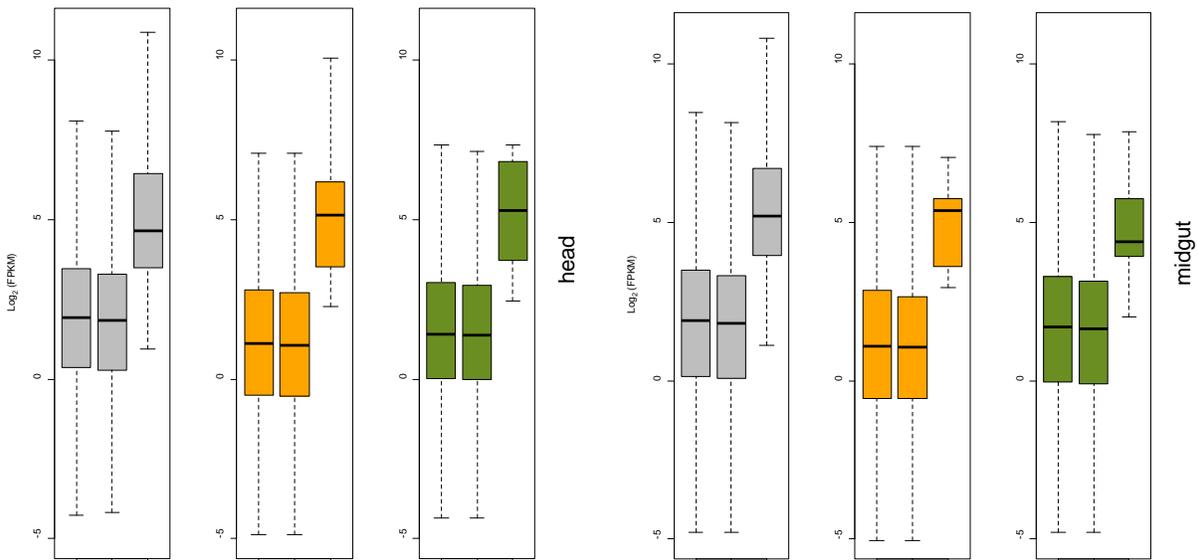


Fig. 16

Expression levels (\log_2 transformed FPKM) of subset of genes in each tissue by chromosomal linkage. Grey: autosomal genes; orange: ancl-Z genes; green: neo-Z genes. 3 boxplots of each chromosome location are (from left to right): all expressed genes

(FPKM>0); all expressed genes with tissue-specific genes removed; tissue-specific genes.

Each plot box represents the median and interquartile range of expression levels. The whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range.



(Fig. 16 continued)

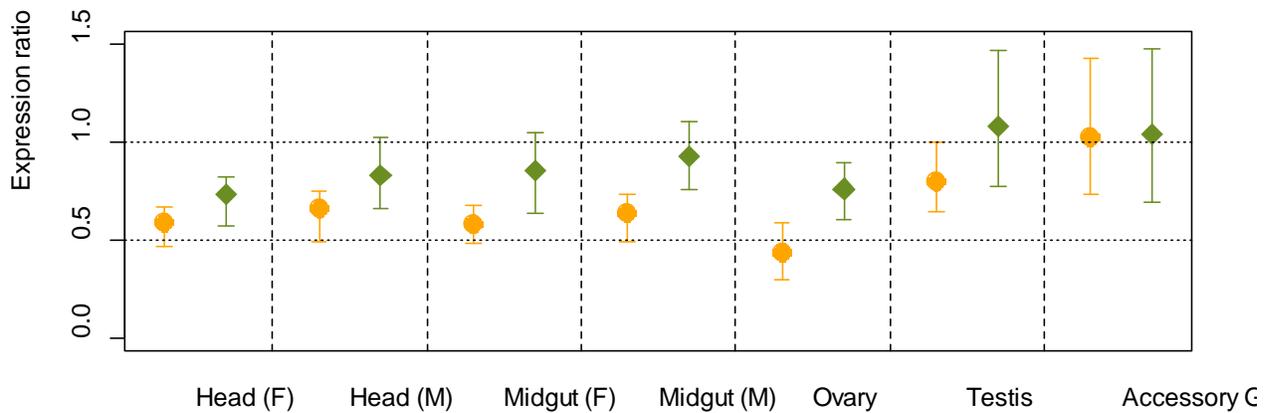


Fig. 17

Z(Z):AA median expression ratios in *C. pomonella* tissues disregarding all tissue-specific genes. Orange circles: ancl-Z; green diamond: neo-Z. Error bars show 95% confidence intervals estimated by 1000 bootstrap replicates. The horizontal dotted lines denote the reference ratios: 0.5 suggests equal Z(Z) expression to monoallelic autosomal expression (A); 1 suggests equal Z(Z) expression to biallelic autosomal expression (AA). Note the pattern is comparable with Fig. 9 in Chapter III.

Sex-biased gene expression is typically associated with sexual dimorphism. In particular, a substantial fraction of sex-biased genes are predominantly expressed in the germline (Ellegren and Parsch 2007). Therefore, germline-specific genes represent a substantial fraction of sex-biased genes. Sexually antagonistic selection is the leading theory to explain the nonrandom distribution of sex-biased genes between the sex chromosome and autosomes (Rice 1984, Ellegren and Parsch 2007). Compared to autosomes, sex chromosomes have a special mode of inheritance and organization between male and female, and hence confer different cost-benefit scenarios associated with sex linkage. The sex chromosome is favored by at least partially dominant antagonistic mutations that are beneficial to the homogametic sex and detrimental to the heterogametic sex, while disfavored by those beneficial to the heterogametic sex and detrimental to the homogametic sex. The reverse holds true for recessive mutations. For female heterogametic systems (ZZ/ZW) such as *C. pomonella*, the Z chromosome would be predicted to have an excess of male-beneficial genes and a deficit of female-beneficial genes, provided that dominant mutations occur more often than recessive mutations. The overrepresentation of testis-specific, and possibly accessory gland-specific (both male-biased) genes on *C. pomonella* Z chromosome is consistent with this prediction. However, the distribution pattern of ovary-specific (female-biased) genes is not.

Sexually antagonistic selection, however, cannot explain the under-representation of *C. pomonella* Z-linked genes that are predominantly expressed in the non-reproductive somatic tissues of either sex because these genes do not involve sex-biased expression. Hence I hypothesize that insufficient up-regulation of Z-linked genes (Z~A compensation)

would make the Z chromosome a suboptimal environment for genes that require high expression in non sex-specific soma and drive selection for their relocation to the autosomes. In *Drosophila* the general paucity on the X chromosome of tissue-biased genes that lack an apparent role in reproduction is also attributed to dosage compensation (Mikhaylova and Nurminsky 2011). On the contrary, the situation would be different for sex-specific genes that are primarily expressed in the reproductive tissues. A much larger proportion of genes expressed in the germline should not require similar expression between sexes, thus reducing the selective pressure on dosage compensation in gonadal tissues. Similarly, the small repertoire of genes expressed in the accessory gland carries specialized functions and shares no counterpart in the female, which obviates selection for dosage compensation as well. These explanations are consistent with the absence of dosage compensation in the reproductive tissues shown in Chapter III and further confirmed in this chapter. Conversely, absence of dosage compensation in male-specific tissues would also drive the accumulation of genes that require high expression in these tissues, most of which are appreciably tissue-specific genes. Therefore, the remarkably high bias of testis-specific genes on the Z chromosome in *C. pomonella* could reflect the synergistic effects of both sexually antagonistic selection and the absence of dosage compensation in the male reproductive tissues.

The neo-Z chromosome appeared to follow the same pattern as the ancl-Z chromosome with respect to the distribution of tissues-specific genes but to a lesser and statistically insignificant extent. It is noteworthy that I used edgeR to identify DE genes under stringent FDR and p values, so the number of tissue-specific genes retrieved would be estimated

conservatively. The pattern of distribution of accessory gland-specific genes in the genome was also not statistically significant, possibly owing to the much smaller number of total expressed genes in the accessory gland *per se*, which would compromise the statistical resolution.

The comparison of tissues-specific gene distribution between the neo-Z and ancl-Z chromosomal segments not only suggests the impact of sex-linkage on tissue-biased expression but also provides further evidence that the shaping of gene content on the sex chromosomes is an evolutionary process (Vicoso and Charlesworth 2006). Curiously, the neo-Z segment exhibited the same pattern as the ancl-Z segment only for soma-specific (head, midgut and accessory gland) but not gonad-specific (ovary and testis) genes. This implies that the selection for tissue-biased expression may be weaker in the gonads than in the soma.

Chapter VI

Summary and Conclusions

Dosage compensation in Lepidoptera

The *de novo* transcriptome assembly from RNAseq data showed that in *C. pomonella* the pattern of dosage compensation in the head and midgut exhibited equalized Z-linked gene expression between sexes (complete Z~ZZ compensation) but overall reduced Z gene expression relative to the autosomes ($Z \approx ZZ < AA$). This pattern was in accord with that found in 5 somatic tissues of *B. mori* (Walters and Hardcastle 2011), the head of *M. sexta* (Smith, Chen et al. 2014) and the whole body without abdomen of two *Heliconius* species (Walters et al, unpublished data).

The overall reduction of transcriptional output from homozygotic Z chromosomes in males indicates that Lepidopterans achieve complete Z~ZZ compensation via down-regulation of Z-linked gene expression in males. This is consistent with a recent study investigating sex determination in *B. mori* (Kiuchi, Koga et al. 2014). In embryos injected with short interfering RNA targeting the *Masc* locus, Z chromosome-derived transcripts showed higher expression in males than in females, indicating that the Masc protein is involved in global repression of Z-linked gene expression during the embryonic stage. I suggest that the most parsimonious way to downregulate the bi-allelic Z expression in males to match the mono-allelic Z expression in females would be to silence one of the two Z chromosome copies in males.

In *C. pomonella*, the extent of Z~A compensation varied among tissues and was generally more complete on the neo-Z chromosomal segment than on the ancl-Z segment. In particular, ancl-Z expression in all three female tissues resembles mono-allelic autosomal expression, suggesting a lack of Z~A compensation. Comparative transcriptome analysis using outgroup species further confirmed the imperfect Z~A compensation on the neo-Z chromosome in the head as well as complete Z~ZZ compensation. The transcriptional output of the neo-Z segment was reduced by ~30% compared to that of its inferred autosomal progenitor. Comparison of distributions of expression levels further revealed that only a minority of neo-Z genes, most of which are highly expressed, are fully compensated. The expression of ancl-Z-linked genes was also reduced compared to that in more derived species. Z(Z):AA expression on the ancl-Z segment in *C. pomonella* was lower than that of the more derived species. These observations suggest that, in contrast to Z~ZZ compensation, Z~A compensation may be an ongoing evolutionary process in Lepidoptera. Furthermore, in *C. pomonella* the deficit of head- and midgut-specific genes on the ancl-Z segment implies an alternative strategy for circumventing global Z~A expression, whereas the similar trend to a lesser extent on the neo-Z segment also implies the evolutionary process.

In *C. pomonella* male reproductive tissues, both ancl-Z and neo-Z genes were expressed constitutively at levels comparable to those of autosomal genes. The high levels of ancl-Z and neo-Z expression were not accounted for by the excess of testis-specific genes in this tissue. This was also consistent with the lack of selective pressure for dosage compensation suggested by the distinct expression profiles in reproductive tissues. Similarly, in *B. mori*

the ZZ:AA median ratio is also 1 in the testis (Walters and Hardcastle 2011). Together these lines of evidence support the lack of dosage compensation in the reproductive tissues of lepidopterans.

The dosage compensation pattern ($Z < ZZ \approx AA$) reported in *P. interpunctella* (Harrison, Mank et al. 2012) does not agree with the other Lepidopteran species examined so far. However, the conclusions drawn in that study are subject to question primarily because the authors did not isolate the reproductive tissues. Like in other taxa, the mechanism of dosage compensation is expected also to be conserved in light of the conservation of the Z chromosome in Lepidoptera. Nevertheless, a broader sampling among moth and butterfly lineages will give a more conclusive answer.

In conclusion, Lepidopterans globally equalize sex-linked gene expression between sexes, a phenomenon previously described only in XX/XY species. In particular, dosage compensation in lepidopterans are intriguingly analogous to that in mammals, i.e. Z inactivation of one Z chromosome copy in male soma which is restored in male reproductive tissues, accompanied by imperfect Z~A compensation.

Differences between Lepidoptera and other independently evolved ZZ/ZW taxa

The anecdotal claim of incomplete dosage compensation in the Lepidoptera from two previous studies (Zha, Xia et al. 2009, Harrison, Mank et al. 2012) has led to the popular speculation that incomplete dosage compensation is associated with female heterogamety (Graves and Disteché 2007, Mank 2009, Vicoso and Bachtrog 2009, Naurin, Hansson et al.

2010, Wolf and Bryk 2011, Mank 2013). This generalization is not supported by the increasing body of evidence provided in this study and elsewhere ((Walters and Hardcastle 2011, Kiuchi, Koga et al. 2014, Smith, Chen et al. 2014); Walters et al., unpublished data), which points to the departure of the Lepidoptera from the other ZZ/ZW systems studied so far. In fact, careful scrutiny reveals striking differences apart from the shared female heterogamety between the Lepidoptera and the other ZZ/ZW taxa.

Abundant evidence supports the theory that degeneration of the heterogametic sex chromosome is a progressive process that occurs over an extended period of time (Charlesworth, Charlesworth et al. 2005). Different stages of W chromosome degeneration are manifested in different avian and serpent lineages. For instance, the basal ratite birds, such as the emu, display one extreme in which little chromosomal differentiation has occurred. By contrast in most bird species, such as the chicken, the W chromosome is almost completely degenerate (Graves 2014). Among snakes, the Boidae have entirely homomorphic sex chromosomes and the Viperidae have completely heteromorphic sex chromosomes, whereas the Colubridae show partial differentiation (Becak and Becak 1969, Vicoso, Emerson et al. 2013). Furthermore, no bird or snake species has been reported to lack a W chromosome (Marshall Graves 2008, Graves 2014). Whole genome sequencing of the fish *C. semilaevis* revealed that the nascent ZW chromosomes are in the initial stages of sex chromosome differentiation and share the same ancestry with the ZW chromosomes of birds (Chen, Zhang et al. 2014). In contrast, degeneration of the sex chromosomes is far more advanced in their vertebrate relative, the XY mammals. No mammal shows minimal XY differentiation comparable to the homomorphic ZW of ratite birds and boid snakes.

Some rare rodent groups have even completely lost the Y chromosome (Marshall Graves 2008, Graves 2014). In addition to the vertebrate ZZ/ZW systems, the trematode *S. mansoni* also shows minimal differentiation between the Z-W pair: the pseudoautosomal region (PAR) takes up 70 to 90% of the Z-W pair, and the W-specific region is composed of solely 36 satellite repeat families (Lepesant, Cosseau et al. 2012). An early karyotype survey suggests a progression of sex chromosome differentiation within the family Schistosomatidae (Grossman, Short et al. 1981), similar to that seen in birds and snakes.

The Lepidoptera represent the extreme end of the continuum of sex chromosome differentiation and sex chromosome heterogamety. All lepidopteran species in the basal lineages that have been examined exhibit Z/ZZ systems, whereas the W chromosome in more derived species is likely to be a secondary acquisition (Sahara, Yoshido et al. 2012). For those species possessing a W chromosome, there is neither genetic exchange between the Z-W pair nor an extant PAR. Information on genes located on the lepidopteran W chromosome is generally scarce. In *B. mori*, the W chromosome is replete with repetitive elements and lacks identified functional genes. The W chromosome is hypothesized to have originated either from an autosome whose homolog was fused to the ancestral Z chromosome (Traut and Marec 1996), or from a B chromosome, a supernumerary small chromosome, which was incorporated in the sex chromosome system by acquiring pairing partnership with the Z chromosome (Lukhtanov 2000). In light of the absence of any trace of degeneration of the W chromosome, the latter hypothesis seems more likely. By comparison, the *D. melanogaster* Y chromosome is small and entirely heterochromatic (Bridges 1916, Brosseau 1960), but all of the Y-linked genes identified were acquired from

autosomes, rather than from a degenerate X homolog (Carvalho 2002). Thus the *Drosophila* Y chromosome was also suggested to derive from a B chromosome, rather than correspond to a degenerated X chromosome (Hackstein, Hochstenbach et al. 1996, Bernardo Carvalho, Koerich et al. 2009).

Chromosomal translocations are typically rare among most ZZ/ZW taxa, with so far only one reported autosome-Z fusion in birds (Pala, Naurin et al. 2012). In contrast, both chromosome fusion and fission events are quite common in Lepidoptera, resulting in substantial karyotype variation ($n = 5-223$) from the inferred ancestral haploid chromosomes number of 31 (Heliconius Genome 2012, Sahara, Yoshido et al. 2012, Ahola, Lehtonen et al. 2014). Accordingly, various sex chromosome constitutions also occur throughout lepidopteran lineages (Marec, Sahara et al. 2010, Sahara, Yoshido et al. 2012). In addition to the neo-Z chromosomes, in *C. pomonella*, $W1W2Z/ZZ$ and $WZ1Z2/Z1Z1Z2Z2$ systems also exist, in some cases among different populations of the same species. Chromosome translocation is probably facilitated by the dispersed kinetochores associated with their holocentric chromosomes (Wolf 1996, Marec, Sahara et al. 2010). It is therefore not surprising that Lepidopterans have a systemic and efficient dosage compensation system to accommodate the variable sex chromosomes schemes. The dosage compensation of the neo-Z chromosome in *C. pomonella*, which is described in this study, provides a prime case in point. A similar example is also found within the much younger *Drosophila* genus, where autosomes have also repeatedly fused to the ancestral sex chromosomes, independently giving rise to the neo-X chromosomes (Kaiser and Bachtrog 2010). Among those, the neo-X chromosome of *D. pseudoobscura* has acquired full dosage compensation

by the same molecular mechanisms used by the ancestral *Drosophila* X chromosome (Abraham and Lucchesi 1974, Bone and Kuroda 1996, Marin, Franke et al. 1996, Nozawa, Fukuda et al. 2014).

In conclusion, what fundamentally sets Lepidoptera apart from the other ZZ/ZW systems is that its members are generally (if not all) in the terminal stage/phase of sex chromosome differentiation to such an extent that the W chromosome, the pairing partner of the Z chromosome is even lost in many lineages. Similarly, XX/XY taxa with complete dosage compensation are also typically associated with highly differentiated heterogametic sex chromosomes. The efficient global dosage compensation mechanism in the Lepidoptera, similar to that in the genus *Drosophila*, may facilitate the translocations involving sex chromosomes among the lineages. The uniqueness of the Lepidoptera among the other ZZ/ZW systems with regard to dosage compensation lends further support to a recent suggestion that dosage compensation is the prerequisite for, rather than just the consequence of, sex chromosome evolution (Adolfsson and Ellegren 2013).

References

Abraham, I. and J. C. Lucchesi (1974). "Dosage compensation of genes on the left and right arms of the X chromosome of *Drosophila pseudoobscura* and *Drosophila willistoni*." Genetics **78**(4): 1119-1126.

Adolfsson, S. and H. Ellegren (2013). "Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches." Mol Biol Evol **30**(4): 806-810.

Ahola, V., R. Lehtonen, P. Somervuo, L. Salmela, P. Koskinen, P. Rastas, N. Valimaki, L. Paulin, J. Kvist, N. Wahlberg, J. Tanskanen, E. A. Hornett, L. C. Ferguson, S. Luo, Z. Cao, M. A. de Jong, A. Duploux, O. P. Smolander, H. Vogel, R. C. McCoy, K. Qian, W. S. Chong, Q. Zhang, F. Ahmad, J. K. Haukka, A. Joshi, J. Salojarvi, C. W. Wheat, E. Grosse-Wilde, D. Hughes, R. Katainen, E. Pitkanen, J. Ylinen, R. M. Waterhouse, M. Turunen, A. Vaharautio, S. P. Ojanen, A. H. Schulman, M. Taipale, D. Lawson, E. Ukkonen, V. Makinen, M. R. Goldsmith, L. Holm, P. Auvinen, M. J. Frilander and I. Hanski (2014). "The *Glanville fritillary* genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera." Nat Commun **5**: 4737.

Albritton, S. E., A. L. Kranz, P. Rao, M. Kramer, C. Dieterich and S. Ercan (2014). "Sex-Biased Gene Expression and Evolution of the X Chromosome in Nematodes." Genetics **197**(3): 865-883.

Arunkumar, K. P., K. Mita and J. Nagaraju (2009). "The silkworm Z chromosome is enriched in testis-specific genes." Genetics **182**(2): 493-501.

Avila, F. W., L. K. Sirot, B. A. LaFlamme, C. D. Rubinstein and M. F. Wolfner (2011). "Insect seminal fluid proteins: identification and function." Annu Rev Entomol **56**: 21-40.

Baxter, S. W., J. W. Davey, J. S. Johnston, A. M. Shelton, D. G. Heckel, C. D. Jiggins and M. L. Blaxter (2011). "Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism." PLoS One **6**(4): e19315.

Bean, C. J., C. E. Schaner and W. G. Kelly (2004). "Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*." Nat Genet **36**(1): 100-105.

Becak, W. and M. L. Becak (1969). "Cytotaxonomy and chromosomal evolution in Serpentes." Cytogenetics **8**(4): 247-262.

Beldade, P., S. V. Saenko, N. Pul and A. D. Long (2009). "A gene-based linkage map for *Bicyclus anynana* butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome." PLoS Genet **5**(2): e1000366.

Berlin, S. and H. Ellegren (2006). "Fast accumulation of nonsynonymous mutations on the female-specific W chromosome in birds." J Mol Evol **62**(1): 66-72.

Berlin, S., D. Tomaras and B. Charlesworth (2007). "Low mitochondrial variability in birds may indicate Hill-Robertson effects on the W chromosome." Heredity (Edinb) **99**(4): 389-396.

Bernardo Carvalho, A., L. B. Koerich and A. G. Clark (2009). "Origin and evolution of Y chromosomes: *Drosophila* tales." Trends Genet **25**(6): 270-277.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Bone, J. R. and M. I. Kuroda (1996). "Dosage compensation regulatory proteins and the evolution of sex chromosomes in *Drosophila*." Genetics **144**(2): 705-713.

Bridges, C. B. (1916). "Non-disjunction as proof of the chromosome theory of heredity." Genetics **1**(3): 309-309.

Brosseau, G. E. (1960). "Genetic Analysis of the Male Fertility Factors on the Y Chromosome of *Drosophila Melanogaster*." Genetics **45**(3): 257-274.

Carvalho, A. B. (2002). "Origin and evolution of the *Drosophila* Y chromosome." Curr Opin Genet Dev **12**(6): 664-668.

Charlesworth, B. (1978). "Model for evolution of Y chromosomes and dosage compensation." Proc Natl Acad Sci U S A **75**(11): 5618-5622.

Charlesworth, D., B. Charlesworth and G. Marais (2005). "Steps in the evolution of heteromorphic sex chromosomes." Heredity (Edinb) **95**(2): 118-128.

Chen, S., G. Zhang, C. Shao, Q. Huang, G. Liu, P. Zhang, W. Song, N. An, D. Chalopin, J. N. Volff, Y. Hong, Q. Li, Z. Sha, H. Zhou, M. Xie, Q. Yu, Y. Liu, H. Xiang, N. Wang, K. Wu, C. Yang, Q. Zhou, X. Liao, L. Yang, Q. Hu, J. Zhang, L. Meng, L. Jin, Y. Tian, J. Lian, J. Yang, G. Miao, S. Liu, Z. Liang,

F. Yan, Y. Li, B. Sun, H. Zhang, J. Zhang, Y. Zhu, M. Du, Y. Zhao, M. Schartl, Q. Tang and J. Wang (2014). "Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle." Nat Genet **46**(3): 253-260.

Civetta, A. and R. S. Singh (1995). "High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species." J Mol Evol **41**(6): 1085-1095.

Conrad, T. and A. Akhtar (2012). "Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription." Nat Rev Genet **13**(2): 123-134.

d'Alencon, E., H. Sezutsu, F. Legeai, E. Permal, S. Bernard-Samain, S. Gimenez, C. Gagneur, F. Cousserans, M. Shimomura, A. Brun-Barale, T. Flutre, A. Couloux, P. East, K. Gordon, K. Mita, H. Quesneville, P. Fournier and R. Feyereisen (2010). "Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements." Proc Natl Acad Sci U S A **107**(17): 7680-7685.

Deng, X., J. B. Hiatt, D. K. Nguyen, S. Ercan, D. Sturgill, L. W. Hillier, F. Schlesinger, C. A. Davis, V. J. Reinke, T. R. Gingeras, J. Shendure, R. H. Waterston, B. Oliver, J. D. Lieb and C. M. Disteche (2011). "Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*." Nat Genet **43**(12): 1179-1185.

Ellegren, H., L. Hultin-Rosenberg, B. Brunstrom, L. Dencker, K. Kultima and B. Scholz (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." Nat Rev Genet **8**(9): 689-698.

Ferrari, F., A. A. Alekseyenko, P. J. Park and M. I. Kuroda (2014). "Transcriptional control of a whole chromosome: emerging models for dosage compensation." Nat Struct Mol Biol **21**(2): 118-125.

Fukova, I., W. Traut, M. Vitkova, P. Nguyen, S. Kubickova and F. Marec (2007). "Probing the W chromosome of the codling moth, *Cydia pomonella*, with sequences from microdissected sex chromatin." Chromosoma **116**(2): 135-145.

Gelbart, M. E. and M. I. Kuroda (2009). "*Drosophila* dosage compensation: a complex voyage to the X chromosome." Development **136**(9): 1399-1410.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev (2011). "Full-length transcriptome assembly from RNA-Seq data without a reference genome." Nat Biotechnol **29**(7): 644-652.

Graves, J. A. (2014). "Avian sex, sex chromosomes, and dosage compensation in the age of genomics." Chromosome Res **22**(1): 45-57.

Graves, J. A. and C. M. Disteché (2007). "Does gene dosage really matter?" J Biol **6**(1): 1.

Grossman, A. I., R. B. Short and G. D. Cain (1981). "Karyotype evolution and sex chromosome differentiation in *Schistosomes* (Trematoda, Schistosomatidae)." Chromosoma **84**(3): 413-430.

Guioli, S., R. Lovell-Badge and J. M. Turner (2012). "Error-prone ZW pairing and no evidence for meiotic sex chromosome inactivation in the chicken germ line." PLoS Genet **8**(3): e1002560.

Gupta, V., M. Parisi, D. Sturgill, R. Nuttall, M. Doctolero, O. K. Dudko, J. D. Malley, P. S. Eastman and B. Oliver (2006). "Global analysis of X-chromosome dosage compensation." J Biol **5**(1): 3.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman and A. Regev (2013). "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis." Nat Protoc **8**(8): 1494-1512.

Hackstein, J. H., R. Hochstenbach, E. Hauschteck-Jungen and L. W. Beukeboom (1996). "Is the Y chromosome of *Drosophila* an evolved supernumerary chromosome?" Bioessays **18**(4): 317-323.

Haerty, W., S. Jagadeeshan, R. J. Kulathinal, A. Wong, K. Ravi Ram, L. K. Sirot, L. Levesque, C. G. Artieri, M. F. Wolfner, A. Civetta and R. S. Singh (2007). "Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*." Genetics **177**(3): 1321-1335.

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**(11): 1118-1126.

He, X., X. Chen, Y. Xiong, Z. Chen, X. Wang, S. Shi, X. Wang and J. Zhang (2011). "He et al. reply." Nature Genetics **43**(12): 1171-1172.

Heliconius Genome, C. (2012). "Butterfly genome reveals promiscuous exchange of mimicry adaptations among species." Nature **487**(7405): 94-98.

International Silkworm Genome, C. (2008). "The genome of a lepidopteran model insect, the silkworm *Bombyx mori*." Insect Biochem Mol Biol **38**(12): 1036-1045.

Itoh, Y., E. Melamed, X. Yang, K. Kampf, S. Wang, N. Yehya, A. Van Nas, K. Replogle, M. R.

Band, D. F. Clayton, E. E. Schadt, A. J. Lusis and A. P. Arnold (2007). "Dosage compensation is less effective in birds than in mammals." J Biol **6**(1): 2.

Itoh, Y., K. Replogle, Y. H. Kim, J. Wade, D. F. Clayton and A. P. Arnold (2010). "Sex bias and dosage compensation in the zebra finch versus chicken genomes: general and specialized patterns among birds." Genome Res **20**(4): 512-518.

Jue, N. K., M. B. Murphy, S. D. Kasowitz, S. M. Qureshi, C. J. Oberfell, S. Elsis, R. J. Foley, R. J. O'Neill and M. J. O'Neill (2013). "Determination of dosage compensation of the mammalian X chromosome by RNA-seq is dependent on analytical approach." BMC Genomics **14**: 150.

Julien, P., D. Brawand, M. Soumillon, A. Necșulea, A. Liechti, F. Schutz, T. Daish, F. Grutzner and H. Kaessmann (2012). "Mechanisms and evolutionary patterns of mammalian and avian dosage compensation." PLoS Biol **10**(5): e1001328.

Kaiser, V. B. and D. Bachtrog (2010). "Evolution of sex chromosomes in insects." Annu Rev Genet **44**: 91-112.

Kelly, W. G., C. E. Schaner, A. F. Dernburg, M. H. Lee, S. K. Kim, A. M. Villeneuve and V. Reinke (2002). "X-chromosome silencing in the germline of *C. elegans*." Development **129**(2): 479-492.

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**(12): 1167-1169; author reply 1171-1162.

Khil, P. P., N. A. Smirnova, P. J. Romanienko and R. D. Camerini-Otero (2004). "The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation." Nat Genet **36**(6): 642-646.

Kiuchi, T., H. Koga, M. Kawamoto, K. Shoji, H. Sakai, Y. Arai, G. Ishihara, S. Kawaoka, S. Sugano, T. Shimada, Y. Suzuki, M. G. Suzuki and S. Katsuma (2014). "A single female-specific piRNA is the primary determiner of sex in the silkworm." Nature **509**(7502): 633-636.

Labandeira, C., D. Dilcher, D. Davis and D. Wagner (1994). "Ninety-seven million years of angiosperm-insect association: paleobiological insights into the meaning of coevolution." Proceedings of the National Academy of Sciences **91**(25): 12278-12282.

Lavery, C., J. Lucci and A. Akhtar (2010). "The MSL complex: X chromosome and beyond." Curr Opin Genet Dev **20**(2): 171-178.

Lepesant, J. M., C. Cosseau, J. Boissier, M. Freitag, J. Portela, D. Climent, C. Perrin, A. Zerlotini and C. Grunau (2012). "Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence." Genome Biol **13**(2): R14.

Lercher, M. J., A. O. Urrutia and L. D. Hurst (2003). "Evidence that the human X chromosome is enriched for male-specific but not female-specific genes." Mol Biol Evol **20**(7): 1113-1116.

Lin, F., K. Xing, J. Zhang and X. He (2012). "Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation." Proc Natl Acad Sci U S A **109**(29): 11752-11757.

Lin, H., V. Gupta, M. D. Vermilyea, F. Falciani, J. T. Lee, L. P. O'Neill and B. M. Turner (2007). "Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes." PLoS Biol **5**(12): e326.

Lin, H., J. A. Halsall, P. Antczak, L. P. O'Neill, F. Falciani and B. M. Turner (2011). "Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis." Nat Genet **43**(12): 1169-1170; author reply 1171-1162.

Lukhtanov, V. (2000). "Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta)." Journal of Zoological Systematics and Evolutionary Research **38**(2): 73-79.

Lyon, M. F. (1961). "Gene action in the X-chromosome of the mouse (*Mus musculus* L.)." Nature **190**: 372-373.

Mank, J. E. (2009). "The W, X, Y and Z of sex-chromosome dosage compensation." Trends Genet **25**(5): 226-233.

Mank, J. E. (2013). "Sex chromosome dosage compensation: definitely not for everyone." Trends Genet **29**(12): 677-683.

Marec, F., K. Sahara and W. Traut (2010). "Rise and fall of the W chromosome in Lepidoptera." Molecular biology and genetics of the lepidoptera **3**: 49-63.

Marin, I., A. Franke, G. J. Bashaw and B. S. Baker (1996). "The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes." Nature **383**(6596): 160-163.

Marshall Graves, J. A. (2008). "Weird animal genomes and the evolution of vertebrate sex and sex chromosomes." Annu Rev Genet **42**: 565-586.

Meiklejohn, C. D., E. L. Landeen, J. M. Cook, S. B. Kingan and D. C. Presgraves (2011). "Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation." PLoS Biol **9**(8): e1001126.

Meyer, B. J. (2000). "Sex in the worm: counting and compensating X-chromosome dose." Trends Genet **16**(6): 247-253.

Meyer, B. J. (2010). "Targeting X chromosomes for repression." Curr Opin Genet Dev **20**(2): 179-189.

Meyer, M. and M. Kircher (2010). "Illumina sequencing library preparation for highly multiplexed target capture and sequencing." Cold Spring Harb Protoc **2010**(6): pdb prot5448.

Mikhaylova, L. M. and D. I. Nurminsky (2011). "Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the *Drosophila X* chromosome." BMC Biol **9**: 29.

Monesi, V. (1965). "Differential rate of ribonucleic acid synthesis in the autosomes and sex chromosomes during male meiosis in the mouse." Chromosoma **17**(1): 11-21.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods **5**(7): 621-628.

Naurin, S., B. Hansson, S. Bensch and D. Hasselquist (2010). "Why does dosage compensation differ between XY and ZW taxa?" Trends Genet **26**(1): 15-20.

Nguyen, D. K. and C. M. Disteche (2006). "Dosage compensation of the active X chromosome in mammals." Nat Genet **38**(1): 47-53.

Nguyen, P., M. Sykorova, J. Sichova, V. Kuta, M. Dalikova, R. Capkova Frydrychova, L. G. Neven, K. Sahara and F. Marec (2013). "Neo-sex chromosomes and adaptive potential in tortricid pests." Proc Natl Acad Sci U S A **110**(17): 6931-6936.

Nozawa, M., N. Fukuda, K. Ikeo and T. Gojobori (2014). "Tissue- and stage-dependent dosage compensation on the neo-X chromosome in *Drosophila pseudoobscura*." Mol Biol Evol **31**(3): 614-624.

Ohno, S. (1967). Sex chromosomes and sex-linked genes, Springer.

Ohno, S., W. D. Kaplan and R. Kinoshita (1959). "Formation of the sex chromatin by a single X-chromosome in liver cells of *Rattus norvegicus*." Exp Cell Res **18**: 415-418.

Pala, I., S. Naurin, M. Stervander, D. Hasselquist, S. Bensch and B. Hansson (2012). "Evidence of a neo-sex chromosome in birds." Heredity (Edinb) **108**(3): 264-272.

Parisi, M., R. Nuttall, D. Naiman, G. Bouffard, J. Malley, J. Andrews, S. Eastman and B. Oliver (2003). "Paucity of genes on the *Drosophila* X chromosome showing male-biased expression." Science **299**(5607): 697-700.

Pessia, E., T. Makino, M. Bailly-Bechet, A. McLysaght and G. A. Marais (2012). "Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome." Proc Natl Acad Sci U S A **109**(14): 5346-5351.

Pringle, E. G., S. W. Baxter, C. L. Webster, A. Papanicolaou, S. F. Lee and C. D. Jiggins (2007). "Synteny and chromosome evolution in the lepidoptera: evidence from mapping in *Heliconius melpomene*." Genetics **177**(1): 417-426.

Reinke, V., I. S. Gil, S. Ward and K. Kazmer (2004). "Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*." Development **131**(2): 311-323.

Rice, W. R. (1984). "Sex chromosomes and the evolution of sexual dimorphism." Evolution: 735-742.

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**(1): 139-140.

Robinson, M. D. and A. Oshlack (2010). "A scaling normalization method for differential expression analysis of RNA-seq data." Genome Biol **11**(3): R25.

Sahara, K., A. Yoshido and W. Traut (2012). "Sex chromosome evolution in moths and butterflies." Chromosome Res **20**(1): 83-94.

Saifi, G. M. and H. S. Chandra (1999). "An apparent excess of sex- and reproduction-related genes on the human X chromosome." Proceedings of the Royal Society B: Biological Sciences **266**(1415): 203-209.

Schoenmakers, S., E. Wassenaar, J. W. Hoogerbrugge, J. S. Laven, J. A. Grootegoed and W. M. Baarends (2009). "Female meiotic sex chromosome inactivation in chicken." PLoS Genet **5**(5): e1000466.

Schulz, E. G. and E. Heard (2013). "Role and control of X chromosome dosage in mammalian development." Curr Opin Genet Dev **23**(2): 109-115.

Sichova, J., P. Nguyen, M. Dalikova and F. Marec (2013). "Chromosomal evolution in tortricid moths: conserved karyotypes with diverged features." PLoS One **8**(5): e64520.

Singh, N. D., A. M. Larracuente, T. B. Sackton and A. G. Clark (2009). "Comparative genomics on the *Drosophila* phylogenetic tree." Annual Review of Ecology, Evolution, and Systematics **40**(1): 459-480.

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**(3): 526-537.

Storey, J. D. and R. Tibshirani (2003). "Statistical significance for genomewide studies." Proc Natl Acad Sci U S A **100**(16): 9440-9445.

Sugimoto, M. and K. Abe (2007). "X chromosome reactivation initiates in nascent primordial germ cells in mice." PLoS Genet **3**(7): e116.

Talebizadeh, Z., S. D. Simon and M. G. Butler (2006). "X chromosome gene expression in human tissues: male and female comparisons." Genomics **88**(6): 675-681.

Tarazona, S., F. García-Alcalde, J. Dopazo, A. Ferrer and A. Conesa (2011). "Differential expression in RNA-seq: a matter of depth." Genome research **21**(12): 2213-2223.

Thomas, S. and R. S. Singh (1992). "A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. VII. Varying rates of genic divergence as revealed by two-dimensional electrophoresis." Mol Biol Evol **9**(3): 507-525.

Traut, W. and F. Marec (1996). "Sex chromatin in lepidoptera." Q Rev Biol **71**(2): 239-256.

Turner, J. M. (2007). "Meiotic sex chromosome inactivation." Development **134**(10): 1823-1831.

Uebbing, S., A. Kunstner, H. Makinen and H. Ellegren (2013). "Transcriptome sequencing reveals the character of incomplete dosage compensation across multiple tissues in flycatchers." Genome Biol Evol **5**(8): 1555-1566.

Van't Hof, A. E., P. Nguyen, M. Dalikova, N. Edmonds, F. Marec and I. J. Saccheri (2013). "Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism." Heredity (Edinb) **110**(3): 283-295.

Vibrantovski, M. D. (2014). "Meiotic sex chromosome inactivation in *Drosophila*." J Genomics **2**: 104-117.

Vicoso, B. and D. Bachtrog (2009). "Progress and prospects toward our understanding of the evolution of dosage compensation." Chromosome Res **17**(5): 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. and B. Charlesworth (2006). "Evolution on the X chromosome: unusual patterns and processes." Nat Rev Genet **7**(8): 645-653.

Vicoso, B., J. J. Emerson, Y. Zektser, S. Mahajan and D. Bachtrog (2013). "Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation." PLoS Biol **11**(8): e1001643.

Vitkova, M., I. Fukova, S. Kubickova and F. Marec (2007). "Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera)." Chromosome Res **15**(7): 917-930.

Wahlberg, N., C. W. Wheat and C. Peña (2013). "Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths)." PloS one **8**(11): e80875.

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.

Wolf, J. B. and J. Bryk (2011). "General lack of global dosage compensation in ZZ/ZW systems? Broadening the perspective with RNA-seq." BMC Genomics **12**: 91.

Wolf, K. W. (1996). "The structure of condensed chromosomes in mitosis and meiosis of insects." International Journal of Insect Morphology and Embryology **25**(1-2): 37-62.

Xia, Q., Z. Zhou, C. Lu, D. Cheng, F. Dai, B. Li, P. Zhao, X. Zha, T. Cheng, C. Chai, G. Pan, J. Xu, C. Liu, Y. Lin, J. Qian, Y. Hou, Z. Wu, G. Li, M. Pan, C. Li, Y. Shen, X. Lan, L. Yuan, T. Li, H. Xu, G. Yang, Y. Wan, Y. Zhu, M. Yu, W. Shen, D. Wu, Z. Xiang, J. Yu, J. Wang, R. Li, J. Shi, H. Li, G. Li, J. Su, X. Wang, G. Li, Z. Zhang, Q. Wu, J. Li, Q. Zhang, N. Wei, J. Xu, H. Sun, L. Dong, D. Liu, S. Zhao, X. Zhao, Q. Meng, F. Lan, X. Huang, Y. Li, L. Fang, C. Li, D. Li, Y. Sun, Z. Zhang, Z. Yang, Y. Huang, Y. Xi, Q. Qi, D. He, H. Huang, X. Zhang, Z. Wang, W. Li, Y. Cao, Y. Yu, H. Yu, J. Li, J. Ye, H.

Chen, Y. Zhou, B. Liu, J. Wang, J. Ye, H. Ji, S. Li, P. Ni, J. Zhang, Y. Zhang, H. Zheng, B. Mao, W. Wang, C. Ye, S. Li, J. Wang, G. K. Wong and H. Yang (2004). "A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*)."
Science **306**(5703): 1937-1940.

Xiong, Y., X. Chen, Z. Chen, X. Wang, S. Shi, X. Wang, J. Zhang and X. He (2010). "RNA sequencing shows no dosage compensation of the active X-chromosome."
Nat Genet **42**(12): 1043-1047.

Yasukochi, Y., L. A. Ashakumary, K. Baba, A. Yoshido and K. Sahara (2006). "A second-generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects."
Genetics **173**(3): 1319-1328.

Yasukochi, Y., M. Tanaka-Okuyama, F. Shibata, A. Yoshido, F. Marec, C. Wu, H. Zhang, M. R. Goldsmith and K. Sahara (2009). "Extensive conserved synteny of genes between the karyotypes of *Manduca sexta* and *Bombyx mori* revealed by BAC-FISH mapping."
PLoS One **4**(10): e7465.

Yildirim, E., R. I. Sadreyev, S. F. Pinter and J. T. Lee (2012). "X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription."
Nat Struct Mol Biol **19**(1): 56-61.

You, M., Z. Yue, W. He, X. Yang, G. Yang, M. Xie, D. Zhan, S. W. Baxter, L. Vasseur, G. M. Gurr, C. J. Douglas, J. Bai, P. Wang, K. Cui, S. Huang, X. Li, Q. Zhou, Z. Wu, Q. Chen, C. Liu, B. Wang, X. Li, X. Xu, C. Lu, M. Hu, J. W. Davey, S. M. Smith, M. Chen, X. Xia, W. Tang, F. Ke, D. Zheng, Y. Hu, F. Song, Y. You, X. Ma, L. Peng, Y. Zheng, Y. Liang, Y. Chen, L. Yu, Y. Zhang, Y. Liu, G. Li, L. Fang, J. Li, X. Zhou, Y. Luo, C. Gou, J. Wang, J. Wang, H. Yang and J. Wang (2013). "A heterozygous

moth genome provides insights into herbivory and detoxification." Nat Genet **45**(2): 220-225.

Zha, X., Q. Xia, J. Duan, C. Wang, N. He and Z. Xiang (2009). "Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*." Insect Biochem Mol Biol **39**(5-6): 315-321.

Zhan, S., C. Merlin, J. L. Boore and S. M. Reppert (2011). "The monarch butterfly genome yields insights into long-distance migration." Cell **147**(5): 1171-1185.

Zhang, Y. E., M. D. Vibranovski, P. Landback, G. A. Marais and M. Long (2010). "Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome." PLoS Biol **8**(10).